nortelnetworks.com

Non Stop Routingの実装と課題

MPLS JAPAN 2004

ノーテルネットワークス株式会社
近藤 卓司
tkondo@nortelnetworks.com

はじめに

NORTEL NETWORKS
BUSINESS WITHOUT BOUNDARIES

- ルータのノードレベルのHigh Availability実現技術の一つとして…
- コントロールプレーンがリスタートする際の影響をいかにして最小化するか? …
- 2つのアプローチがある
  - いかにしてフォワーディング処理を継続するか？
    → Graceful Restart (Non Stop Forwarding)
  - いかにしてルーティング処理を(もちろんフォワーディング処理も)継続するか?
    → Non Stop Routing

アプローチ比較

NORTEL NETWORKS
BUSINESS WITHOUT BOUNDARIES

Graceful Restart (Non Stop Forwarding)

- フォワーディング処理継続
- ルータ/スイッチ的発想
- 隣接交渉型
- 標準およびドラフトに基づいた一般的な実装
- 割と簡単

Non Stop Routing

- ルーティング処理継続
- 交換機的発想
- 自己解決型
- 一部ベンダ(ノーテルなど)独自の実装
- かなり難しい

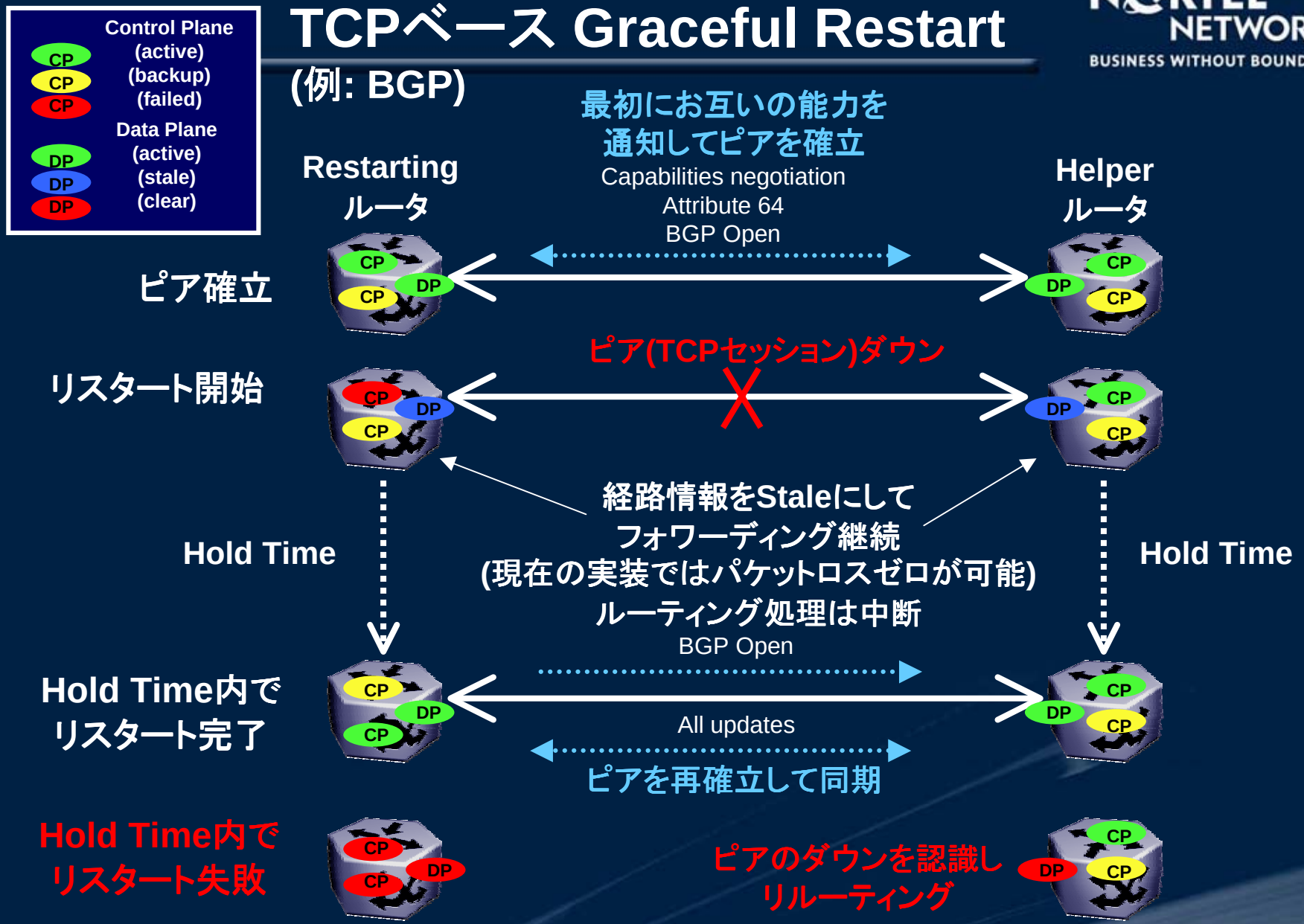
Graceful Restart

NORTEL NETWORKS
BUSINESS WITHOUT BOUNDARIES

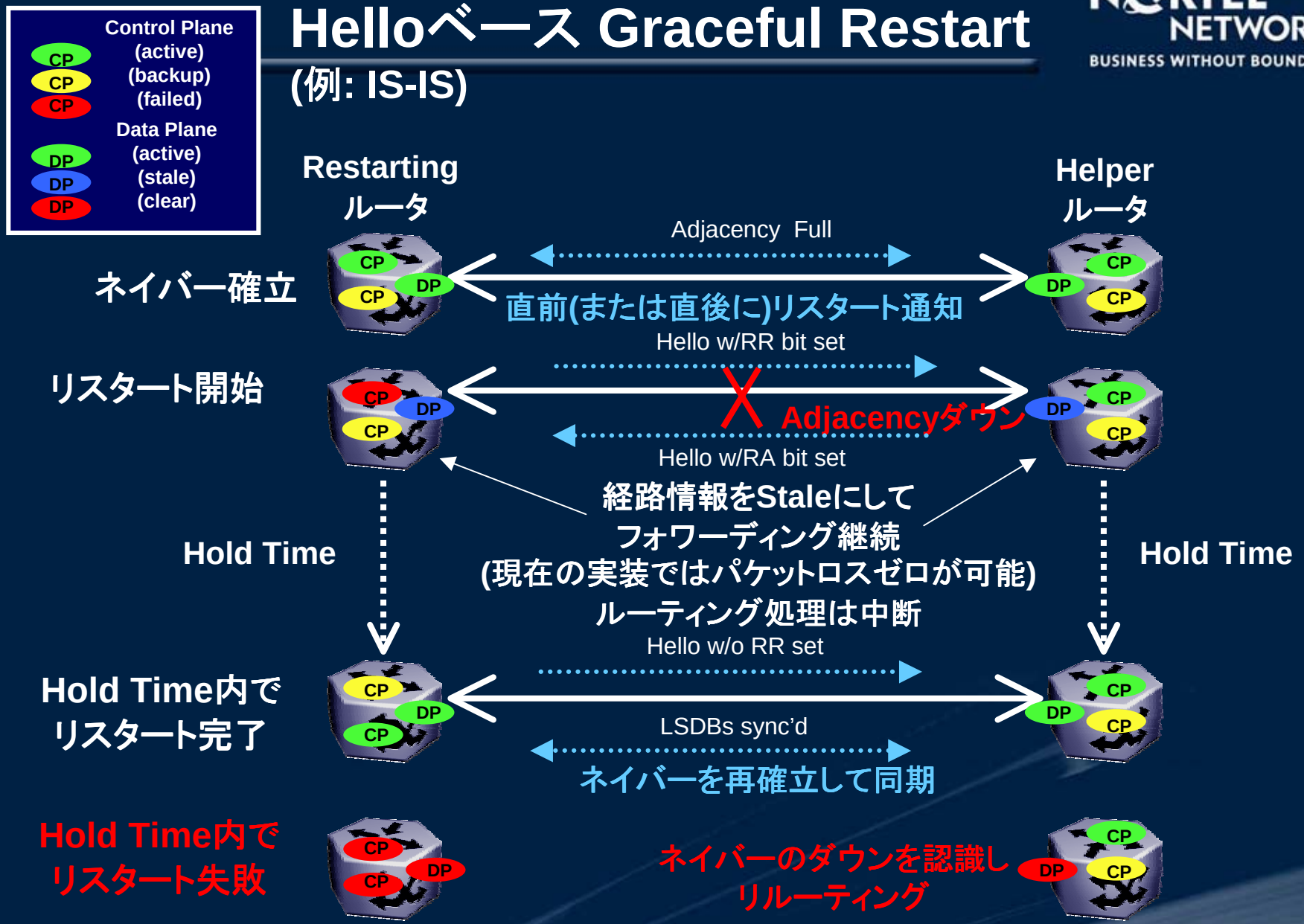
- 基本動作
  - 隣接ルータに事前にリスタートすると言う通知を出してリスタート
  - リスタートルータ(Localルータ)とそれを助ける隣接ルータ(Helperルータ)はその間のルーティング処理は行わず、直前のFIB情報に基づいてStale (古い、Freshでない)フォワーディングを継続
  - リスタート完了後にルーティング処理を再開し同期
- Graceful Restartに対応するためにプロトコルごとの拡張が必要
  - 現時点での対応はBGP、IS-IS、OSPF、LDP、REVP-TE
- プロトコルごとに動作が異なる
  - 大きく分けると…
  - TCPベースは事前通知型、Helloベースは直前(または直後)通知型

TCPベース Graceful Restart

(例: BGP)



Helloベース Graceful Restart
(例: IS-IS)



※RR: Restart Request、RA: Restart ACK

Graceful Restartの課題

NORTEL NETWORKS
BUSINESS WITHOUT BOUNDARIES

- Non Stop Forwardingと言うと聞こえは良いが、実際には長い時間 Headlessフォワーディングをする
  - リスタート中にルーティングループやブラックホールを起こす可能性がある
- タイマー値の設計が要求される
  - HoldタイマーはノーテルのDefaultではBGP 360秒、IS-IS 90秒、OSPF 30秒、LDP 180秒、RSVP-TE 30秒
  - 最適なタイマー値はネットワークのトポロジや規模にも依存する
  - 短すぎるとリスタートできない間にStaleフォワーディングを中断してしまい意味がなくなり、長すぎるとStaleフォワーディング状態が継続し危険が多い
  - 実はコントロールプレーンが冗長化されてなくても使えるが、リスタート時間がより長くなるケースが多いのでタイマー値には注意が必要
  - オペレーションに負担がかかる

Graceful Restartの課題

NORTEL NETWORKS
BUSINESS WITHOUT BOUNDARIES

- Graceful Restart動作はピアやネイバーに完全に依存
  - 隣接ルータにGraceful Restart Helperモード サポートが必須
  - 異なるベンダ間の場合は相互接続性の問題も発生
  - IP-VPNの場合、ユーザのCEルータに実装と相互接続性を要求することになる
- 基本的にはUnplanned リスタートには向いていない
  - RFC3623 “Graceful OSPF Restart”にも記述があるようにUnplannedにも使えるが推奨しない
  - オペレータ側に準備ができているPlannedと異なり、コントロールプレーンの故障(障害)や予期せぬスイッチオーバなどのUnplannedのStaleフォワーディングには危険が多い

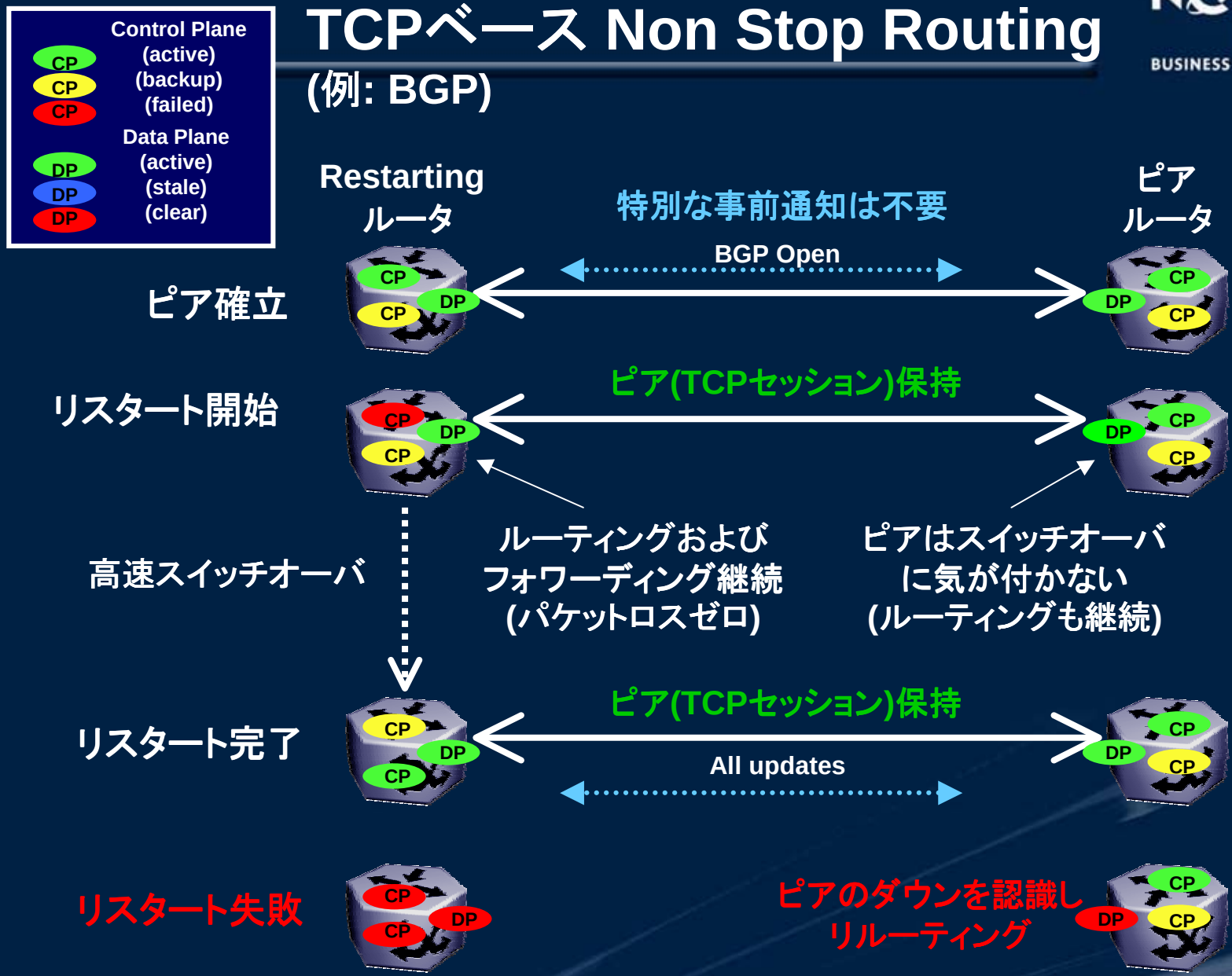
Non Stop Routing

NORTEL NETWORKS
BUSINESS WITHOUT BOUNDARIES

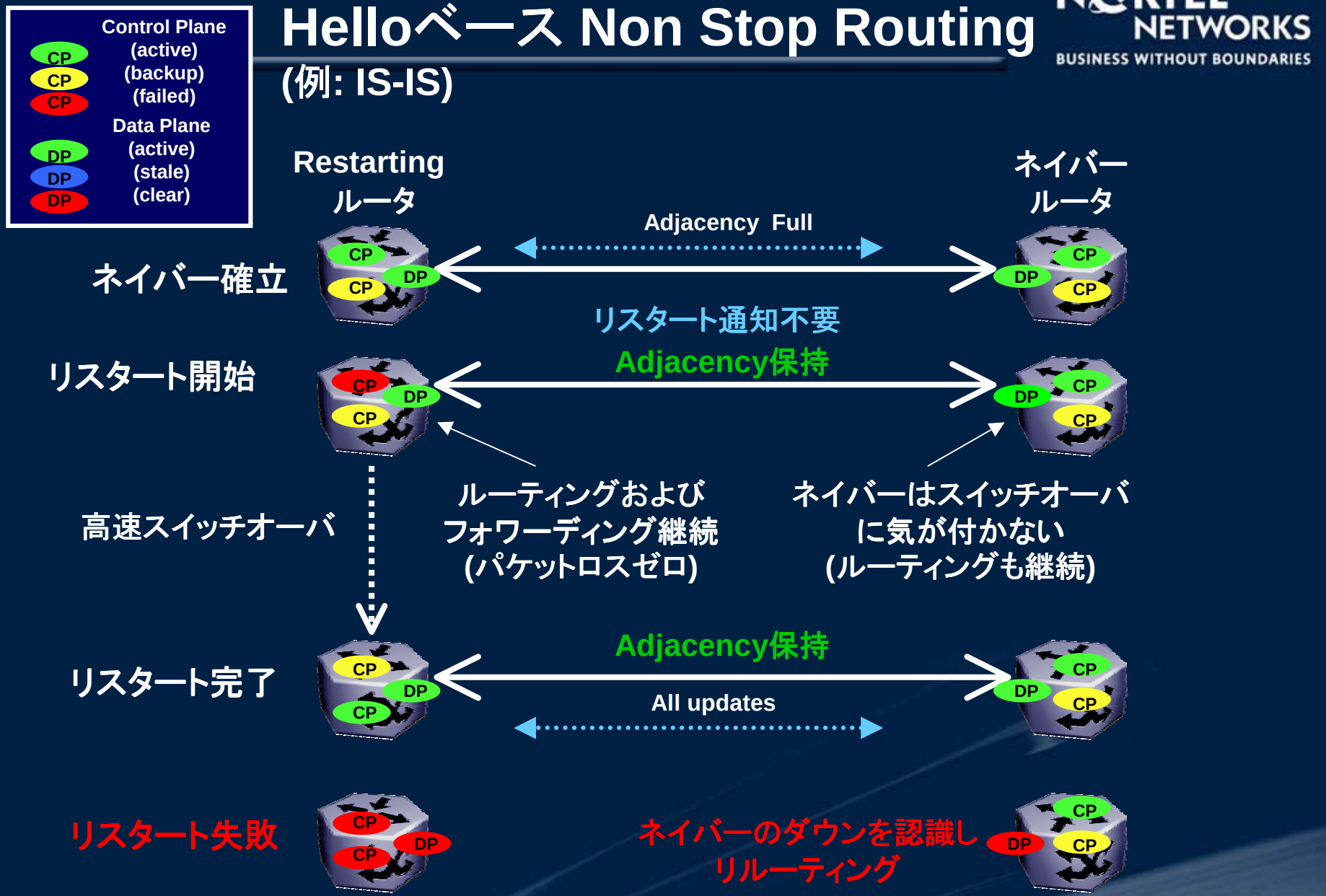
- 基本動作
  - ルータ内にActiveとBackupの2つの環境を作りその間を完全に同期化
  - Activeの環境に障害が発生するとすぐにBackup環境に切り替えてルーティング処理を継続
  - もちろんフォワーディング処理も継続
- プロトコルごとに基本動作に違いはない
- 自己解決型であり、ピアやネイバーに実装を必要としない
- タイマー値の設計が不要
- Unplannedリスタートにも向いている
- ただし、冗長化されたコントロールプレーンが前提

TCPベース Non Stop Routing

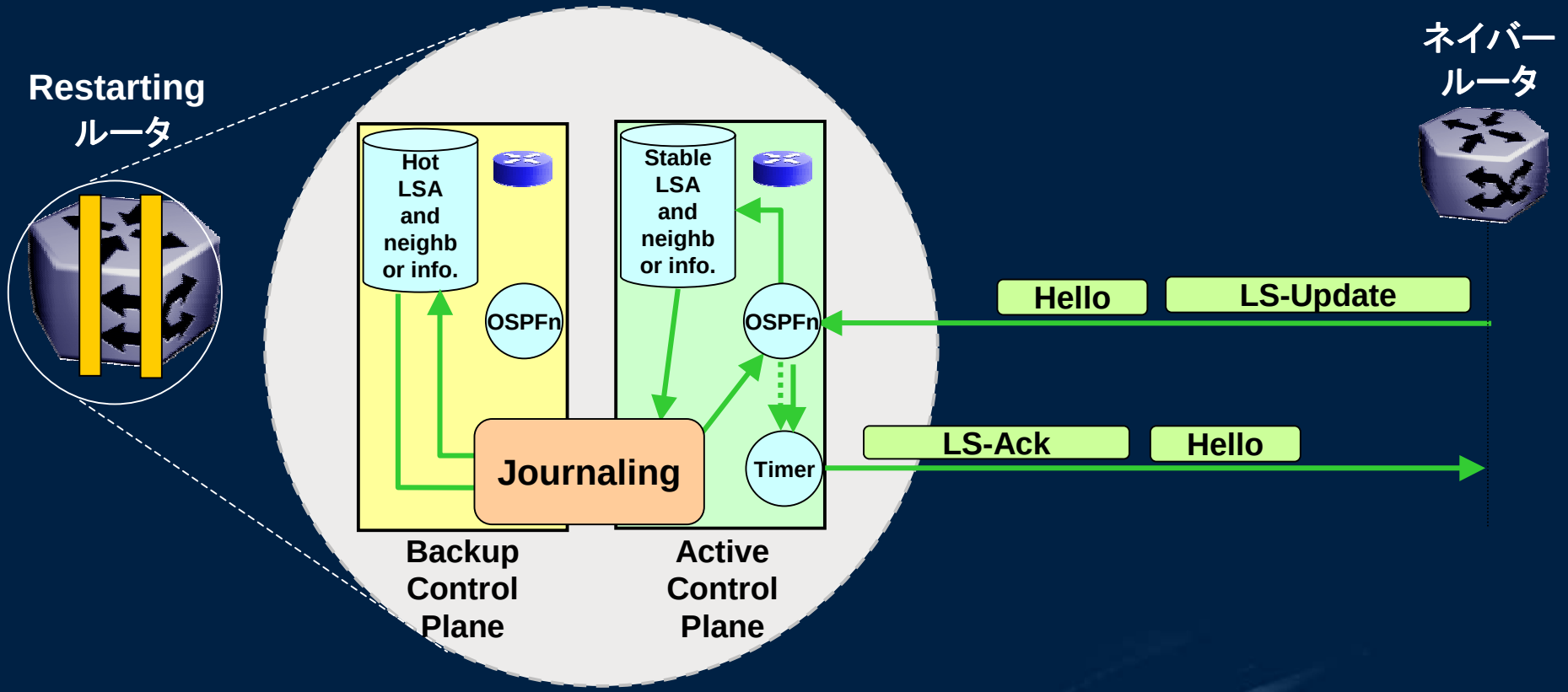
(例: BGP)



Helloベース Non Stop Routing
(例: IS-IS)

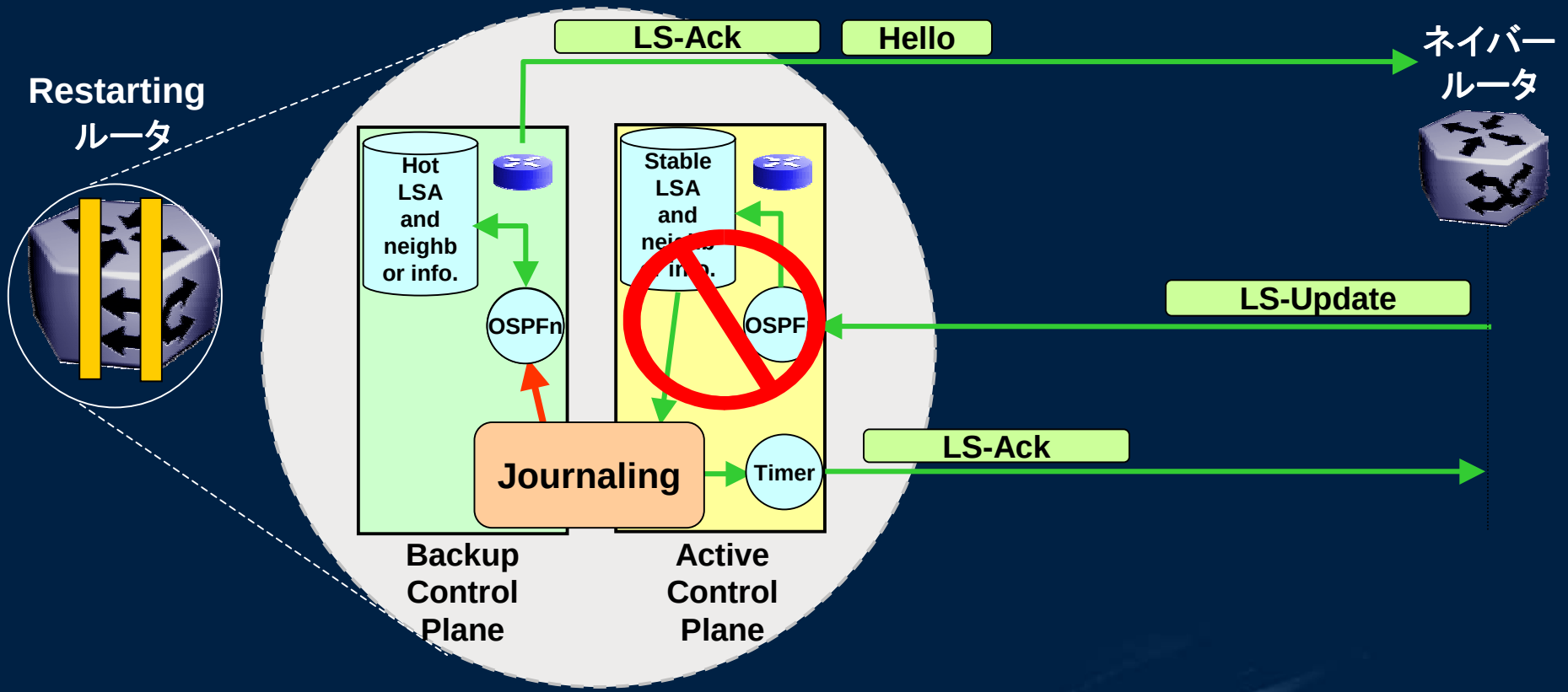


Non Stop Routing 内部動作
(例: OSPF)



- 通常時すべてのOSPF情報(ネイバー AdjacenciesやLSAデータベース)はJournalingによりActiveプロセスからBackupプロセスへ同期化
- Backup側はこのOSPF情報をインタフェース情報と合わせてショーテストパスの計算やRIBの生成に使用し、Active側と全く同じ状態を保持

Non Stop Routing 内部動作
(例: OSPF)



- PlannedおよびUnplannedリスタート時はActive側からBackup側へ高速切り替えを実施してルーティングとFreshなフォワーディングを継続
- ただし、現状の実装では障害検出は数msで可能だが、切り替え(主に障害解析や通知情報作成などの準備)に1~2秒程度かかる
(実際にはその間のルーティング処理は中断しStaleフォワーディング)

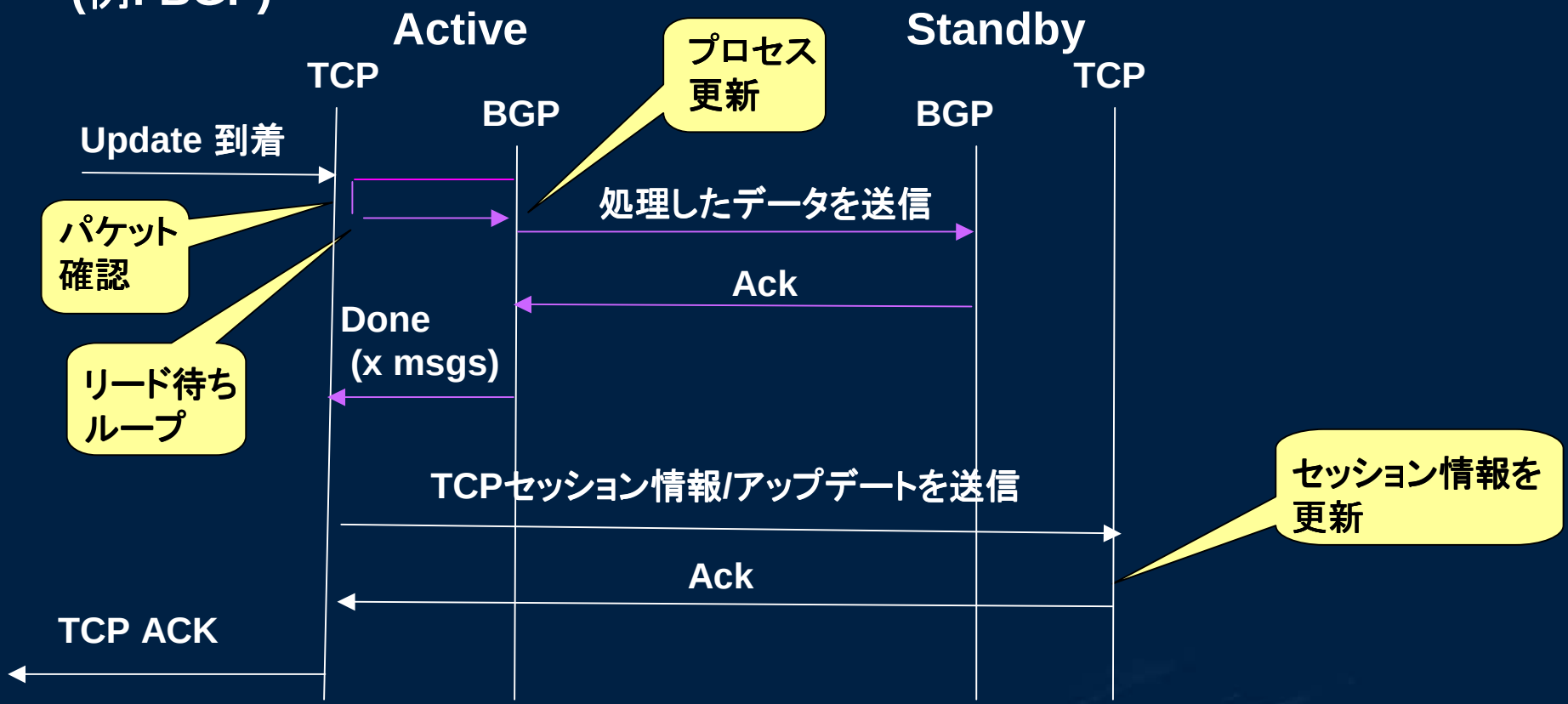
Non Stop Routingの課題

NORTEL NETWORKS
BUSINESS WITHOUT BOUNDARIES

- 実装には高度で複雑なHW/SW環境が必要
  - コントロールプレーン間のハイキャパシティ メッセージング環境やマルチCPUなどのハードウェアが必要
  - JournalingはOSのカーネルレベルで組み込まないと難しい
  - 既存のルータにプロトコル的に後付けで載せて動くような技術ではない
- TCPベースのプロトコルには現時点では拡張性に課題がある
  - 例えばBGPはTCPを用いるがアプリケーションレベルのACKがない
  - TCPセッション状態を同期するにはアプリケーションの同期の後にTCPの同期を取る必要がありTCP ACKに遅延が生じる

TCP ACKの遅延問題

(例: BGP)



- TCP ACKは以下が完了するまでの遅延が発生:
  - Active BGPがデータを処理
  - Standby BGPが処理されたデータを受信
  - Standby TCPがセッション情報を受信

Non Stop Routingの実装戦略

NORTEL NETWORKS
BUSINESS WITHOUT BOUNDARIES

Protocol	Graceful Restart	Non Stop Routing
OSPF	✓	✓
IS-IS	✓	✓
BGP	✓	Planned
LDP	✓	Planned
RSVP-TE	✓	✓

※BGPにはIP-VPNのためのMP-BGP、
LDPにはPWE3のためのE-LDPを含む

- 他社との接続のためGraceful RestartのHelperモードは必須実装
- vsではなく両方式を実装
- Non Stop Routingの実装に時間がかかるプロトコルはGraceful Restartでカバー
- 両方式の柔軟な設定を提供
  - 例えばIP-VPNなどではVRFごとに選択設定可能

Non Stop Routingの設定例

(例: OSPF)

```
routing {
…
    ospf {
            non-stop-routing | no-non-stop-routing;
            graceful-restart {
                    disable;
                    no-unplanned;
                    recovery-time <seconds>;
…
services {
    vrf {
            vrf-instance-name vrf1 {
                    …
                    ospf {
                            non-stop-routing | no-non-stop-routing;
                            graceful-restart {
                                    disable;
                                    no-unplanned;
                                    recovery-time <seconds>;
```

Non Stop Routingの状態表示例

(例: OSPF)

```
mpe> show routing carrier-grade
…
OSPF
   Service State: Active
    Non-Stop-Routing: Enabled
      Sparing State: In-Sync
      Spare Objects:   Count    Adds    Deletes    Updates
                LSAs:    2455    2469         14        154
          Interfaces:      14      18          4          2
         Adjacencies:      11      15          4          0
   Graceful-Restart: Enabled
          Local-Mode: Disabled
…
```

Non Stop Routingの今後

- TCP ACKの遅延問題の解決
  - 遅延を生じさせず同期を取る独自技術を開発中
- 切り替え時間の更なる短縮
  - ルーティング処理断ゼロ切り替えは難しいが、少なくとも50ms以内の切り替えが目標値
- Graceful Restartと同様にマルチキャストとIPv6プロトコルへも対応
- BRASアプリケーションへの応用
  - Session Manager (PPPoX、L2TP)、AAA (DHCP、Radius)などのNon Stop Routing (Hot Standby)