



P2MP TE Label
Switched Multicast
の実装における課題



Tetsuya Murakami
tmurakam@cisco.com
October 10 2007



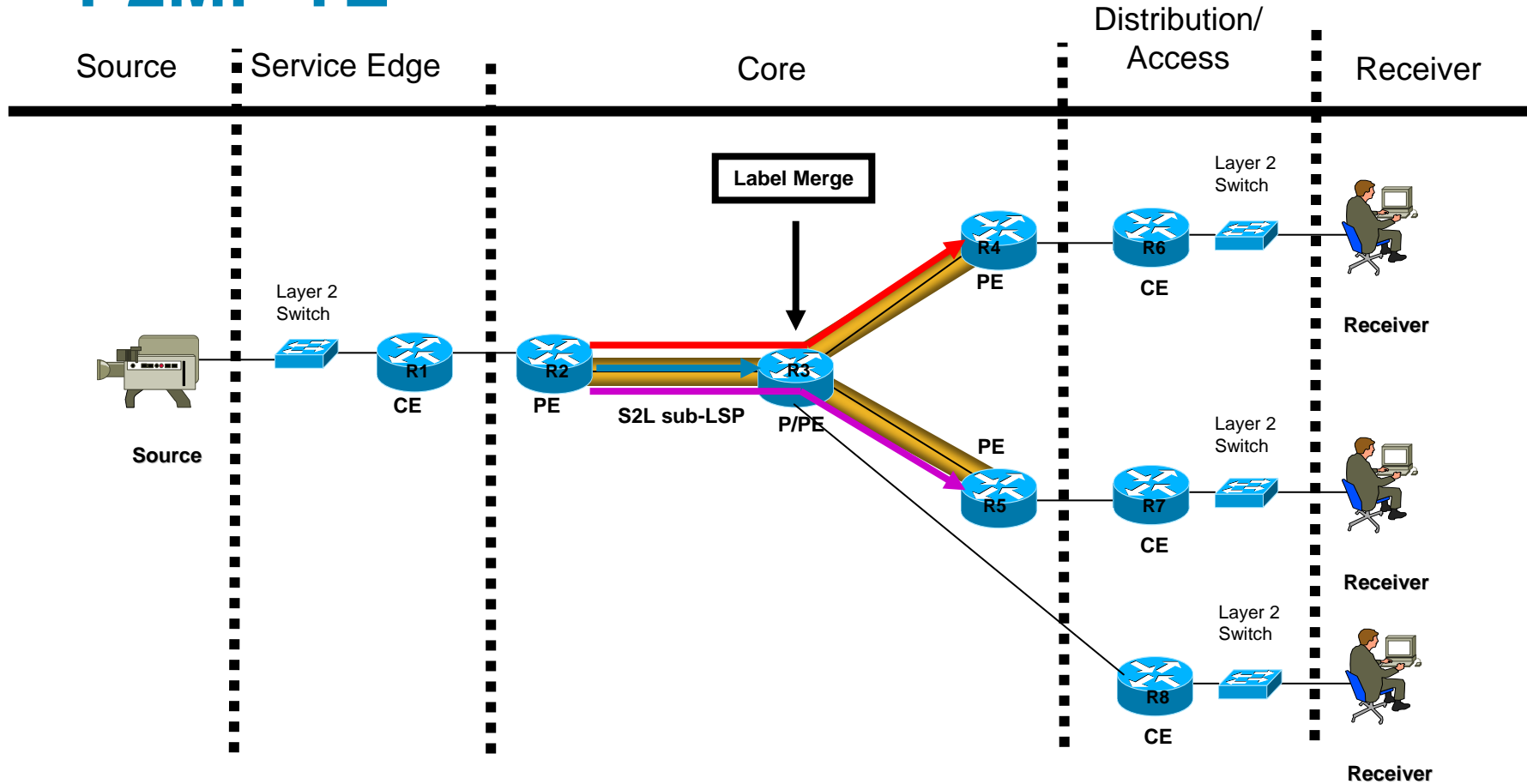
Agenda

- JDC (Japan Development Centre)
- P2MP TE LSP
- PHP
- Payload Discrimination
- Forwarding plane

Japan Development Centreとは。

- 2005/2 JDC設立
- 現在、10人
- IOSおよびIOS-XRの開発に従事(コードかいてます！！)
 - IOS
 - Metro-Ethernet (current)
 - IPv6 Proxy Mobile
 - IOS-XR
 - P2MP TE (current)
 - P2MP Ping/Trace (current)
 - BFD (current)
 - DHCPv6

P2MP TE



- one S2L sub-LSP signaling, multiple S2L sub-LSP signalingの2つのsignaling方法が存在する。
- Headendは、各Leaf nodeに対し個別にsignalingを行い、S2L(Source-to-Leaf) sub-LSPをset upする。
- Replication Nodeでは、複数のS2L sub-LSPに対し、1つのlabelのみallocationする (Label Merge)。
- Tailendとして動作するS2L sub-LSPとMidnodeとして動作するS2L sub-LSPを同時にもつ場合、Bud nodeとして動作する (Label forwardingとMulticast forwardingを同時に行う)。

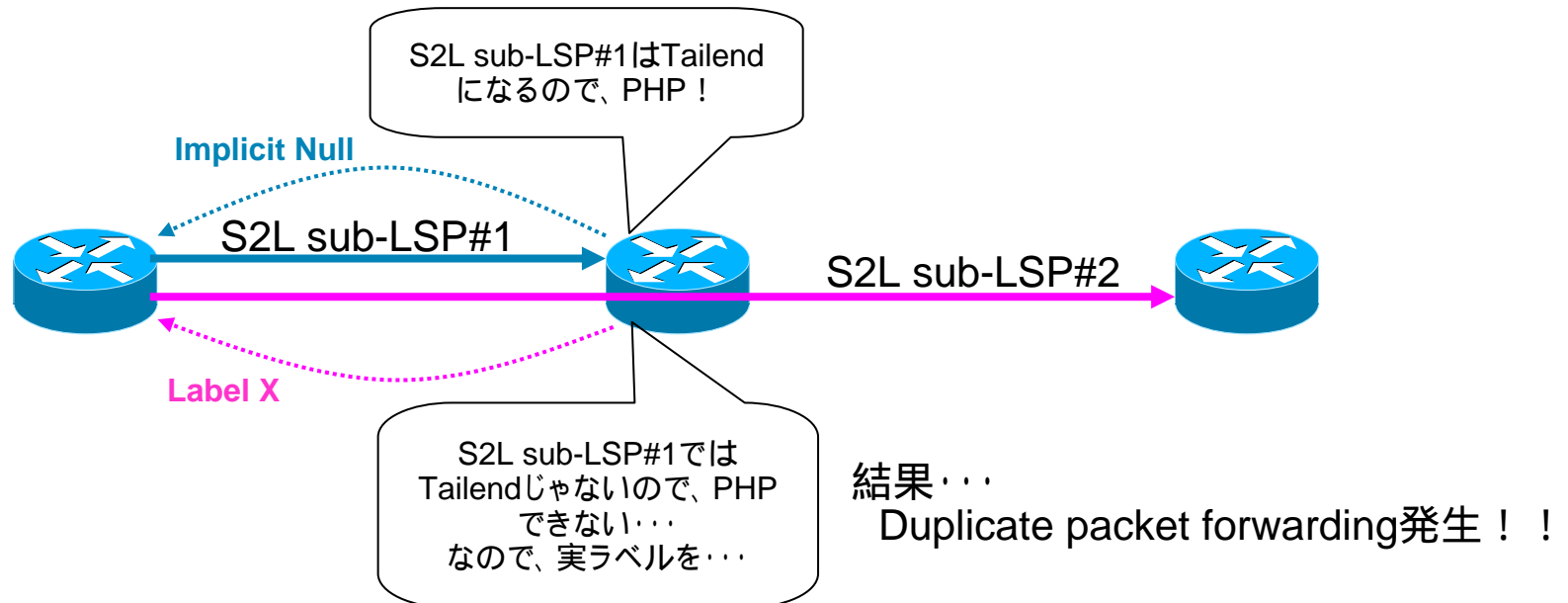
PHP (1)

- PHPはしてもいい??

- 弊社ルータは、通常PHPを行う。
- P2MP TE LSPの場合もOK?

- 答えはNo!

- 各S2L sub-LSPは独立してsignalingされるため、Tailendとして動作するS2L sub-LSPをsignaling後、Mid nodeとして動作するS2L sub-LSPがsignalingされた場合、



PHP (2)

- Leaf nodeにおいてRPF checkを行うためには、
 - Incoming情報が何か必要
 - PHPされるとどのLSPではいつてきたかも分からなくなる・・・
 - incoming labelがあれば、どのLSPから入ってきたか分かる。
 - PHP (implicit null) をすると、input packetはmulticast packetとなり、incoming interfaceにおいてもmulticastを有効にする必要がある。

結果、PHPはしない(できない)。

デメリットは、

**Leaf NodeにてLabel switching & multicast routingが必要
lookup (label, IP) が2回となり、forwarding性能に影響。**

Payload Discrimination (1)

- Label Packetのpayloadは、IPv4 or IPv6??

Leaf NodeにてLabel Packet受信時、Label Lookup後IP forwarding処理を行うが、IPv4 or IPv6のどちらとして処理するか？

IPv4のforwarding処理とIPv6のforwarding処理はプロセスが別のため、事前に判断しなければならない。

2つ考えました。

- Snoopingして判断
- labelから判断

Payload Discrimination (2)

- Snoopingして判断

Label POP後にpayload部分のIPとしてversionをチェック

- IP以外のframeがpayloadにあると思わぬ動作になることが。。
- Label forwardingでIP version checkをすることは、開発エンジニアから反対が。。
Label forwardingでIP headerをみるのは、やだ??

というわけで、採用されず。

Payload Discrimination (3)

- Labelから判断

受信したlabel frameのinner labelにて判断

- labelとpayload typeをどのように結びつけるか？
 - L3PIDを使用
 - additional labelを使用

- L3PIDの場合

RSVP-TEのPath messageに含まれるLabel Request ObjectにL3PIDが定義されている。

L3PIDとは、そのpathを使用するLayer3 protocolを指定するためのもの。

L3PIDの値として、standard ether typeが使用される。

これにより、LSP上で転送されるLayer3 protocolを判断可能。

Single stack (P2MP Tunnel Labelのみ) でforwardingされる場合は有効。

ただし、各LSPがIPv4 multicast or IPv6 multicast専用になってしまう。

(同一LSP上にIPv4, IPv6の両方のtrafficを同時には転送できない)

サポート決定。

Payload Discrimination (4)

2. additional labelを使用する場合

- MPLS VPNのように他プロトコル(BGP等)でmulticast routeとlabel情報の交換が行われれば、そのlabelにより判断可能
- Headendにて単にmulticast traffic (VPNではなく)をP2MP LSPにstatic mappingするような場合にもout-band signalingを使うのはoverheadが大きくなりすぎる。

Headendにて、inner labelとしてexplicit null labelをpushし、Core内はTop label (P2MP Tunnel Label)にて転送処理が行われる。

Leaf nodeにてinner labelであるexplicit null labelからpayloadのprotocolを判断する。

ただし、大きな反対が・・・、理由は、

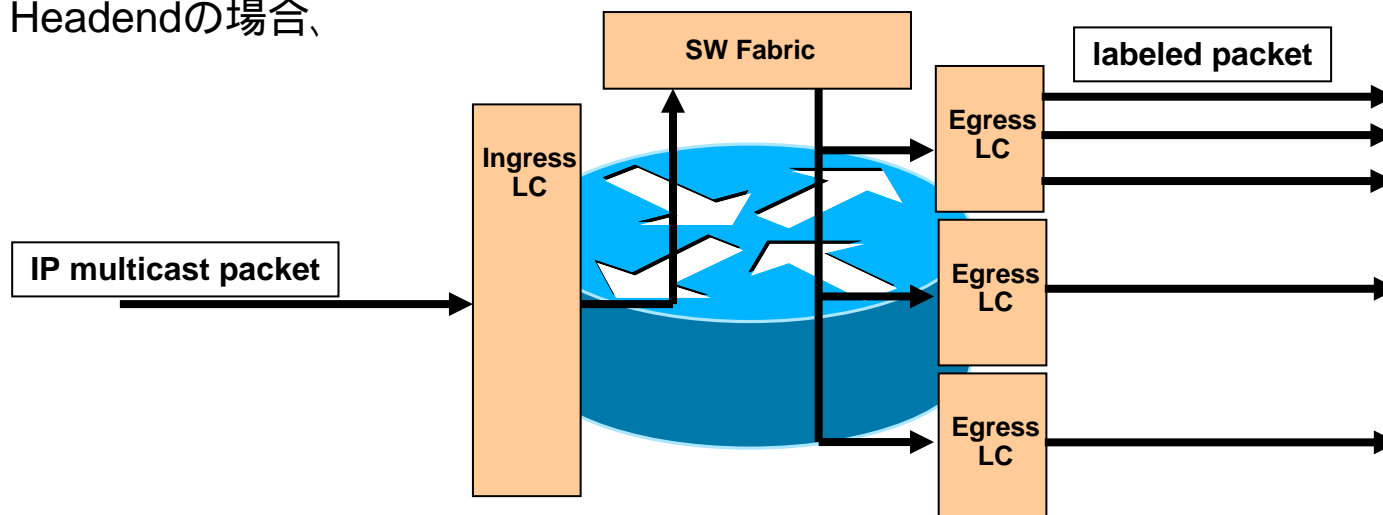
Explicit Null Labelは、「UH-PH」間でのみ使われるものであり、「Ingress-Egress」間のようなmultihopで使用することは、本来のlabel encoding standardと異なり、protocolを識別するために安易に使うべきではない。

といっても、static mappingのような場合に、additional labelで判断するとなるとexplicit nullが便利(out band signalingがいらない)。あと、他社とのinteroperabilityのためにもサポートが必須。

サポート決定。

Forwarding Plane (1)

- Headendの場合、



- 誰がpacket replicationを行うか？

- Ingress LC?

- Fabricにて処理するpacketが増加する Fabricの性能Down...

- FabricおよびEgress LC

- FabricにてLC単位のpacket replicationを行う。

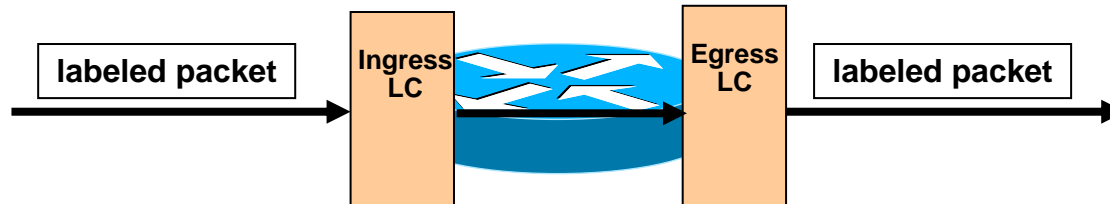
Ingress LCはEgress LC setを示す特殊なIDと共にpacketをFabricに転送

Fabricでは、IDに従い指定されたEgress LCに対しpacketを複製し、転送

- Egress LCにて、interface単位(LC内の)のpacket replicationを行う。

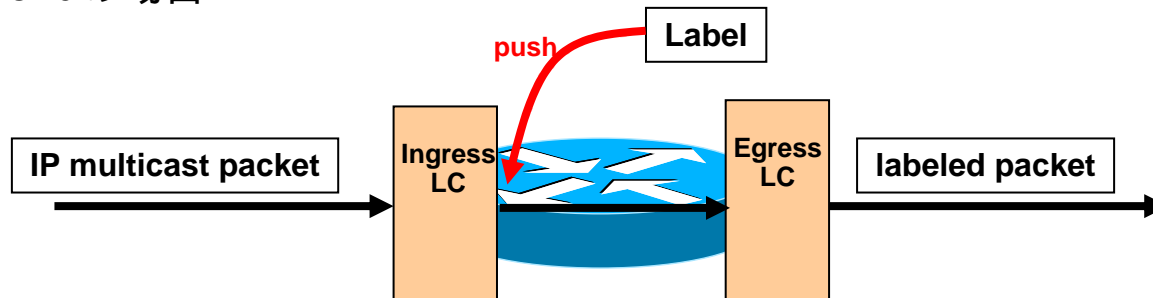
Forwarding Plane (2)

- Mid/Leaf nodeの場合



- Egress LCでは、FabricからLabeled Packetを受け、Labeled Packetを転送 (Egress LC)
Label forwarding IP forwarding(必要であれば)

- Headendの場合



- このままだと、Egress LCでは、FabricからIP Multicast packetを受け、Labeled Packetを転送
- Egress LCの動作が異なる。
(Egress LC)
~~IP forwarding~~ Label forwarding Label forwarding IP forwarding(必要であれば)

Ingress LC にて label(local label) pushした後、labeled packetとしてFabricに転送
Egress LCでは、Fabricからlabeled packetを受け取るので、Mid/Leaf nodeの場合と同じforwarding scheme

ご清聴ありがとうございました。

