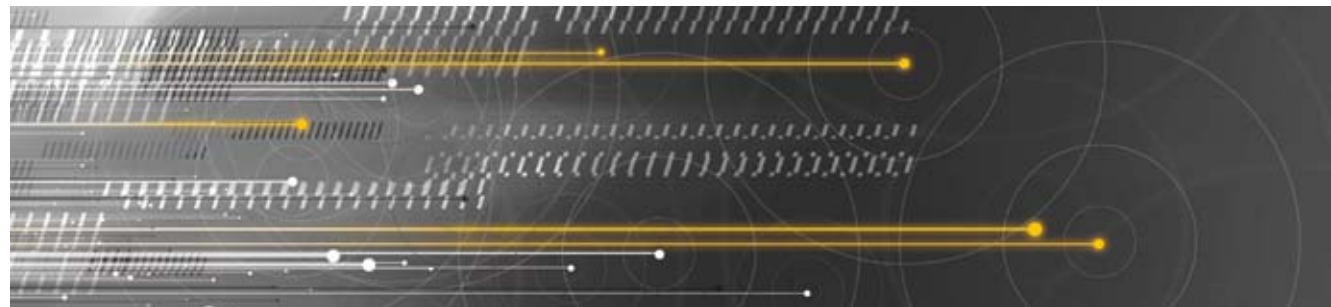


Multi-Segment Pseudo-Wire の 運用上の課題



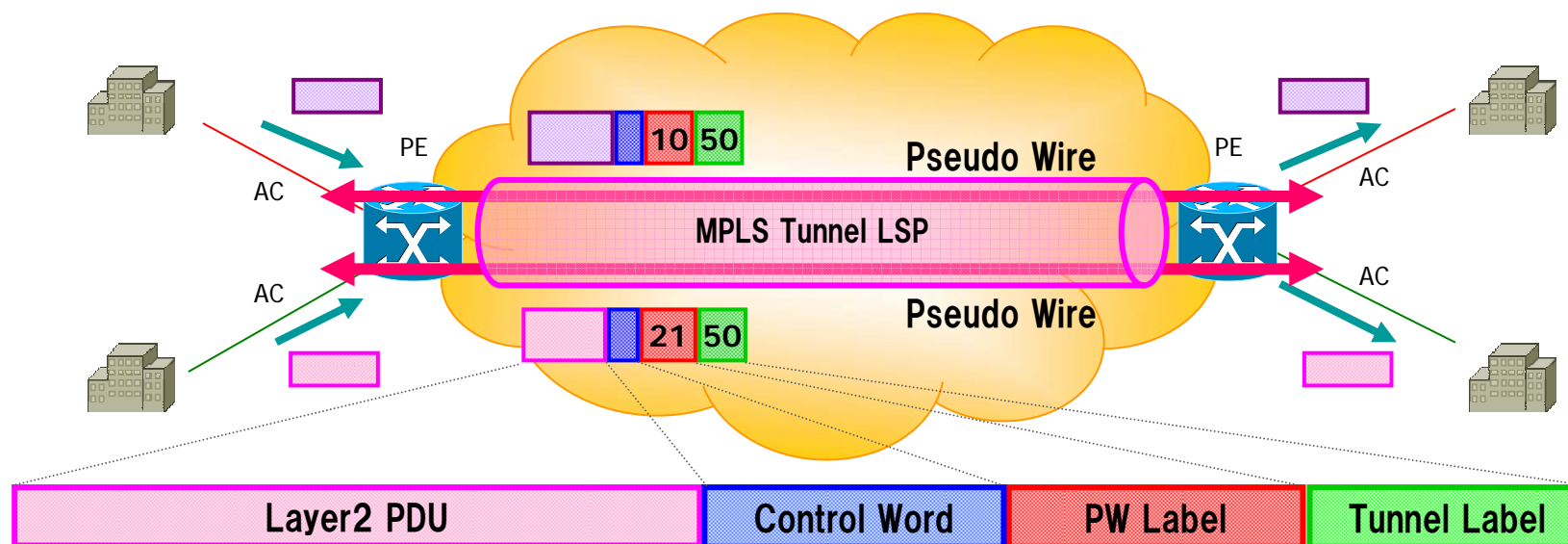
ソフトバンクテレコム株式会社
大矢 晃之
teruyuki.oya@tm.softbank.co.jp
MPLS JAPAN 2008 — Oct 28, 2008

For Your Networking Universe

Copyrights Reserved ©SOFTBANK TELECOM

1. Pseudo Wireとは
2. Pseudo Wire導入実績
3. Multi Segment Pseudo Wire
4. MS-PWの課題①OAM機能の充実
5. MS-PWの課題②信頼性・高可用性の確保
6. MS-PWの課題③運用負荷の増加

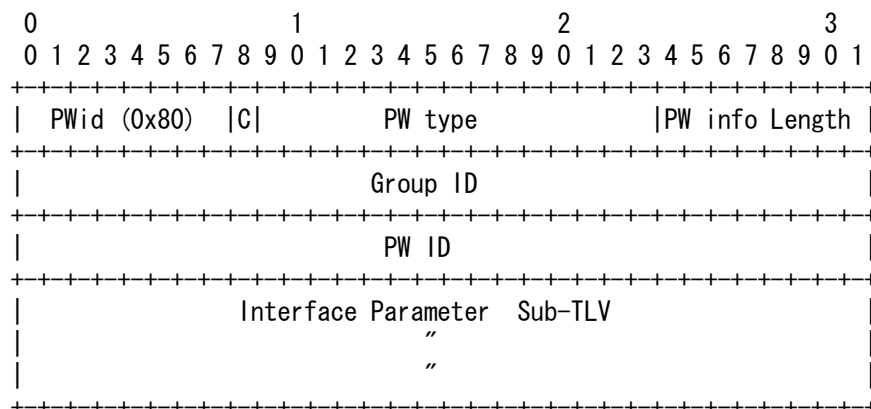
- ・ 専用線/ATM/FR/EtherなどのL2 PDUをMPLSヘッダでカプセル化し、IP/MPLS網を透過する技術
- ・ RFC3985 :Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture などで標準化



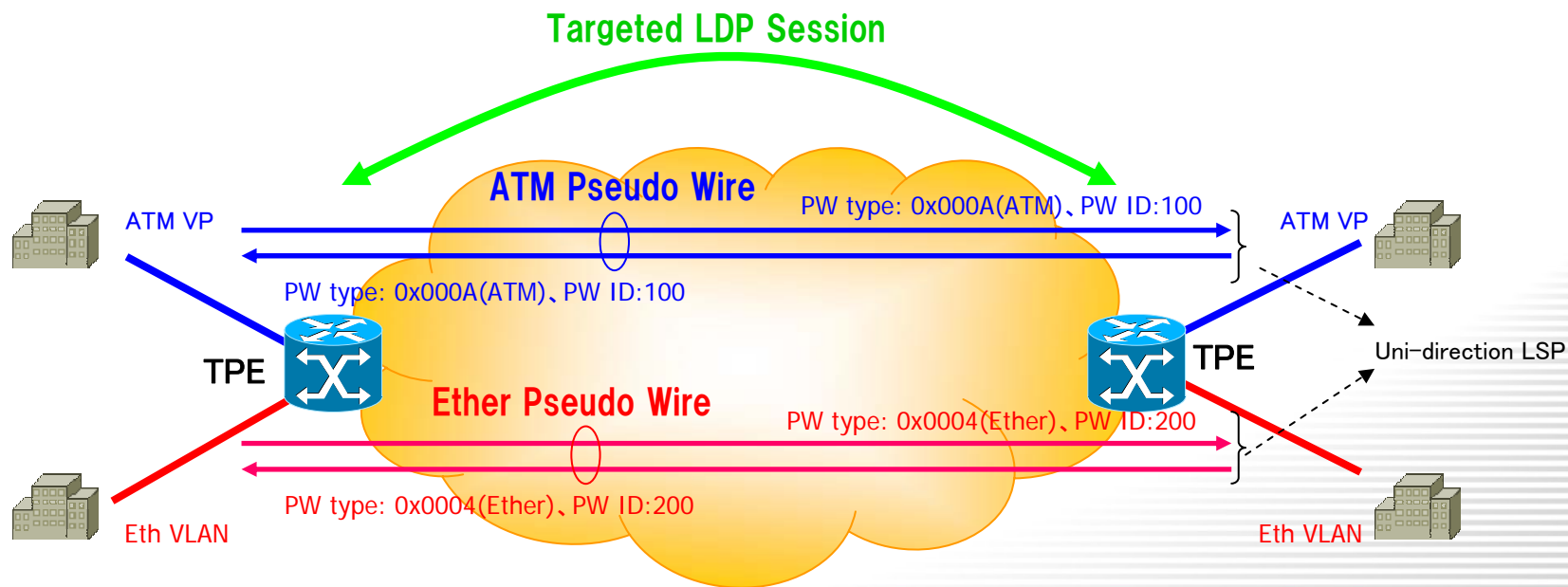
AC: Attachment Circuit

PWを終端するPhysical portや、FR DLCI, ATM VPI/VCI, Ethernet VLANなどのEndpoint

PWid FEC TLV (type 0x80)



PseudoWireの両端のTPEは Targeted LDP Sessionを張り、PW Label をMappingする。
 両端で同じPW Type、同じPW IDが必要。
 これをキーに、往復の片方向LSP2本を1本のPseudo Wireとして束ねる。



RFC3985

Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture

RFC4447

Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)

RFC4446

IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)

RFC5085

Pseudowire Virtual Circuit Connectivity Verification (VCCV):A Control Channel for Pseudowires

RFC4448

Encapsulation Methods for Transport of Ethernet over MPLS Networks

RFC4553

Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)

RFC4618

Encapsulation Methods for Transport of PPP/High-Level Data Link Control (HDLC) over MPLS Networks

RFC4619

Encapsulation Methods for Transport of Frame Relay over Multiprotocol Label Switching (MPLS) Networks

RFC4717

Encapsulation Methods for Transport of Asynchronous Transfer Mode (ATM) over MPLS Networks

RFC4842

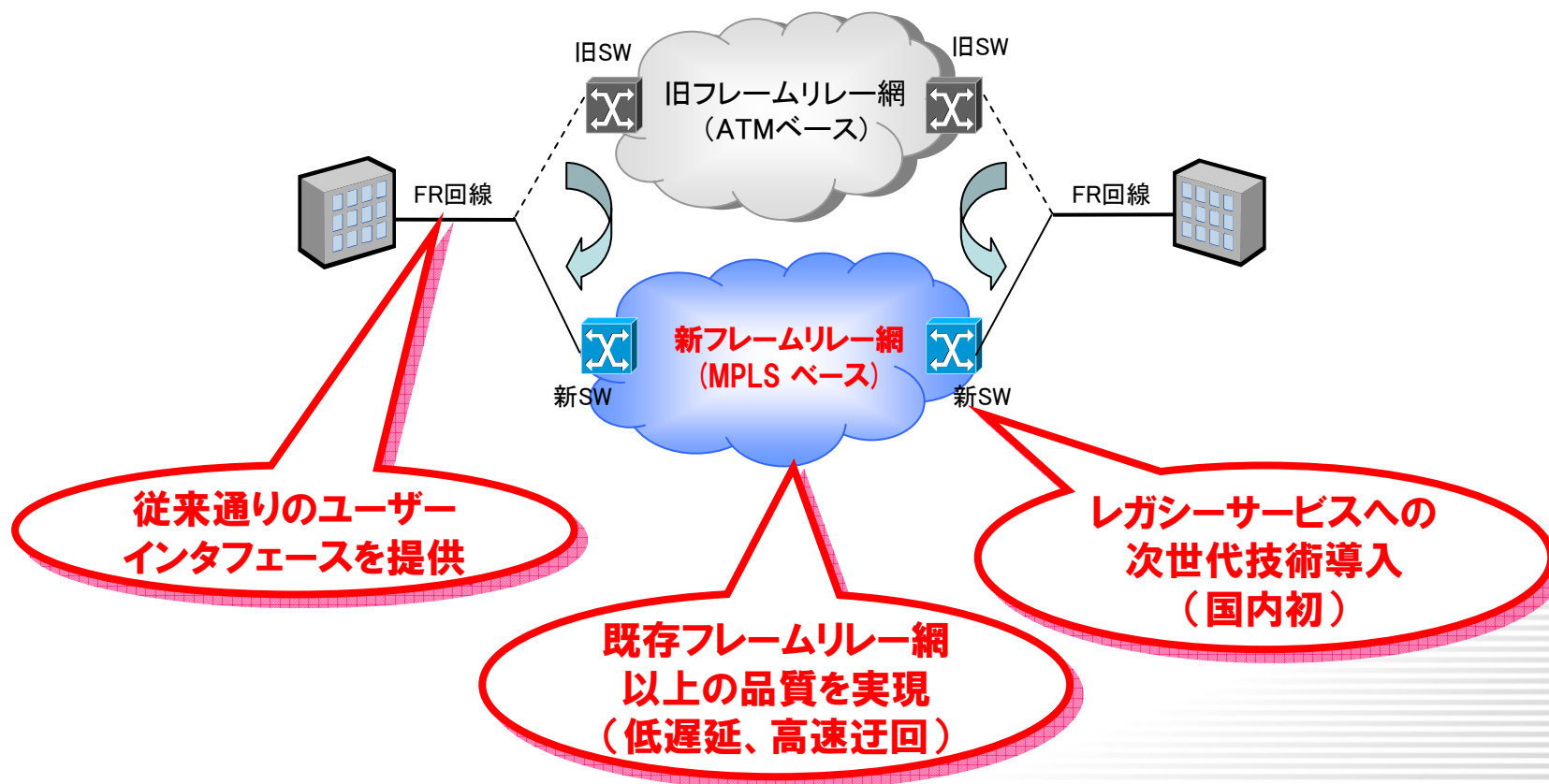
Synchronous Optical Network/Synchronous Digital Hierarchy (SONET/SDH) Circuit Emulation over Packet (CEP)

RFC5086

Structure-Aware Time Division Multiplexed (TDM) Circuit Emulation Service over Packet Switched Network (CESoPSN)

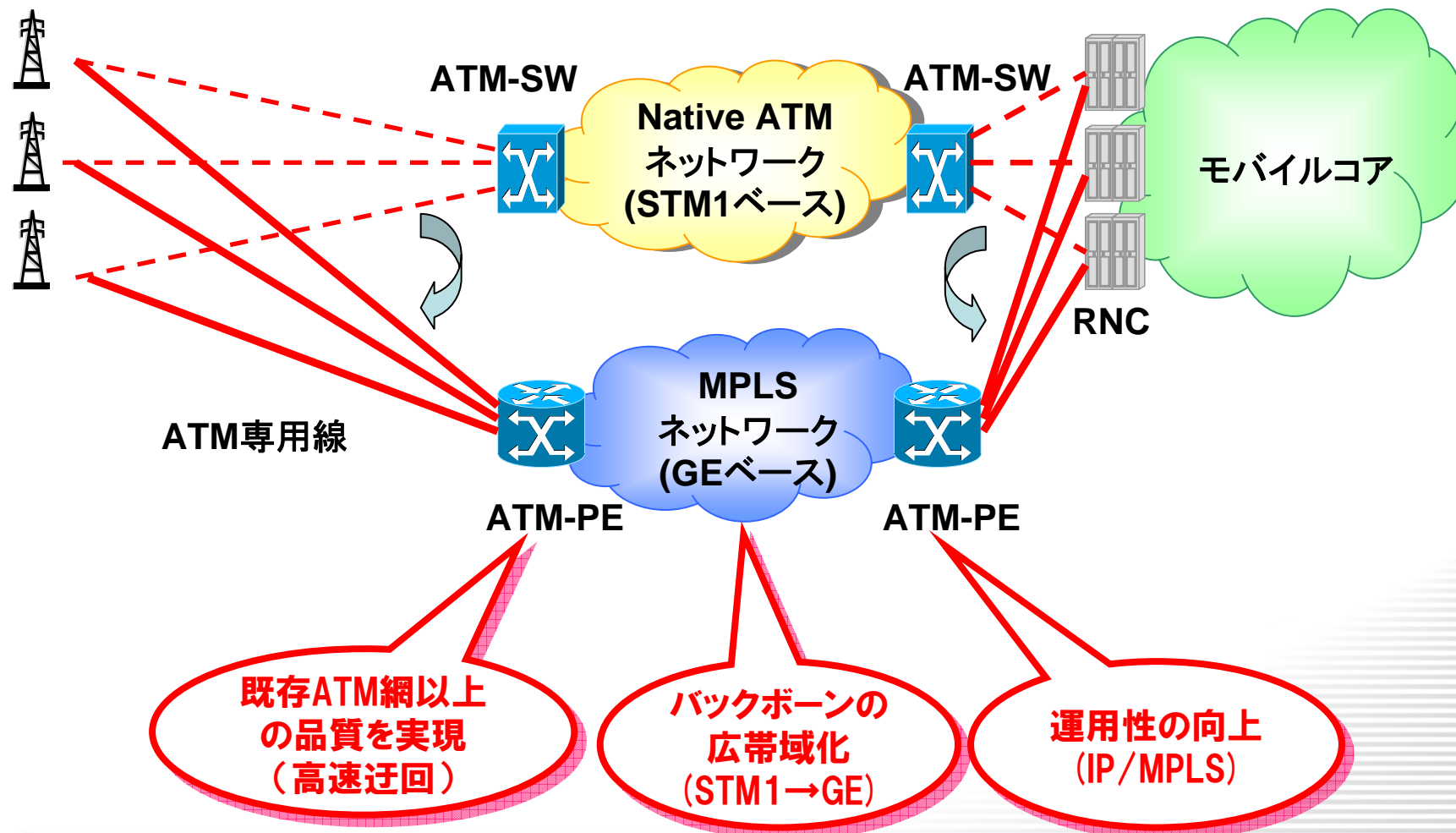
FRネットワークのMPLS化

設備老朽化に伴うリスク回避、及びお客様の投資保護を目的として、次世代ネットワークによるレガシーサービスの継続的な提供を実現。

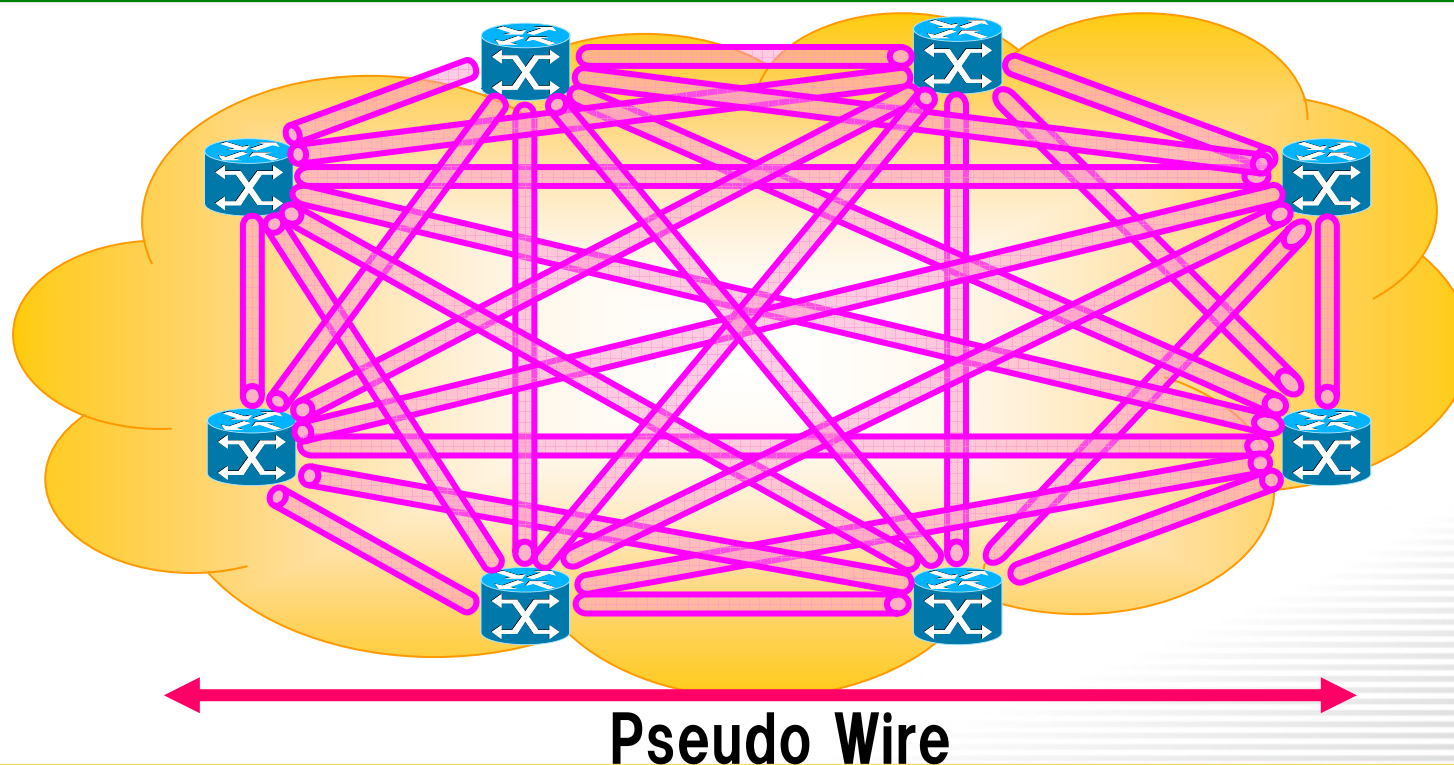


ATMネットワークのMPLS化

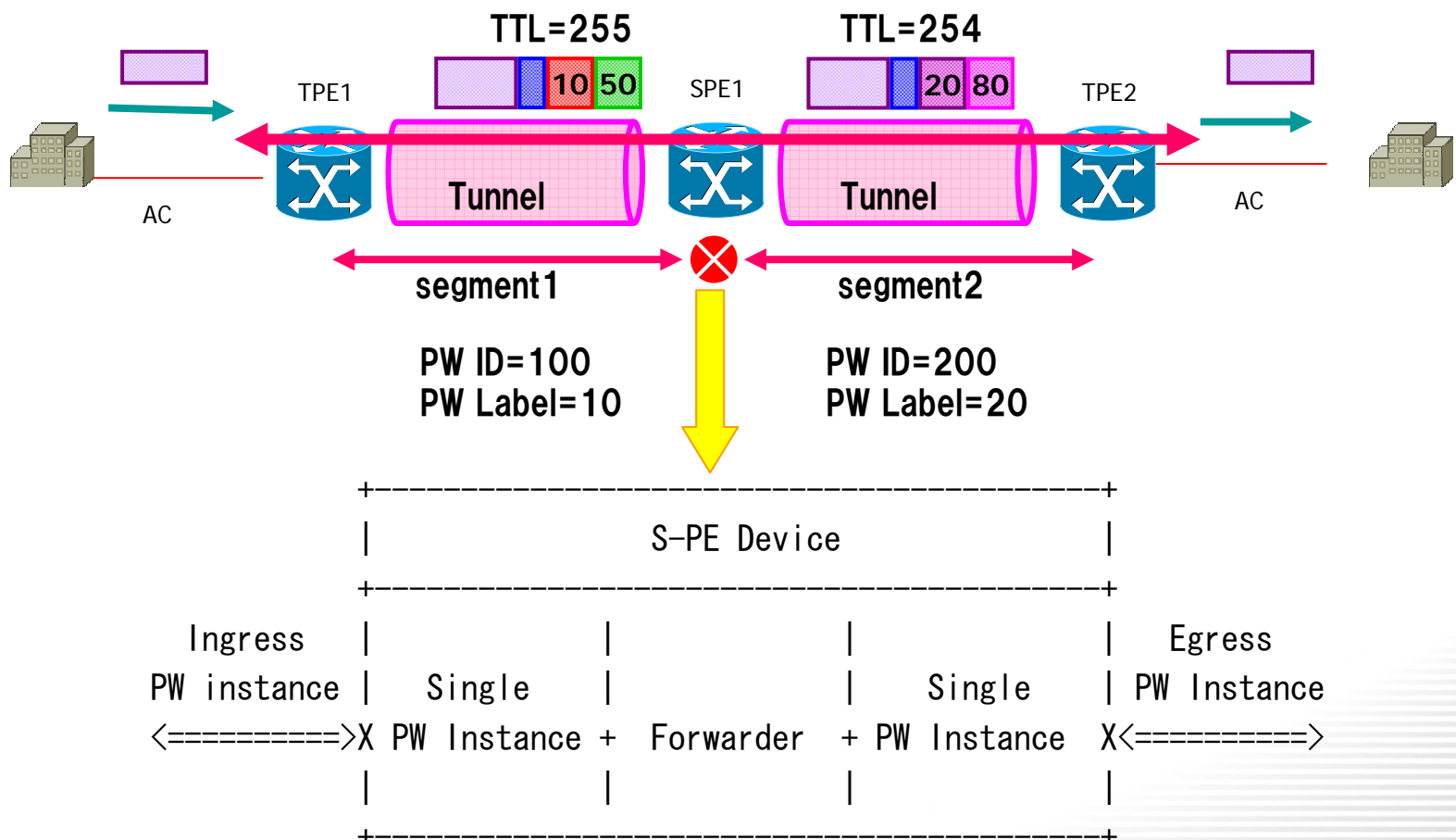
増え続けるモバイルトラフィックへの対応と信頼性・運用性向上を目的として、ATMネットワークのATM-PWによるMigrationを推進。



- Scalability Issue
 - >> 数100～数1000node規模のNW
 - >> PE-to-PE MPLS Tunnelが数万～数百万本 (Edge間フルメッシュ)
 - >> 同一Link内LSPが増え、RSVP-TEで帯域が足りなくなり、LSPが張れない
- 実際のキャリアネットワーク = Multi Domainの階層構造
 - >> Inter Area/ Inter AS TEは難しい



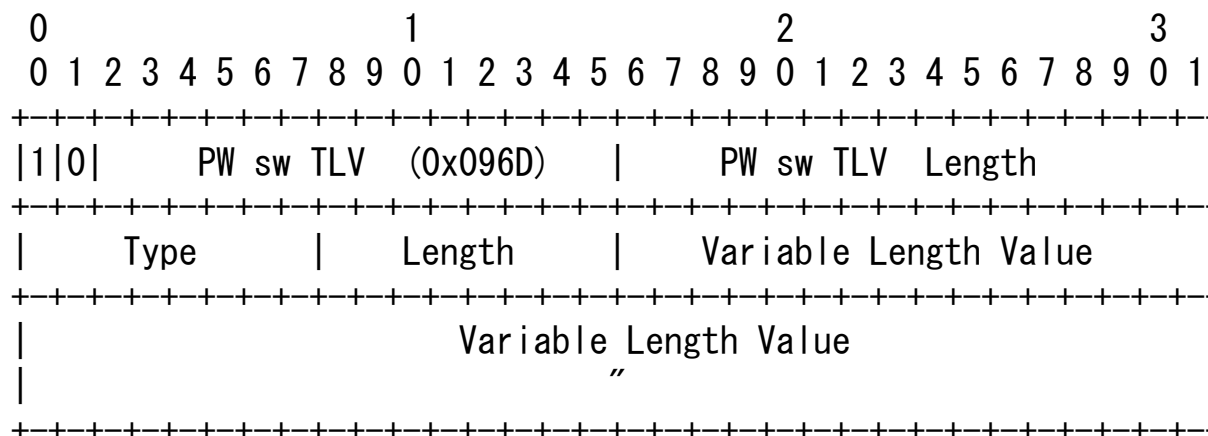
SPEで2つのSS-PWをクロスコネクトし、つなぎ合わせる。
 SPEは何段あってもよい。
 SPEを経由するごとにTTLをdecrementする。(十分大きいTTLで送信:例TTL=255)



draft-ietf-pwe3-ms-pw-arch-05.txt

draft-ietf-pwe3-segmented-pw-09.txt

PW switching point TLV (Optional)が新たに定義されている。
SPEでFEC TLVに追加される。経由したSPEの情報がわかるようになる。

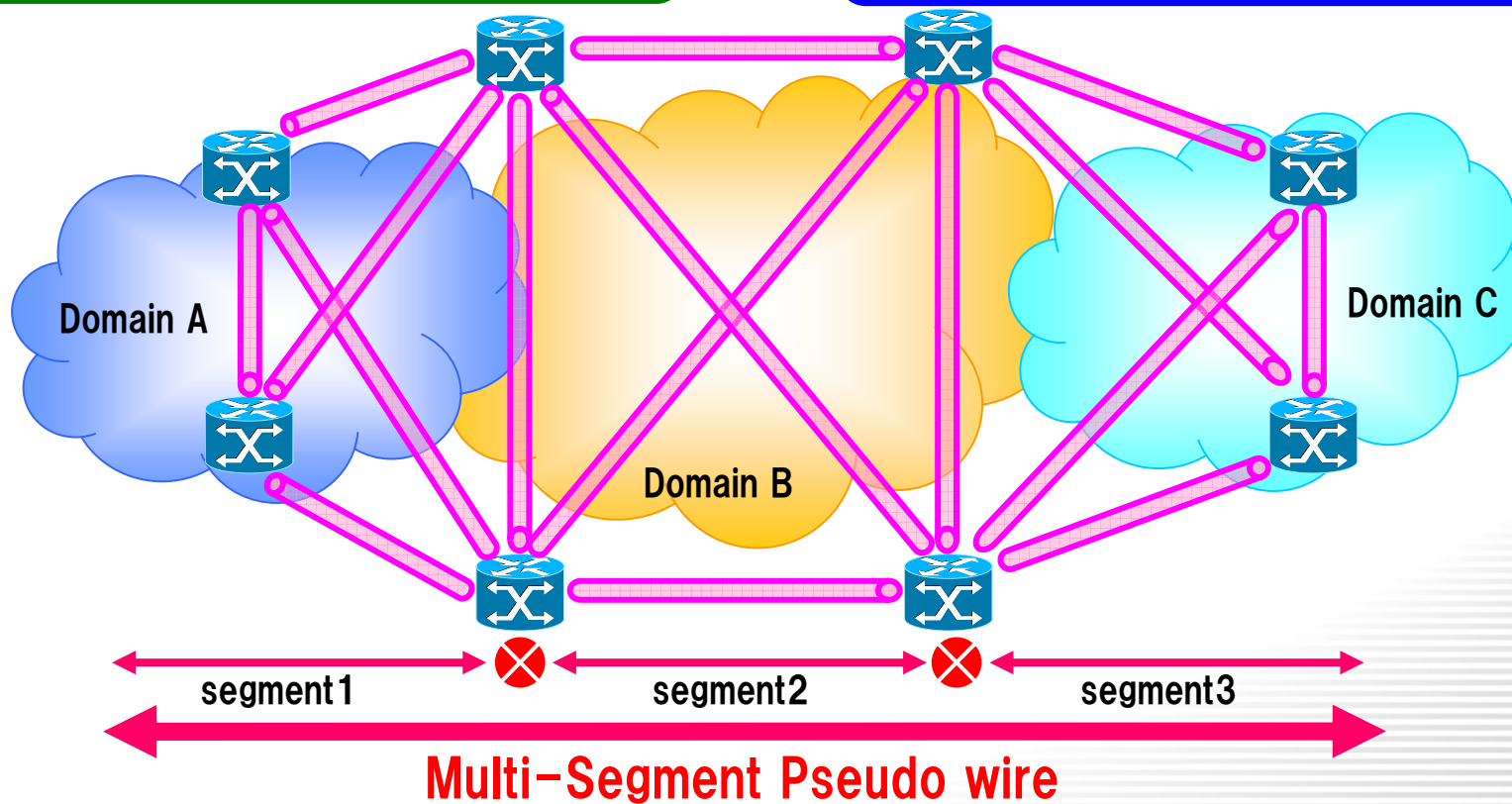


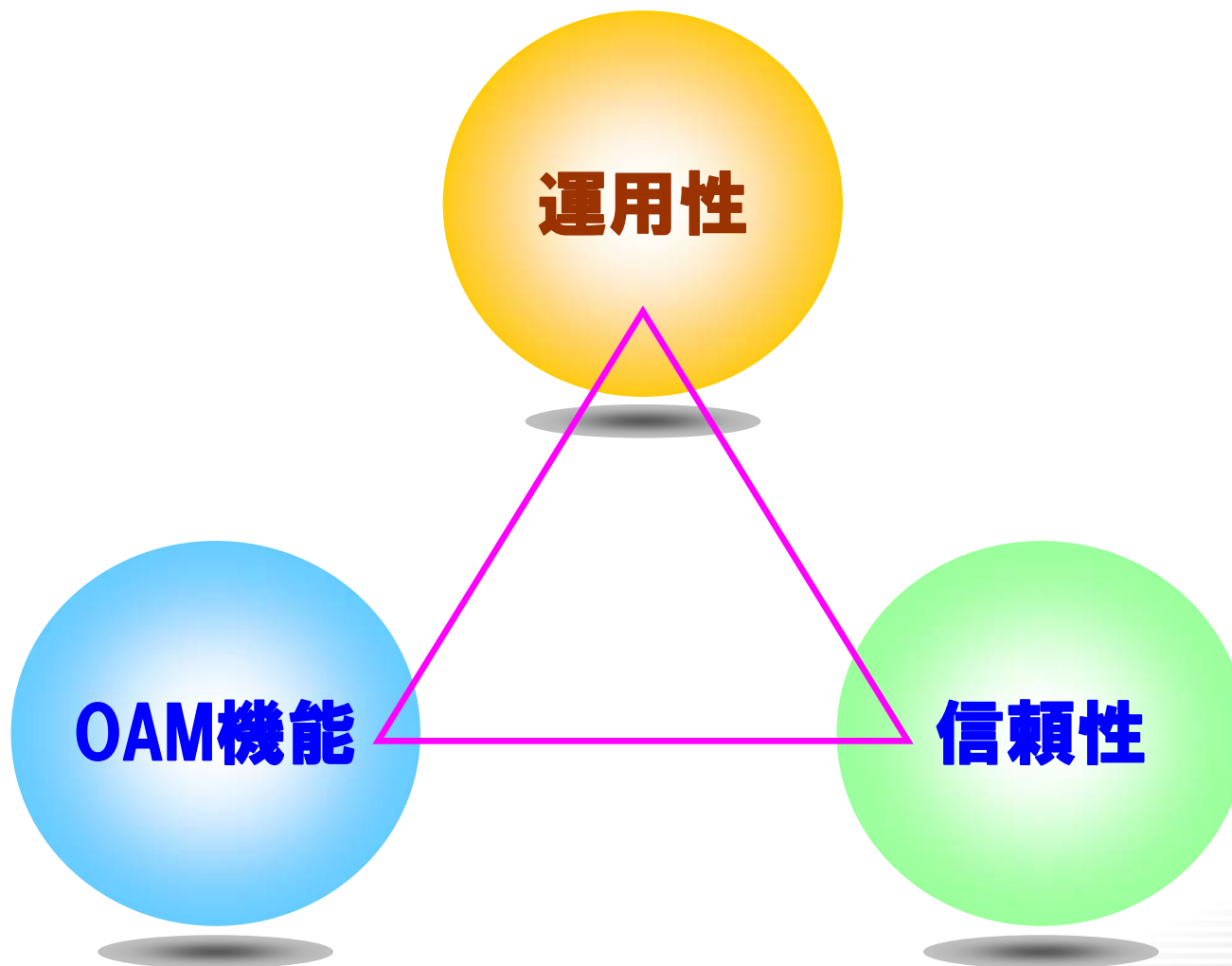
Type	Length	Description
0x01	4	PW ID of last PW segment traversed
0x02	variable	PW Switching Point description string
0x03	4/16	Local IP address of PW Switching Point
0x04	4/16	Remote IP address of last PW Switching Point traversed or of the T-PE
0x05	variable	AI of last PW segment traversed
0x06	10	L2 PW address of PW Switching Point

Multi-Segment Pseudo wire



- ・スケーラビリティ問題の解消
- ・実稼働中Multi Domain NWへの適用



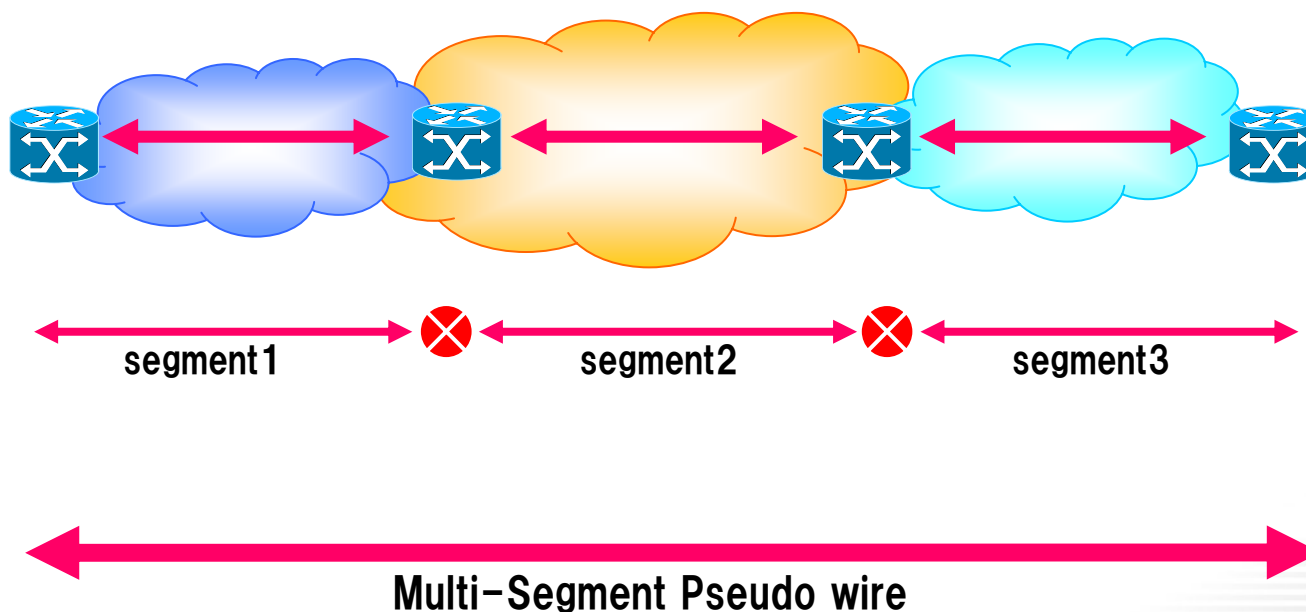


課題①

- ・ OAM機能の充実
- > 回線単位でEnd-to-Endの試験機能
- > 警報転送



- ・ Pseudowire Ping/Tracerouteの拡張 (VCCV:RFC5085)
- ・ PW Status TLVの実装
- ・ PW OAM Message Mapping の拡張



課題①

・ OAM機能の充実

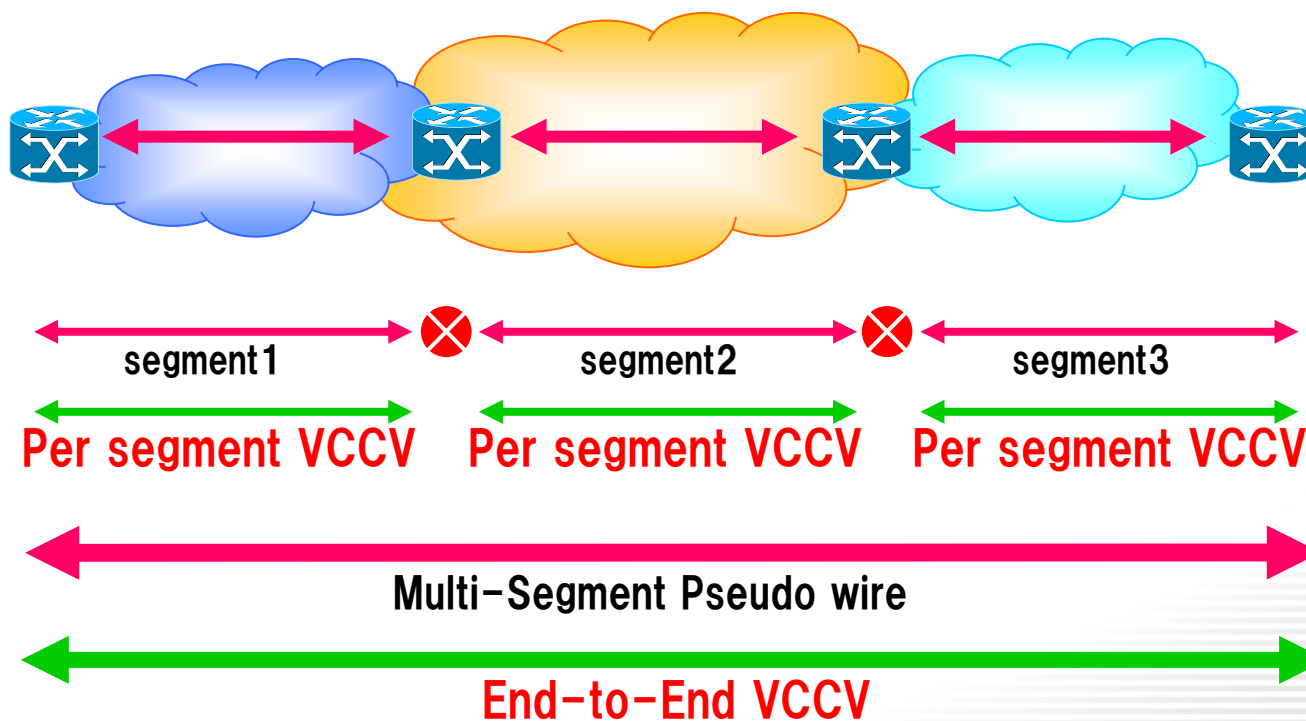
> 回線単位でEnd-to-Endの試験機能

> 警報転送

・Pseudowire Ping/Tracerouteの拡張
(VCCV:RFC5085)

・PW Status TLVの実装

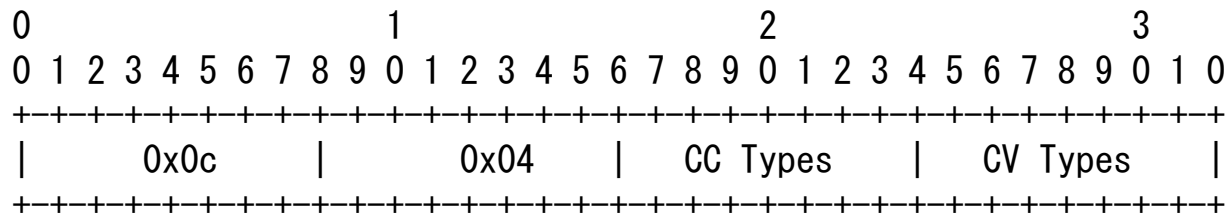
・PW OAM Message Mapping の拡張



The VCCV parameter ID is defined as follows in [RFC4446]:

Parameter ID	Length	Description
0x0c	4	VCCV

The format of the VCCV parameter field is as follows:



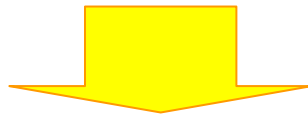
MPLS Control Channel (CC) Types:

Bit (Value)	Description
0x01	Type 1: PWE3 control word with 0001b as first nibble as defined in [RFC4385].
0x02	Type 2: MPLS Router Alert Label.
0x04	Type 3: MPLS PW De-multiplexor Label TTL = 1 (Type 3).

MPLS Connectivity Verification (CV) Types:

Bit (Value)	Description
Bit 0 (0x01)	- ICMP Ping
Bit 1 (0x02)	- LSP Ping

CC Type1: VCCV via Control ChannelがMS-PWでも使えれば、Data Planeの正常性も確認できてよい、が、MS-PWを構成するTPE、SPE全てがControl Wordをサポートする必要がある。



Control WordはPWの種類によってはMandatoryではないため、既存の機器ではサポートされていないことが多い。

PW encapsulation	Control Word
FR	Required
ATM (N:1 cell mode)	Optional
Ethernet	Must (初期のインプリはCWなし)

SS-PWでは最も一般的にサポートされているCC Type2は
MS-PWではSPEを越えられないので使えない。

CC Type2は、Negotiationの段階で、SPEでCapabilityがremoveされる。

draft-ietf-pwe3-segmented-pw-09.txt

9.4.2. MS-PW and VCCV CC type 2

VCCV CC type 2 is not supported for MS-PWs and MUST be removed form a VCCV parameter field by the S-PE.

Control Wordをサポートしていない場合の手段として、
CC Type3でTTLを対向TPEまでのSPE hop数にsetする拡張が検討されている。
(SS-PWではCC type3ではTTL=1)

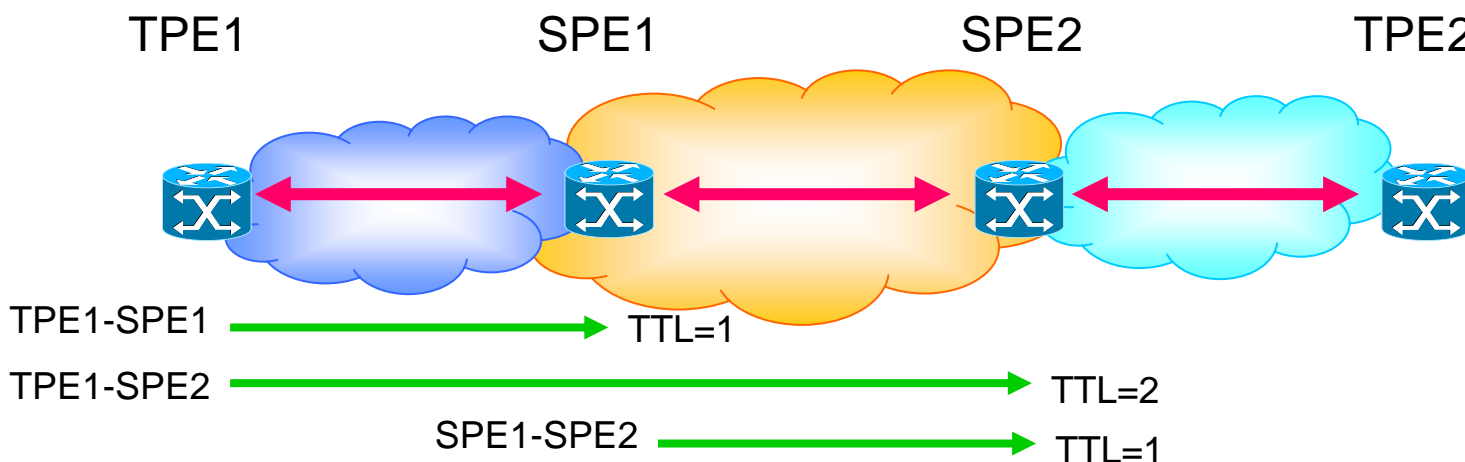


どうやってTPEまでのhop数を認識するか。
PW switching point TLVをInspectする方法があるが、
PW switching point TLV は完全にoptional

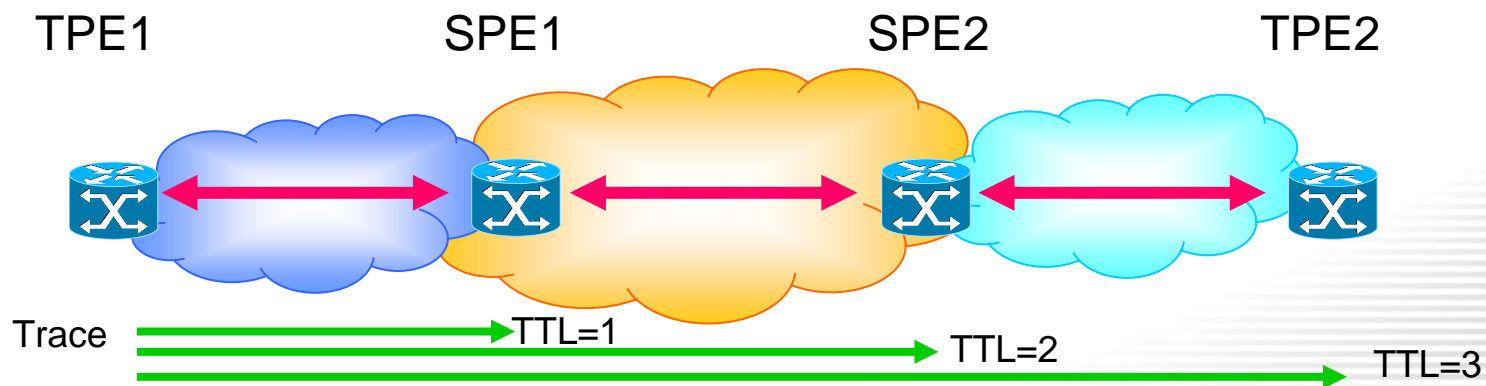
draft-ietf-pwe3-segmented-pw-09.txt

9.4.3. MS-PW and VCCV CC type 3

VCCV CC type 3 can be used for MS-PWs, however if the CW is enabled VCCV type 1 is preferred according to the rules in [RFC5085]. Note that for using the VCCV type 3, TTL method, the PE will set the PW label TTL to the appropriate value necessary to reach the target PE, otherwise the VCCV packet might be forwarded over the AC to the CPE.

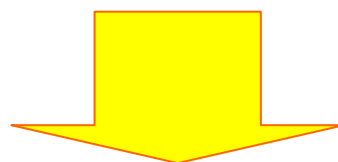


TPE to SPEや、SPE to SPEのVCCV Pingは、TTLを1や2などにsetし、TargetのSPEでTTL expireさせることで実現する。



Traceも同様。

CC type	課題
1 : Control Channel	Control Wordサポートしない機器がある
2 : Router Alert Label	MS-PWでは使用不可
3 : TTL method	適切なTTL値を決める方法が複雑

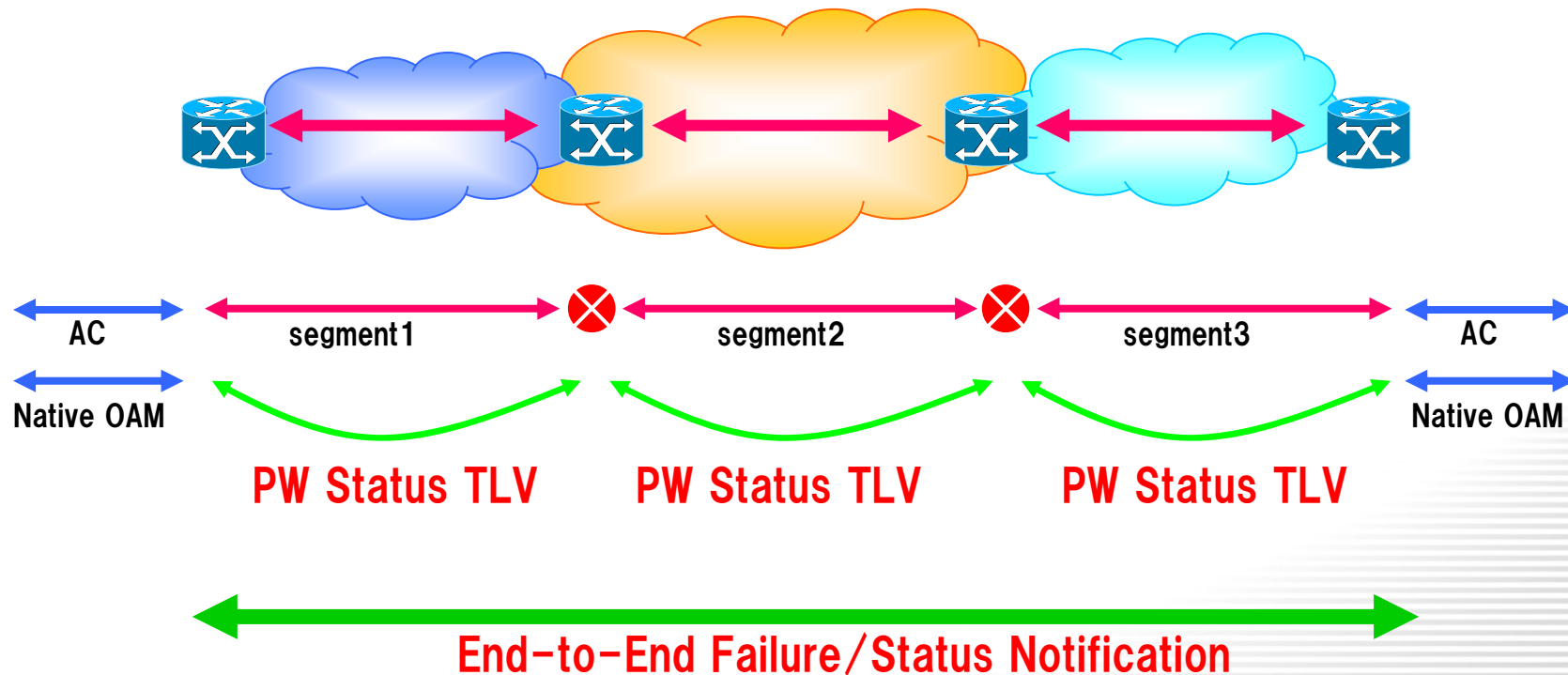


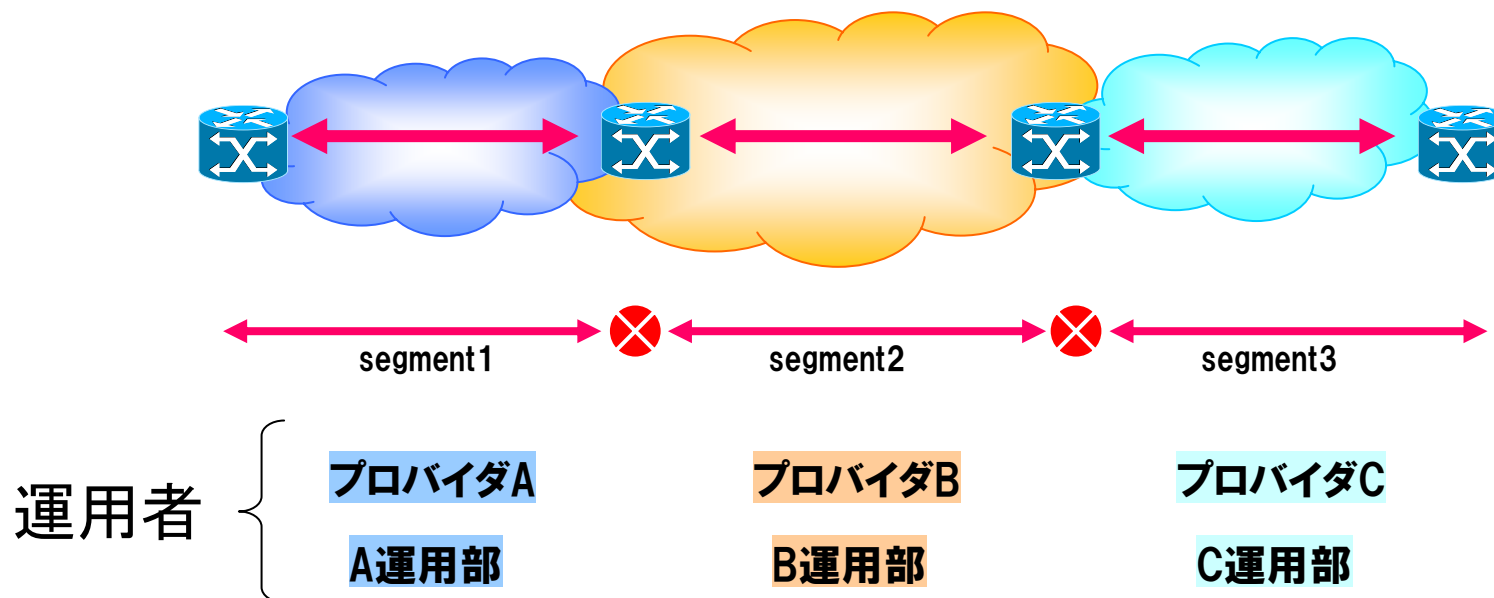
まだフルサポートしているベンダーがない。
標準化の推進とともに、CC Type1/3の実装を急ぐ必要あり

課題①

- ・ OAM機能の充実
 - > 回線単位でEnd-to-Endの試験機能
 - > 警報転送

- ・Pseudowire Ping/Tracerouteの拡張 (VCCV:RFC5085)
- ・PW Status TLVの実装
- ・PW OAM Message Mapping の拡張





MS-PWを監視・運用していくには、
各segmentでのPW/ACの状態・警報を、
共通の方法で伝達し、識別する必要あり



PW Status TLVのサポートは必須

しかし、いまだにLabel Withdrawを使ったり、
Native OAMをData Planeに挿入することで伝達する装置もある。

標準化では、PW StatusがDefaultとされている。

draft-ietf-pwe3-oam-msg-map-07.txt

Overview of fault notifications

For a MPLS PSN and a IP PSN using MPLS-in-IP [RFC4023], a PW that are established and maintained using LDP **SHOULD use LDP status signaling messages as the default mechanism for AC, PW status and defect notification [RFC4447].**

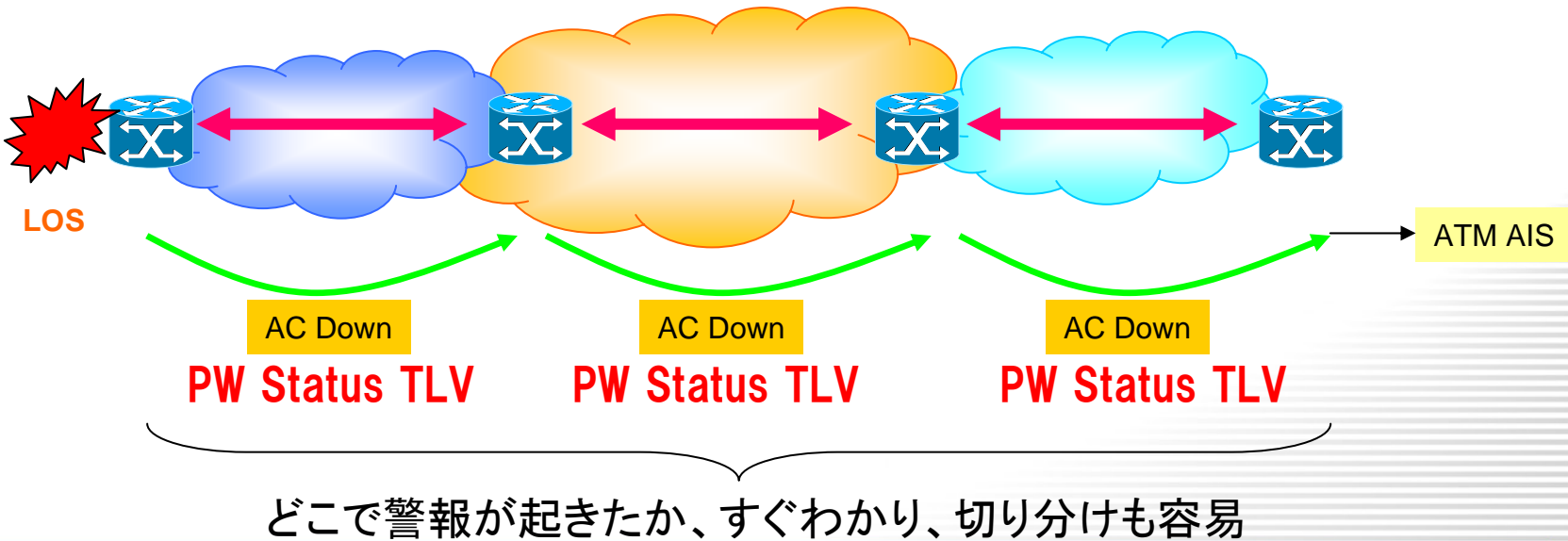
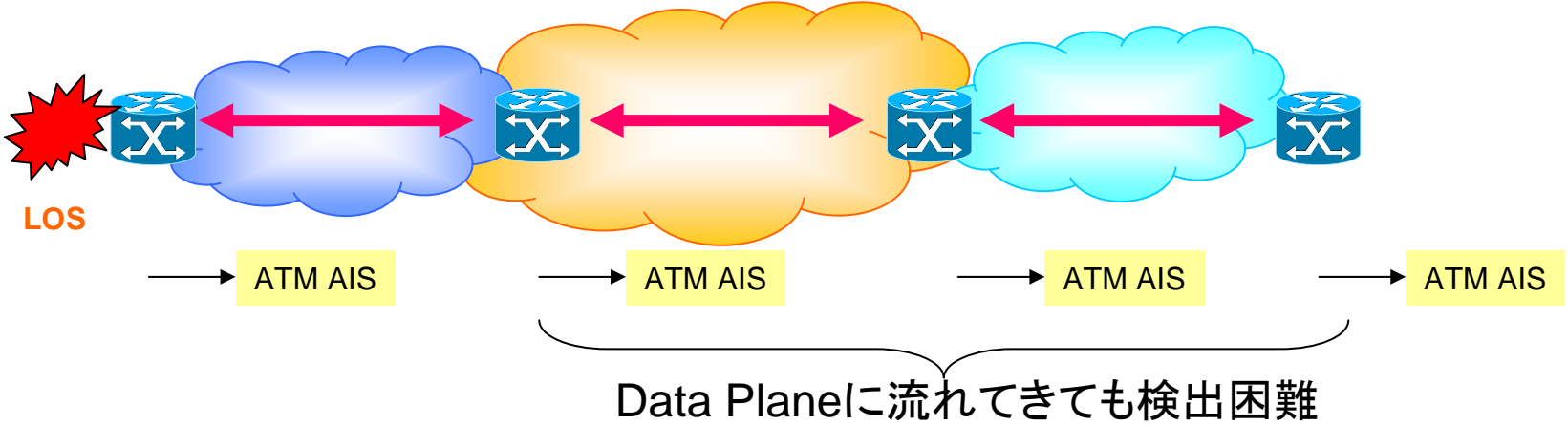
しかし、このDraft、かなり複雑で、多方面で混乱を招いているのが現状。

例として、ATM-PWでは、PW Statusはoptionalで、Native ATM OAM cellをPW Data Plane上に流すのがDefaultとされている。

Use of Native Service notifications

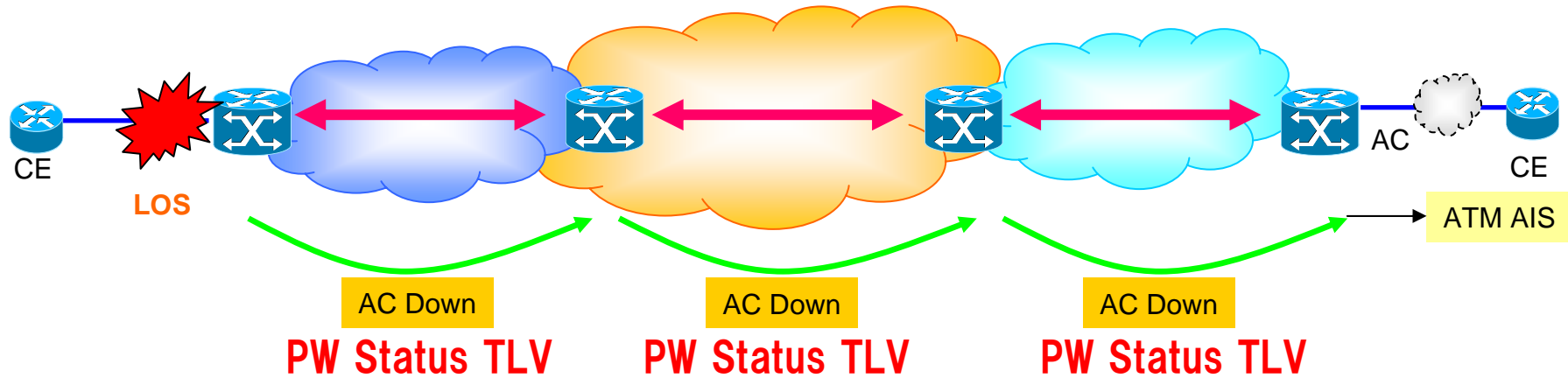
ATM PWs may optionally also use PW specific notification mechanisms. in-band ATM OAM over PW and is the default method.

InbandでData Planeに送ってこられても・・・

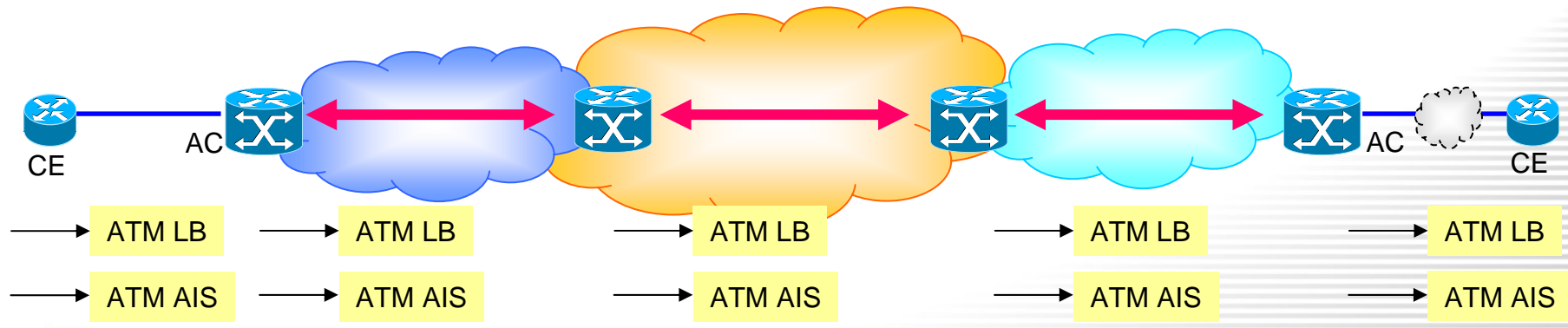


PW OAM Message Mappingのあるべき姿

基本はすべてのPW typeで、PW Status に統一。
そうしないとPW Redundancyもできなくなる(後述)



CE-CE間のNative OAM等、ACのData Planeから入ってくるOAMは
PWのData Plane上でTransparentに転送する。



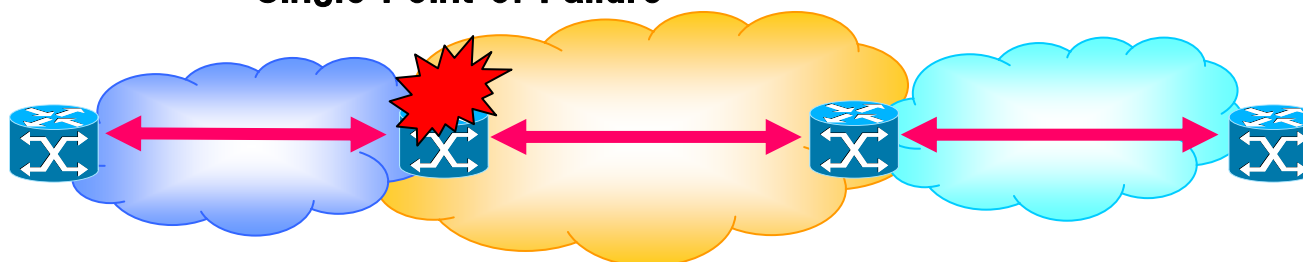
課題②

- ・信頼性・高可用性の確保
- > Single Point of Failureの解消

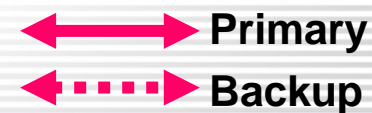


Pseudowire Redundancy

Single Point of Failure



PW Redundancy with MS-PW



PW Redundancy: SS-PW

draft-ietf-pwe3-redundancy-bit-01.txt (Preferential Forwarding Status bit definition)
 draft-ietf-pwe3-redundancy-01.txt (Pseudowire (PW) Redundancy)
 でいくつかのmodelをベースに定義されている。

Case1:

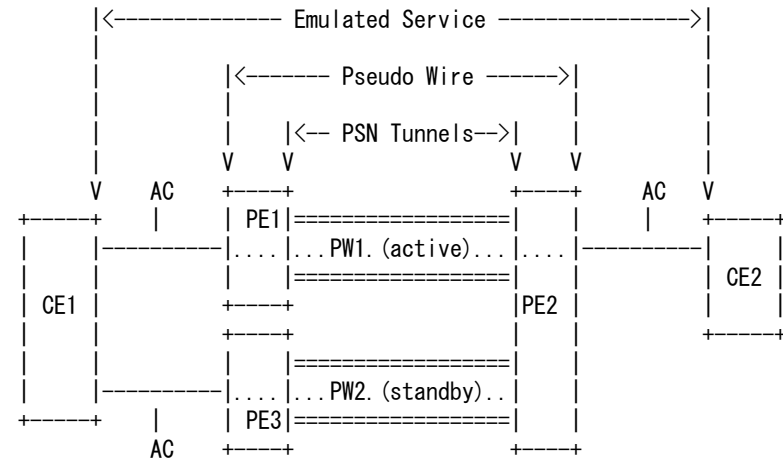
One Multi-homed CE with single SS-PW redundancy

標準的なreference model。

PE1/PE3をノード冗長化。

PE2から、PE1とPE3へRedundancy PWを設定。

PE1/PE3側はPE2向けに通常のPWを設定。



Case2:

Multiple Multi-homed CEs with single SS-PW redundancy

両端のPEを冗長化したmodel。

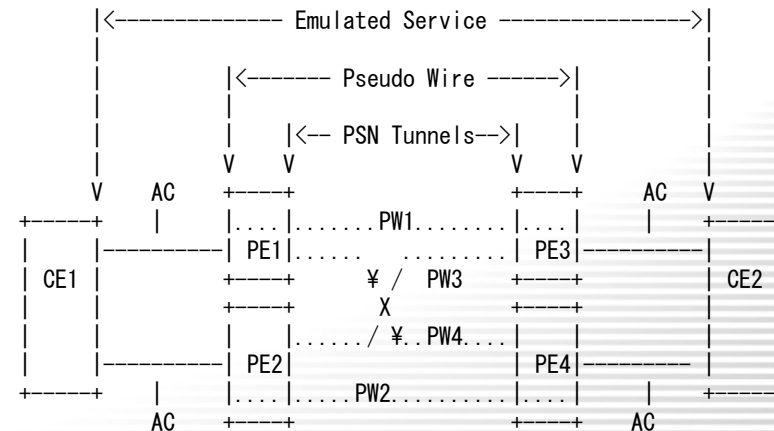
PE1から、PE3とPE4へ

PE2から、PE3とPE3へ

PE3から、PE1とPE2へ

PE4から、PE1とPE2へ

それぞれRedundancy PWを設定。

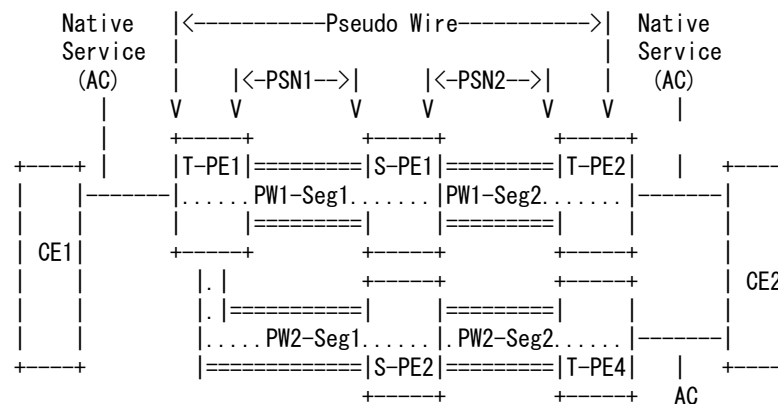


PW Redundancy: MS-PW

Case3:

Multi-homed CE with MS-PW redundancy

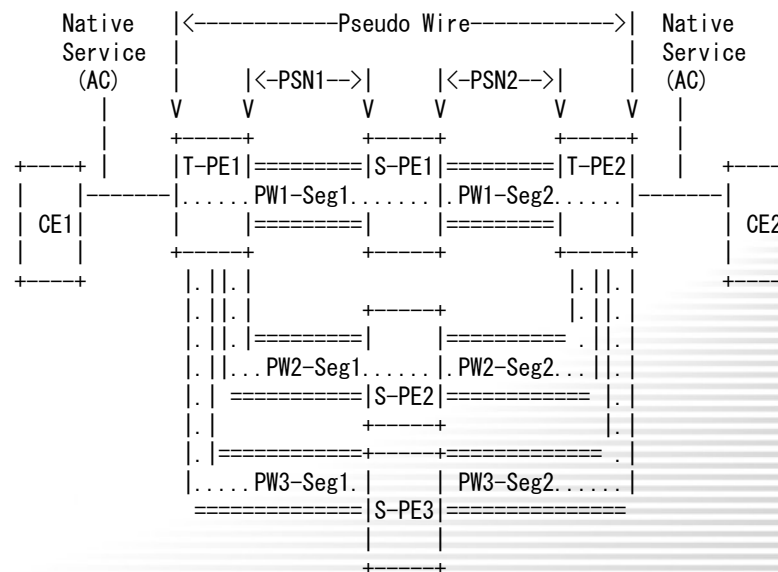
- ①の標準modelのMS-PW版。
- SPE1/SPE2とTPE2/PTE4をノード冗長化。
- TPE1から、SPE1とSPE2へRedundancy PWを設定。
- SPE1/SPE2から、TPE1、TPE2/TPE4へは通常のPWを設定。
- TPE2/TPE4から、SPE1/SPE1へ通常のPWを設定。



Case4:

Single Homed CE with MS-PW redundancy

- SPEを冗長化したmodel。
- TPE1から、SPE1/SPE2/SPE3へ
- TPE2から、SPE1/SPE2/SPE3へ
- それぞれRedundancy PWを設定。
- SPE1/SPE2/SPE3から、TPE1/TPE2へは通常のPWを設定。
- Primary PWは1つだけ、Secondary PWは複数設定可能。



0x00000000 - Pseudowire forwarding (clear all failures)

0x00000001 - Pseudowire Not Forwarding

0x00000002 - Local Attachment Circuit (ingress) Receive Fault

0x00000004 - Local Attachment Circuit (egress) Transmit Fault

0x00000008 - Local PSN-facing PW (ingress) Receive Fault

0x00000010 - Local PSN-facing PW (egress) Transmit Fault

0x00000020 - PW Preferential Forwarding Status

When the bit is set, it indicates "PW forwarding standby".

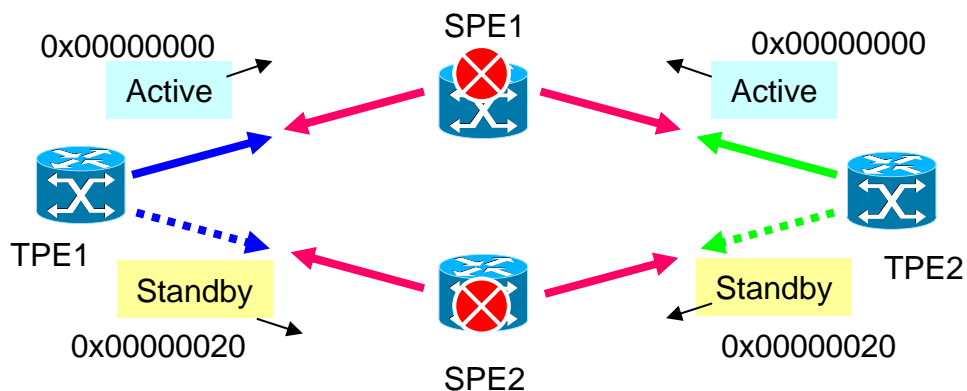
When the bit is cleared, it indicates "PW forwarding active".

0x00000040 - PW Request Switchover Status

When the bit is set, it represents "Request switchover to this PW".

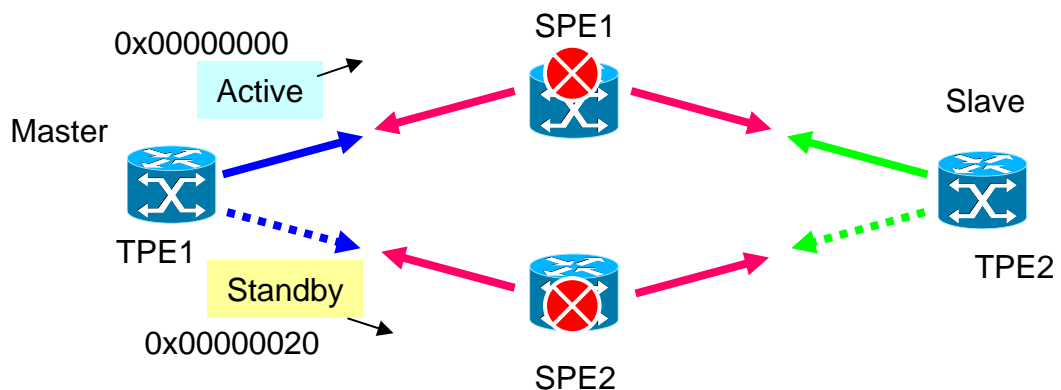
When the bit is cleared, it represents no specific action.

Independent Mode



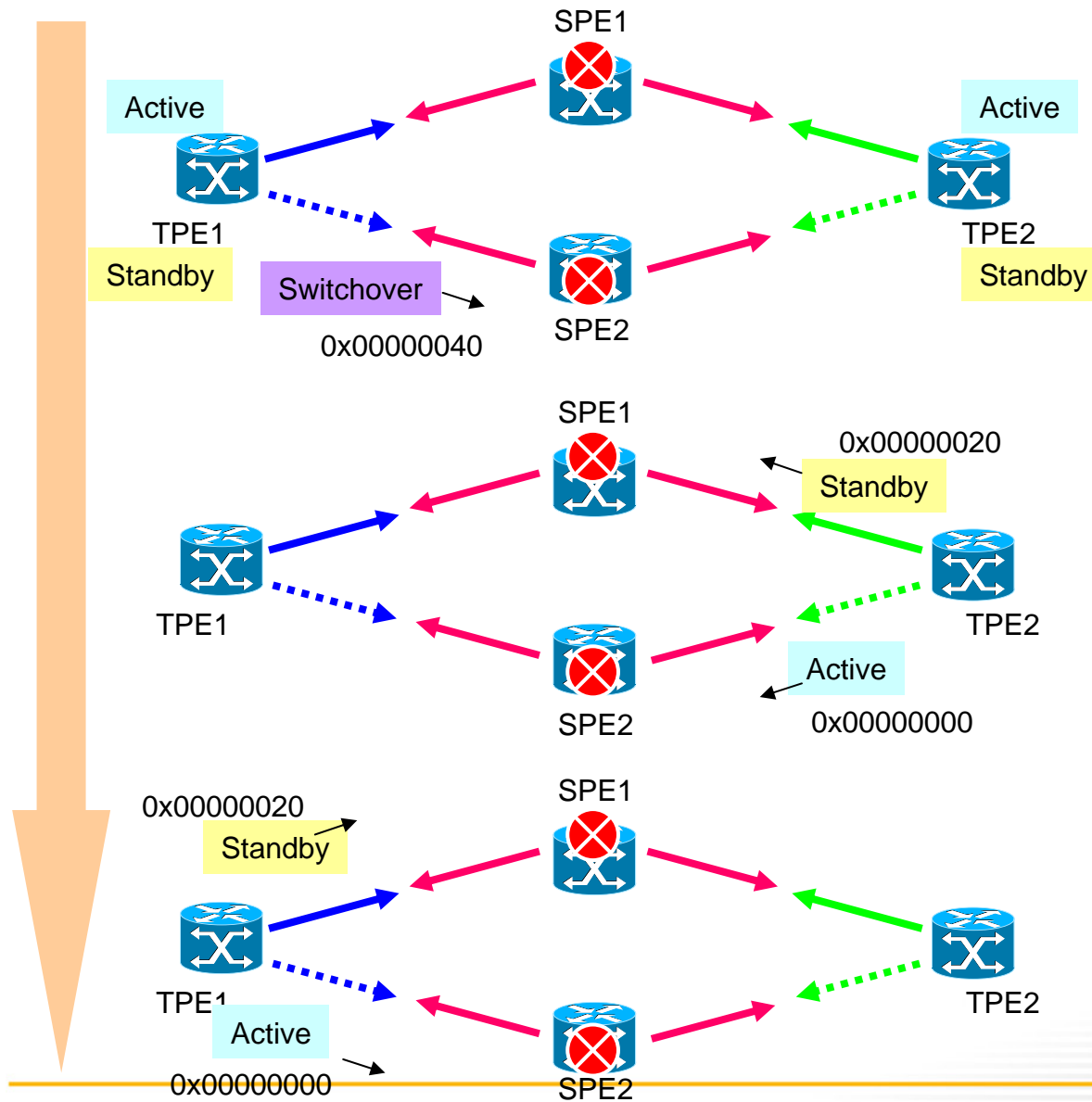
それぞれのTPEでLocalでPWごとのPrimary/Secondaryを設定。Primary PWにActive(0x00000000)、Secondary PWにStandby(0x00000020)を送信。送信(Local)と受信(Remote)の両方がActiveで一致するとPWはActive Stateになり、TrafficがForwardingされる。

Master/Slave Mode



Slave側TPEはMaster側から受信したPW Statusに従い、Activeを受信したPWへTrafficをForwardingする。

Request Switchover (Independent Mode)



ActivateしたいPWに
Request Switchoverを送信

Request Switchoverを受信したら、
受信したPWにActive、その他のPWに
Standbyを送信

Request Switchoverを送信したPWで
Activeを受信したらACKとみなして、
Active受信したPWにActive、
その他のPWにStandbyを送信

- ✓ まだ標準方式で実装しているベンダーがない
 - みんな独自方式で実装
- ✓ 障害によるSwitchoverのトリガが整理されていない
 - Not forwardingとか、AC Downとか、PW Downとか、withdrawとか、Tunnel Downとか、、、
 - VCCVで監視して、連続でVCCV落ちたらSwitchというのもあり？
- ✓ Mismatchを防止・解消する手段があまり考慮されていない
 - Local PrecedenceをSignalingでやりとりするのもあり？
- ✓ Flapping防止の規定がない
 - Signaling Delayとか、両系Downとか
- ✓ Standby PWでのTraffic受信の規定がない
 - Standby PWへのTraffic送信はMUST NOTになっているが、Traffic受信はoutside the scope
 - MismatchやSignaling Delayを考慮して、Standby PWでも受信できるほうが望ましい
- ✓ Scalabilityが心配
 - ✓ LDP Sessionが増えたり、PW数が増えたりすると、かなりSwitchoverに時間かかるのでは？

課題③

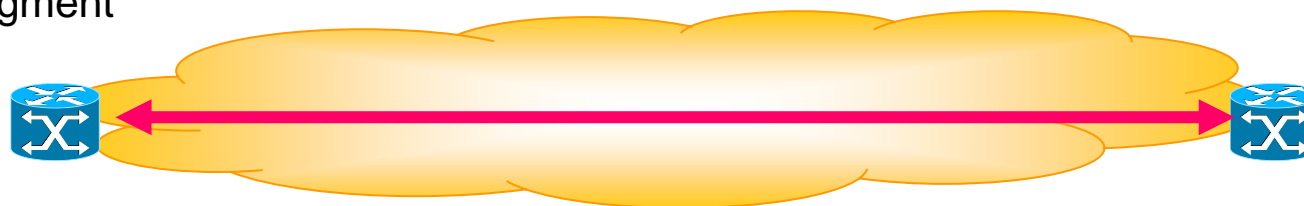
- ・運用負荷の増加
 - >管理elementの増加
 - >回線構成の複雑化



OSSの開発

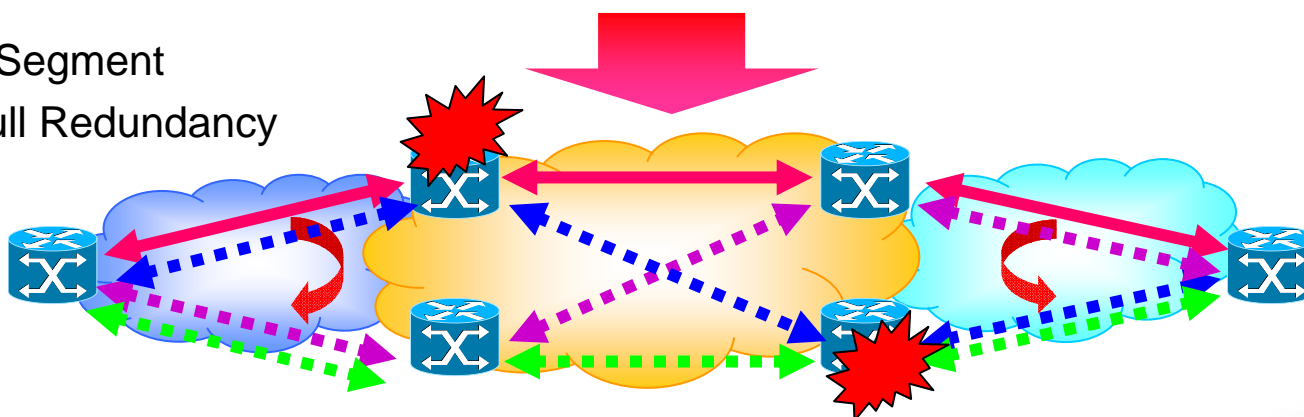
- ・オペレーションの自動化
- ・回線管理の高度化

1 Segment



設定ノード:2
回線:1

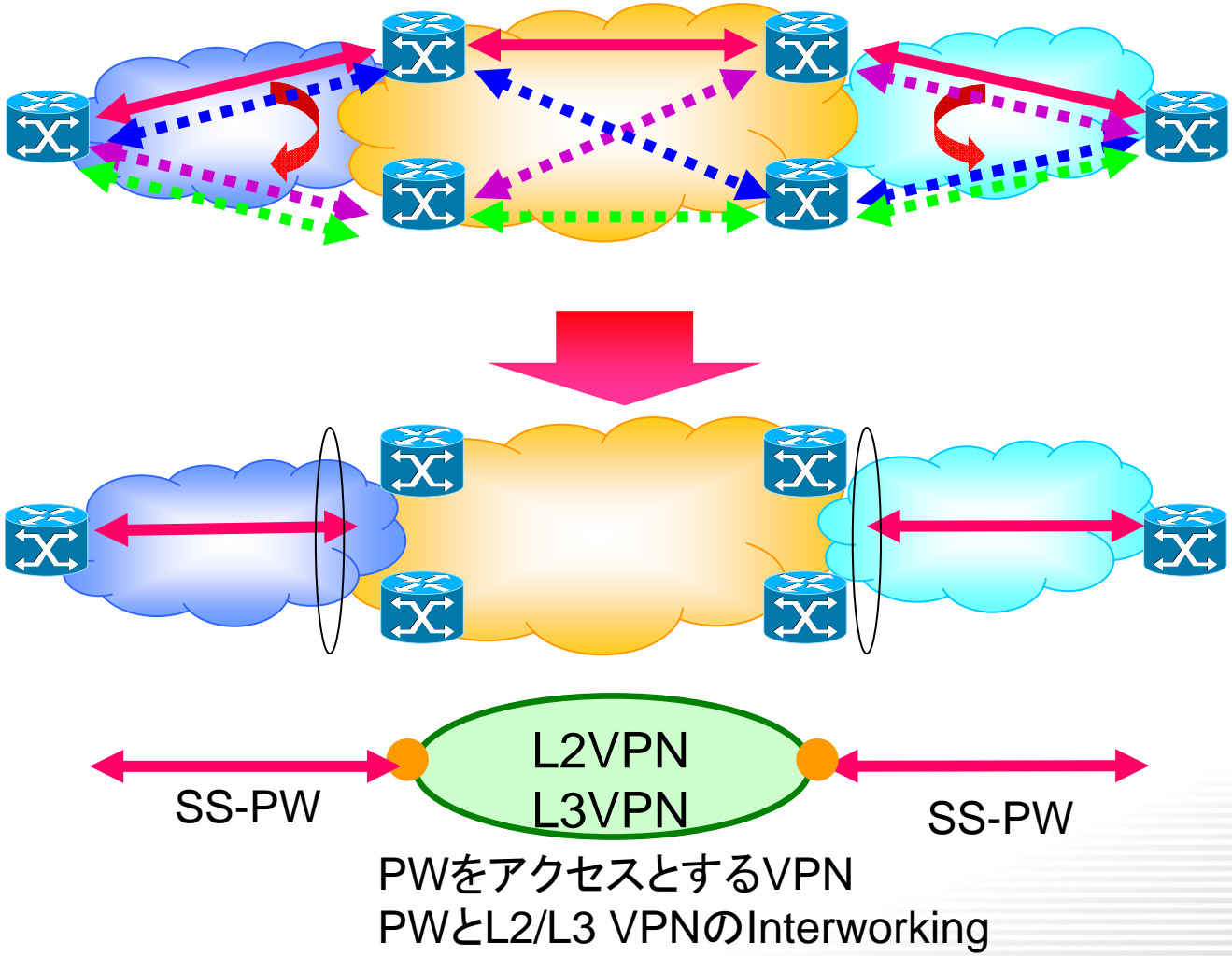
3 Segment
Full Redundancy

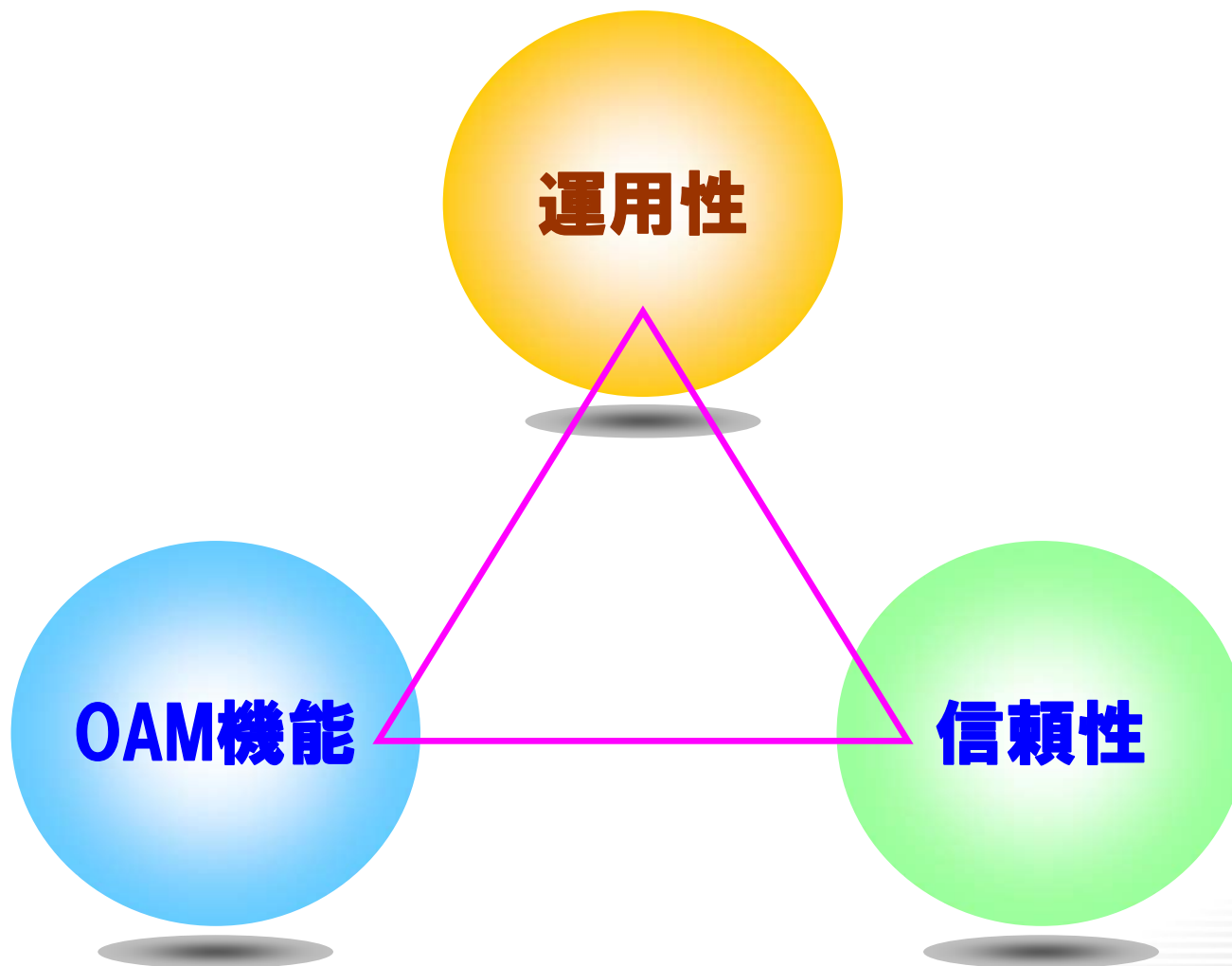


設定ノード:6
回線:12

Primary
Backup

MS-PW環境下でのSPEの冗長性確保を迫ると、
飛躍的に回線構成は複雑化し、運用負荷は増加する。





Thank you !