Cisco Plus
Japan

Session 2D-2

# FabricPathによる レイヤ2ネットワークの革新 --- 技術解説とデモンストレーション

シスコシステムズ合同会社
基盤技術チーム
コンサルティング
システムズエンジニア
山下 薫

cisco

アニメーションあり

# Cisco FabricPath とは?

- FabricPath ≒ TRILL + α

  IETF TRILL (RFC 6325他) と同等以上の機能を持つ、
  スパニングツリー(STP)に代わるプロトコル　※ STPとの共存もサポート
  レイヤ2 におけるルーティング
- 高信頼性の実現、STPの限界の打破
  - ルーティングには IS-ISを使用 → TRILLと同じ
  - 3つ以上の隣接機器へのロードバランス (マルチパス, 最大16)の実現
  - ルーティングなので、任意のトポロジを構成可能
  - ルーティングテーブル更新時の断時間が皆無 or 非常に短い

    障害 → 切り替わり、切り戻り、片寄せ、増設 etc.
- Cisco Nexus 5500 で年内に正式サポート予定

  βテストがほぼ完了
  Nexus 7000 では既にサポート (Production Networkでの実績あり)
  ハードウェアは 5500、7000ともに TRILLに対応

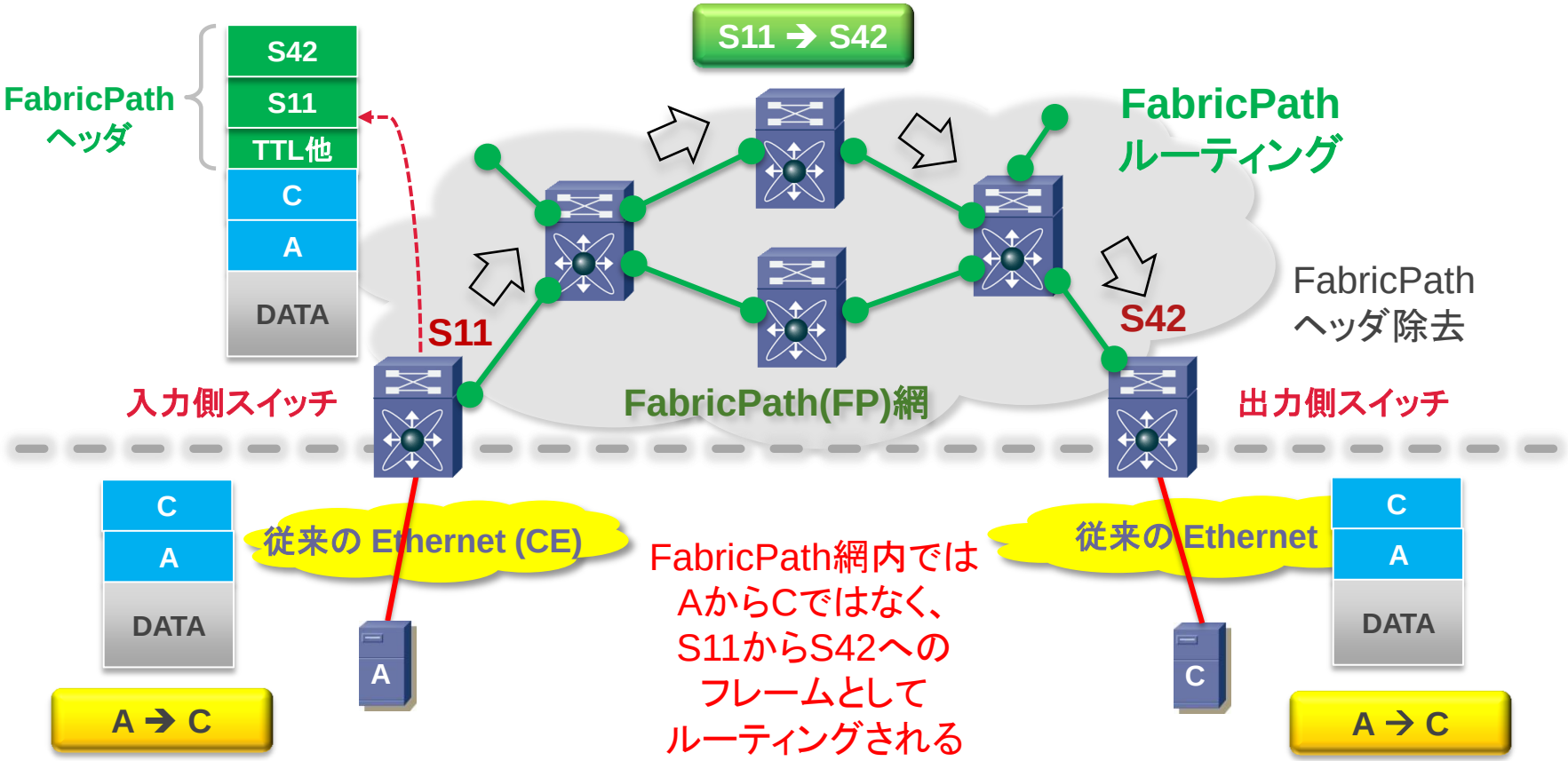
# Cisco FabricPath の仕組み

アニメーションあり

# FabricPath の動作概要

- 入力側スイッチにおいて FabricPathヘッダが付加される
- 入力側スイッチと出力側スイッチに割り当てられた “Switch ID (SWID)” (ここでは S11, S42) を用いて「ルーティング」を行う
- FabricPath網内では、MACアドレス学習は不要

※ ルーティング
テーブルは後述



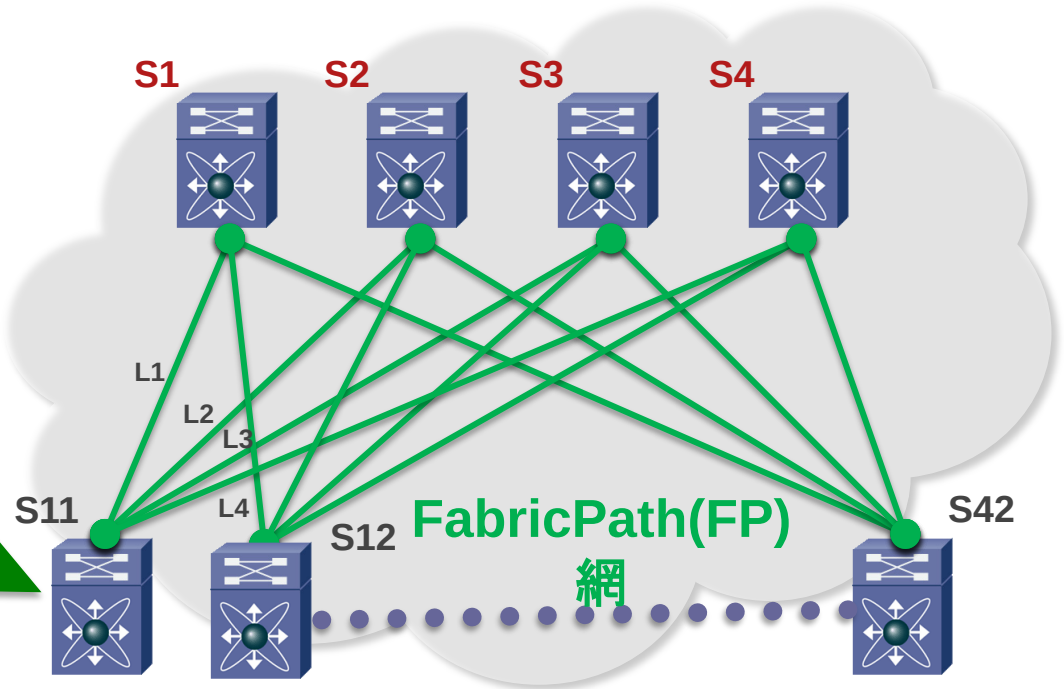
アニメーションあり

# FabricPath ルーティングテーブルと等コストマルチパス

- FabricPathが有効な全てのスイッチに、自動的にSWIDが割り振られる
  (デフォルト。ですが、手動で割り振ることをお勧めします)
- IS-ISによって最短のパス (スイッチとI/Fのペア、ルーティングテーブル) を計算する
- 複数のFabricPathスイッチを経由する等コストマルチパス(ECMP) → 最大16
  S11 から見ると S12, S42へはそれぞれ4つのパスを経由して到達可能

**FabricPath ルーティングテーブル**

| Switch | I/F |
|---|---|
| S1 | L1 |
| S2 | L2 |
| S3 | L3 |
| S4 | L4 |
| S12 | L1, L2, L3, L4 |
| … | … |
| S42 | L1, L2, L3, L4 |

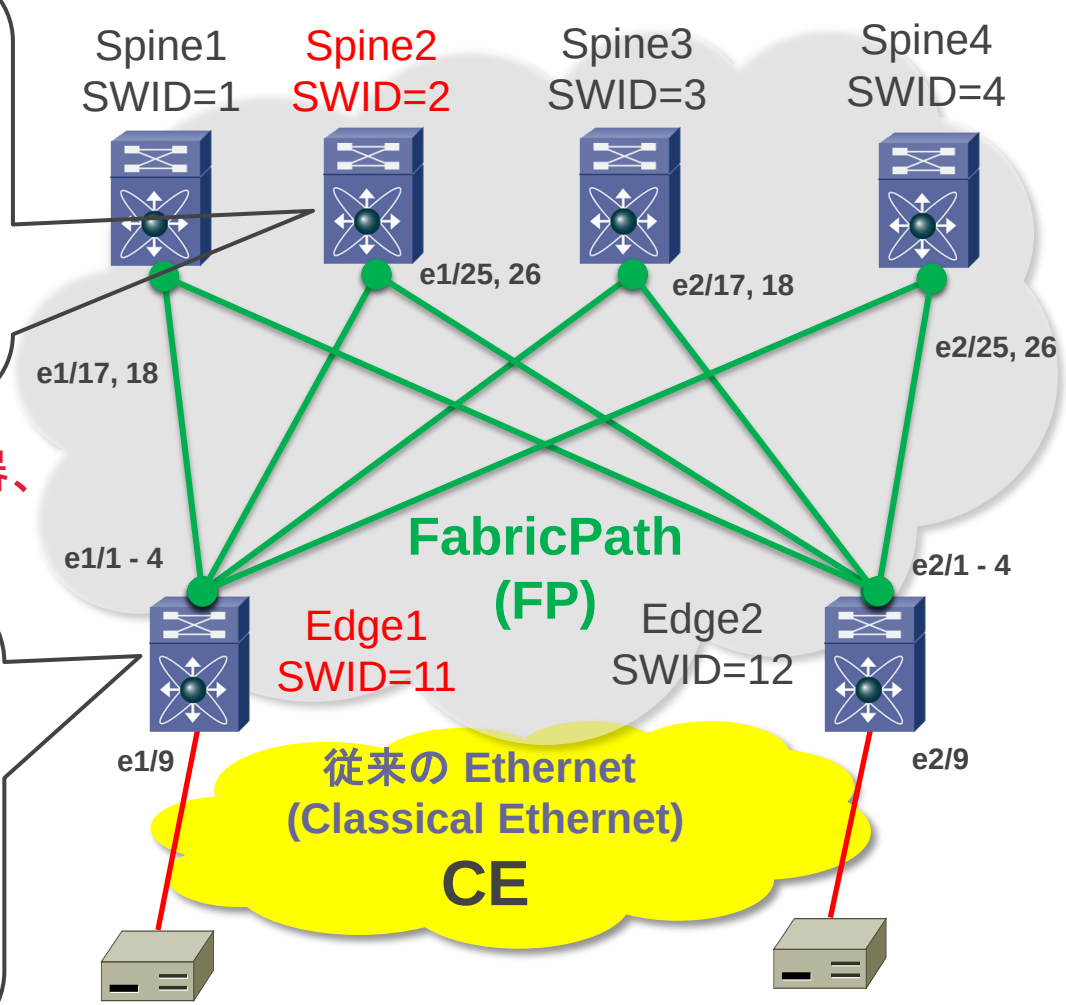


アニメーションあり

# FabricPath Config例
# IS-ISを意識する必要なし

10G (FP)
10G (CE)

```
Spine2# conf t
Spine2(config)# feature-set fabricpath
Spine2(config)# fabricpath switch-id 2
Spine2(config)# int e1/25-26
Spine2(config-if-range)# switchport
mode fabricpath
Spine2(config-if-range)# exit
Spine2(config)# vlan 1-64
Spine2(config-vlan)# mode fabricpath
Spine2(config-vlan)# end
Spine2#
```




“Spine” (背骨) は CEに接続しない機器、
“Edge” は FP – CE 境界にある機器

```
Edge1# conf t
Edge1(config)# feature-set fabricpath
Edge1(config)# fabricpath switch-id 11
Edge1(config)# int e1/1-4
Edge1(config-if-range)# switchport
mode fabricpath
Edge1(config-if-range)# exit
Edge1(config)# vlan 1-64
Edge1(config-vlan)# mode fabricpath
Edge1(config-vlan)# end
Edge1#
```

Nexus 7000 Default VDC, Nexus 5500 では “**install feature-set fabricpath**” の投入が最初に必要

アニメーションあり

# 実際の FabricPath ルーティングテーブル @ Edge1, SWID = 11

ftag : 複数のトポロジを識別するID
トポロジは現在一つなので、
ここで表示される ftag は常に ‘1’
(自分自身の場合は ‘0’ )

metric : 10Gbps の場合 “40”
1Gbps だと “400”
(∵ Reference Bandwidth = 400Gbps)

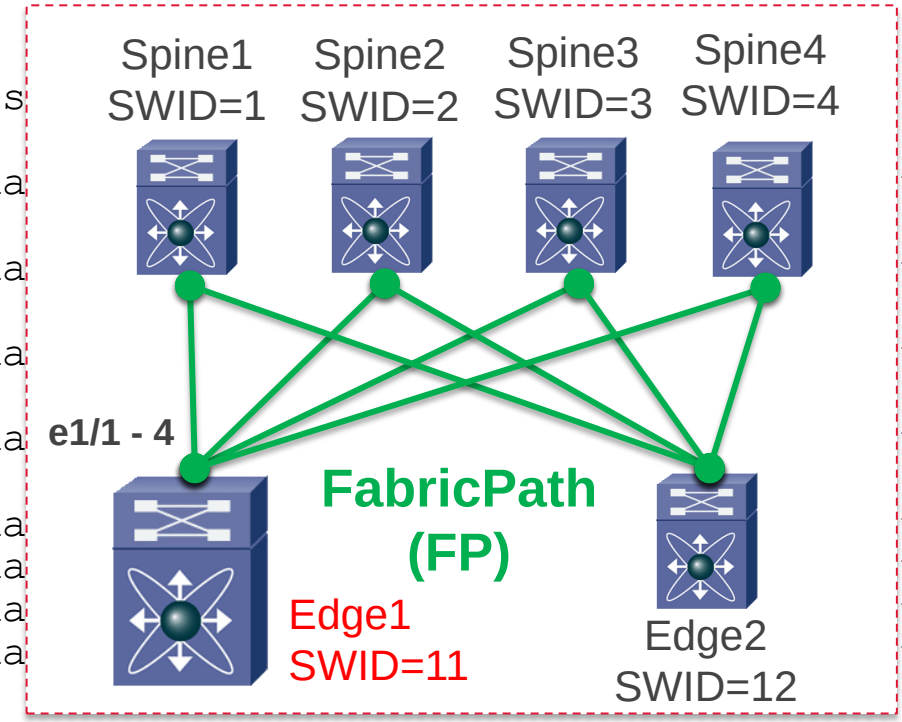
```
Edge1# show fabricpath route
FabricPath Unicast Route Table
'a/b/c' denotes ftag/switch-id/subswitch-id
'[x/y]' denotes [admin distance/metric]
ftag 0 is local ftag
subswitch-id 0 is default subswitch-id

FabricPath Unicast Route Table for Topology-Default

0/11/0, number of next-hops: 0
        via ----  , [60/0], 0 day/s
1/1/0, number of next-hops: 1
        via Eth1/1, [115/40], 0 da
1/2/0, number of next-hops: 1
        via Eth1/2, [115/40], 0 da
1/3/0, number of next-hops: 1
        via Eth1/3, [115/40], 0 da
1/4/0, number of next-hops: 1
        via Eth1/4, [115/40], 0 da
1/12/0, number of next-hops: 4
        via Eth1/1, [115/80], 0 da
        via Eth1/2, [115/80], 0 da
        via Eth1/3, [115/80], 0 da
        via Eth1/4, [115/80], 0 da
```

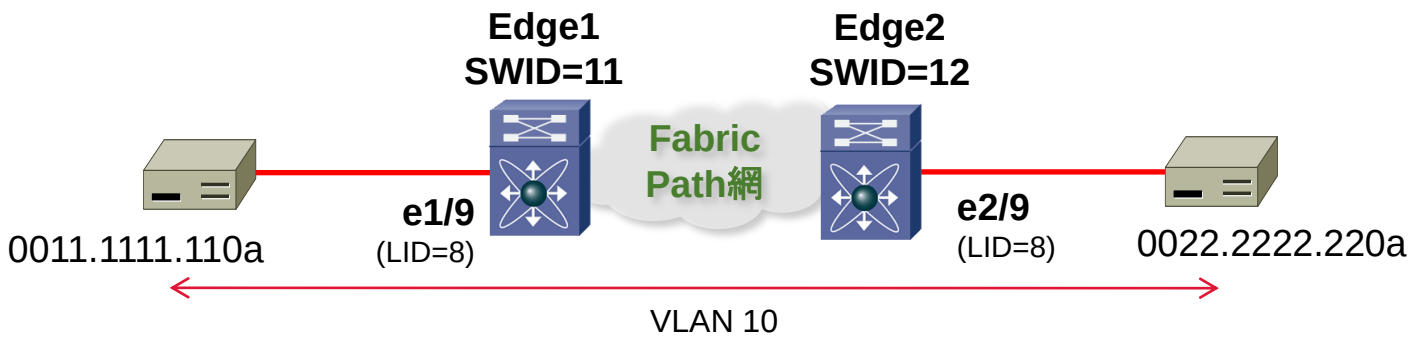
ult
ult
ult
ult
ult
ult
ult
ult

宛先 SWID



アニメーションあり

# FabricPath MACアドレステーブル

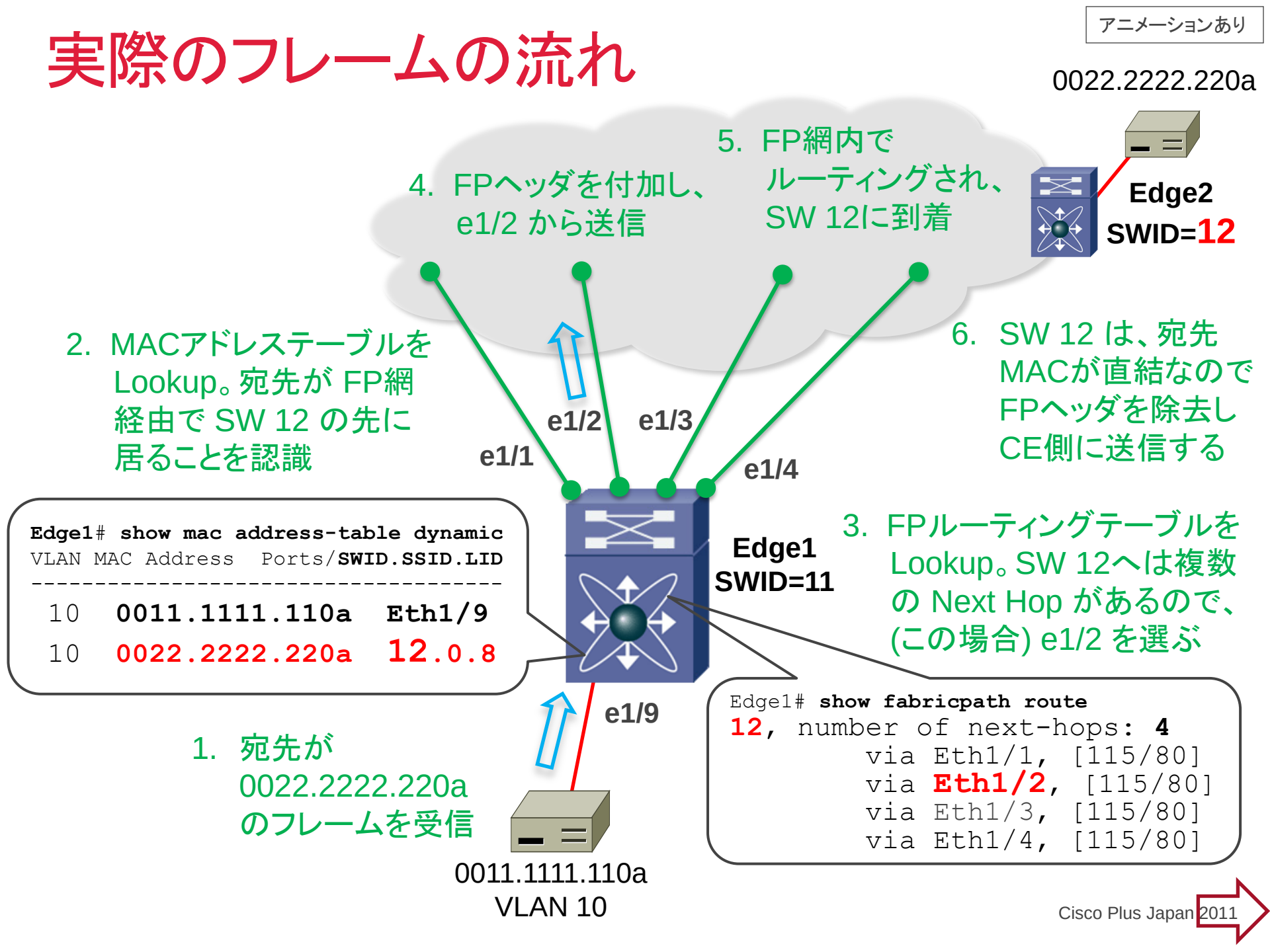



```
Edge1# show mac address-table dynamic
Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link
   VLAN     MAC Address      Type      age     Secure NTFY Ports/SWID.SSID.LID
---------+-----------------+--------+---------+------+----+------------------
* 10        0011.1111.110a    dynamic   0          F    F  Eth1/9
  10        0022.2222.220a    dynamic   0          F    F  12.0.8
```

```
Edge2# show mac address-table dynamic
Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link
   VLAN     MAC Address      Type      age     Secure NTFY Ports/SWID.SSID.LID
---------+-----------------+--------+---------+------+----+------------------
  10        0011.1111.110a    dynamic   0          F    F  11.0.8
* 10        0022.2222.220a    dynamic   0          F    F  Eth2/9
```

アニメーションあり

# 実際のフレームの流れ




```
Edge1# show mac address-table dynamic
VLAN MAC Address  Ports/SWID.SSID.LID
---------------------------------------
  10   0011.1111.110a   Eth1/9
  10   0022.2222.220a   12.0.8
```

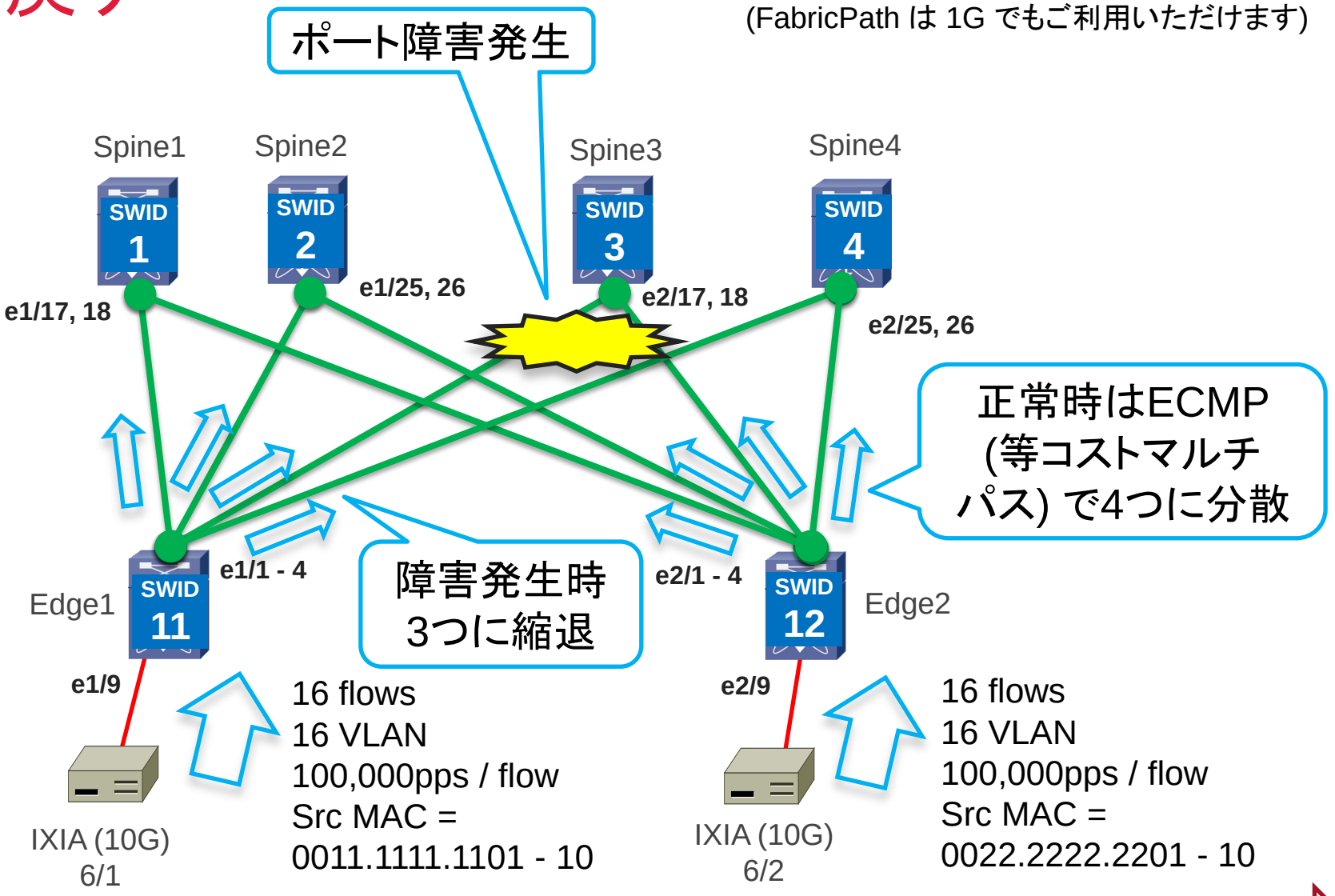
```
Edge1# show fabricpath route
12, number of next-hops: 4
        via Eth1/1, [115/80]
        via Eth1/2, [115/80]
        via Eth1/3, [115/80]
        via Eth1/4, [115/80]
```

# FabricPath デモンストレーション (解説 + ビデオ: 約10分)

アニメーションあり

# ECMP、ポート障害と切り戻り

10G (FP)
10G (CE)
(FabricPath は 1G でもご利用いただけます)



# デモビデオに切り替えます

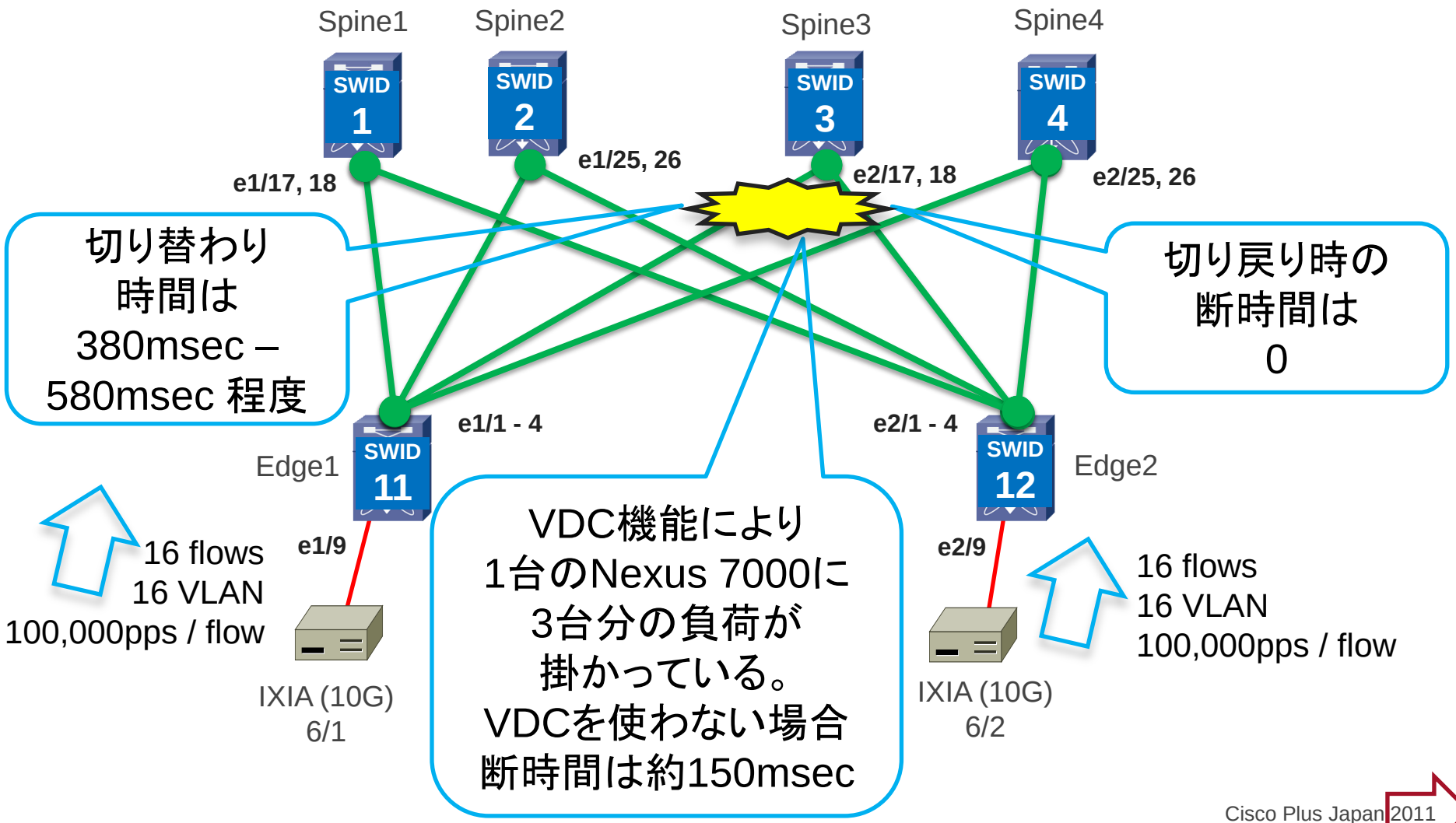
(お待ちください)

アニメーションあり

# ポート障害と切り戻り : 結果

10G (FP)
10G (CE)

(FabricPath は 1G でもご利用いただけます)



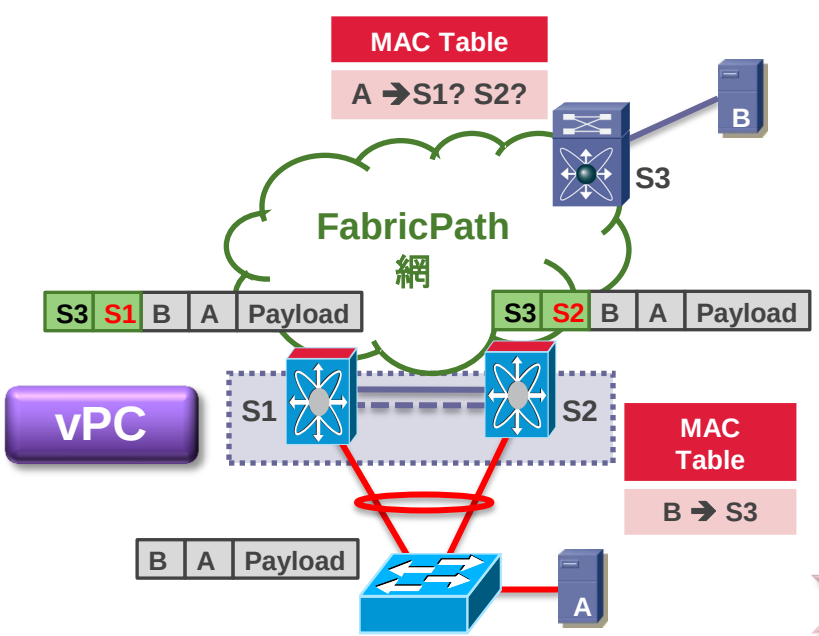
# Cisco FabricPath のアドバンテージと利用例

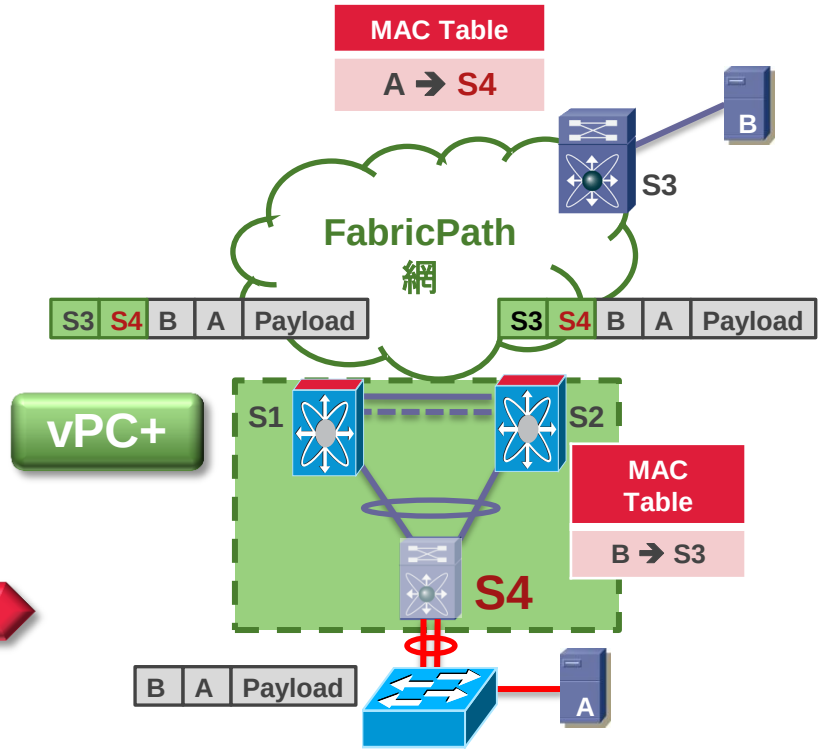
# 高信頼、高速性

- 最短 150ミリ秒での切り替わり
  - リンク障害 → ECMP Nexthop が減る場合
- リンク障害からの切り戻り、ノード増設、片寄せなどの際に、トラフィックへの影響が無い (あっても数十ミリ秒オーダー)
  - IS-IS Metric (OSPFのCostに相当)を変更して、IPルーティングと同様のルーティングテーブルの操作が可能
  - サーバ/アプリケーションへの影響が最小化できる
  - 最初は Spine x 2台で稼働させ、順次増設することも可能
- SUP冗長化時の切り替わりの際は無停止 (Nexus 7000)
- MACアドレス学習、更新(移動) は全てハードウェアベース
  ソフトウェアによるスイッチ間 MACアドレス同期が不要

アニメーションあり

# vPC+：FabricPath非対応機器との アクティブ－アクティブ接続



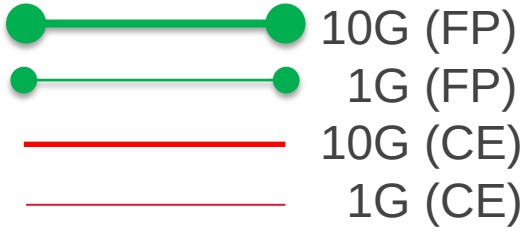

## FabricPath網とCEの境界(Edge)スイッチとその配下のFP非対応スイッチのための機能

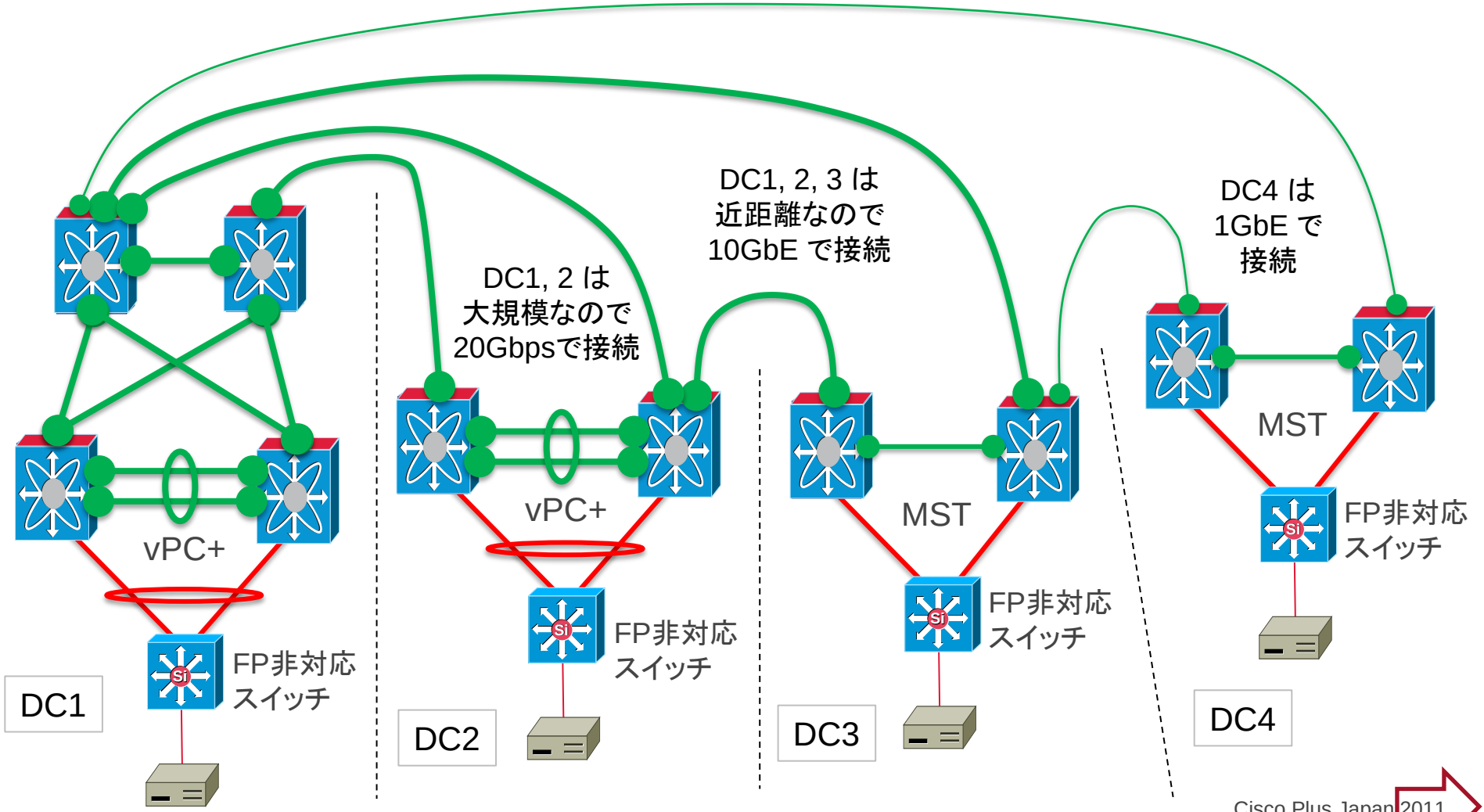
- CE網からFabricPath非対応機器(スイッチ)経由でFabricPath網へActive-Active接続するには、vPCが必要になる
- しかし、MACテーブルではMACとSwitchIDは1対1にしかマッピングできない (図のS3 からはAが S1とS2の両方の配下に見えてしまう)

- それぞれのvPCドメイン(図のS1とS2)はFabricPath網に対して仮想的なFabricPathスイッチ(Emulated Switch、図のS4)として見える
- Emulated SwitchのSwitchID が、FabricPathカプセル化時のソースアドレスとして用いられる。

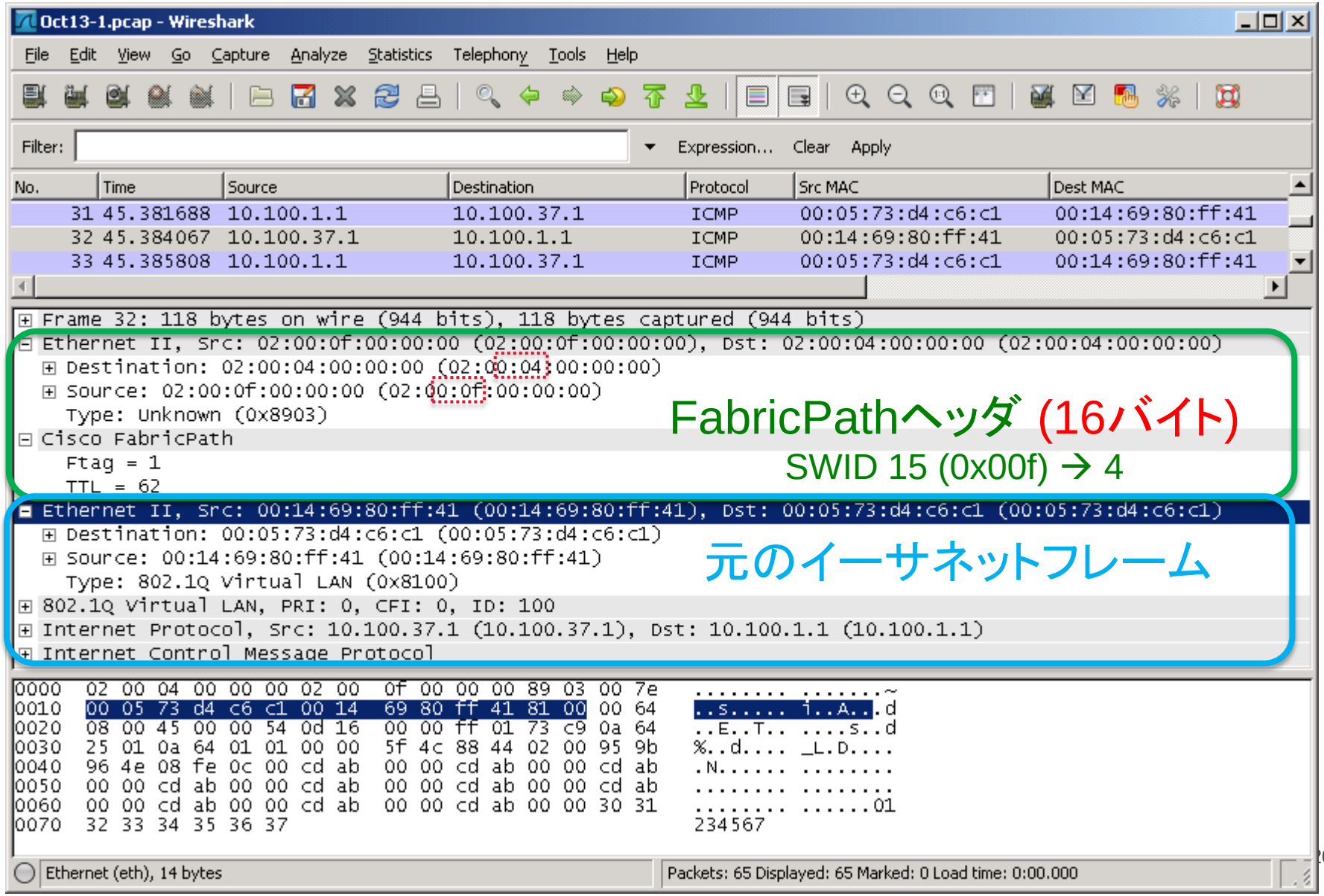
アニメーションあり

# FabricPath によるデータセンタ間レイヤ2 接続 (DCI) 例



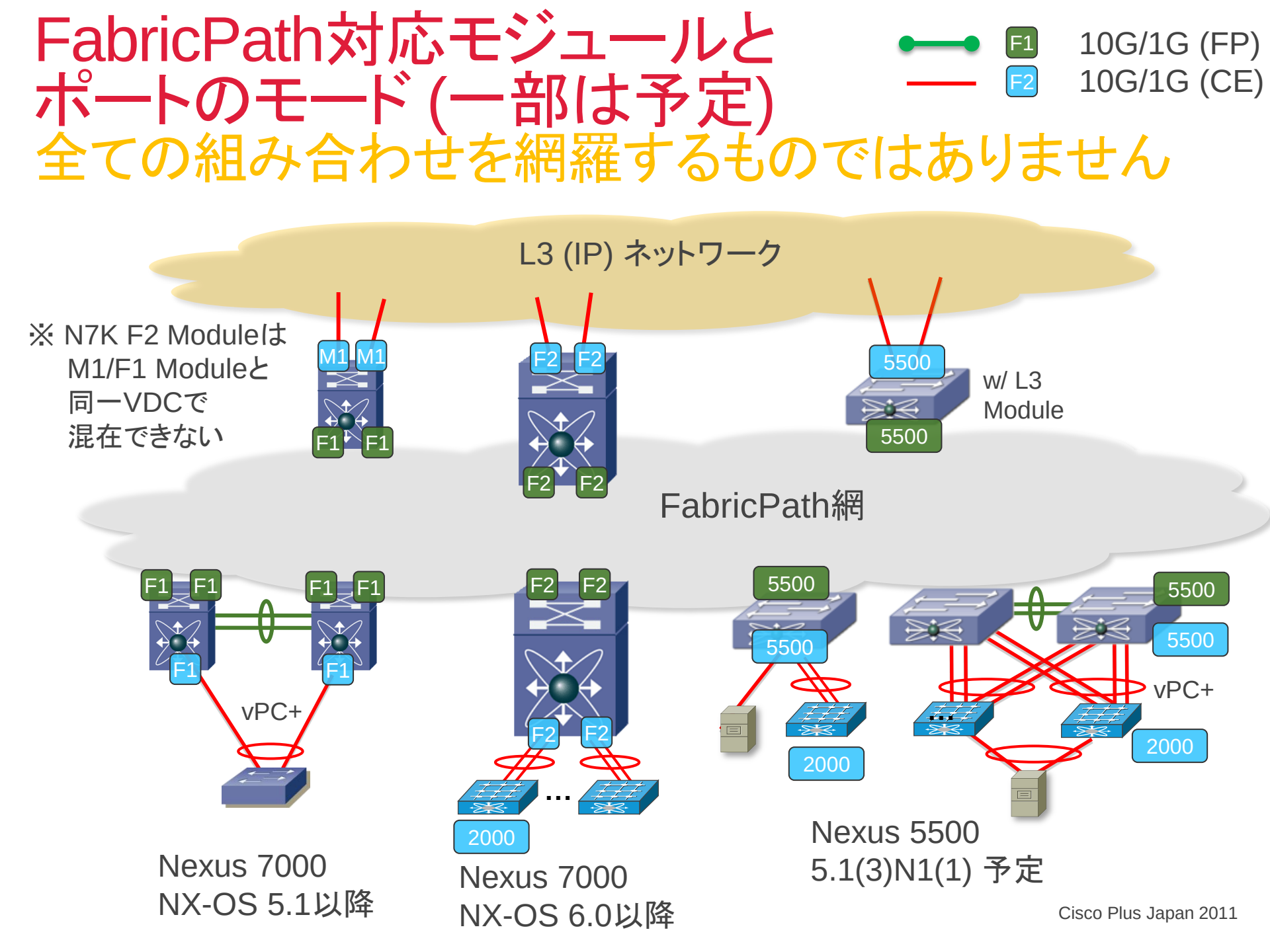



# FabricPathフレーム (Nexus 5500 β版)



# FabricPath対応モジュールとポートのモード (一部は予定)

## 全ての組み合わせを網羅するものではありません




Nexus 7000
NX-OS 5.1以降

Nexus 7000
NX-OS 6.0以降

Nexus 5500
5.1(3)N1(1) 予定

# まとめ: 各コンポーネントが「ゆるく」結合したスケーラブルなスイッチ/ファブリックの実現

- レイヤ2 でルーティングするので、トポロジの制約がない
  フルメッシュからリングまで、自由に構成可能。10G / 1G も混在できる
- 自律分散制御なので、単一障害点がない
  一つの制御ノードに過負荷が掛かることもない
- FabricPathスイッチ(ノード)の逐次増設 → スケールアウトが容易
  接続、config投入だけで新しいスイッチが原則無停止で網に組み込まれる
- STPとの相互運用性
  STP側からは、FabricPath網は1台の大きな Root Bridgeに見える
- レイヤ3ネットワークとシームレスに接続
  HSRP, VRRP, GLBP (N7Kのみ) と連携。IPv4マルチキャストにも対応
- vPC+ により、FabricPath非対応機器とアクティブ – アクティブ接続
- ワイヤレートでの MAC注入、移動にも耐えられる
  ソフトウェアベースでは不可能

# P.S.

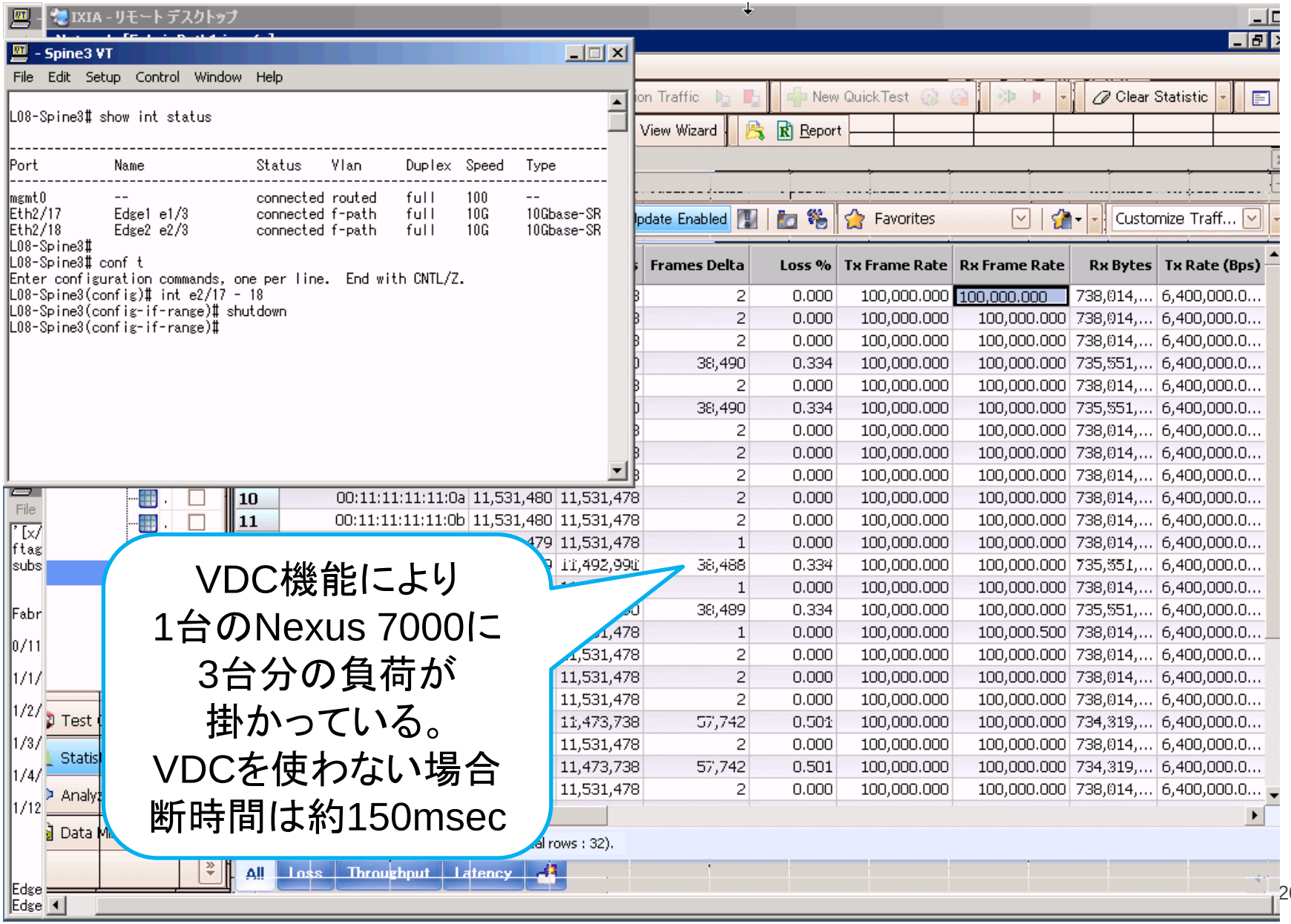
- FabricPath の実機デモンストレーション(リモート)を、このフロアの展示エリア “A” Booth にて実施しております。
- FabricPath に関するご質問、ご要望がおありでしたら、上記 Booth にお越しいただくか、アンケートにご記入ください。


CISCO

# 以降は参考スライドです

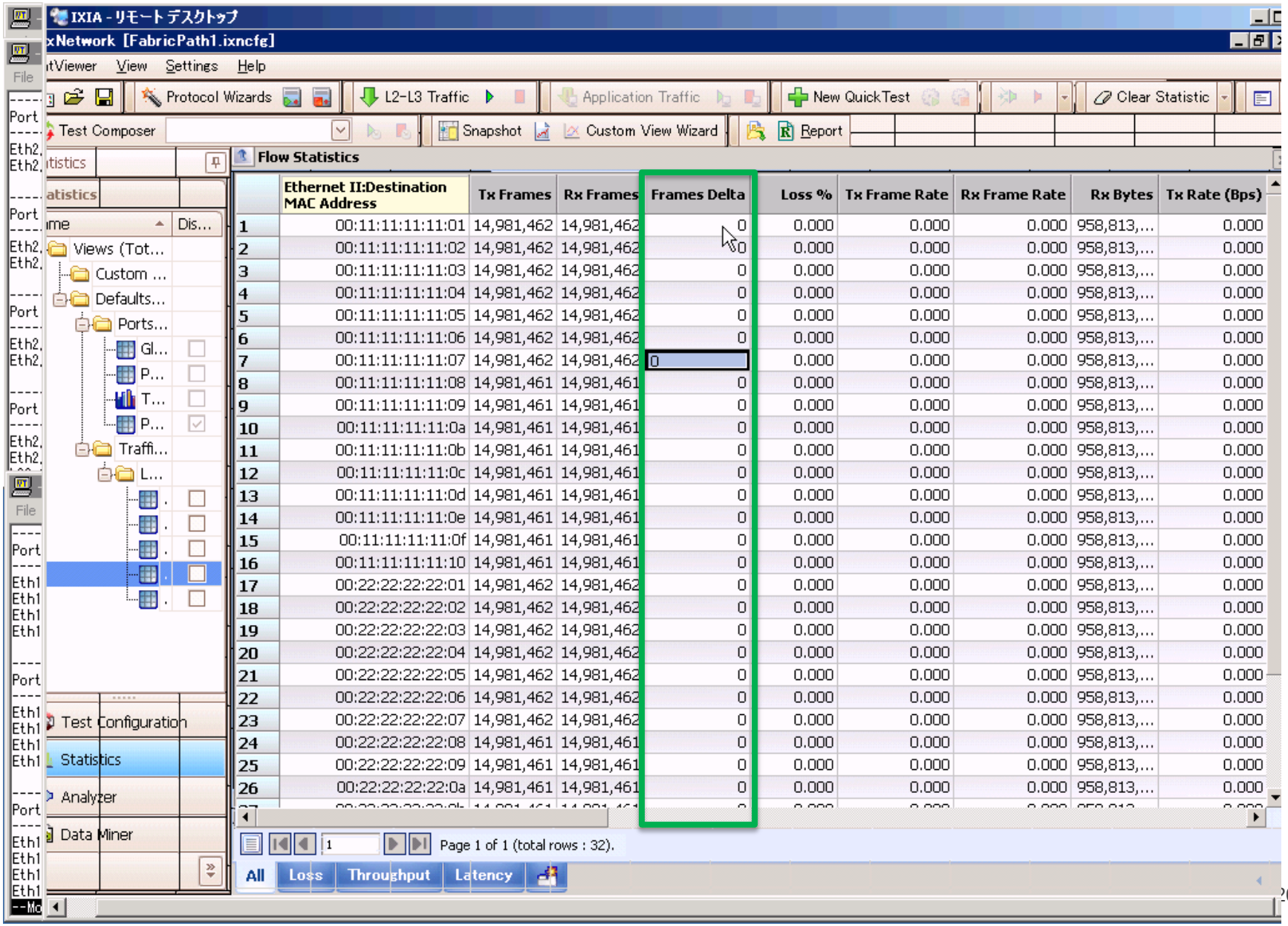
[デモ画面のスナップショット]

# 切り替わり直後 (断時間 = 約380 - 580msec)



[デモ画面のスナップショット]

# 切り戻り (断時間 = 0)




| | Ethernet II:Destination MAC Address | Tx Frames | Rx Frames | Frames Delta | Loss % | Tx Frame Rate | Rx Frame Rate | Rx Bytes | Tx Rate (Bps) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 00:11:11:11:11:01 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 2 | 00:11:11:11:11:02 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 3 | 00:11:11:11:11:03 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 4 | 00:11:11:11:11:04 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 5 | 00:11:11:11:11:05 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 6 | 00:11:11:11:11:06 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 7 | 00:11:11:11:11:07 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 8 | 00:11:11:11:11:08 | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 9 | 00:11:11:11:11:09 | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 10 | 00:11:11:11:11:0a | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 11 | 00:11:11:11:11:0b | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 12 | 00:11:11:11:11:0c | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 13 | 00:11:11:11:11:0d | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 14 | 00:11:11:11:11:0e | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 15 | 00:11:11:11:11:0f | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 16 | 00:11:11:11:11:10 | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 17 | 00:22:22:22:22:01 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 18 | 00:22:22:22:22:02 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 19 | 00:22:22:22:22:03 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 20 | 00:22:22:22:22:04 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 21 | 00:22:22:22:22:05 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 22 | 00:22:22:22:22:06 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 23 | 00:22:22:22:22:07 | 14,981,462 | 14,981,462 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 24 | 00:22:22:22:22:08 | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 25 | 00:22:22:22:22:09 | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |
| 26 | 00:22:22:22:22:0a | 14,981,461 | 14,981,461 | 0 | 0.000 | 0.000 | 0.000 | 958,813,... | 0.000 |

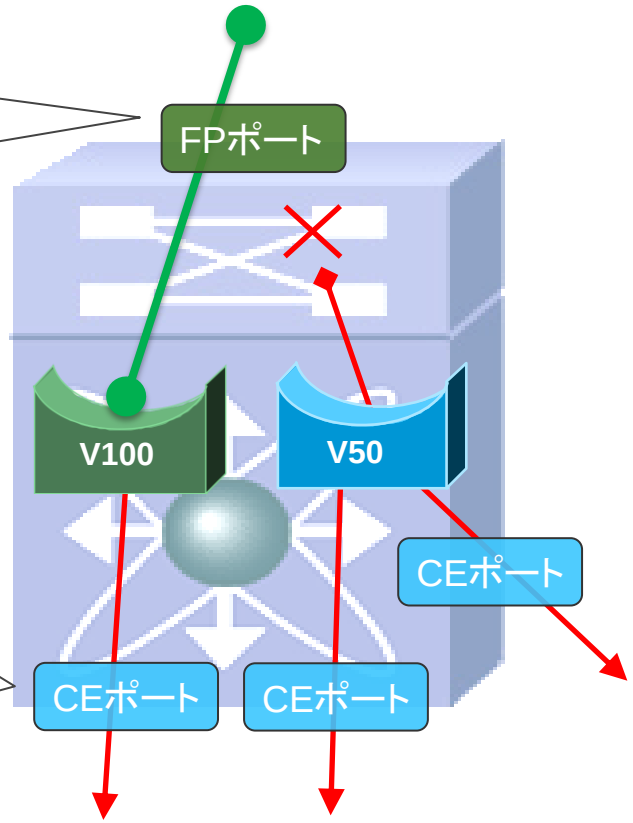
# FabricPath用語

- FP --- FabricPath の略
- CE (Classical Ethernet)
  従来のイーサネット。基本的にSTPが動作する
- SWID (Switch ID)
  FabricPathを喋る機器に一意に割り当てるID
- Spine (背骨)
  FabricPathのみを喋る機器。従来のイーサネット(CE)に接続しない
- Edge
  FabricPathと従来のイーサネット(CE)の境界にある機器
- ECMP (Equal Cost Multi-Path)
  ルーティング時にコスト(FabricPath / IS-IS では「メトリック」)の同じ複数の Next Hopがある場合、それらに分散してトラフィックを流すこと
- vPC+ (virtual Port Channel Plus)
  FabricPath Edgeで vPC (匡体跨ぎEtherChannel / LAG)を組み、FabricPath非対応機器をActive-Active接続できる機能
- IS-IS (Intermediate System to Intermediate System)
  ルーティングプロトコルの一つ

# “FabricPath VLAN” モード

FPポート(switchport mode fabricpath) には、FPモードのVLANのみが通る (Trunkされる)

FPモードVLANを出せるCEポートは、Nexus 7000 F1, F2, F2+FEX あるいは Nexus 5500, 5500+FEX のいずれか





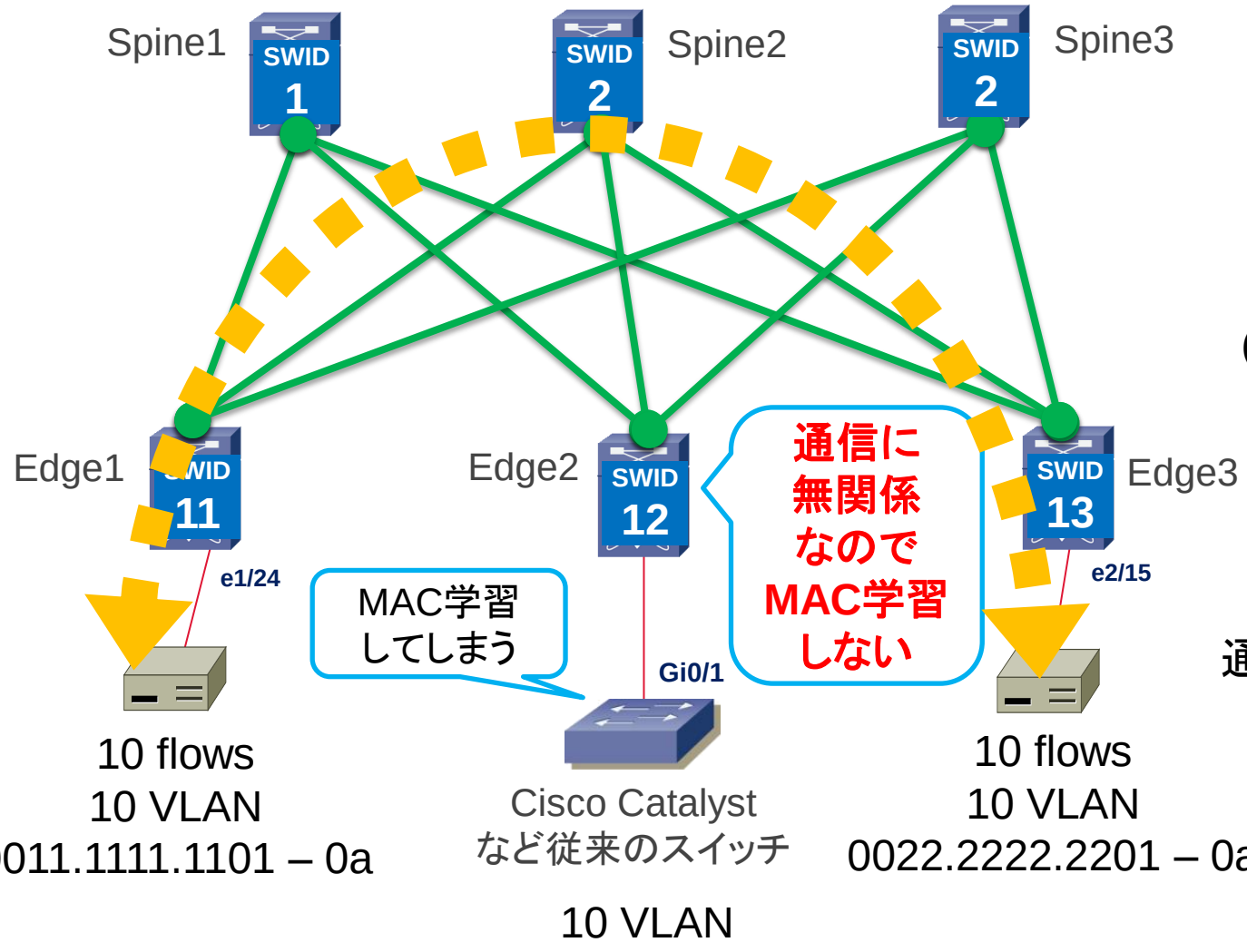
CEモードVLAN
(従来のブリッジ)


FabricPath
モードVLAN
(TRILLのRBridge
に相当)

```
Edge1(config)# vlan 100
Edge1(config-vlan)# mode ?
  ce          Classical Ethernet VLAN mode
  fabricpath  Fabricpath VLAN mode

Edge1(config-vlan)# mode fabricpath
Edge1(config-vlan)#
```

# “Conversational” MAC学習
## 最初から必要なMACだけを学習する機能




FP網の「向こう側(リモート)」のMACアドレスを学習する際に、双方向の通信(Conversation) がないと学習しない仕組み

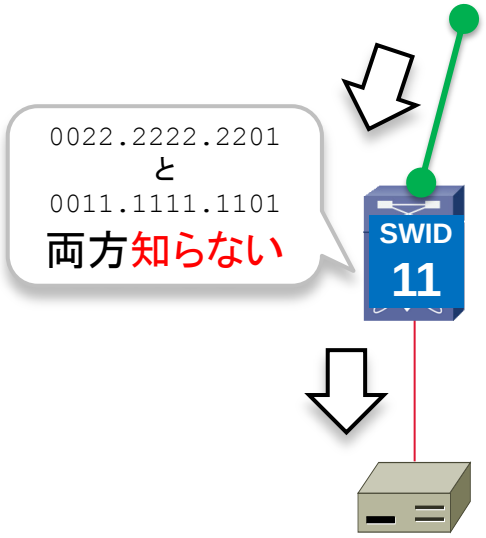
宛先不明でフラッドしても、通信に関係なければ学習せず、MACテーブルが溢れない

# Conversational Learning の実際の動き

## FP網からCE側へフレームが届いた時どうするか

### 1. 初期状態 (MAC学習する前)

0022.2222.2201 @ SW13
→ 0011.1111.1101 @ All SW
または
→ ffff.ffff.ffff @ All SW

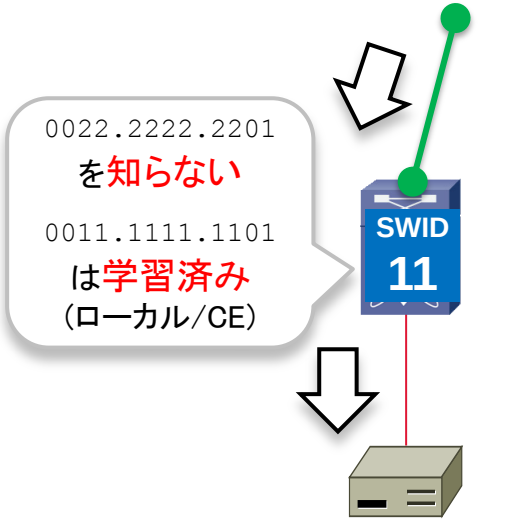



0011.1111.1101

**学習しない**
でもフラッドはする

### 2. CE側のみ学習済みの状態 → FP側を学習

0022.2222.2201 @ SW13
→ 0011.1111.1101 @ SW11

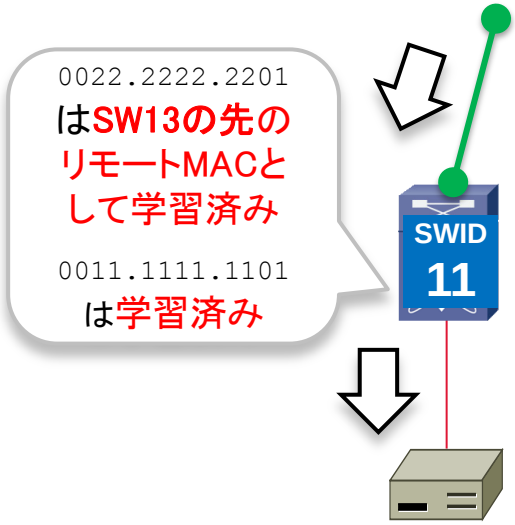



0011.1111.1101

**FP網の向こう(リモート)のMACを学習する**
(0022.2222.2201 @ SW13)

### 3. FP側のMACがSW13から12に移動

0022.2222.2201 @ **SW12**
→ 0011.1111.1101 @ SW11
または
→ ffff.ffff.ffff @ All SW



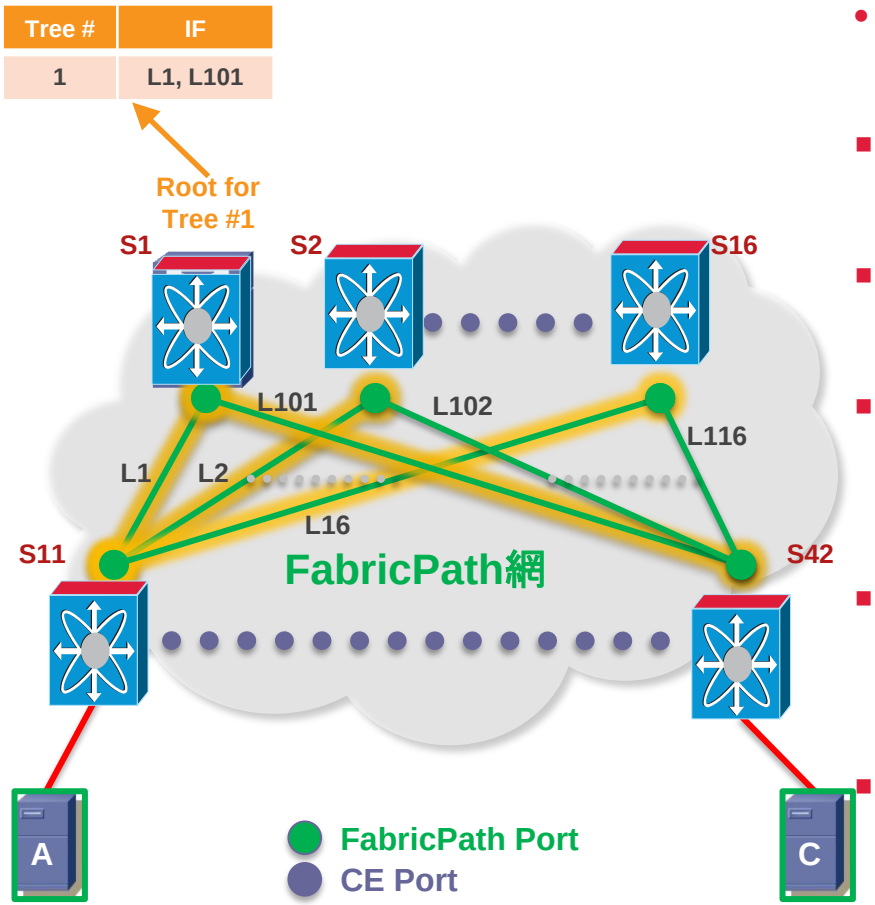

0011.1111.1101

**リモートMACの宛先を更新する**
(0022.2222.2201 @ SW13 → SW12)

アニメーションあり

# FabricPath の「ツリー」

## FabricPath網内での “L2 multi-destination traffic” (Unknown Unicast, Broadcast, and Multicast) の転送に用いられる




- 複数の宛先を持つトラフィックを正しく転送するには、「ツリー」トポロジーが必要になる
- MAC学習できておらず宛先不明のフレームにも用いられる
- 一つ以上の “Root” デバイスが、FabricPath網内で最初に選出される
- それぞれの “Root” から伸びる「ツリー」が形成され、ネットワーク全体で一意なIDが付加される
- 複数の「ツリー」をサポートすることで、FabricPathは複数の宛先を持つトラフィックでもマルチパスを可能にする
- 入力側スイッチは、各トラフィックフローに対して「ツリー」を割り当てる

# ビデオファイルのダウンロードについて

1. http://cisco.webex.com/meet/kyamashi/ にアクセス
2. "Files" タブをクリック
3. "Cisco Plus" の左側の「+」をクリック
4. ファイル名が表示される
5. 右側の "Download" をクリック