BROCADE

IETF TRILL UPDATE

なんで独自実装ばっかで IETF TRILL はさっぱりこないのか

2013年2月22日 ENOG19 Meeting

ブロケードコミュニケーションズシステムズ株式会社

高嶋隆一 〈rtakashi @ brocade.com〉

WHAT IS "TRILL"?

TRANSPARENT INTERCONNECTION OF LOTS OF LINKS

コンセプト

Algorhyme V2, by Ray Perlner from RFC6325

I hope that we shall one day see
A graph more lovely than a tree.

A graph to boost efficiency
While still configuration-free.

- ツリーもうやめよう
- でも複雑な設定はNO!

コンセプト cont.

Algorhyme V2, by Ray Perlner from RFC6325

A network where RBridges can
Route packets to their target LAN.

The paths they find, to our elation,
Are least cost paths to destination!

- 最短経路で必要な所にだけ流したい

コンセプト cont.

Algorhyme V2, by Ray Perlner from RFC6325

With packet hop counts we now see
The network need not be loop-free!

RBridges work transparently,
Without a common spanning tree.

- TTLが欲しい
- ループ構成でもいいじゃない

- エンドを変えずスパツリなしで動かすんだ!

要はアンチSTP!

Photo Credit: iloveportraits via Compfight cc

というかやり直し？

RFC6325 authors

R. Perlman
Intel Labs

D. Eastlake 3rd
Huawei

D. Dutt
S. Gai
Cisco Systems

A. Ghanwani
Brocade

- 802.1D Spanning-Treeの生みの親

From
http://en.wikipedia.org/wiki/Radia_Perlman

どんな感じで?

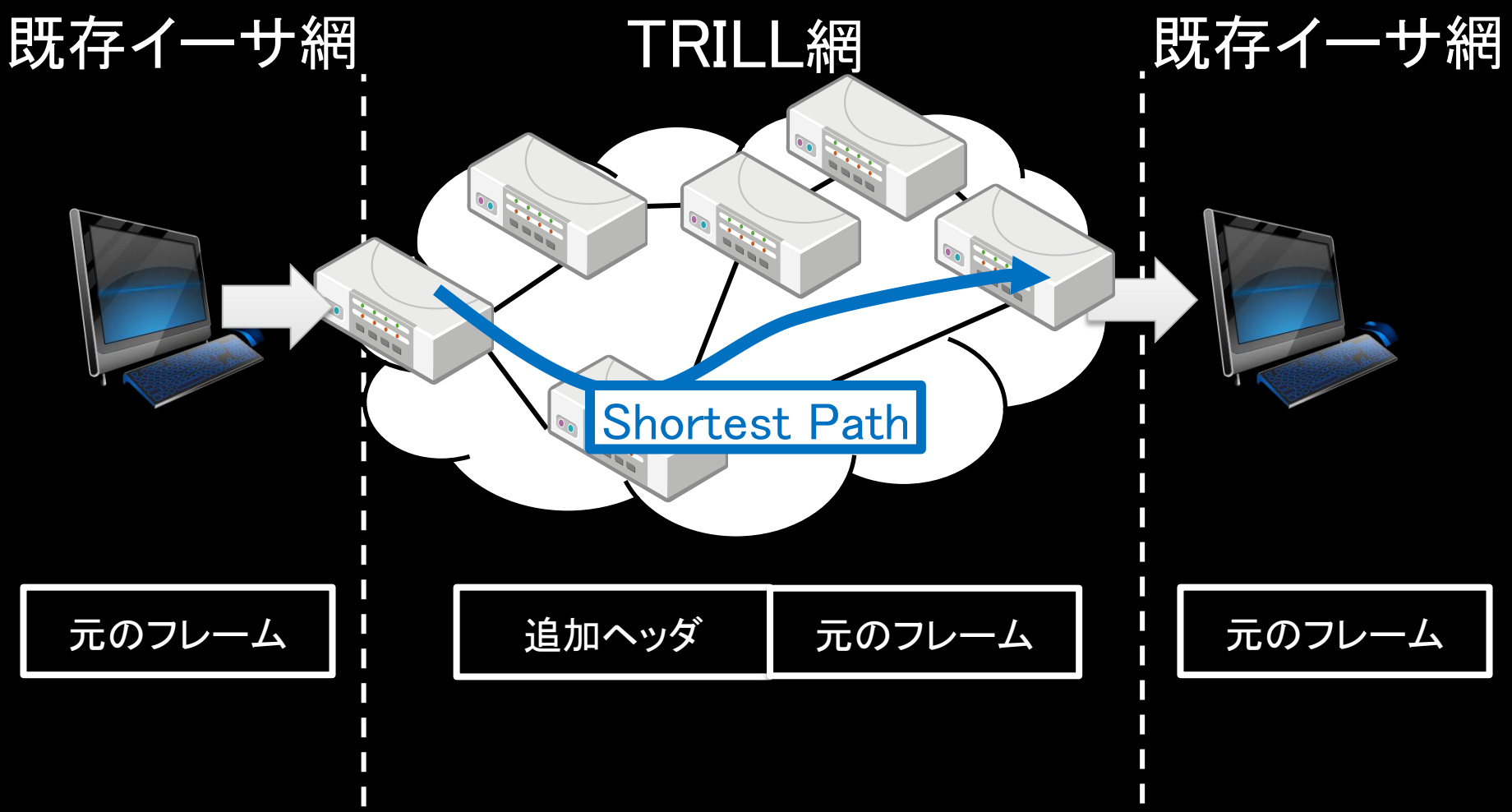
Control Plane

IS-ISによる
ルーティング

Data Plane

追加ヘッダで
Encapsulation

大雑把なイメージ



実はRFC6325のタイトルは“TRILL”ではない

Routing Bridges (RBridges): Base Protocol Specification

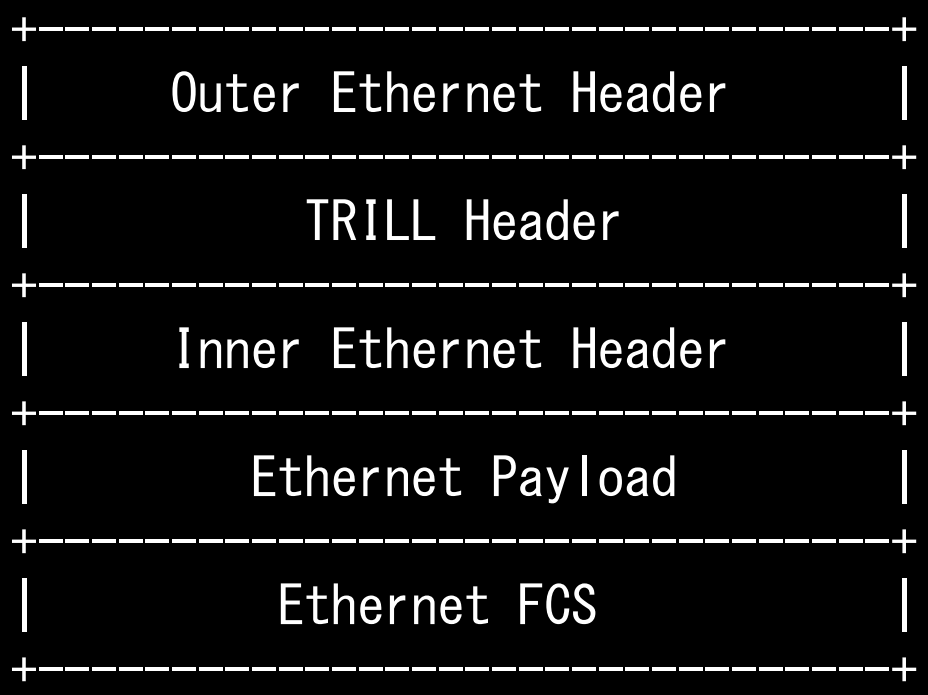
Router（L3）の様な動きをするBridge（L2）なので“Routing Bridge”

※ 以下“Rbridge”“RB”

まずは
Data Plane

TRILL Frame ※

MAC-in-MAC Style Encapsulation



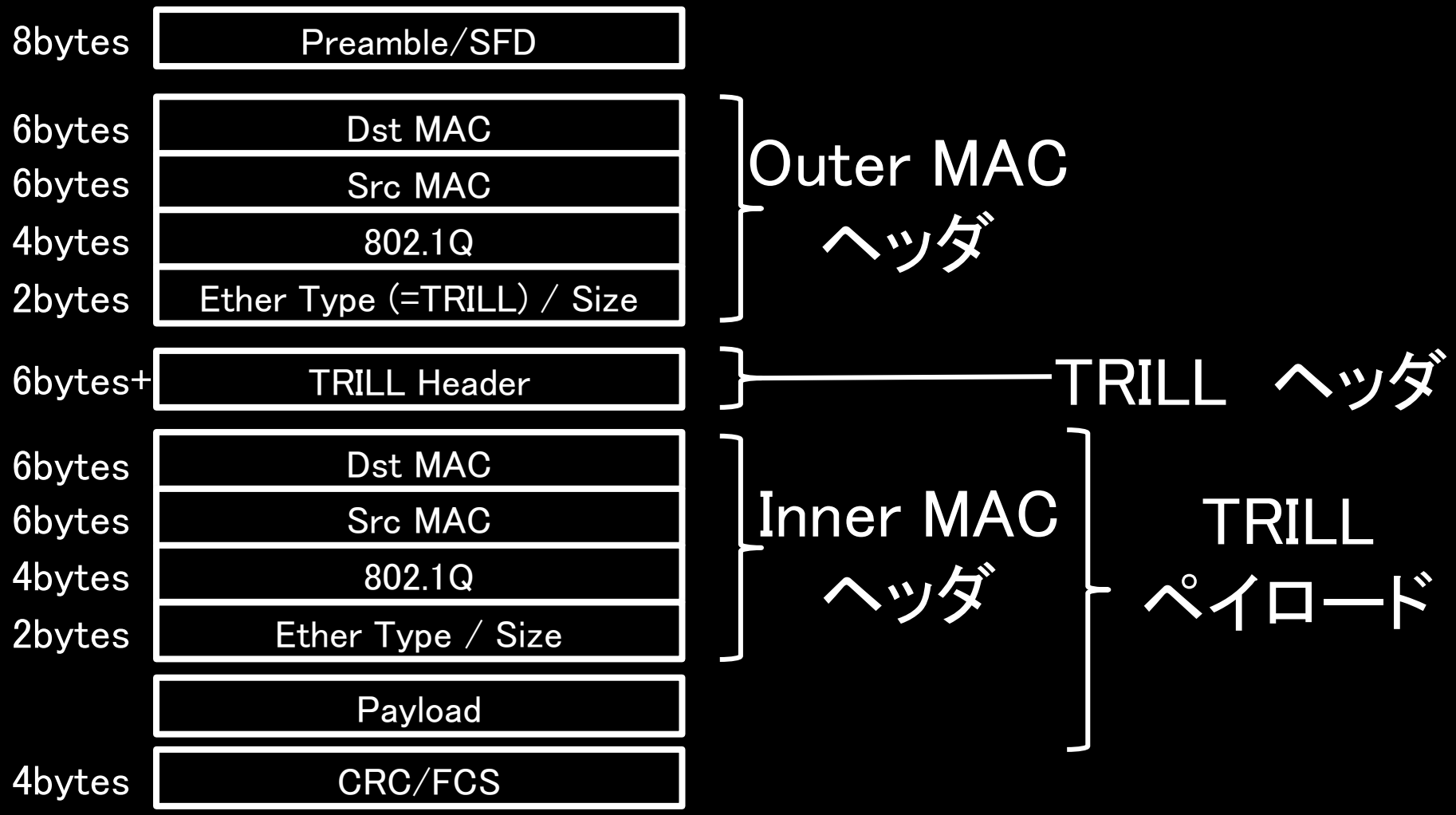
追加ヘッダ

RFC6325 Figure 2: An Ethernet Encapsulated TRILL Frame

※Ethernet以外にPPPもサポートするが今回は省略

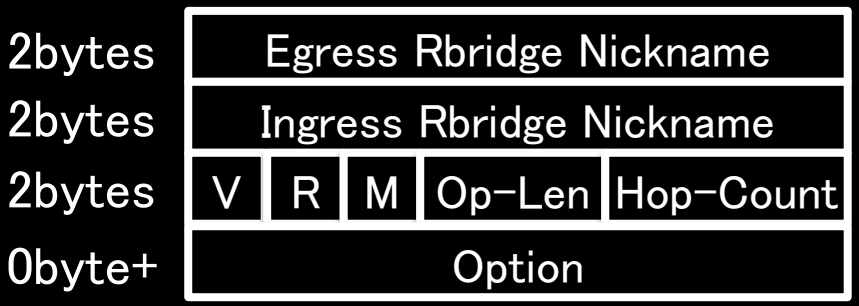
TRILL Frame

もうちょっと詳しく



TRILL Frame

TRILL Header



Egress Rbridge Nickname	宛先のRbridge
Ingress Rbridge Nickname	送信元のRbridge
M	Multi Destination
Hop-Count	TTL

その他

V: Version, R: Reserved, Op-Len: Option Length

次に
Control Plane

L2-IS-IS

IS-IS

- ISIS
  - リンクステート型のルーティングプロトコル
  - L2上で直接動作し、L3設定不要
  - TLV形式のサポートにより、マルチプロトコル拡張が容易
  - SPF (Dijkstra法)による最短経路検索
  - ECMPのサポート

L2-IS-IS

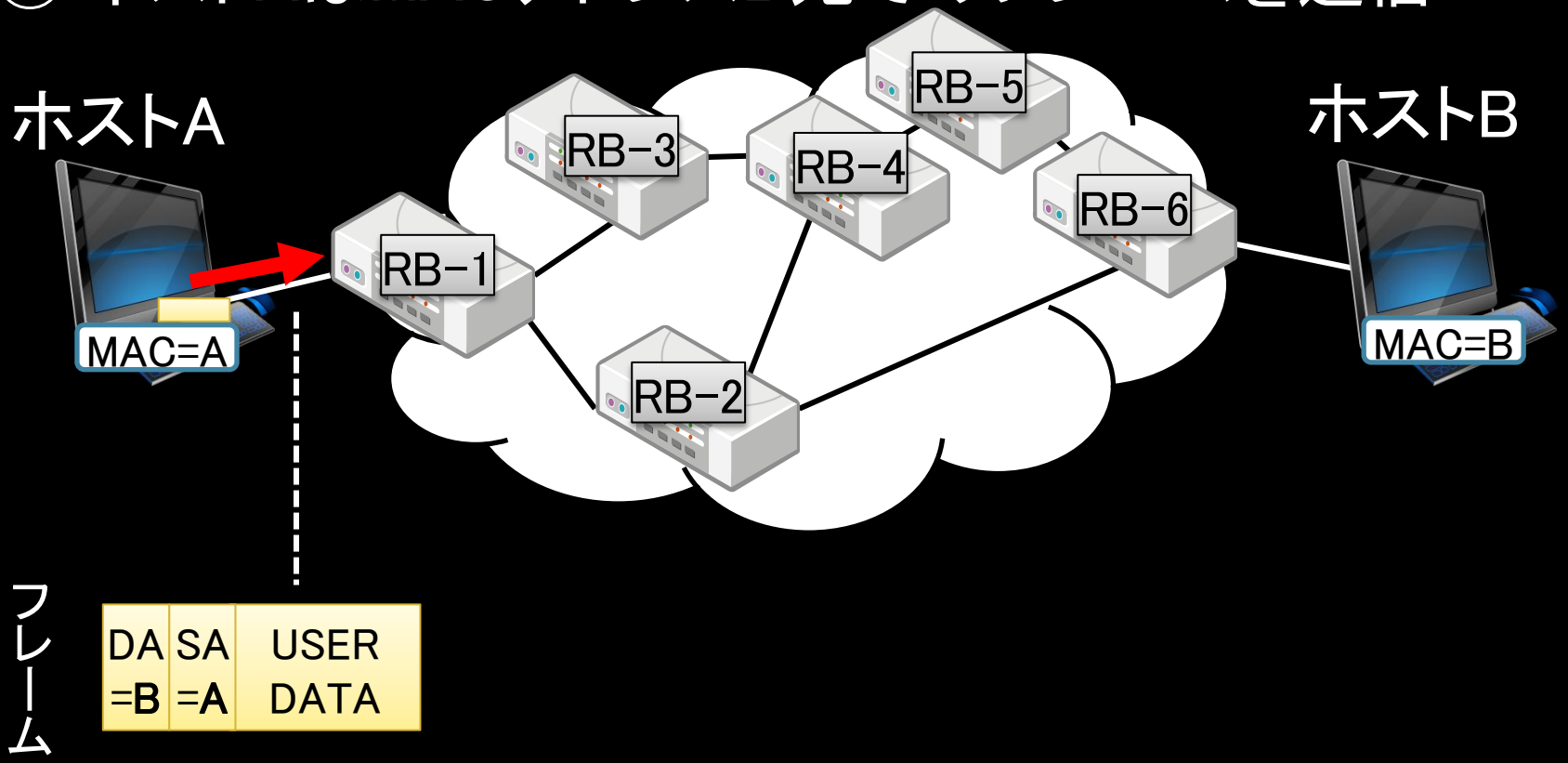
- ✓ L2-IS-IS
  - ➢ L3 IS-ISとは別の Ether Type を持ち、インスタンスも独立
  - ➢ Level-1 のみで構成されるシングルエリアで動作
- ✓ Hello
  - ➢ リンク上のNeighborの確立
  - ➢ リンク上のDesignated Rbridge (DRB)の決定
  - ➢ DRBはTRILLに使用するVLAN ID※を決定

※ Outer MACのVLAN ID, Ethernetでの動作の場合

Unicast

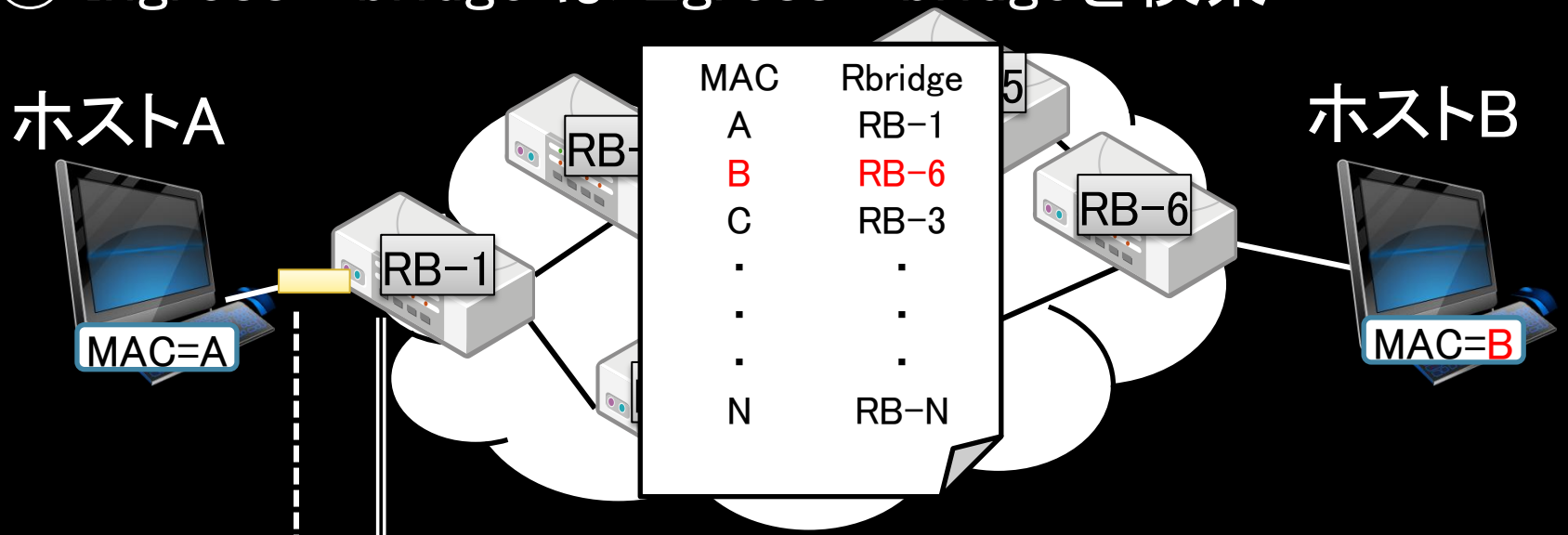
動作イメージ

① ホストAはMACアドレスB宛てのフレームを送信



動作イメージ

② Ingress Rbridge は Egress Rbridgeを検索



フレーム

DA	SA	USER
=B	=A	DATA

Q MAC=Bに繋がっているRbridgeはどこか?

A Rbridge-6

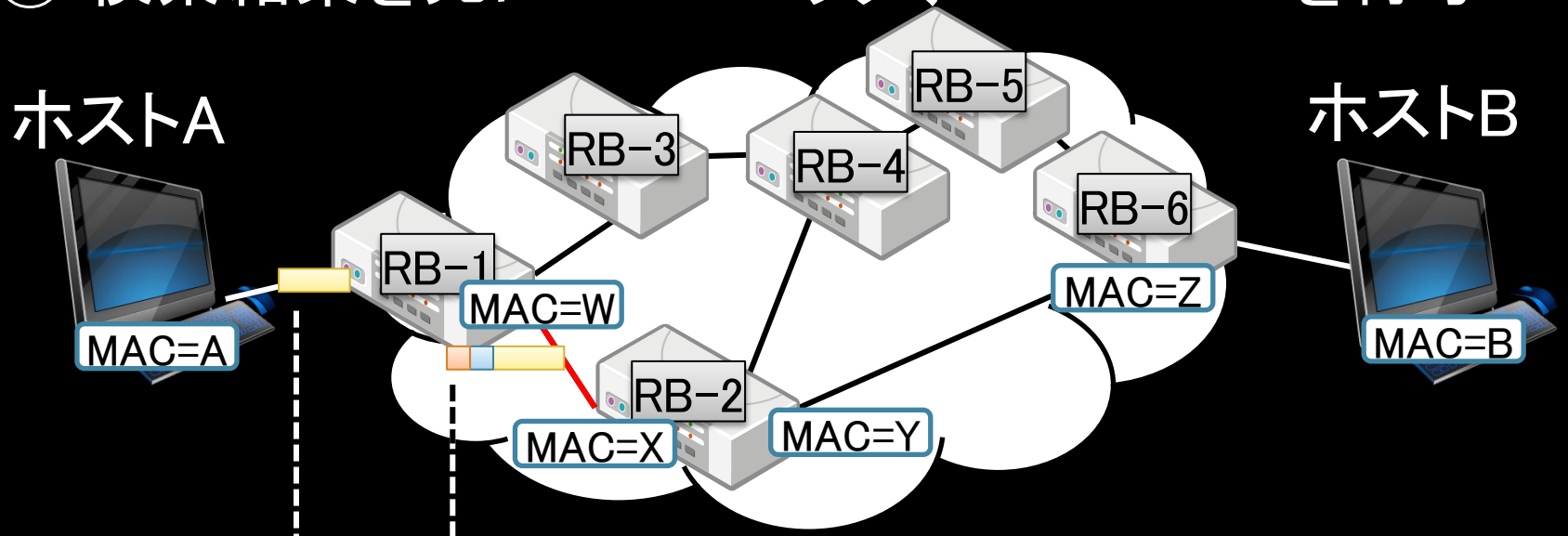
Q どうやって検索するのか

A Rbridge IDとMAC対応表を各Rbridgeが保持

※従来のEgressポートとMAC対応表に相当

動作イメージ

③ 検索結果を元にTRILLヘッダ、Outer MACを付与



フレーム

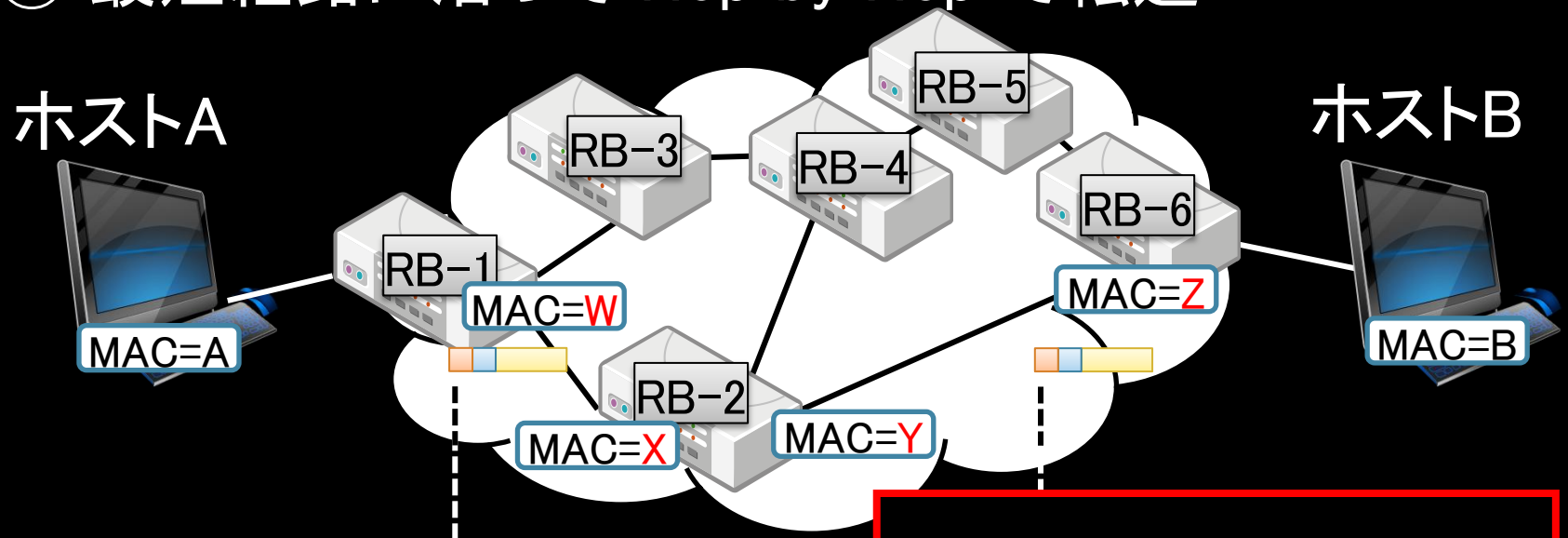
DA	SA	USER
=B	=A	DATA

DA	SA	ER	IR	DA	SA	USER
=X	=W	=6	=1	=B	=A	DATA

Ingress Rbridge = RB-1
Egress Rbridge = RB-6
Source MAC = W
Destination MAC = X
} 最短経路Next-Hopへのリンクの MAC

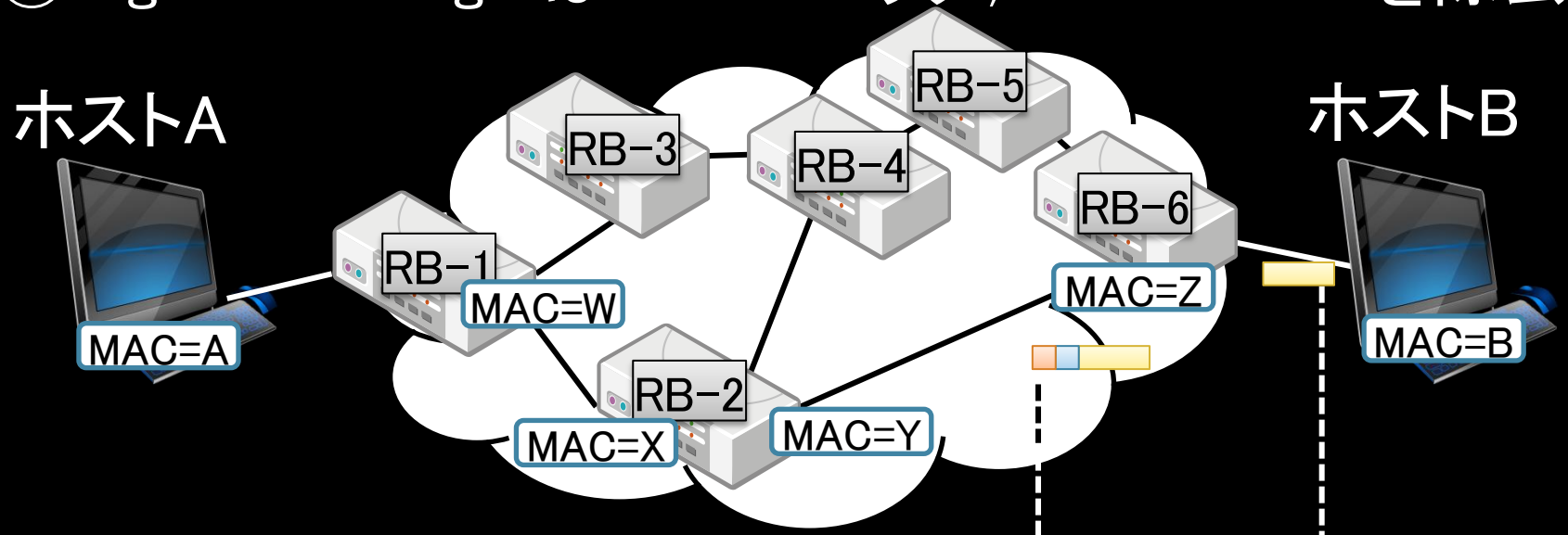
動作イメージ

④ 最短経路に沿って Hop by Hop で転送



動作イメージ

⑤ Egress Rbridge は TRILLヘッダ, Outer MAC を除去



フレーム

DA =B	SA =A	USER DATA

DA =Z	SA =Y	ER =6	IR =1	DA =B	SA =A	USER DATA

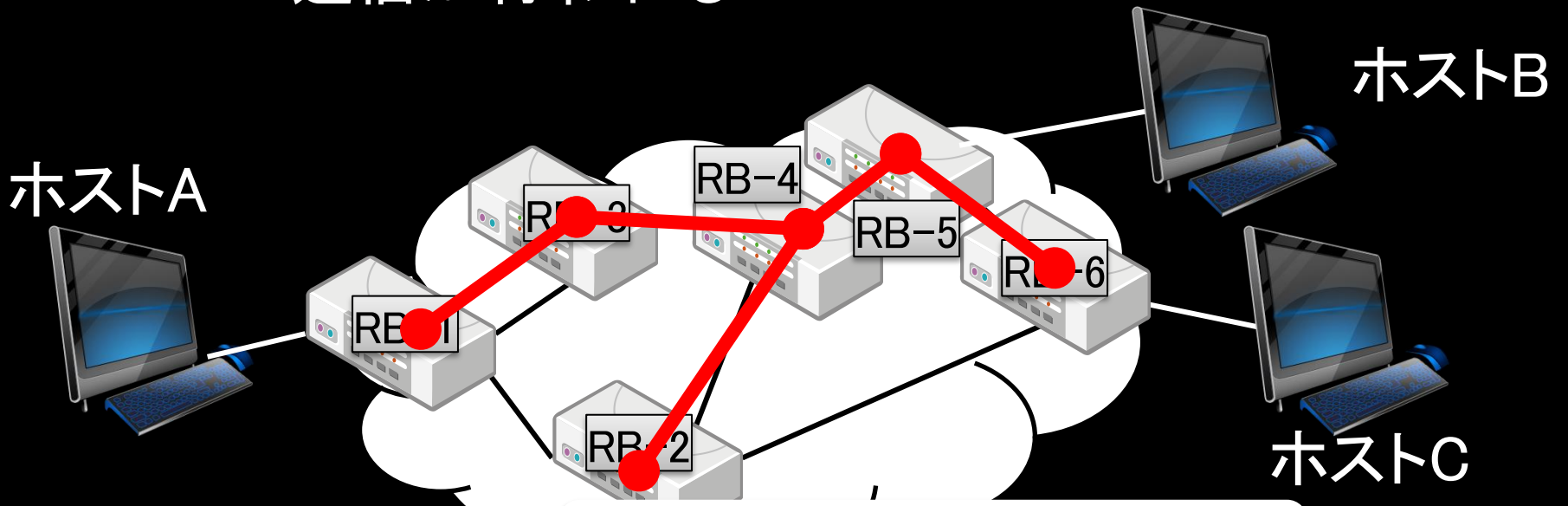
Multi-Destination

BUM
- Broadcast
- Unknown-Unicast
- Multicast

Multi-Destination 概要

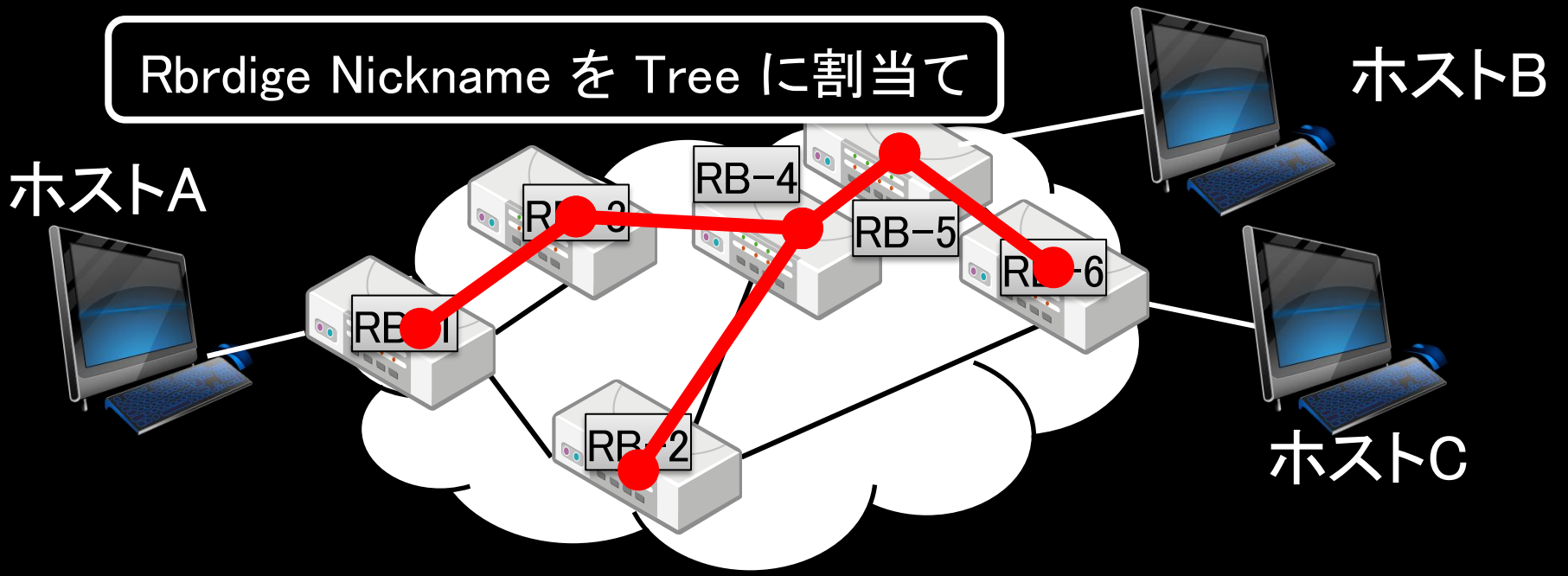
Distribution Tree による配布

- ✓ TRILL網内では IS-IS SPF で作成された双方向 Distribution Tree に沿って BUM 通信が行われる



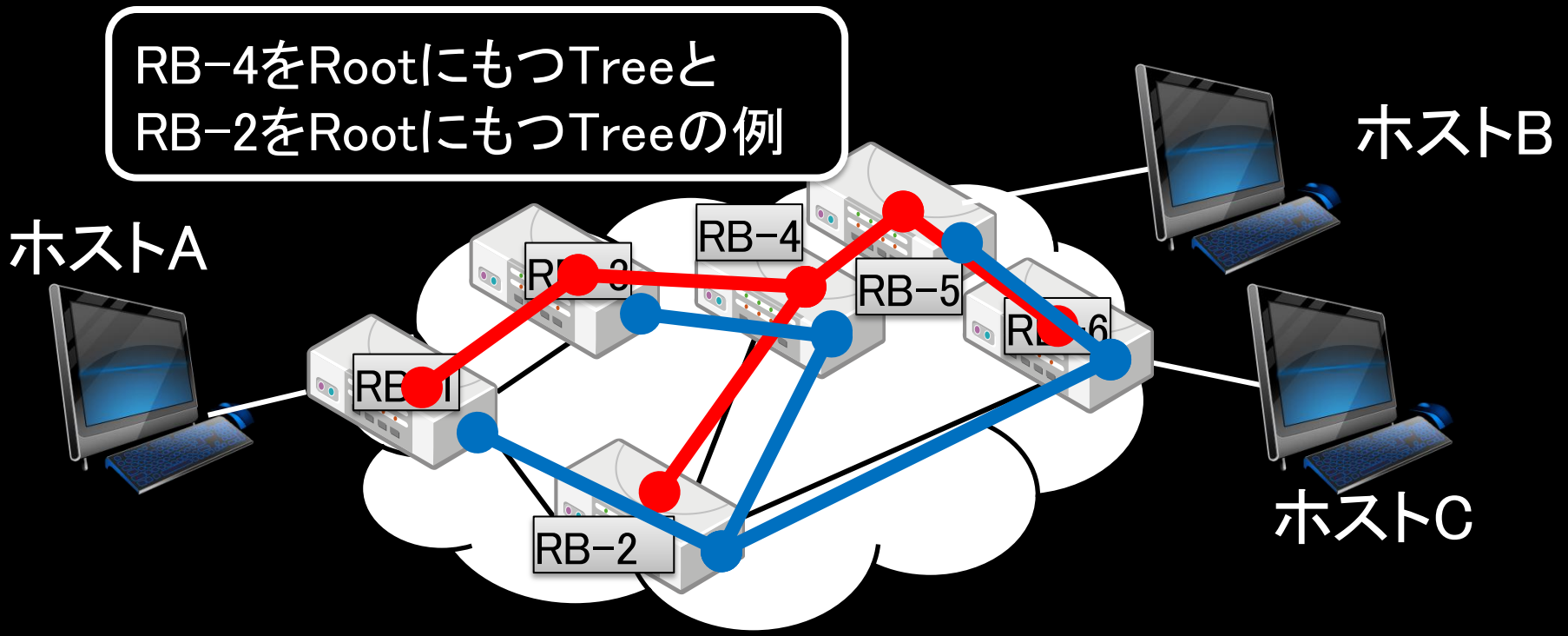
Distribution Tree Basics

- ✓ Distribution TreeにもRbrdige Nicknameが振られる
  - ➢ Egress RbridgeにDistribution Treeを指定
  - ➢ L2/L3 の Broadcast/Multicast Address と同じ



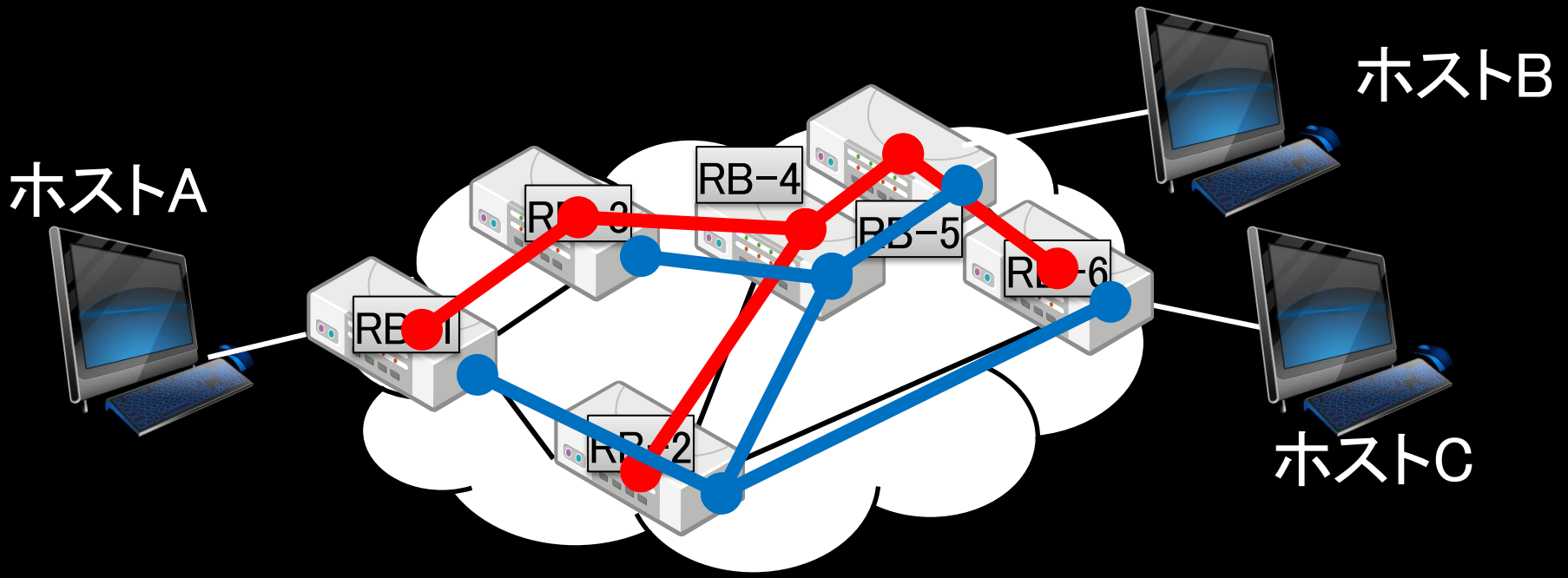
Distribution Tree Basics

- ✓ 複数の Tree があってもよい
  - ➢ Ingress Bridge は最も近いRootを持つTreeを選択



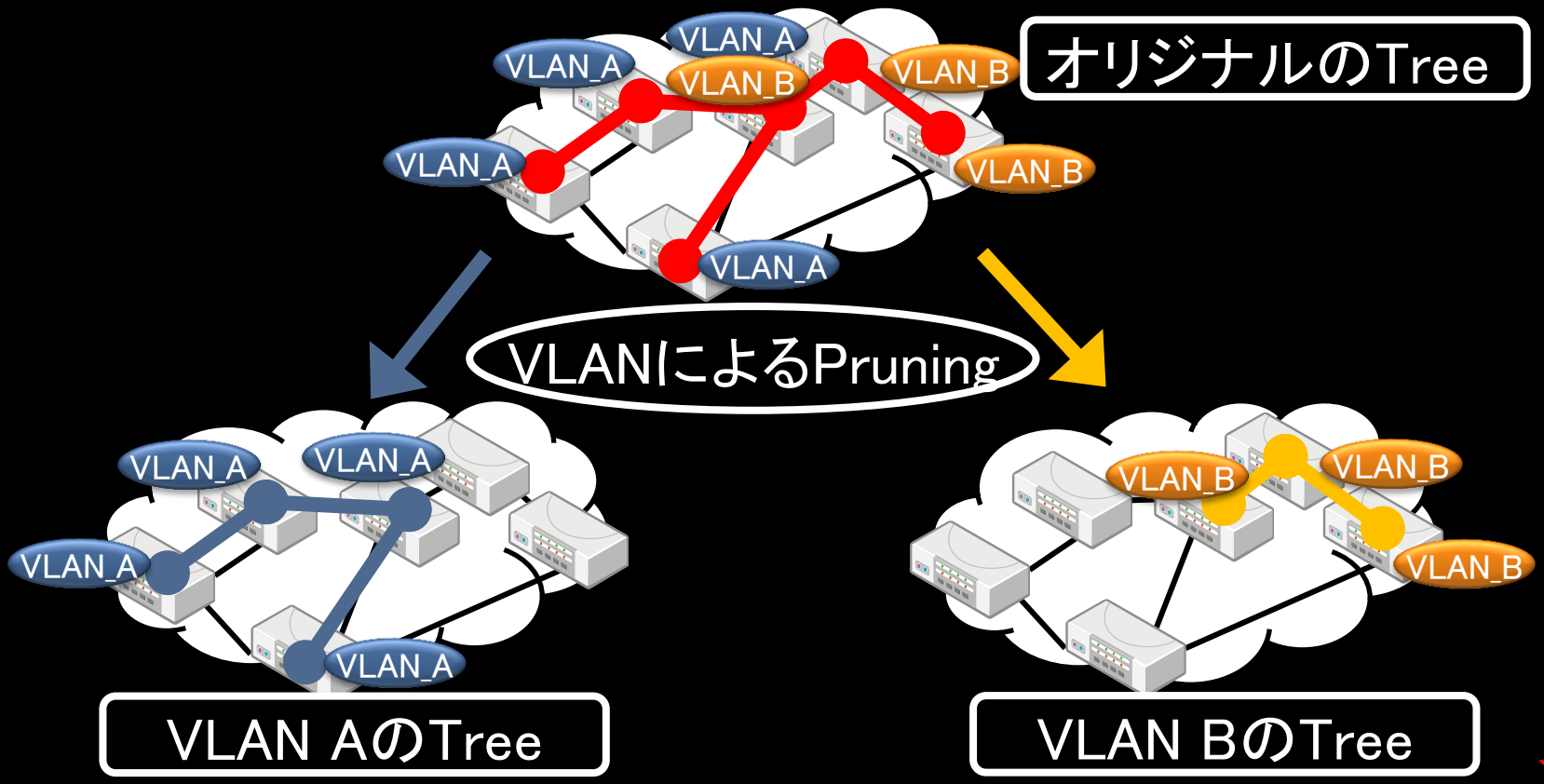
Distribution Tree Basics

- ✓ 複数の Tree は同じ Root をもっていてもよい
  - ➢ Multicast でもロードバランス可能

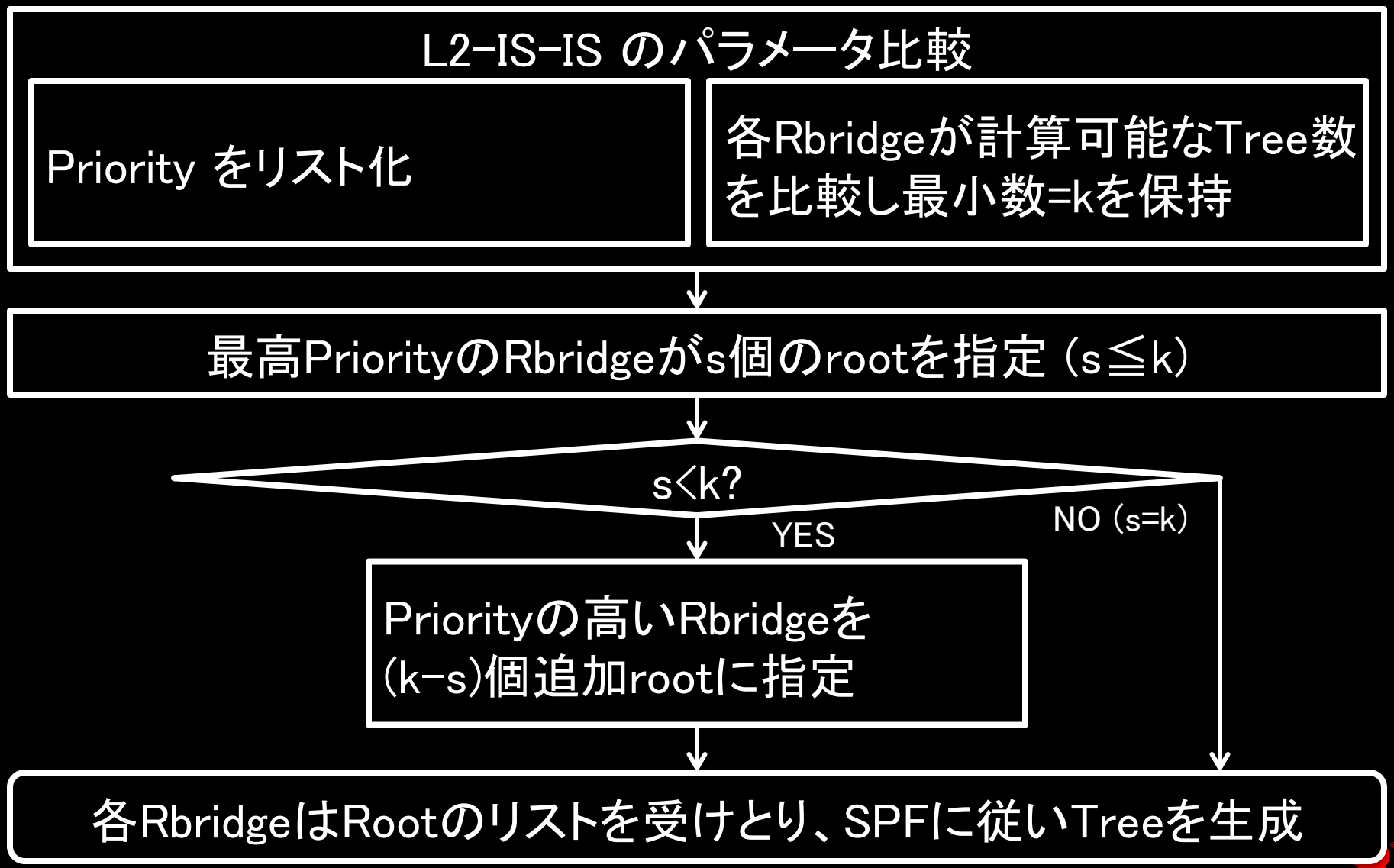


Distribution Tree Basics

- ✓ 各Distribution Tree は用途に応じてPruningされる
  - ➢ VLAN, IP Multicast由来のL2 Multicastアドレス



Tree生成(Root Selection)



MAC学習

5種類のMAC学習方法

① Ingress Rbridgeで学習
- ローカルの”Source Port↔MAC”を学習

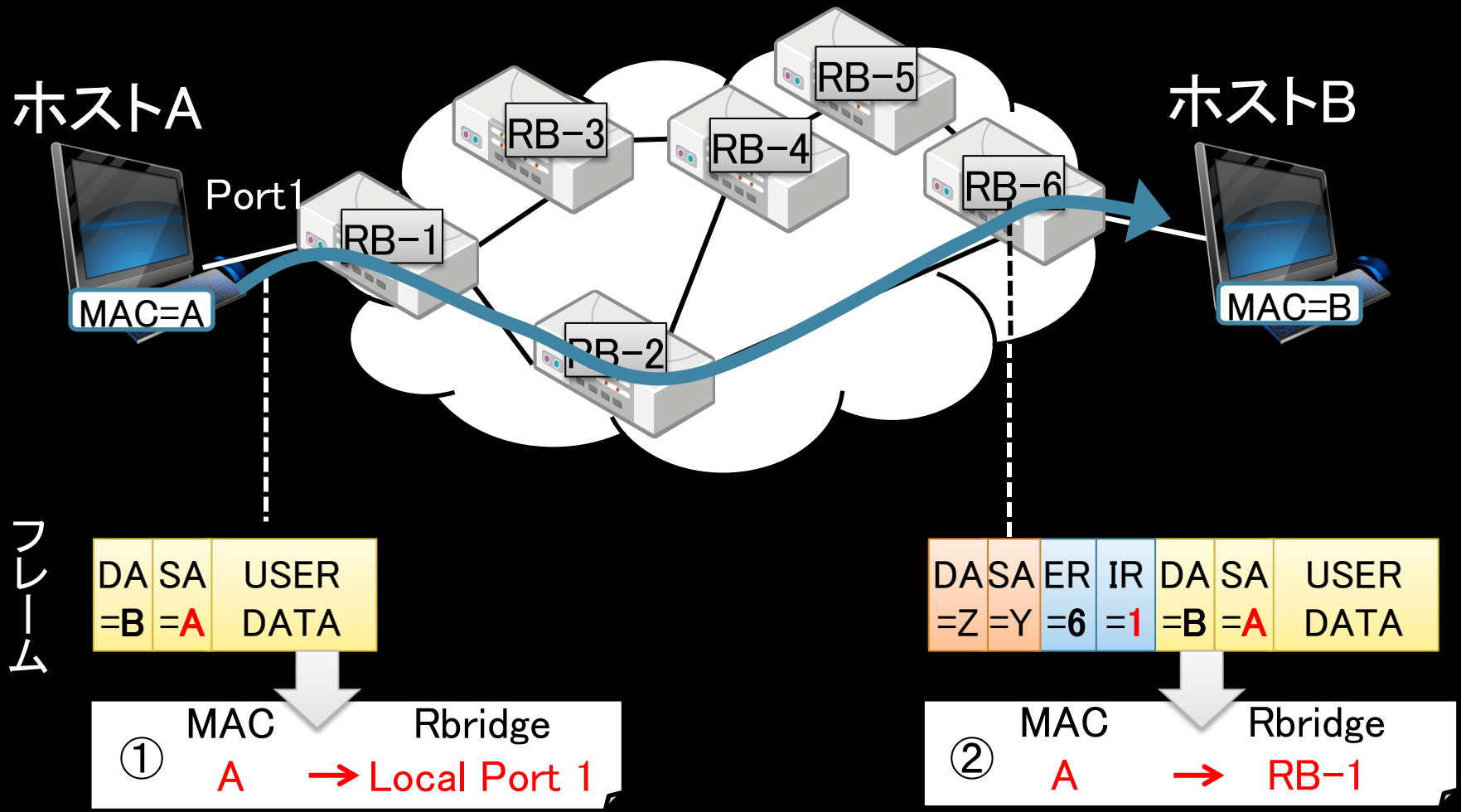
②Egress Rbridgeで学習
- Decap時に”Ingress Rbridge↔MAC”を学習

③L2 Registration プロトコル (IEEE802.11 Association等)

④TRILL ESADI プロトコルによる学習
- Rbridge間で”Rbridge↔MAC”情報を同期

⑤手動設定

流れるフレームから学習



TRILL ESADI Protocol

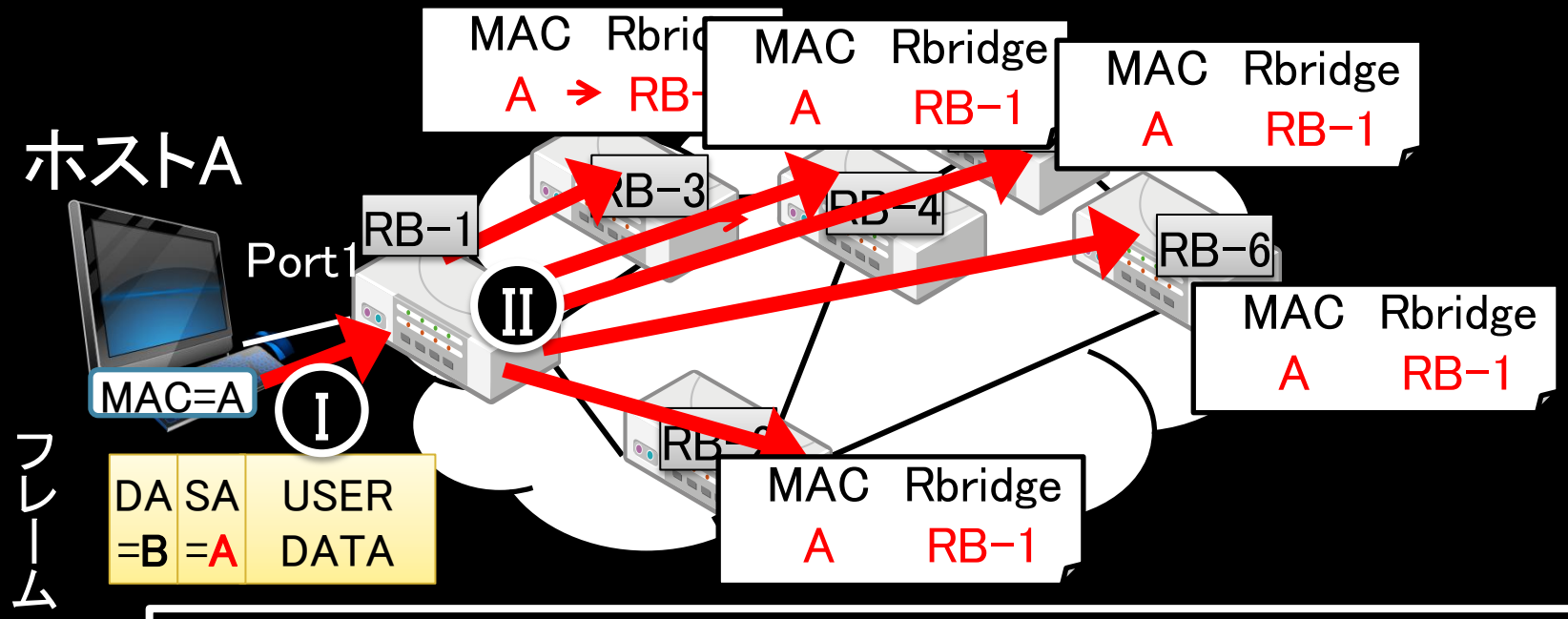
End **S**tation **A**ddress **D**istribution **I**nformation

RFC6325

& draft-ietf-trill-esadi-02 (Expires: 2013-Aug-21)

TRILL ESADI Protocol

動作イメージ



I. いずれかのRbridgeがMACを学習もしくは消去
II. ESADIプロコルにより該当VLANに属する全てのMAC学習テーブルが同期

ESADI Basics

- ✓ ESADIはL2-IS-ISとは独立して動作する
  - ➢ TRILLのデータフレーム扱い
  - ➢ VLAN単位のDistribution Treeを通じて配布
- ✓ ESADIによりRbidge間でのMAC学習テーブルは同期される
  - ➢ BUMの低減
  - ➢ MAC削除時に直接接続されていない Bridge がMAC Age outまでブラックホールとなる状況を回避

ESADI に関する注意

TRILL Working Group
INTERNET-DRAFT
Intended status: Proposed Standard
Updates: 6325

Hongjun Zhai
Fangwei Hu
ZTE
Radia Perlman
Intel Labs
Donald Eastlake
Huawei
Olen Stokes
Extreme Networks

Expires: August 21, 2013

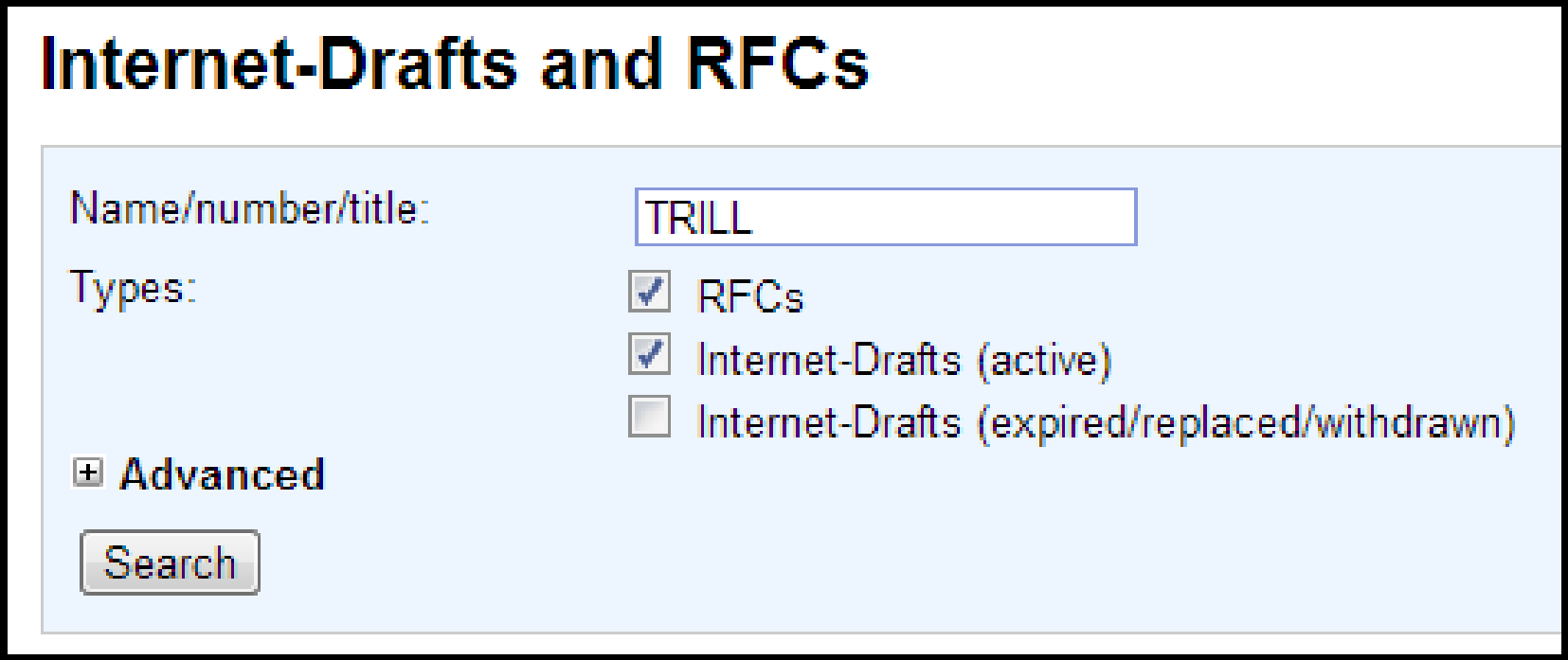
February 22, 2013

TRILL (Transparent Interconnection of Lots of Links):
ESADI (End Station Address Distribution Information) Protocol
<draft-ietf-trill-esadi-02.txt>

概要がRFC6325にあるだけで
まだRFC化されてません

標準化動向

http://datatracker.ietf.org/doc/search/?name=TRILL&rfcs=on&activeDrafts=on&search_submit=



たくさん
ありますが…

Fine-Grained Labeling

TRILL Working Group
INTERNET-DRAFT
Intended status: Proposed Standard
Updates: 6325

Expires: August 12, 2013

Donald Eastlake
Mingui Zhang
Huawei
Puneet Agarwal
Broadcom
Radia Perlman
Intel Labs
Dinesh Dutt
Cumulus Networks
February 13, 2013

TRILL (Transparent Interconnection of Lots of Links):
Fine-Grained Labeling
<draft-ietf-trill-fine-labeling-05.txt>

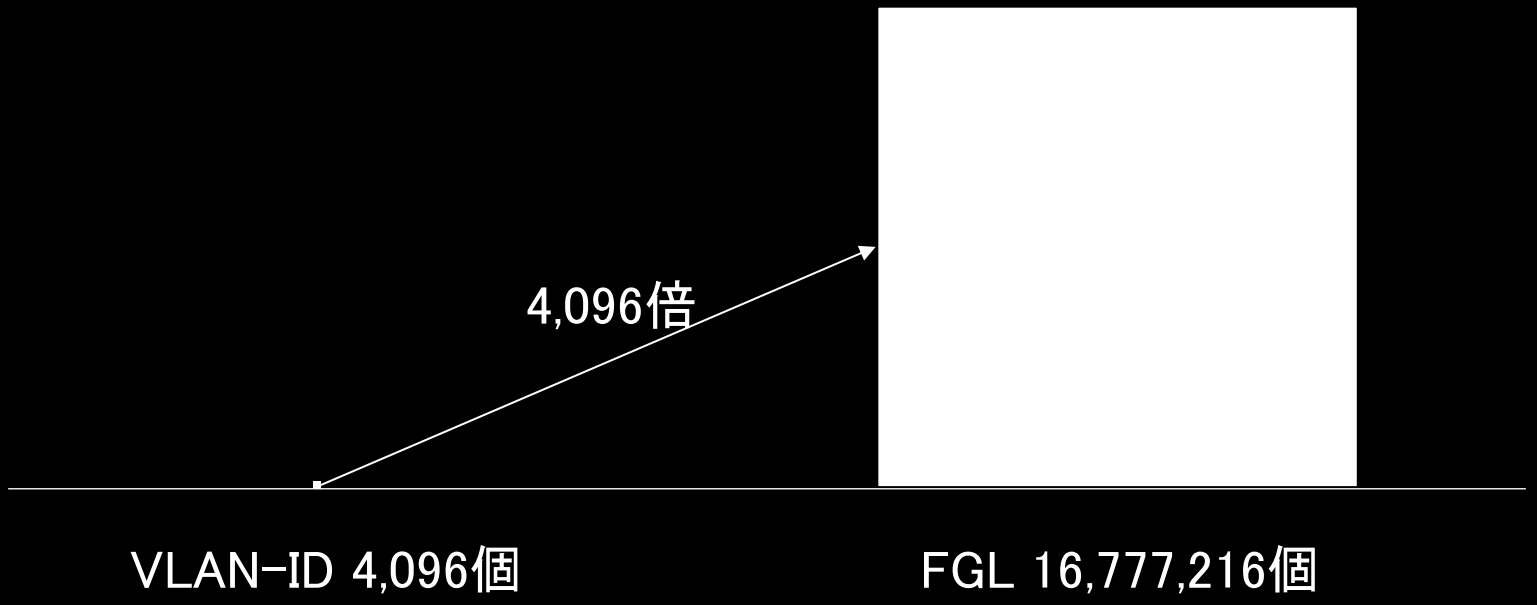
Status: WG Last Call

Fine-Grained Labeling

VLANスケーラビリティの向上

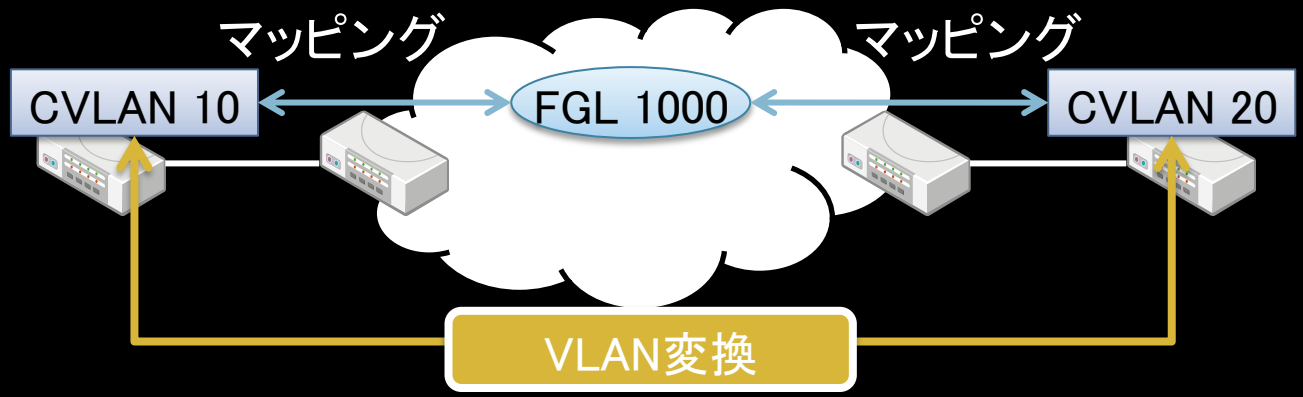
FGL Basics

- ✓ VLAN-ID4096個の壁の突破
  - ➢ 4,096*4,096=16,777,216個のネットワーク
  - ➢ 802.1ad (Q-in-Q)の様に STAG, CTAG の2段構成ではなく16M個をフラットに利用可能



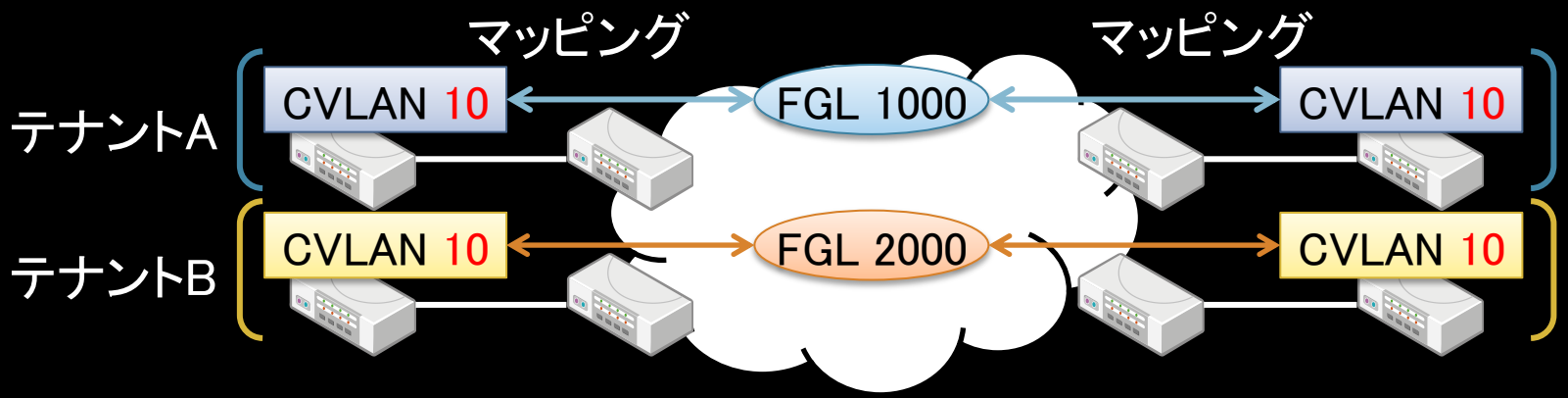
FGL Basics

- ✓ FGLのエッジポート
  - ➢ Untag/802.1QタグをFGLにマッピングし、同一のフラッディングドメインを構成する
  - ➢ 結果としてFGLネットワークはVLANタグ変換機能も持つ事になる
    → VPLSのインスタンスに類似



マルチテナントの考慮

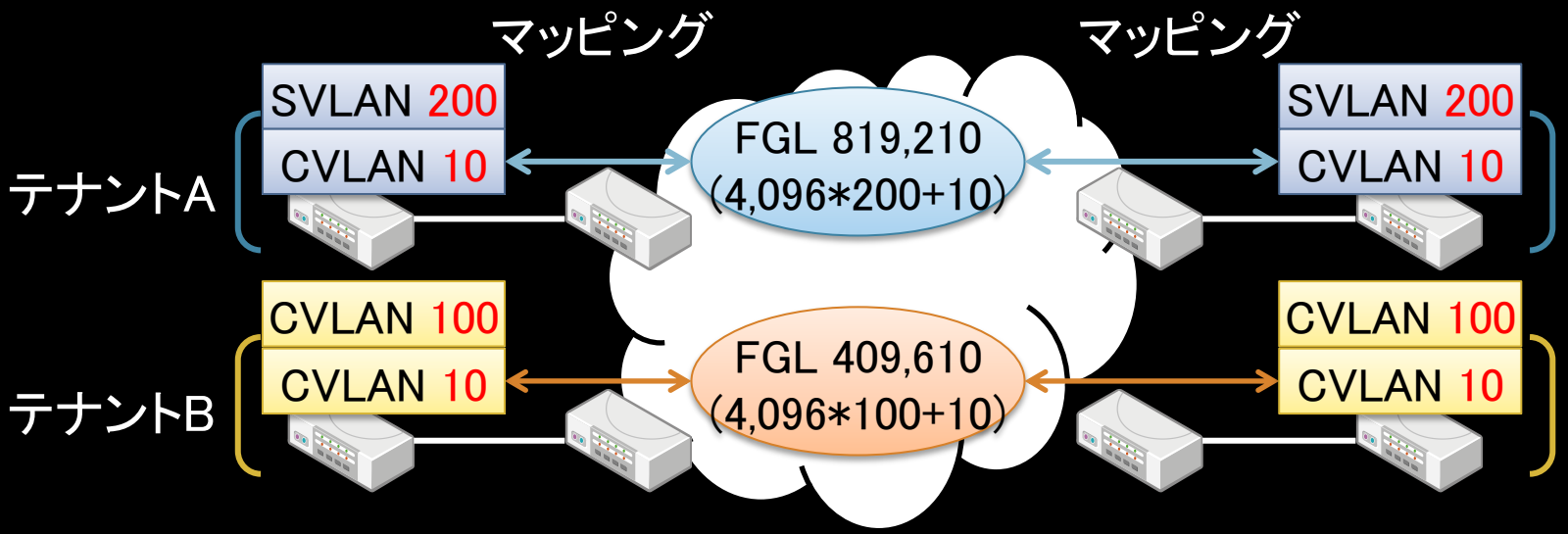
- ✓ VLAN重複の許容



テナントAとテナントBは重複する CVLAN10を利用しているが、
マッピングされているFGLが異なる為、別のネットワークとして動作

802.1adとの共存

- ✓ 相互マッピング
  - ➢ 802.1ad のVLANスペース 4,096x4,096 と同等の空間を持っている為、802.1ad と相互マッピングも可能



単純にCTAGをInner Label Low に、STAGをInner Label Highに割り当てた例

TRILLの実装

IETF標準TRILLの実装

ベンダ	実装	Control Plane	Data Plane
Oracle	Solaris 11 (*2)	OSの標準機能として動作	OSの標準機能として動作
NUST (*1)	Linux (*3)	Quaggaにtrilldを実装	Linux Kernelに拡張

(*1) National University of Sciences & Technology (NUST), Pakistan
(*2) http://docs.oracle.com/cd/E26924_01/html/E25834/rbridgesoverview.html
(*3) http://www.ietf.org/proceedings/84/slides/slides-84-trill-2.pptx

TRILL Like な商用実装

どこが TRILL *Like*?

ベンダ	実装	Control Plane	Data Plane
Cisco	Fabric Path (*1)	Pre Standard(*3), L2MP	Pre Standard(*3), 独自フレーム
Brocade	VCS Fabric (*2)	FibreChannel標準 FSPF(*4)	IETF標準 フレーム

(*1) http://www.cisco.com/web/JP/product/hs/switches/nexus7000/prodlit/white_paper_c11-605488.html
(*2) http://www.brocadejapan.com/solutions-technology/technology/vcs-technology/overview
(*3) TRILLの標準化に先行してI-Dベースの情報とProprietary技術の組合せで実装, IS-IS ベース
(*4) FibreChanel標準 (T11 FC-SW-2) からImport, IS-ISベース
http://www.t11.org/t11/stat.nsf/89dcdcf87b06cf7e85256ebd000a3a96/70d54e897278fd078525660b0071c502?OpenDocument

なんで独自実装ばっかで IETF TRILL はさっぱりこないのか

標準化の遅れ

CY2010Q2

- ✓ Brocade VCS ファブリックのコンセプトを発表
- ✓ Cisco Fabric Pathのコンセプトを発表

CY2010Q4

- ✓ Brocade VCS 対応製品を発売

???

- ✓ Cisco Fabric Path 対応ファームウェアをリリース

CY2011-July

- ✓ TRILLのコアとなるI-DがRFC化
  - RFC6325: Routing Bridges: Base Protocol Specification
  - RFC6326: TRILL Use of IS-IS
  - RFC6327: Routing Bridges: Adjacency
  - RFC6439: Routing Bridges: Appointed Forwarders

CY2013-Feb 現在

- ✓ RFC化されていない基本機能
  - draft-ietf-trill-esadi-02 (MAC学習)
  - draft-ietf-trill-rbridge-channel-08
    (コントロールメッセージの取り扱い)
  - draft-ietf-trill-rbridge-vlan-mapping-08
    (VLAN PCPとのマッピング)

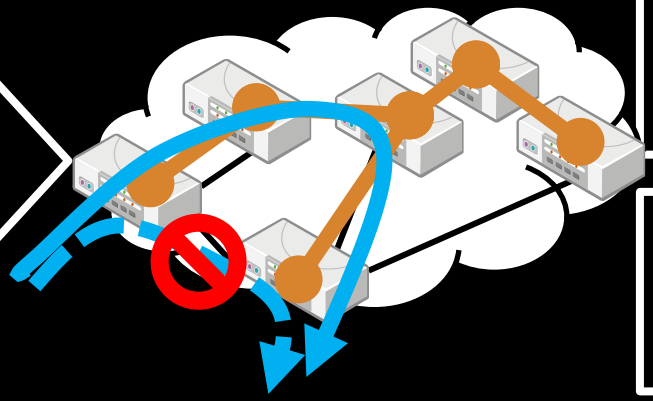
- ✓ コアとなるRFCのリバイス
  - draft-ietf-isis-rfc6326bis-00 (RFC6326)
  - draft-ietf-trill-rbridge-extension-05 (RFC6325)
  - draft-ietf-trill-clear-correct-06 (RFC6325,6327,6439)

WG Draft に限定し、機能拡張を
除いてもこれだけある

実装の難しさ

Multi-Destination

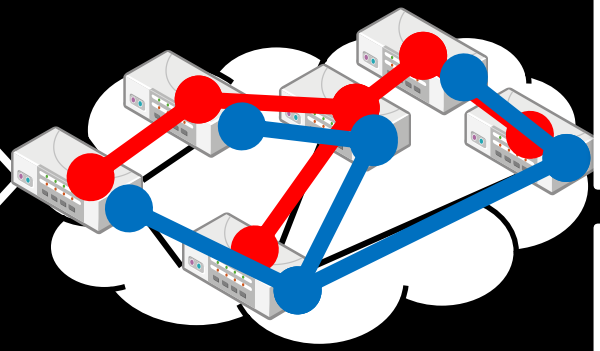
Distribution Tree
による配布



Shortest Path
とは限らない

RootにBUM
が集中する

複数 Tree による
Traffic Engineering?



最適Root配置は?
自動? 手動?

障害時の
取り扱いは?

ターゲットの違い

RFC6325のターゲット

アンチSTP

元々のSTPの領域

キャンパスLAN

商用実装のターゲット

FabricPath および Cisco FabricPath Switching System によるデータセンターの柔軟な拡張

概要

従来のネットワーク アーキテクチャは、静的なアプリケーションに対して高い有用性を提供するように設計
化、きわめてスケーラブルな分散型アプリケーションなど、業界のトレンドは物理データセンター ゾーン間を
Any-to-Any コミュニケーションをサポートする帯域幅のスケーラビリティ向上を必要としています。

http://www.cisco.com/web/JP/product/hs/switches/nexus7000/prodlit/white_paper_c11-605488.html

Brocade VCS Fabric Technology

Brocade VCS Fabric technology helps organizations create efficient data center networks that just work. Ethernet fabrics built on Brocade VCS Fabric technology provide unmatched VM awareness and automation compared to traditional network architectures and competitive fabric offerings. In addition, they increase flexibility and IT agility, enabling organizations to transition smoothly to elastic, mission-critical networks in highly virtualized data centers.

http://www.brocade.com/solutions-technology/technology/vcs-technology/

データセンタネットワーク

データセンタネットワークの要求とTRILLのカバー範囲

	TRILLのカバー範囲
広帯域	○ ECMP ○ ブロックポートの排除
耐障害性向上 断時間の短縮	○ L2ループの排除 △ 中継ノード故障時の高速切替 × マルチシャーシLAG
管理工数の削減	× ロジカルシャーシ × 増設時の設定工数の削減
仮想化アシスト	× ハイパーバイザ連携 × VM Aware QoS
ストレージ トラフィック統合	× 低遅延 × DCB, FCoE, Loss-less iSCSI対応

では残りの要求の実装は?

TRILLでは満たせない要求	実装方法
△ 中継ノード故障時の高速切替 × マルチシャーシLAG	TRILL以外の標準技術
× ロジカルシャーシ × 増設時の設定工数の削減	ベンダ固有拡張
× ハイパーバイザ連携 × VM Aware QoS	
× 低遅延 × DCB,FCoE※,Loss-less iSCSI対応	ハードウェア実装 TRILL以外の標準技術

※BrocadeがFSPFを利用しているのは multihop-FCoE の為という側面もある

では残りの要求の実装は?

現状はベンダ固有拡張
に頼る部分が大きい

× TRILLスイッチを作った

○ DCスイッチにTRILLの一部機能をいれた

データセンタ以外でのTRILLの適用

DC以外のユースケース

■キャンパスLAN

✓元々の要求にはあっているが、コストが見合うか

✓STPとの併用ではあまり意味がない ※

➢エンドユーザからの需要が増えればあり得る

※TRILL自体はSTPとの相互接続を考慮している

■メトロエリアイーサネット

✓既存のPB/PBB、MPLS/VPLSに対して優位性があるか

✓OAMやBUM制御等の機能が必要

➢ 機能の大幅追加がないと厳しい

どちらも今すぐには難しそう

まとめ

- ✓ TRILLはSTPにかわるループフリーのL2制御プロトコル
- ✓ 標準化は基本的な機能を含めてまだまだまだ進行中
- ✓ 商用実装はRFC準拠機能に加え、独自機能を実装して製品化されている
- ✓ 現状はデータセンタでの利用がメイン

BROCADE

Questions?