INTERNATIONAL TELECOMMUNICATION UNION

TELECOMMUNICATION STANDARDIZATION SECTOR

STUDY PERIOD 2022-2024

SG11-TD747/GEN

STUDY GROUP 11

Original: English

Question(s):	13/11	Geneva, 10-20 October 2023

TD

Source: Editors

Title: Consent – draft Recommendation ITU-T Q.3962 (ex. Q.joint_tr): Requirements and Reference Model for optimized traceroute of joint Internet Protocol/Multi-Protocol Label Switching (Geneva, 10-20 October 2023)

Contact:	Cancan Huang China Telecom China	Tel: +86 20 38639366 Fax: +86 20 38639489 Email: huangcanc@chinatelecom.cn
Contact:	Minrui Shi China Telecom China	Tel: +86 18918588657 Email: shimr@chinatelecom.cn
Contact:	Yongsheng Liu China Unicom China	Tel: +86 10 18601106253 Email: liuys170@chinaunicom.cn
Contact:	Fangzheng Nie State Grid of China China	Tel: +86 10 13604001445 Email: m13604001445@163.com

Abstract: This document is the output text of draft Recommendation ITU-T Q.3962 (ex. Q.joint_tr) “Requirements and Reference model for optimized traceroute of joint Internet Protocol/Multi-Protocol Label Switching”. It includes the discussion results in the Q13/11 meeting held at Geneva, 10-20 October 2023. It is proposed for Consent.

This document is the draft recommendation for Q.3962 (ex. Q.joint_tr) "Requirements and Reference model for optimized traceroute of joint Internet Protocol/Multi-Protocol Label Switching".

The following table shows discussion results for contributions.

Document Number	Source	Title	Meeting results
C319	China Telecom China Unicom	ITU-T Q.joint_tr Requirements and Reference Model for optimized traceroute of joint IPMPLS-Proposal for editorial modification-for consent	Accepted with modifications: (1) all of the “vrf” changes to “VRF” (2) modifying the format of Figure 6-2 and Figure 8-1 for better reading (3) all of the “substitute changed to substituted” (4) changing the serial number of each steps for clause 8; (5) changing the P1 to Pn in clause 8. Pn represents P1, P2 and other P router if they are existed in practice. (6) Modifying “XXX in is shown in the figure 6-2”.

Draft new Recommendation ITU-T Q.3962 (ex. Q.joint_tr)

Requirements and Reference Model for optimized traceroute of joint Internet Protocol/Multi-Protocol Label Switching

Summary

This Recommendation aims to solves the problems of wrong failure location and performance information which brought by the traditional isolated traceroute tools in joint Internet Protocol /Multi-Protocol Label Switching (IP/MPLS) scenario. This Recommendation describes the requirements and reference model for optimized traceroute for joint IP/MPLS.

Keywords

ICMP traceroute, IP/MPLS.

Table of Contents

Draft Recommendation ITU-T Q.3962 (ex. Q.joint_tr)

Requirements and Reference Model for Optimized Traceroute for joint Internet Protocol/Multi-Protocol Label Switching

1. Scope

The scope of this Recommendation consists of:

(1) Requirements of route tracing of joint IP/MPLS;

(2) Methods of optimized traceroute of joint IP/MPLS;

(3) Reference Model for optimized traceroute of joint IP/MPLS.

2. References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

None

3. Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

None

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 **Substituted IP address**: An IP address that replaces the original IP address of the device during a test scenario that uses ping or traceroute.

Note: The substituted IP addresses are only used for the purpose of network security, and should not be assigned to the customers.

4. Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

CE	Customer Equipment
ICMP	Internet Control Message Protocol
IP	Internet Protocol
LFIB	Label Forwarding Information Base
MPLS	Multiple Protocol Label Switch
PE	Provider Edge
TTL	Time To Live
VPN	Virtual Private Network

VRF Virtual Routing Forwarding

5. Conventions

In this Recommendation:

The keywords "**is required** " indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.

6. Background

There are several service scenarios using joint Internet Protocol IP/Multi-Protocol Label Switching (IP/MPLS). ”Joint” here means that the end-to-end path is jointed by several sections using different technologies. Figure 6-1 shows a typical end-to-end enterprise network. In this scenario, customer equipment (CE) accesses to the carrier network using IP protocol. Within the carrier’s network, the carrier uses MPLS protocol to transfer enterprise’s packets. So, it is a typical service of jointly using of different protocols (IP and MPLS).

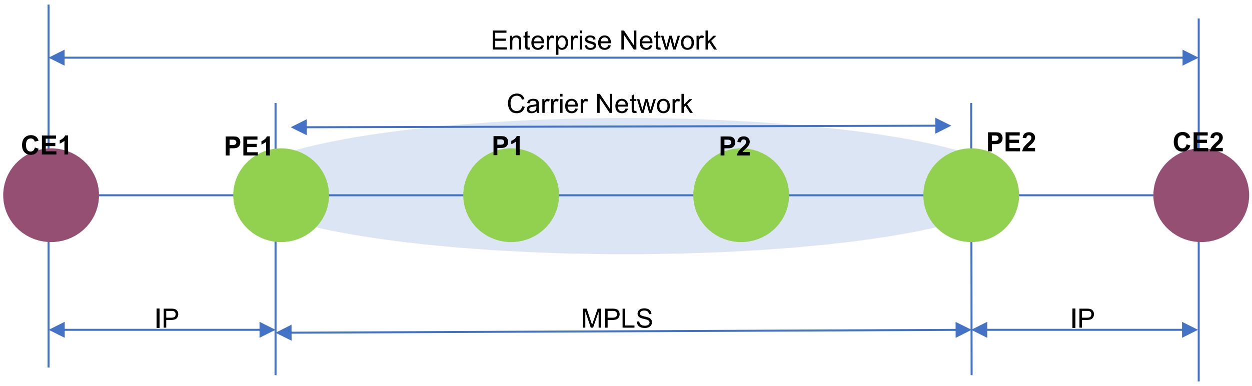

Figure 6-1 Enterprise IP Network carried by service provider’s MPLS network

In such scenario, the current route tracing technologies are IP traceroute technology and MPLS traceroute technology. Those tools are running perfectly in separate IP or MPLS environments. But when they are put together, because of they are run in different layers and cannot communicate with each other, it is incapable to find out the breakdown points in an IP and MPLS joint network effectively.

Here is a simple example which explains the behavior when route trace is triggered from CE1 to remote CE2 through IP and MPLS domain.

(1) The first packet is sent with Time To Live(TTL) value of 1 in IP header from CE1 to CE2:

This is a normal IP packet and when it arrives at Provider Edge 1(PE1), the TTL value decreases to 0. PE1 generates the Internet Control Message Protocol(ICMP) error message and sends it directly to CE1.

(2) The second packet sent with TTL=2 in IP header from CE1 to CE2:

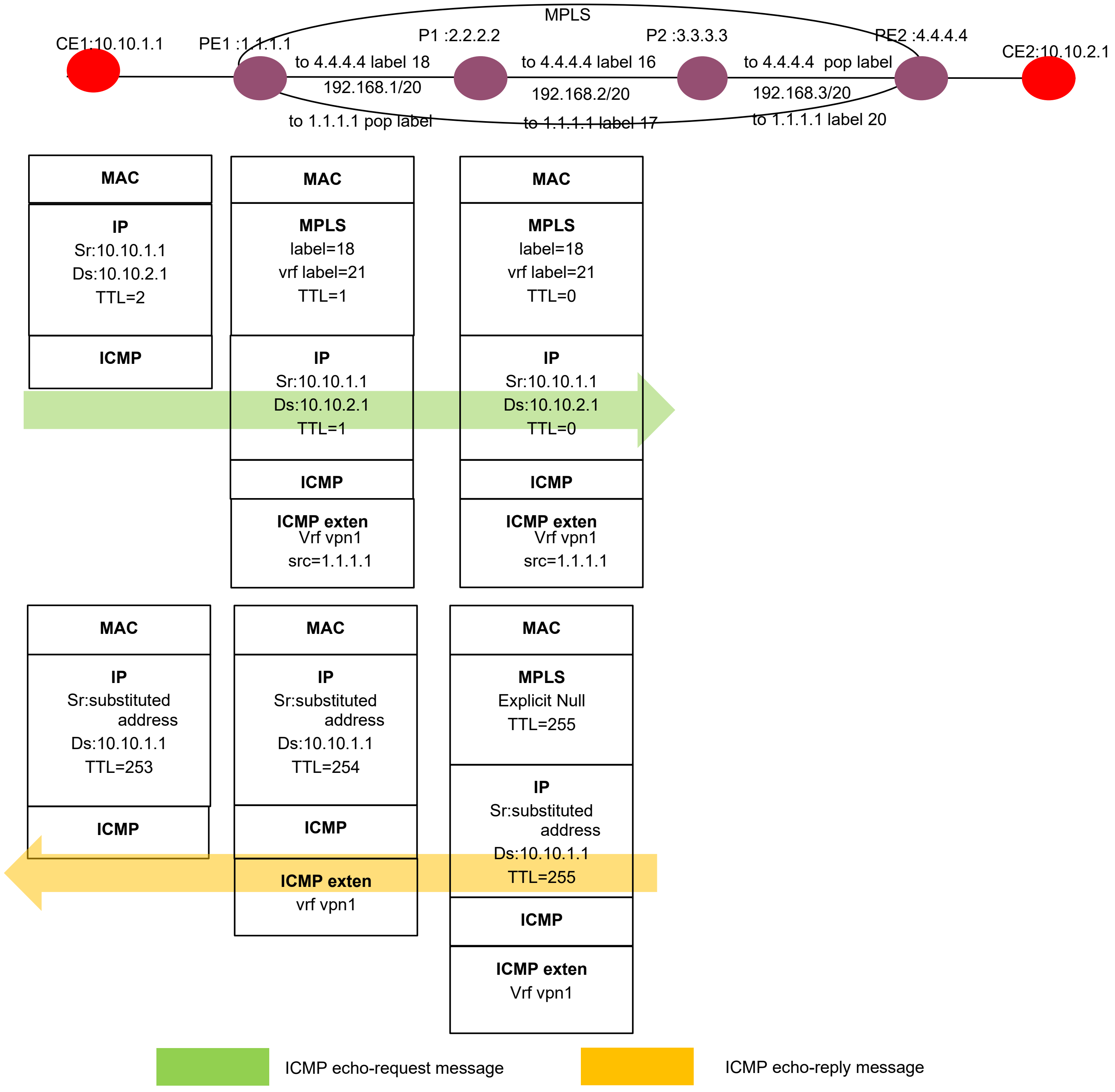

Figure 6-2 -example of traceroute for End-to-End Enterprise Network

The example of TTL=2 traceroute for end-to-end enterprise network is shown in the figure 6-2. After the traceroute, packet arrives at provider edge (PE) 1, the TTL of IP header decreases to 1. The PE1 adds the two MPLS layer tags(outside label=18;inside Virtual Routing Forwarding(VRF) label=21) to the packet header and sends the packet to PE2.The traceroute operation within the MPLS domain is using MPLS traceroute tool. In MPLS scenarios, the traceroute packet is switched based on the MPLS tag values, not destination IP addresses of CE2 .
There are two different options when PE1 transforms the IP traceroute packet to MPLS traceroute packets:

A. One option is not copying the TTL value to the MPLS header from the IP header in PE1.

Usually, it is forbidden to leak the service providers' network information. For example, it is forbidden to leak the IP address of the routers to the customers. So, the TTL value of the IP header will not be copied to the MPLS header and the service provider's network is transparent to the customers. In this situation, when there are network failures between CE1 and CE2, the customer has no opportunities to know where the network failure is happened in the service providers' network or in customer network.

B. The other one is copying the TTL value to the MPLS header from the IP header in PE1.
In this way, PE1 re-generates the traceroute packet with MPLS header of TTL value =1 which is copied from IP header. When packet arrives at P1, the MPLS TTL value is decreased to 0. P1 buffers the label stack and generates ICMP error message and includes the incoming label stack from the buffer in ICMP payload. It further populates the IP header with source address from incoming interface (192.168.1.1) of the labeled packet, destination address as the source of the labeled packet(10.10.1.1). The TTL value is set to 255. It now pushes the label stack from the buffer and consults the label forwarding information base (LFIB) table for forwarding action on top label. In the above topology for example, the received label stack is {18, 21}. On performing a lookup in LFIB table for top label, 18 will be swapped with label 16 and will be forwarded towards next hop P2. P2 in turn will pop the top label and forward the packet to PE2. PE2 will use the VRF label 21 to identify the VRF and forward the packet back towards CE1(outer label =20,VRF label=35).To conclude, the ICMP packet is not sent back to CE1 directly from P1. Instead, this ICMP packet generated from P2 is firstly steered to the end point of MPLS Virtual Private Network(VPN) tunnel PE2 and then steered back to CE1 from PE2.
The drawbacks of this method are obvious:

- The packet takes a long journey with a big circle to reach the destination. It not only cost a large time-delay but also waste the bandwidth from P2 to PE2.
- Most seriously, the statistic of delay reflected from P2 to CE1 is totally incorrect. It cannot provide the correct time cost between P2 and CE1 and consequently it is meaningless.

In summary, if the TTL value is not copied from IP header, the customer has no opportunity to find out the location of failure points if they are located in the service provider's network. And if the TTL value is copied, the customer has the opportunity to find the failure point, but the delay/loss/jitter information of this failure point is not correct.

7. Requirements of route tracing of enterprise network

Since the enterprise network is constructed by different technologies and different network operators, to coordinate all of these factors to find out the failures in real time, the traceroute technology has following requirements:

- It is required that the TTL value of IP header should be copied to TTL field of MPLS header. In this way, the customer is capable to identify the faults no matter the faults are located in the MPLS domain of service provider or in the IP domain of the enterprise itself;
- It is required that the service providers' network information won't be leaked to the customer;
- It is required that the delay/packet loss/jitter information of the failure point are correct.

8 Optimized method of route tracing of enterprise network

It is essential to directly steer the ICMP echo-reply message from tested node to the source node.

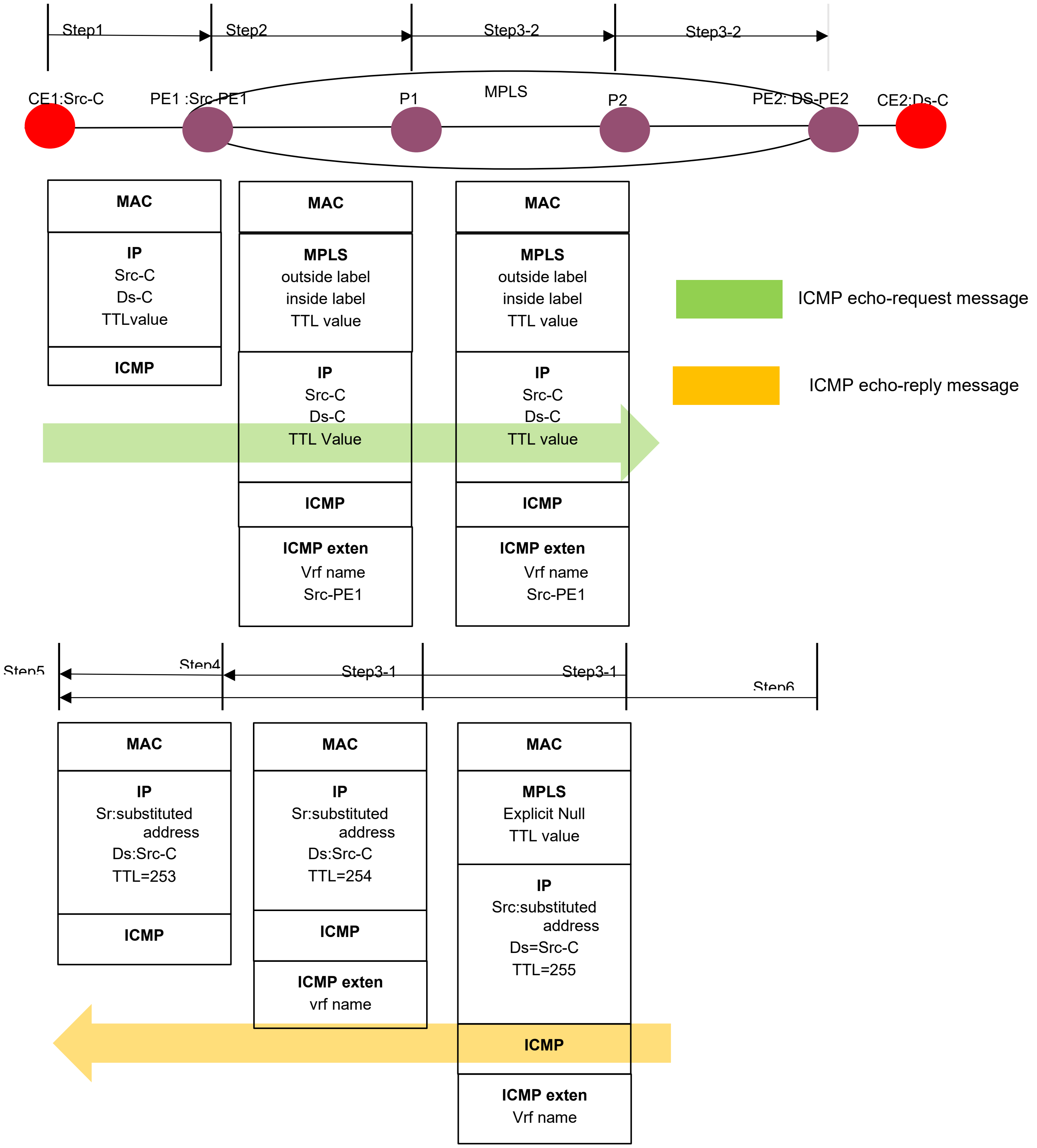
Step1
Step2
Step3-2
Step3-2
CE1:Src-C
PE1 :Src-PE1
P1
MPLS
P2
PE2: DS-PE2
CE2:Ds-C
MAC
IP
Src-C
Ds-C
TTLvalue
ICMP
MAC
MPLS
outside label
inside label
TTL value
IP
Src-C
Ds-C
TTL Value
ICMP
ICMP exten
Vrf name
Src-PE1
MAC
MPLS
outside label
inside label
TTL value
IP
Src-C
Ds-C
TTL value
ICMP
ICMP exten
Vrf name
Src-PE1
ICMP echo-request message
ICMP echo-reply message
Step5
Step4
Step3-1
Step3-1
Step6
MAC
IP
Sr:substituted address
Ds:Src-C
TTL=253
ICMP
MAC
IP
Sr:substituted address
Ds:Src-C
TTL=254
ICMP
ICMP exten
vrf name
MAC
MPLS
Explicit Null
TTL value
IP
Src:substituted address
Ds=Src-C
TTL=255
ICMP
ICMP exten
Vrf name

Figure 8-1 - Optimized traceroute for joint IP/MPLS

NOTE: The detailed example of this optimized method of route tracing of enterprise network is illustrated in Appendix I.

Step1: In the topology of figure 8-1, ICMP traceroute packet (ICMP echo-request packets) which TTL value=n in IP header is triggered from CE1 to CE2.

Step2: When ICMP echo-request packets arrives at PE1, PE1 will do the following actions:

Phase 1: To deal with the received IP packet:

a) Minus the TTL value by 1 of the IP header;

b) Decapsulating the source IP address Src-C from the IP header;

Phase 2: To reconstruct the outgoing ICMP echo-request packet:

a) finding out the mapped VRF name by checking the source IP address from VRF routing table;

b) finding out the VRF name related outside outgoing label;

c) finding out the VRF name related inside outgoing label;

d) finding out the VRF name related inside incoming label;

e) Adding the outside label to MPLS header;

f) Adding the inside outgoing VRF label to MPLS header;

g) Copying the TTL value from IP header to MPLS header;

h) Remaining the TTL value of the IP header;

i) Adding the VRF name to the ICMP payload;

j) Adding the source end of MPLS VPN tunnel (Src-PE1) to the ICMP payload;

k) Sending the restructured ICMP echo-request packets to next node.

Step3: When the restructured packets arrives at next Pn node (P1 or P2 in figure 8-1), Pn will do the following actions:

Minusing the TTL of MPLS layer. If TTL value=0, go to Step3-1. If the TTL value>0, go to Step3-2.

Step3-1: The TTL value=0, it means that the packet is time exceeded. Pn will generate ICMP error message (also known as echo-reply message):

a) Checking the source end of MPLS VPN (Src-PE1)tunnel carried in ICMP payload;

b) Checking the LFIB and find out Src-PE1 mapping to label “Explicit Null”;

c) Generating the MPLS header of ICMP error message:

i). Adding Explicit None to MPLS outside label field;

ii). Setting the TTL=255;

d) Generating the IP header of ICMP error message:

i). Checking out the substituted Pn’s IP address related to the original Pn’s IP address from a dedicated table maintained in Pn which stores the mapping of these two kinds of addresses.

NOTE 1 -To satisfy the second requirements of clause 7 that “It is required that the service providers’ network information won’t be leaked to the customer ”, Pn should acknowledges the substituted IP address instead of the original IP address of itself to the CE1 for the sake of hiding itself.

NOTE 2 - The substituted IP address and original IP address mapping table maintained in Pn is synchronized by the related server administrated by the network service provider.
NOTE 3 - The substituted IP address aims to prevent the DDOS attack from the hackers. If the traceroute packet is initiated by the operator of the network service provider, the operator will check out the original IP address of Pn based on the substituted address from the mapping table and run the regular operations according to the original IP address to figure out the network failures. The second requirement of clause 7 is meet and the routers within the service provider's domain is able to cancel the ping prohibition.

ii). Setting the substituted Pn's IP address as the IP source address in the IP header;

iii). Setting Src-C which is read out from the source IP address field of IP header of the ICMP request packet;

iv). Setting the IP TTL=255.

e) Generating the ICMP payload and remains the VRF name in the ICMP payload.

f) Going to Step 4.

Step3-2: The TTL value>0, Pn reconstructes the ICMP echo-request packet by updating the TTL values of MPLS and send it to the next node. If the next node is a P router, it goes to Step 3. If the next node is PE router, it goes to Step6.

Step4: When PE1 receives the ICMP echo-reply packets from the Pn router, it takes the actions as follows:

a) reading out VRF name from the ICMP payload;

b) reading out the IP destination address(Src-C) from the IP header;

c) checking the FIB of VRF and find the outgoing interface to go to the Src-C and redirect the ICMP the packet to that interface.

d) Going to Step5.

Step5: When CE1 receives the packets from PE1, it reads out the Src-C and find out itself is the termination. It also find out the related time cost from the echo-reply message.

Step6: When PE2 receives the ICMP echo-request message, it set the Src-C as the destination IP address and its loopback IP address as the source IP address in the echo-reply message. It directly sends it the echo-reply message to the CE1.

Hereto, a complete processing of TTL=n traceroute initiated by the enterprise customer is finished.

9 Reference Model of Optimized traceroute for joint IP/MPLS

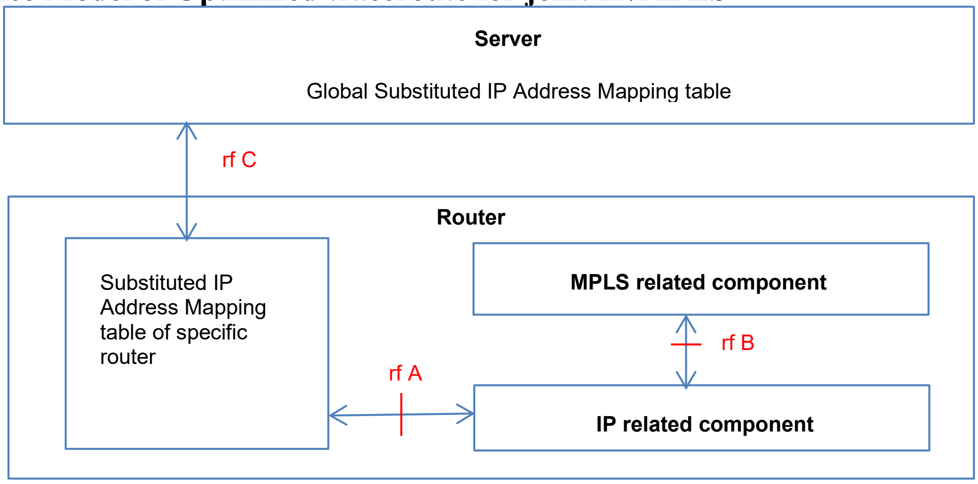
Server
Global Substituted IP Address Mapping table
rf C
Router
Substituted IP Address Mapping table of specific router
MPLS related component
rf B
rf A
IP related component

Figure 9-1 - Reference Model of Optimized traceroute for joint IP/MPLS

As shown in figure 9-1, the reference model of the optimized traceroute for joint IP/MPLS is composed by four components:

(1) MPLS related component. It is responsible for general MPLS operations and additionally copies the IP TTL value to the MPLS TTL value.

(2) IP related component. It is responsible for general IP operations and additionally transferring the IP TTL value to the MPLS related component.

(3) Substituted IP address Mapping table of specific router: It is responsible for maintaining the substituted IP address mapping table of specific router.

(4) Global Substitute IP address Mapping table: It is responsible for maintaining the substituted IP addresses mapping table of each router.

The reference model also includes three reference points:

(1) Reference point rf A: It is located between Substituted IP address Mapping table of specific router component and IP related component. It is responsible for transferring substituted IP address for the specific router to the IP related component for the purpose of further IP header encapsulation.

(2) Reference point rf B:It is located between MPLS related component and IP related component. It is responsible for transferring IP TTL value to the MPLS related component.

(3) Reference point rf C: It is located between Global Substituted IP address Mapping table and Substitute IP Address Mapping table of specific router. It is responsible for transferring the updating and alignment information of substituted IP address of the specific router and the global substituted IP address mapping table.

Appendix I
An example of Optimized Method of Route Tracing of Enterprise Network

(This appendix does not form an integral part of this Recommendation.)

This appendix provides an example based on the rules of optimized method of route tracing of enterprise network which is described in clause 8.

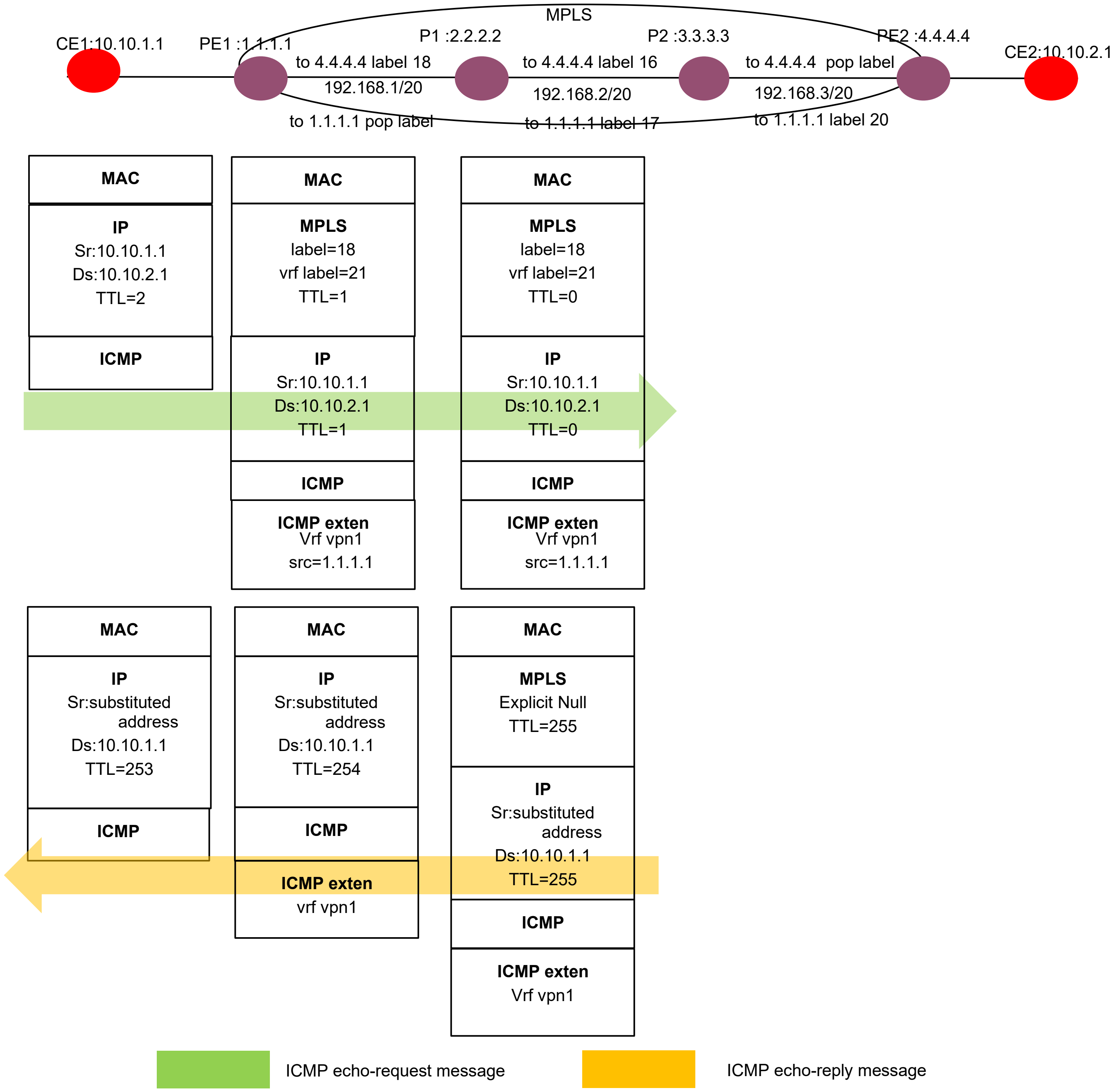

Figure I.1 - The example of Optimized ICMP pattern of TTL=2 for End-to-End Enterprise Network

In topology shown in figure I.1, ICMP traceroute packet which TTL=2 in IP header is triggered from CE1 to CE2.

When ICMP echo-request packets arrives at PE1, PE1 will check the source IP address 10.10.1.1 from VRF routing table, and found it is mapped to VRF "vpn1". Then PE1 will find out:

- The VRF VPN1 endpoint is PE2(4.4.4.4) and the related outside outgoing label =18.
- The VRF vpn1 outgoing label =21. This label is also known as inside label.
- The VRF vpn1 incoming label =35 , This label is also known as inside label.

PE1 encapsulate the ICMP echo-request packets with：

- Adding the outside label=18 to MPLS header;
- Adding the inside outgoing VRF label=21 to MPLS header;
- Copying the TTL=1 from IP header to MPLS header;
- Remaining the TTL=1 of the IP header;
- Adding the VRF name "vpn1", inside to the ICMP payload;
- Adding the source end of MPLS VPN tunnel 1.1.1.1(PE1) to the ICMP payload;

When the restructured packets arrive at P1:

- The TTL of MPLS layer will minus 1 and decrease to 0, so P1 will generate ICMP error message(echo-reply message);
- P1 first check the source end of mpls vpn tunnel 1.1.1.1 carried in ICMP payload. Then P1 check the LFIB and find out 1.1.1.1 mapping to label "Explicit Null";
- P1 generates the MPLS header of ICMP error message :
  a) It adds Explicit None to MPLS outside label field
  b) It sets the TTL=255.
- P2 generates the IP header of ICMP error message :
  a) It checks out the substituted P1's IP address (eg.127.0.0.1) related to the original P1's IP address 2.2.2.2 from a dedicated table maintained in P1 which stores the mapping of these two kinds of addresses;
  b) It sets the substituted P1's IP address as the IP source address in the IP header of ICMP error message,
  c) It sets destination IP address to 10.10.1.1 which is read out from the IP header of the ICMP request packet.
  d) It sets the IP TTL=255
- P2 generates the ICMP payload, it remains the VRF name "vpn1" in the ICMP payload.

When PE1 receives the packets from P1, it will take the actions as follows:

- It reads out vpn name "vpn1" from the ICMP payload;
- It reads out the IP destination address 10.10.1.1 from the IP header;
- It checks the FIB of "vpn1" and find the outgoing interface to go to the destination 10.10.1.1 and redirect the ICMP the packet to that interface.

When CE1 receives the packets from PE1, it reads out the "source IP address"=127.0.0.1 and the related time cost from this node.

Hereto, a complete processing of TTL=2 traceroute initiated by the enterprise customer is finished.

Through this optimized method, P1 is able to generate the ICMP packets in a better way to make a direct throw between tested node and source node. Consequently, an accurate traceroute statistic could be collected, and the network resources (like bandwidth) are saved.