The basic principle and forwarding process of a three-layer Ethernet switch

This article briefly introduces the two-three layer forwarding mechanism of the three-layer Ethernet switch, mainly aiming to help readers further understand the basic principles and forwarding process of the switch, so as to be beneficial for better device maintenance work and establish an index for further learning.

这种方法被称为“健身”。

This method is called "fitness".

The forwarding mechanism of the three-layer Ethernet switch is mainly divided into two parts: Layer 2 forwarding and Layer 3 switching.

Two-Layer Forwarding Process

MAC Address Introduction

MAC Address is a 48 bit binary address, such as: 00-e0-fc-00-00-06.

It can be divided into unicast address, multicast address and broadcast address.

Unicast Address: The lowest bit of the first byte is 0, e.g.: 00-e0-fc-00-00-06

Multicast address: The lowest bit of the first byte is 1, e.g.: 01-e0-fc-00-00-06

Broadcast address: 48 bits all 1, e.g.: ff-ff-ff-ff-ff-ff

本文仅作参考

Note: This article is for reference only.

1) The MAC address of the network card or router device routing interface must be a unicast MAC address to ensure its interoperability with other devices.

2) MAC address is the basis for an Ethernet network device to operate on the network, and also the starting point for the realization of link layer functions.

Two-Layer Forwarding Introduction

The forwarding feature of the switch two-layer is compliant with the 802.1D bridge protocol standard.

The two key threads involved in the Layer 2 forwarding of a switch are the address learning thread and the packet forwarding thread.

The learning threads are as follows:

Switch receives all data frames from the subnet and builds a MAC address table based on the source MAC address in the received data frames.

2) Port Mobility Mechanism: If the switch finds that the incoming port of a packet is different from the source MAC address port, port mobility will be generated and the MAC address will be re-learned to the new port.

3) Address Aging Mechanism: If the switch does not receive a packet from a host for a long period of time, the MAC address corresponding to that host will be deleted, and it will be re-learned when the next packet arrives.

Note: Aging is also done according to the source MAC address.

Message Forwarding Thread

The switch looks for the destination MAC address in the MAC address table of the data frame. If found, the data frame is sent to the corresponding port, if not found, it is sent to all ports.

If the source MAC address and the destination MAC address in the packet received by the switch are on the same port, the packet is discarded.

3) The switch forwards broadcast packets to all ports other than the ingress port.

VLAN Layer 2 Forwarding Introduction

Message forwarding thread

After introducing VLAN, the following impacts were generated to the packet forwarding thread of the two-layer switch:

The switch looks up the destination MAC address in the MAC address table of the data frame. If it is found (and also making sure that the incoming VLAN and outgoing VLAN are consistent), it will send the data frame to the corresponding port. If it is not found, it will be sent to all ports (in the VLAN).

If the source MAC address and destination MAC address in the packet received by the switch are on the same port, the packet is discarded.

3) The switch forwards broadcast messages to all ports outside (VLAN) of the inbound port.

zh-CN: 你有没有收到我昨天发给你的邮件？

en: Did you receive the email I sent you yesterday?

By introducing VLANs on Ethernet switches, the following benefits are brought:

2）把网络分为两个部分，其中一部分用来处理数据，另一部分用来存储数据。

1）Limiting local network traffic can improve the overall network processing capabilities to some degree.

2）Divide the network into two parts, one part is used to process data and the other part is used to store data.

2) Virtual workgroups, through flexible VLAN settings, divide different users into the workgroup.

3) Security - users within a VLAN cannot access other VLANs, thus improving security.

zh-CN: 我们应该珍惜时间

en: We should cherish our time.

In addition, there are two common concepts of VLAN termination and transmission. From the literal meaning, we can easily understand these two concepts. The so-called VLAN transmission means that a VLAN is not only effective on one switch, but also needs to be extended to other Ethernet switches in some way, and is still effective on other devices; the meaning of termination is relative, and the effective domain of a VLAN cannot be extended to other equipment, or cannot be extended to other equipment through a certain link.

VLAN pass-through can be implemented using 802.1Q technology, and VLAN termination can be done using PVLAN technology.

IEEE802.1Q protocol is the technical standard of VLAN, mainly modifying the standard frame header and adding a tag field which contains VLAN ID and other VLAN information. Here we won't discuss the implementation in detail. If you are interested, you can read related standards and materials.

Note: When forwarding packets on a Trunk port, if the VLAN Tag of the packet is equal to the default VLAN ID configured on the port, then the Tag of the packet should be removed. After the other end receives the packet without the Tag information, it obtains the VLAN information of the packet from the PVID of the port. Therefore, when configuring, it is necessary to ensure that the PVID settings of the two ends of a Trunk link between two switches are the same.

Why go to Tag?

Doing this ensures that general users can still communicate normally after they are plugged into the Trunk, because ordinary users cannot recognize packets with 802.1Q Vlan information.

802.1Q technology can be used to effectively implement VLAN transmission, but sometimes it is necessary to terminate VLAN, that is, where the boundary of this VLAN terminates. PVLAN technology can well achieve this function and achieve the purpose of saving VLAN. Cisco's PVLAN means private vlan, while our PVLAN means primary vlan.

There are two types of VLANs here: Primary VLAN and Secondary VLAN (Sub VLAN).

The isolation of user two-layer packets has been realized, and the packets issued by the upper-level switch can be received by each user, simplifying configuration and saving VLAN resources. Specific implementation is not discussed here. If interested, please refer to relevant materials.

zh-CN: 我的朋友们都很喜欢购物。

en: My friends all like shopping.

Let's talk about the three-tier switching process.

Segmenting with VLANs isolates communication between VLANs, and a router supporting VLANs (a three-tier device) can establish communication between VLANs. But using a router to interconnect different VLANs in an enterprise campus network is obviously not in line with the trend of the times. Because we can use three-layer switching to achieve this.

Difference 1 (Performance): Traditional routers forward packets based on micro-processors, while three-layer switches forward packets via ASIC hardware, leading to a big performance difference.

Difference 2 (Interface Type): The interfaces of three-layer switches are mostly Ethernet interfaces, and there are no rich router interface types.

Difference 3: A three-layer switch can also operate in a two-layer mode, directly exchanging packet files that do not require routing, whereas a router does not have two-layer functions.

Let's first look at the process of device interoperability.

zh-CN: 她把手机放在桌子上。

en: She put her phone on the table.

As shown in the figure: two VLANs are divided on the switch, and routing interfaces are configured on VLAN1 and VLAN2 to realize interconnection between VLAN1 and VLAN2.

Interconnectivity between A and B (taking A sending a ping request to B as an example):

A checks the destination IP address of the packet and finds that it is in the same network segment as itself.

A-->B: ARP Request packet, broadcasted within VLAN1.

3) B--->A ARP Response Message.

A-->B: ICMP Request

5) B--->A: ICMP Reply

zh-CN: 我非常喜欢这件衣服。

en: I really like this piece of clothing.

Interconnection between A and C (taking A initiating a ping request to C as an example):

A checks the destination IP address of the packet, and finds that it is not in the same network segment as itself.

2) A-->switch (int vlan 1) ARP request packet, the packet is broadcasted within VLAN1.

3) Gateway --> A ARP Response Packet.

4) A---> switch icmp request (destination MAC is the MAC of int vlan 1, source MAC is the MAC of A, destination IP is C, source IP is A).

5) After receiving the packet, the switch determines that it is a three-layer packet. Check the destination IP address of the packet and find that it is in its own directly connected network segment.

Switch (int VLAN 2) --> C ARP request packet, the packet broadcasts within VLAN2.

7) C--->switch (int vlan 2) ARP reply packet;

Switch (int VLAN 2) --> C ICMP Request (destination MAC is C's MAC, source MAC is int VLAN 2's MAC, destination IP is C, source IP is A) Compared to the packet in Step 4), the MAC header has been re-packaged, while the fields above the IP layer remain mostly unchanged.

9) C-->A: The subsequent processing is basically the same as that of the previous ICMP request.

If there is already an entry in the ARP table, no ARP request message will be sent to the other party in the above processing steps.

这是一个美好的时光

This is a wonderful time.

How to differentiate between two and three layer data flow?

The 3526 product is a three-layer Ethernet switch, which includes both two-layer processing functions and three-layer processing functions in its processing process.

Distinguish the basic models of two- and three-layer forwarding.

zh-CN: 今天是我的生日

en: Today is my birthday.

As shown in the figure:

A three-layer switch has divided two VLANs. The communication between A and B is completed within one VLAN, which is a two-layer data flow for the switch. The communication between A and C needs to cross VLANs and is a three-layer data flow.

The above mentioned are macro approaches. On the micro level, how does a switch device differentiate between a layer two packet and a layer three packet after it has entered the port?

The destination MAC address of the packet from A to B is the MAC address of host B since they are in the same VLAN. However, for the packet from A to C which crosses VLANs, the destination MAC address is the MAC address of the virtual interface VLAN1.

Therefore, the standard for a switch to distinguish between layer 2 and layer 3 packets is to check if the destination MAC address of the packet is equal to the MAC address of the switch's virtual interface.

zh-CN: 我可以帮你吗？

en: Can I help you?

接收、转发和发送。

For example, with the S3526 switch, the whole processing process of the three-layer switch is divided into three parts: reception, forwarding and sending.

Platform Software Protocol Stack Component

The key features of this section are:

Run routing protocols and maintain routing information tables.

The IP protocol stack functions play an important role in the processing process of the whole system. When the hardware cannot complete the packet forwarding, this part can replace the hardware to complete the three-layer forwarding of the packet. In addition, the data flows of telnet, ping, ftp, snmp for switches are all processed in this part.

For example.

Show IP Route

路由表：

Routing tables are data structures used by network-enabled devices to determine the best path for a given data packet to travel from its source to its destination. Routing tables are used by routers to determine the most efficient route for a packet to take through a network. Each entry in the routing table contains information about the destination of a data packet, the route to take to reach it, and the cost associated with taking that route.

Destination/Mask Proto Pre Metric Nexthop Interface

0.0.0.0/0 static 0 10.0.1.3 eth0

0.0.0.0/0 Static 60 0 10.110.255.9 VLAN-Interface2

10.110.48.0/21 Direct 0 0 10.110.48.1 VLAN-Interface1

10.110.48.1/32 Direct 0 0 127.0.0.1 In LoopBack0

10.110.255.8/30 Direct 0 0 10.110.255.10 VLAN-Interface2

10.110.255.10/32 Direct 0 0 127.0.0.1 InLoopBack0

127.0.0.0/8 Direct 0 0 127.0.0.1 InLoopBack0

127.0.0.1/32 Direct 0 0 127.0.0.1 InLoopBack0

zh-CN: 你想要什么？

en: What do you want?

Maintain ARP table

Show ARP

IP Address MAC Address VLAN ID Port Name Type

Yes

10.110.255.9 00e0.fc00.5518 2 GigabitEthernet2/1 Dynamic Yes

F

10.110.51.75 0010.b555.f039 1 Ethernet0/9 Dynamic F

10.110.54.30 0800.20aa.f41d 1 Ethernet0/10 Dynamic

10.110.51.137 0010.a4aa.fce6 1 Ethernet0/12 Dynamic

10.110.50.90 0010.b555.e04f 1 Ethernet0/8 Dynamic

zh-CN: 我们应该坚持自己的原则。

EN: We should adhere to our principles.

：

Hardware processing process:

The main tables are: two-layer MAC address table and three-layer IP FDB table, which are used to save forwarding information. Under the condition of full forwarding information, the forwarding and processing of packets are all handled by hardware without the intervention of software. These two tables are independent of each other and have no relationship because as soon as a packet enters the switch, the hardware will distinguish whether the packet is two-layer or three-layer. It's either one or the other.

For example: show mac all:

MAC ADDR: 00-00-5E-00-53-01

VLAN ID: 10

STATE: LEARNING

PORT INDEX: 5

AGING TIME(s): 300

MAC ADDR: 00-00-5E-00-53-01

VLAN ID: 10

STATE: LEARNING

PORT INDEX: 5

AGING TIME(s): 300

0000.21cf.73f4 1 Learned Ethernet0/19 266

.0.0.1

Learned: Ethernet0/12: 225.0.0.1

0004.7673.0b38 1 Learned Ethernet0/9 262

.15.14.0

0005.5d04.9648 1 Learned Ethernet0/16 232.15.14.0

0005.5d04.9648: The MAC address of the device

1: The table ID

Learned: The entry type

Ethernet0/16: The port number

232.15.14.0: The IP address

0005.5df5.9f64 1 Learned Ethernet0/16 300

MAC Address table is an exact match IVL method, where the key parameters are: Vlan ID, Port Index.

zh-CN: 我们应该努力做最好的自己

En: We should strive to be the best version of ourselves.

For example: show ipfdb all:

0: System 1: Learned 2: UsrCfg Age 3: UsrCfg noAge Other: Error

0: System 1: Learned 2: User Configuration Age 3: User Configuration noAge Other: Error

IP Address RTIF Vtag VTValid Port MAC Status

1 Dynamic

10.11.83.77 2 2 Invalid GigabitEthernet2/1 00-e0-fc-00-55-18 1 1 Dynamic

1

10.11.198.28 2 2 Invalid GigabitEthernet2/1 00-e0-fc-00-55-18 1 1

1

10.63.32.2 2 2 Invalid GigabitEthernet2/1 00-e0-fc-00-55-18 1 1

IP Address: 10.63.32.2; Age: 2; Interface: GigabitEthernet2/1; MAC Address: 00-e0-fc-00-55-18; Type: Invalid; VLAN: 1; State: 1

10.72.255.100 2 2 Invalid GigabitEthernet2/1 00-e0-fc-00-55-18 2

10

10.75.35.103 2 2 Invalid GigabitEthernet2/1 00-e0-fc-00-55-18 2 10

10.75.35.106 2 2 Invalid GigabitEthernet2/1 00-e0-fc-00-55-18 2

RtIf (Routing Interface Index): This index is used to determine which forwarding table entry is located under which routing interface. For the 3526 product, the number of supported routing interfaces is 32.

Vlan tag: This value is used to indicate the VLAN it is in, which is corresponding to the routing interface.

VLAN tag valid (VTValid): used to identify whether VLAN tag needs to be inserted into the forwarded packets.

Port Index: Used to indicate the outgoing port of the forwarding table entry.

Next hop MAC: After each hop of forwarding from a three-layer device, the MAC header in the packet will be re-encapsulated. Generally, the hardware ASIC chip will encapsulate the packet header based on the value in this field.

zh-CN: 对那些受过良好教育的人来说，学习新技能并不容易。

en: For those who have had a good education, learning new skills is not easy.

尊重和公平

Two important concepts: Respect and Fairness

Parsing, unparsed, each time a message is received, ASIC will extract the source and destination addresses from it and search them in the MAC Table or IP Fdb Table. If the address can be found in the forwarding table, it is considered parsed. If not found, it is considered unparsed. According to whether the address is the source or destination, there can be combinations such as source parsing and destination unparsed, etc.

For two-layer unparsed packets, hardware can broadcast the packet within the VLAN, but for three-layer packets without address resolution, hardware does not process the packet at all, resulting in a CPU interrupt, and software is used for processing.

硬件处理器在计算机系统中负责处理输入的信息并将其转化为有用的输出。

The hardware part of the processing can be described with this phrase: the hardware processor is responsible for processing the input information and transforming it into useful output in the computer system.

After receiving the packet, judge whether it is a two- or three-layer packet, then check if the source and destination addresses have been parsed. If they have been parsed, the hardware will complete the forwarding of the packet. If the addresses are not parsed, a CPU interrupt will be generated and the software will learn the unparsed address.

zh-CN: 你想做什么？

en: What do you want to do?

Driver code section

The key cores are:

Address resolution task: In this task, the unresearched address reported is studied, so that the hardware can complete the subsequent message forwarding without software intervention.

Address Management Task: in order to facilitate the management and maintenance of software, the Software Department has kept a copy of the address table that is the same as the forwarding table in the hardware.

FIB (Forwarding Information Base) table: This table's information is derived from the routing information in the IP route table. It is placed in the driver part for the convenience of address resolution tasks when learning the IP address.

For example:

Display Fibonacci Sequence

Destination/Mask NextHop Flag Interface

0.0.0.0/0 10.110.255.9 I VLAN-Interface2

10.110.48.0/21 -> 10.110.48.1 -> D -> VLAN-Interface1

10.110.48.1/32 -> 127.0.0.1, Direction: InLoopBack0

10.110.255.8/30 is a subnet mask, 10.110.255.10 is an IP address, D is a Hop Count, and VLAN-Interface2 is a VLAN Interface.

10.110.255.10/32 127.0.0.1 D InLoopBack0

127.0.0.0/8 - The subnet mask for the subnet comprising the loopback address range of 127.0.0.1 - 127.255.255.254. D - Designates this route as a directly connected route. InLoopBack0 - The name of the interface associated with this route.

zh-CN: 我们需要努力工作

en: We need to work hard.

Three-layer forwarding mainly involves two key threads:

Address learning thread and message forwarding thread, this is similar to the two-layer threads.

The forwarding thread mainly forwards the packets according to the forwarding table (IPFDB table) generated by the address learning thread. If the information in it is enough, the forwarding process will be completed by the hardware. If the information is not enough, the address learning thread will be asked to learn, and the packet hardware will not be able to forward, which will be handed over to the software protocol stack for forwarding.

zh-CN: 这是一个有趣的工作。

en: This is an interesting job.

2) Address learning thread is mainly used to generate hardware forwarding table (IPFDB table).

Actually, an IPFDB table is similar to a Layer 2 MAC address table, but the specific entries in it represent different meanings and functions.

There is a question: In software forwarding engines such as routers, a routing table and an ARP table are consulted for each received packet to find the next hop. However, in a three-layer switch (such as the S3526), no routing table or ARP table is consulted when the packet is forwarded. Does this mean that these two tables are no longer useful?

The answer is of course negative. In the three-layer forwarding process of S3526, the process is usually like this: the first packet cannot be forwarded by hardware and needs to learn the IP address. At the same time, in order to ensure no packet loss, the packet is also forwarded by software. After learning is completed, the second and third packets are all forwarded by hardware. This process can also be summarized vividly by "one routing, multiple switching". In one routing, routing table and ARP table are used to learn IP address and forward the first packet. In the subsequent multiple switching processes, only IPFDB table is needed.