



MPLS Traffic Engineering

日本テレコム(株)
松嶋 聡
<satoru@ft.solteria.net>

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



Traffic Engineeringとは何か？

- 広義には...
 - ネットワーク上に流れるトラフィックを効率的に過不足なく処理するために行うネットワーク設計、作業全般。
 - 必要回線容量の見積もり。
 - 回線使用率、トラフィック量の監視。
 - 回線容量の増強、増速作業など。
 - つまりトラフィックの流れから見て、ネットワーク構成を最適化すること。
- 以後、TEと呼びます。

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

IPネットワークにおけるTE

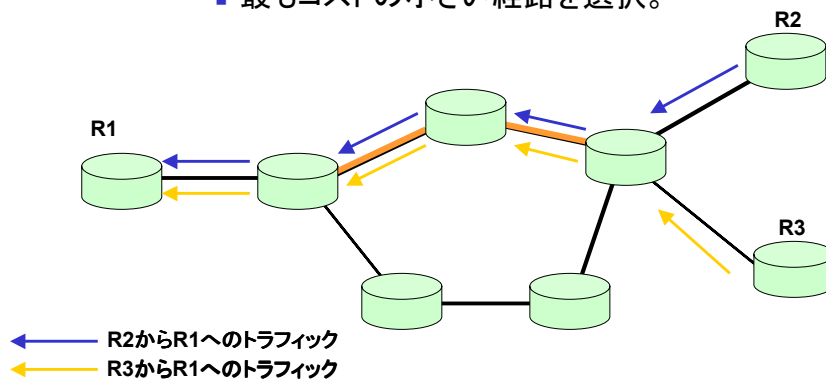
- みんなどんなTEをやっているだろう？
 - 例えば...
 - MRTGでトラフィック量監視。
 - 特定回線のピークトラフィックが50%を超えたあたりから増設計画を立てて、予算確保。
 - 70%~80%あたりで警戒域。急がねば。。。。
 - 100%になると、おそらくクレームが来る。
 - 回線数を増やす。
 - ロードバランスさせる。
 - トラフィックを別回線に移す。
 - IGP(RIPやOSPF)経路のメトリックを変える。
 - などなどが、現在までのソリューション。

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

例題: IPネットワークにおけるTE

- ルータR1へのトラフィックの流れ
 - 最もコストの小さい経路を選択。

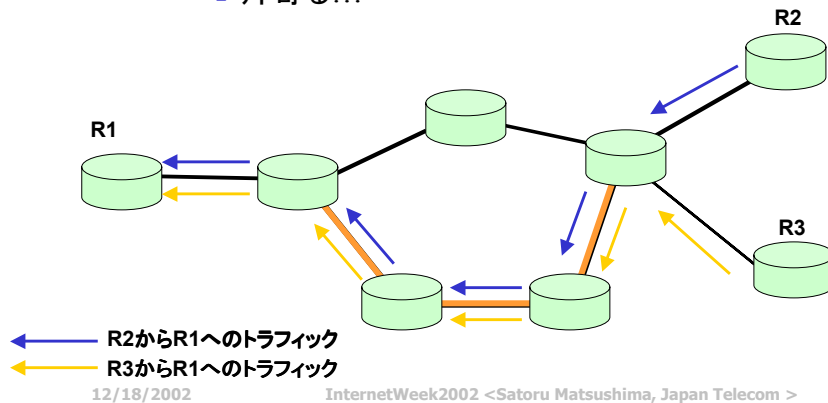


12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

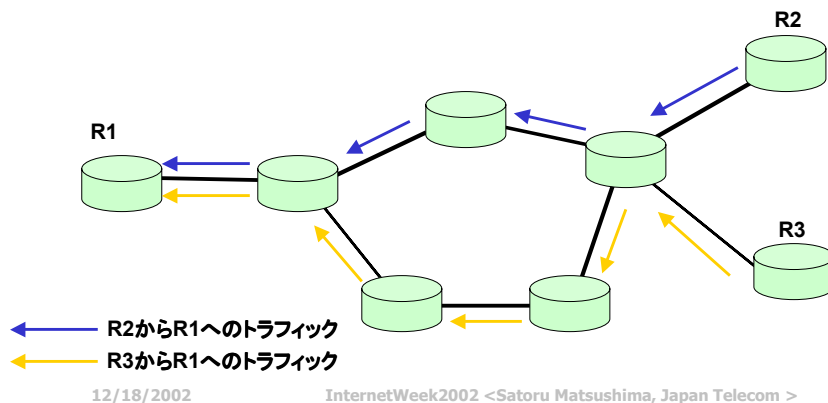
例題: IPネットワークにおけるTE

- リンクのコスト値を変えてみる。
 - 片寄る...



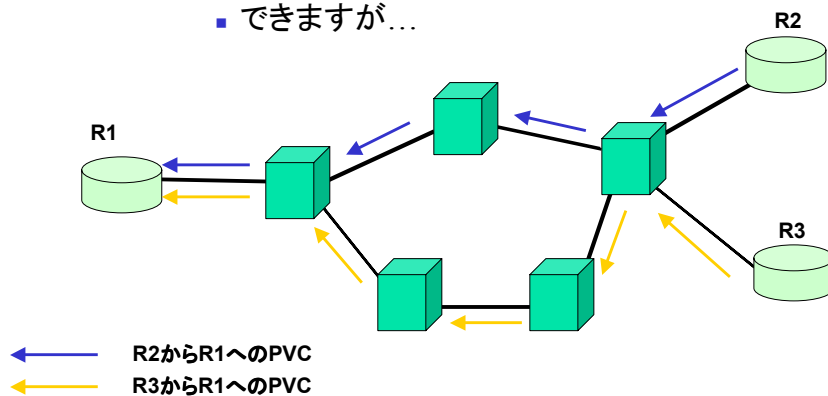
例題: IPネットワークにおけるTE

- 本当はこうしたかったのに！



例題: IPネットワークにおけるTE

- ならばATMやFrame-Relayで...
 - できますが...



12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

ATM/FRをIPTラフィック制御に 用いた時の問題点

- IPネットワークとは別にATM/FRネットワークを用意することになる。
 - ルータ以外のネットワークデバイス。
 - それぞれにおいて運用/監視が必要。
→ 高コスト
- IGPへの負荷が高くなる。
 - メッシュによるAdjacency増。
 - 複雑な論理トポロジー。
- 帯域を圧迫するオーバーヘッド
 - ATM Cell Tax.

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

IPネットワークにおけるトラフィック制御の問題点

- リンクのコストやメトリックによるIGPルーティングはトラフィックの流れから見ると、最適ではない。
 - 利用率が非常に高い回線と低い回線が存在。
 - ネットワーク資源を有効活用できず、しかも高利用率回線の増強を常に余儀なくされる。
→**高コスト!**
 - 利用率(トラフィック量)の高い回線が障害になった時のリスクは非常に大きい。
 - 通常そのような回線は高速/超高速回線。
 - そのような回線が切れたときのバックアップ容量が十分であることを保証するのは困難。
→**信頼性、可用性に欠けたネットワーク。**

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

IPネットワークにおけるトラフィック制御の問題点

- トラフィックの最適化が非常に困難。
 - IGPにトラフィックを最適化する経路変更は行えない。
 - トラフィックフローに対して直接オペレータが介入する手段はほぼ皆無。

**明示的にトラフィックの流れを最適化できる
新たなTE手法が必要**



MPLSによるTraffic Engineering !

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



MPLSによるTraffic Engineering

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



MPLS-TEのメカニズム

- Explicit (明示的) Routing
 - Constraint-based (強制的な)パス選択
 - IGPコストによらず、空いている／要求条件にマッチパスを検索し、選択する。
 - CSPF(Constrained SPF) or Operatorによる設定。
 - IGP(IS-ISやOSPF)を拡張して、ネットワーク資源(リンクの帯域予約情報)やポリシーを運ぶ。
 - 空いている／要求条件にマッチするパスを選択するために使われる。
 - MPLSによるパケットフォワーディング
 - IPアドレスによらないフォワーディングにより、Explicit Routingを実現。

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



Explicit Routing

- RSVP(RFC3209)

“RSVP-TE:Extensions to RSVP for LSP Tunnels”

によりLabel Switched Path(LSP)を確立。

- ネットワーク資源情報などにより、明示的なパスを設定する。
- その他にも
 - LSP tunnelの障害検知
 - LSP tunnelのre-routing管理

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



IGP(IS-IS or OSPF)の拡張

- MPLS-TEに対応したルータは以下の情報を拡張されたIS-IS/OSPFにより伝達する。

- 帯域情報(bandwidth)
 - どれだけの帯域が予約可能か。
- リンク属性(Link Attribute)
 - どのようなポリシーのパスで使用されるのか。
- TEメトリック(TE-specific link metric)
 - リンクの強さ。
- いつ伝達されるのか？
 - 周期的、または予約可能帯域が変化したときなど。

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



IGP(IS-IS or OSPF)の拡張

- OSPFへの拡張方法
 - Opaque LSA (Type10:Intra-area)
(draft-katz-yeung-ospf-traffic-09.txt)
- IS-ISへの拡張方法
 - New wide TLV (draft-ietf-isis-traffic-04.txt)
- なぜIS-ISやOSPFだけなのか？
 - RIPの立場は？
 - それはリンクステート型だから。
 - でないと、リンクに関わる情報を運べない。
 - IS-IS/OSPFでも同一エリア内のみ有効。
 - エリア内にしか全てのリンク情報はない。

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



MPLSによるフォーワーディング

- なぜMPLSでなければならないのか？
 - TEは明示的にトラフィックの流れるパスを指定できなければならない。
 - 宛先IPアドレスによる従来のフォーワーディングではホップバイホップフォーワーディングになってしまう。
→途中のルータが意図したパスにそってパケットを転送するかどうか分からない。
 - MPLSならば、IPアドレスによらないフォーワーディングを実現することができる。
 - 32bit固定長のラベルに基づくフォーワーディング。
 - ラベルにパスを示す値をつけることで、Explicit Routingに従ったフォーワーディングが実現できる。

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



MPLS-TEのセットアップ

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



MPLS-TEのセットアップ

- RSVP拡張を用いたLSP セットアップ
 - LSPをセットアップするルータ(head-end)から起動。
 - LSP宛先ルータ(tail-end)からラベルを割当てる。
 - Downstream On Demand Label Distribution (DoD)
 - LSPに対して帯域を予約する。
 - シグナリング上でのみ。実際に運ばれるトラフィックの帯域を保証しているわけではない。

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

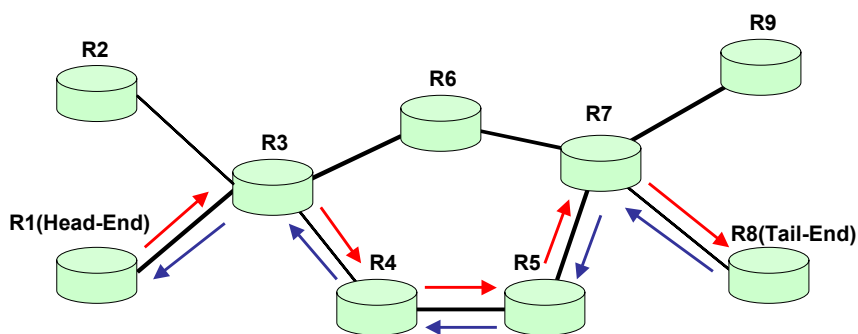
MPLS-TEのための新しいRSVP オブジェクト

- Explicit Route Object (ERO)
 - TEパスを設定するリンクを明示的に連ねる。(PATH)
- Record Route Object (RRO)
 - RSVPメッセージが通過したリンクを記録。(PATH,RESV)
- Label Request Object
 - ラベル割り当て要求(PATH)
- Label Object
 - ラベル割り当て通知(RESV)
- Session Attribute Object(仕様変更)
 - パスのSetup, Holdプライオリティ,
 - Local repairや予約スタイル(SE)などの要求 (PATH)

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

概要:MPLS-TE-LSPセットアップ



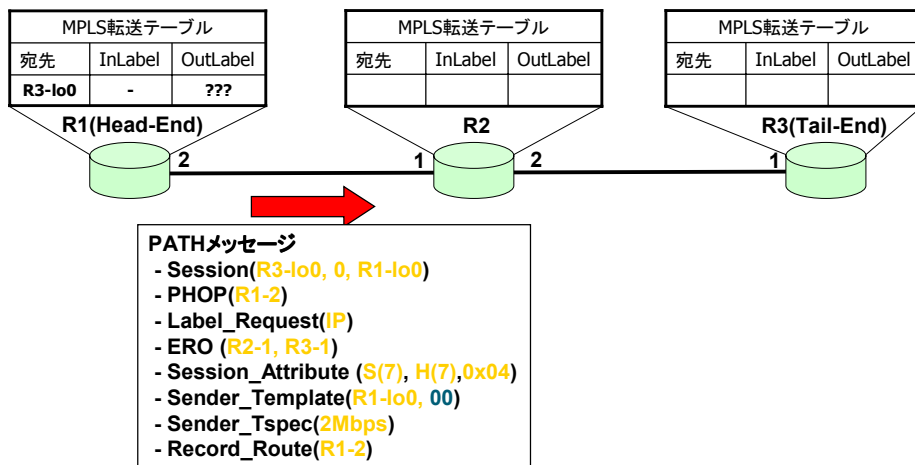
→ R1→R8へのPATHメッセージ (ERO=R1,R3,R4,R5,R7,R8)

← R8→R1へのRESVメッセージ (それぞれのリンクでRRO追加、Label通知)

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

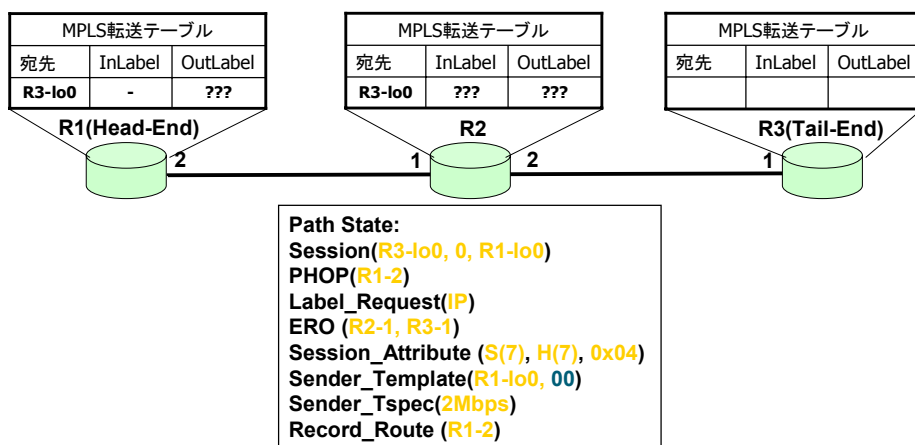
詳細:MPLS-TE-LSPセットアップ



12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

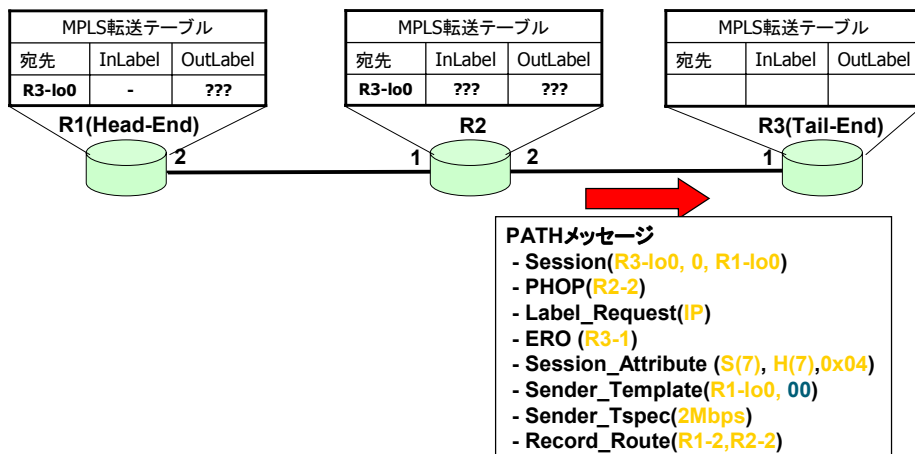
詳細:MPLS-TE-LSPセットアップ



12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

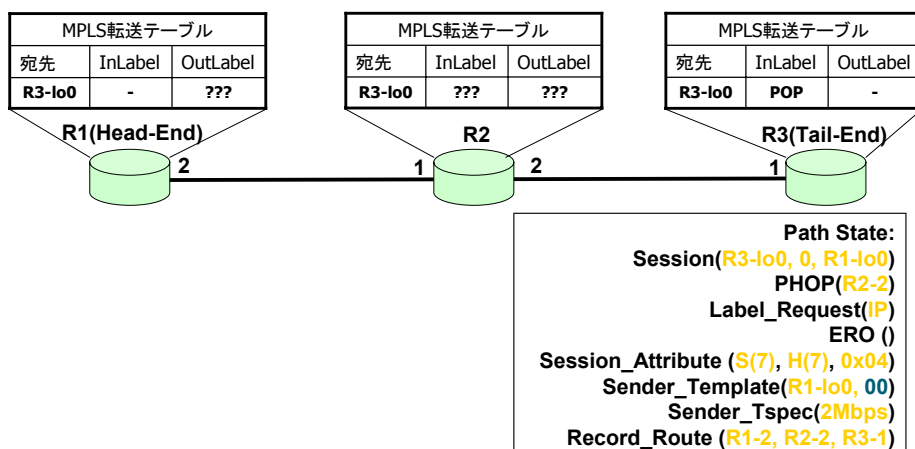
詳細:MPLS-TE-LSPセットアップ



12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

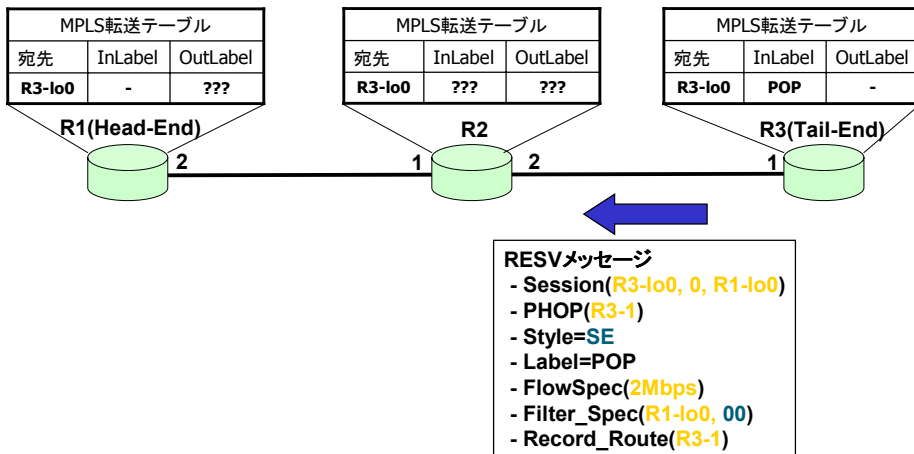
詳細:MPLS-TE-LSPセットアップ



12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

詳細:MPLS-TE-LSPセットアップ

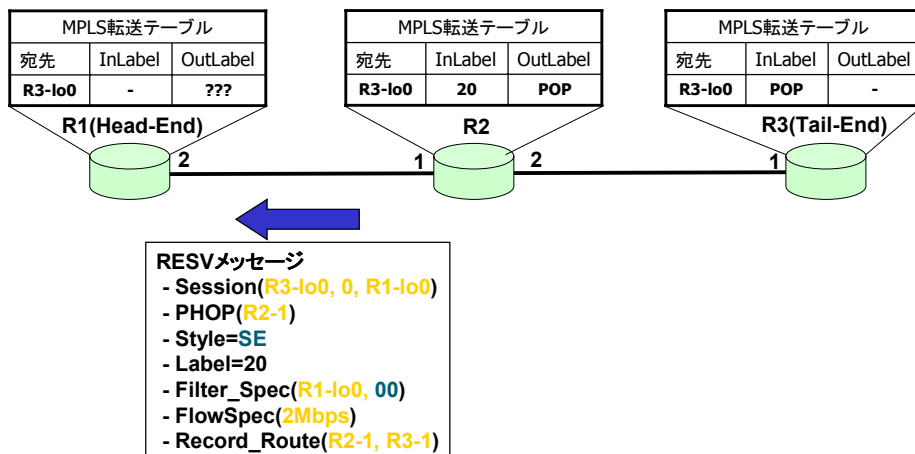


12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



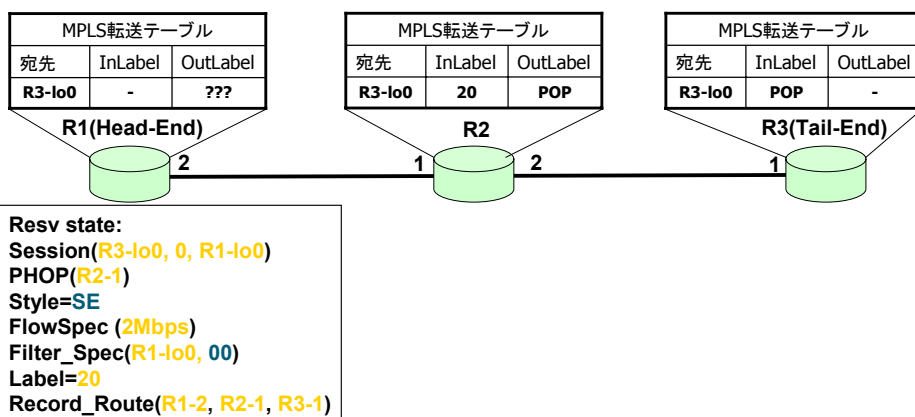
詳細:MPLS-TE-LSPセットアップ



12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

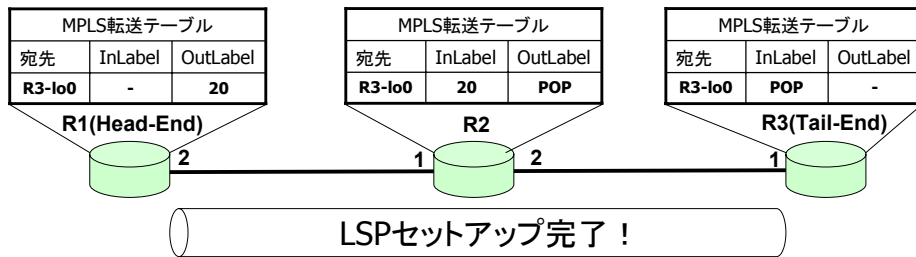
詳細:MPLS-TE-LSPセットアップ



12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

詳細:MPLS-TE-LSPセットアップ



12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

MPLS-TEのメリット

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



MPLS-TEのメリットとは？

- IPネットワーク上でトラフィックを最適化することは今まで述べたとおり。
それ以外には？
 - リンク障害時の負荷分散
 - 高速な障害回復メカニズム(Fast Reroute)の適用が可能
 - 特定POP間のトラフィック管理
 - 負荷分散の調整
- などなど...

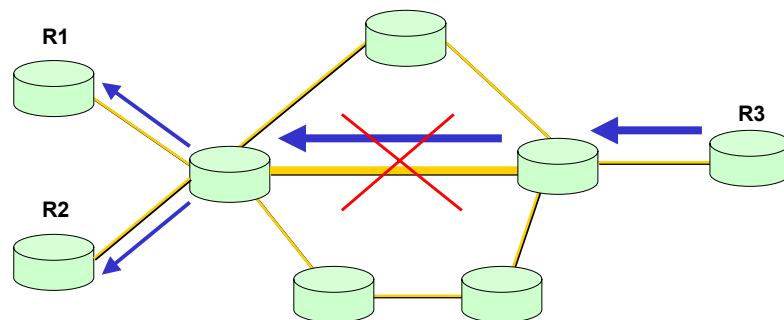
12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



リンク障害時の負荷分散

- 平常時は余裕で運べたトラフィックも...

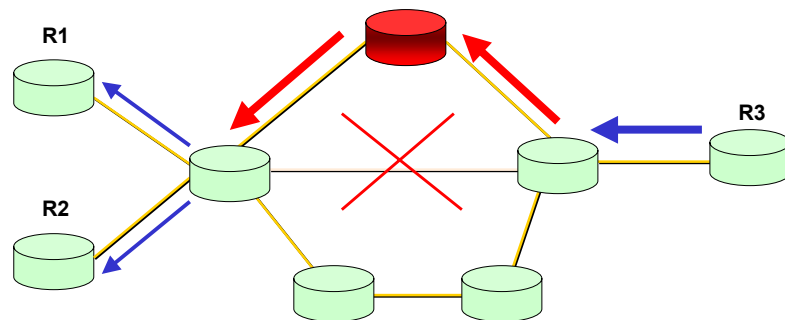


12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

リンク障害時の負荷分散

- コストの都合上、細いリンクに片寄られると苦しい...でも

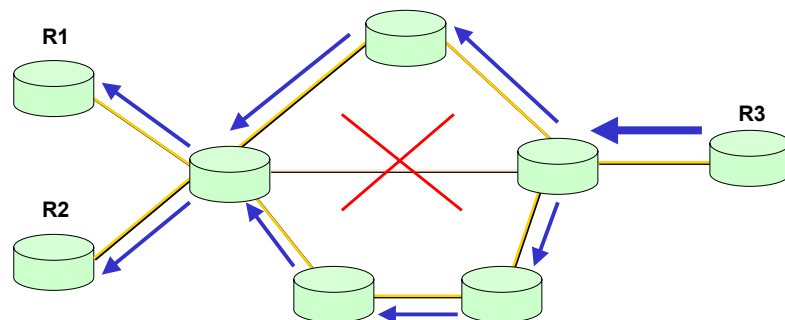


12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

リンク障害時の負荷分散

- MPLS-TEで安心。最適化経路変更。
 - Re-OptimizeでIGPだけよりも高速。(<数Sec)

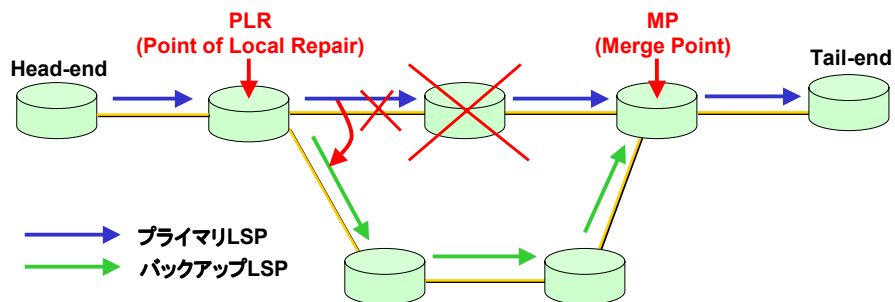


12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

Fast Reroute

- “Local Repair”と呼ばれる高速障害回復手法の1つ。
 - 障害となったリンクまたはノードに隣接し、トラフィックの上流となっているノードによってLSPが高速迂回される。(数msec~数10msec)
 - プライマリLSPのためのバックアップLSPが事前に設定される。
 - 他にはGlobal Repair, Alternative Egress Repair, etc...



Fast Reroute

- 障害検出方法としては、
 - Layer2以下のアラーム(e.g, 光入力断やSONET/SDHアラーム)
 - IGPやRSVPのadjacency/hello断
- バックアップパス確立やトラフィック迂回にいくつか種類がある。
 - One-to-one backup
 - プライマリLSP1本毎にバックアップLSPが設定される。
 - Facility backup
 - 共通のリンクorノードを通過する複数のプライマリLSP毎に、1つのバックアップLSPが設定される。
 - Link protection (注:仕様の中のTerminologyではない)
 - 障害となったリンクを通過するLSPをバックアップLSPへ迂回する。
 - Node Protection
 - 障害となったノードを通過するLSPをバックアップLSPへ迂回する。

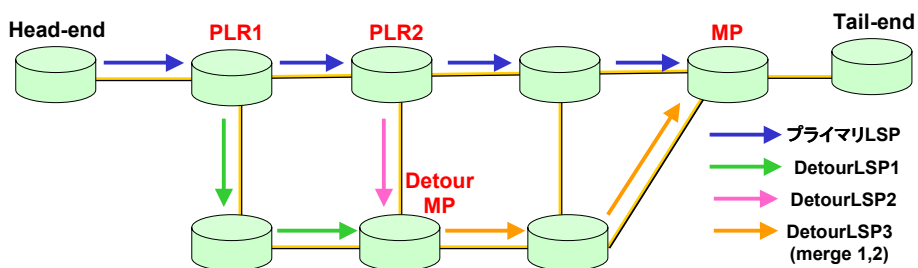
12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

Fast Reroute

One-to-one backup

- Head-endから、
“One-to-one backup desired”(FAST_REROUTE Object)としてProtected LSPを要求。
- PLRでバックアップLSP(Detour LSP)をCSPFにより計算し設定。
 - PLRよりDetour Object付きのPathにより設定される。
- Detour LSPはマージされることがある。



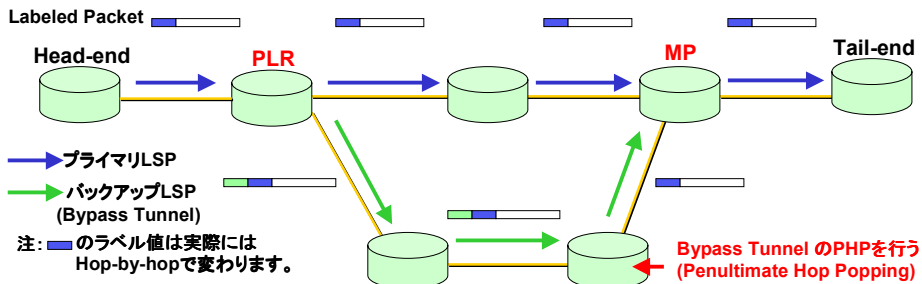
12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

Fast Reroute

Facility backup

- ラベルスタック(ラベルの2段重ね)によって、1バックアップLSP(Bypass Tunnel)で複数LSPのバックアップが可能。
 - One-to-one backupでのDetour LSPはProtected LSPに1対1で対応するので、バックアップのためにラベルスタックは必要ない。
- PLRはMPIにおけるプライマリLSPラベルを学習、変更後バックアップ。
 - Label recordingが必要(Resvメッセージ内RR0IにMPでのラベルが入る)
- MPがGlobal Label Spaceでないと、個別シグナリングが必要となる。



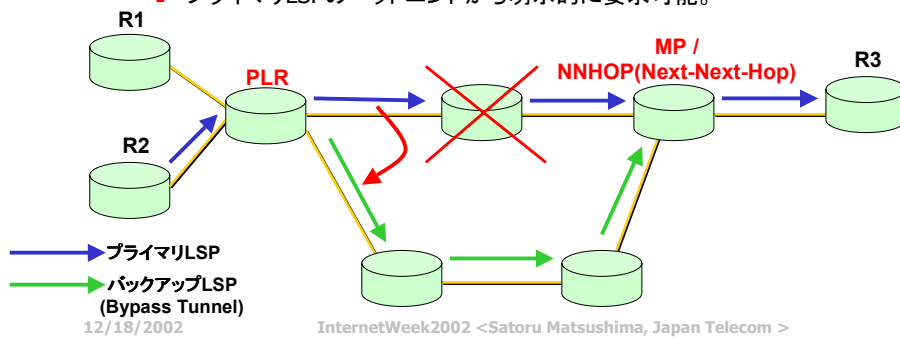
12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

Fast Reroute

Facility backup (ノードプロテクション)

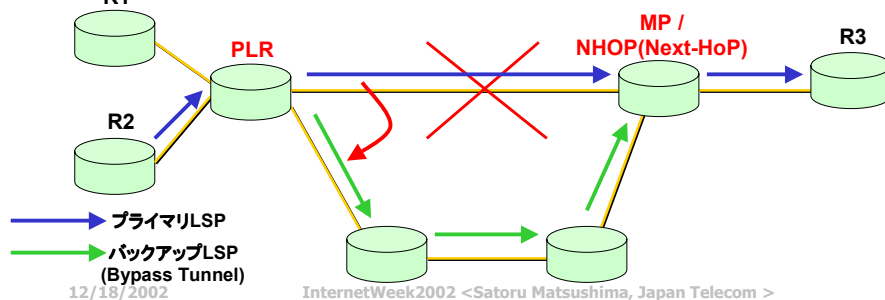
- 隣接下流ノードダウン時に下流隣隣接(Next-Next-HOP)へ高速切替。
 - One-to-one backupの場合は、更に下流ノードがMPとなることが可能。
- PLRからNNHOPへバックアップLSP(Bypass Tunnel)を設定。
- PLRはMPIにおけるプライマリLSPラベルを学習。
- プライマリLSPのヘッドエンドから明示的に要求可能。



Fast Reroute

Facility backupの種類 (リンクプロテクション)

- 下流リンク切断時に隣接ノード(Next-HOP)へ高速切替。
- 仕様上、Facility backupの時のみ可能。
 - One-to-one backupは隣接ノードにDetour LSPを設定しない。
- PLRからNHOPへバックアップLSP(Bypass Tunnel)を設定。
 - ノードプロテクションと特別区別されるものではないが、NHOPへのBypass Tunnelしかなかった場合このような動作となる。
 - MPIにおけるプライマリLSPラベルはそのまま使える。





Fast Reroute 設定例(C)

```
interface Tunnel100
ip unnumbered Loopback0
tunnel destination 100.200.1.1
tunnel mode mpls traffic-eng
tunnel mpls traffic-eng path-option 1 explicit name frr
!
interface POS6/1
ip address 10.1.0.1 255.255.255.0
mpls traffic-eng tunnels
mpls traffic-eng backup-path Tunnel100
pos ais-shut
ip rsvp bandwidth 116250 116250
!
router ospf 1
mpls traffic-eng router-id Loopback0
mpls traffic-eng area 0
!
ip explicit-path name frr enable
next-address 10.1.0.2
```

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



Fast Reroute 設定例(J)

```
interfaces {
  so-0/2/0 {
    unit 0 {
      family inet {
        address 10.0.11.1/24;
      }
      family mpls;
    }
  }
}

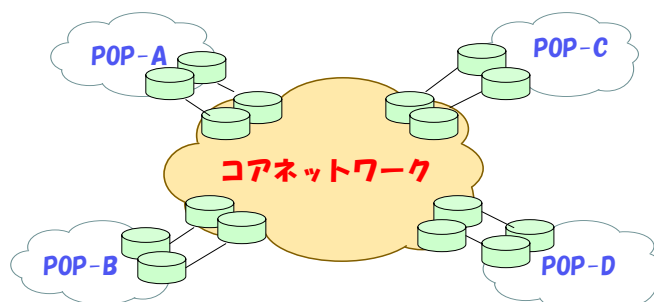
protocol {
  mpls {
    no-propagate-ttl;
    label-switched-path to-hoge {
      from 192.168.254.22;
      to 192.168.254.21;
      fast-reroute;
    }
  }
}
```

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

特定POP間のトラフィック管理

- POP間のトラフィックを知りたい！
 - 下のような構成で特定POP間のトラフィック管理は困難。しかし...

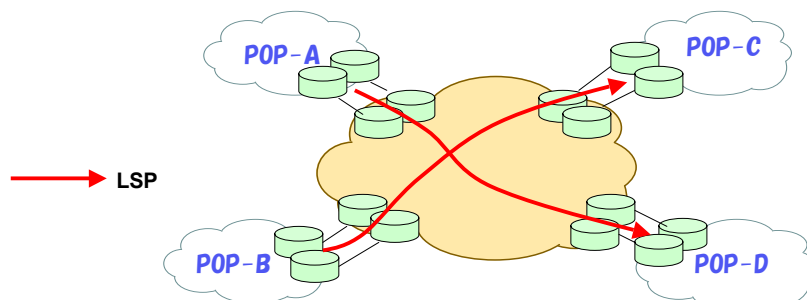


12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >

特定POP間のトラフィック管理

- MPLS-TEで可能になる！
 - あるPOPから他POPへのトラフィックランクに対応するLSPのパケットをカウントできるようになる。



12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



負荷分散の調整

- RIPやOSPFなどのIGPでは、コストやメトリックの足し算の結果の大小で、トラフィックが流れるor流れないという極端な差が現れる。
- MPLS-TEでは、リンクの予約可能帯域量を調整することによりコストに関わらず、それなりのトラフィックをそのリンクにのせることができる。

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



MPLS-TEの課題

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



MPLS-TEの課題は？

- Traffic Engineeringの有効性は十分にあるが、気になるところ。
 - MPLS-TEを使うことでIGPルーティングとトラフィックの流れの関係が疎になる。
 - LSPを設定したルータでないと見えなくなる。
 - Constraint-Pathの元となるパス計算が正しいということをどうやって確認すれば良いのか？
 - LSPの正当さを評価するツールが欲しい。
 - 運用経験
 - どうやって設計すればよいの？
 - やっぱりフルメッシュ？
 - 運用補助ツール／システムの登場
 - MPLSView
(http://http://www.wandl.com/html/mplsview/MPLSview_new.cfm)
etc...

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >



まとめ

- TEとは、トラフィックの流れから見てネットワーク構成を最適化すること。
- MPLS-TEでは以下によりTEを実現。
 - IGP(IS-IS/OSPF)拡張などによる強制的パス選択
 - RSVPIによる明示的パス設定
 - MPLSによるパケットフォワーディング
- FastRerouteなど新しいメリットを生み出す反面、運用ノウハウなどの蓄積が必要。

12/18/2002

InternetWeek2002 <Satoru Matsushima, Japan Telecom >