評価方法

- 中間レポートと、期末レポート
- 出席はとらないが、、、
- 質問やコメントを義務付ける
  - 期中、講義に関する技術的な内容の質問やコメントを最低2回、授業中に行うこと
  - よい質問やコメントは、成績の加点対象
  - 質問者は、講義終了後に名前と学籍番号を申告のこと

インターネットインフラ特論
９．ルーティングとデータリンク層
（IX、ROLC、MPLS、TE）

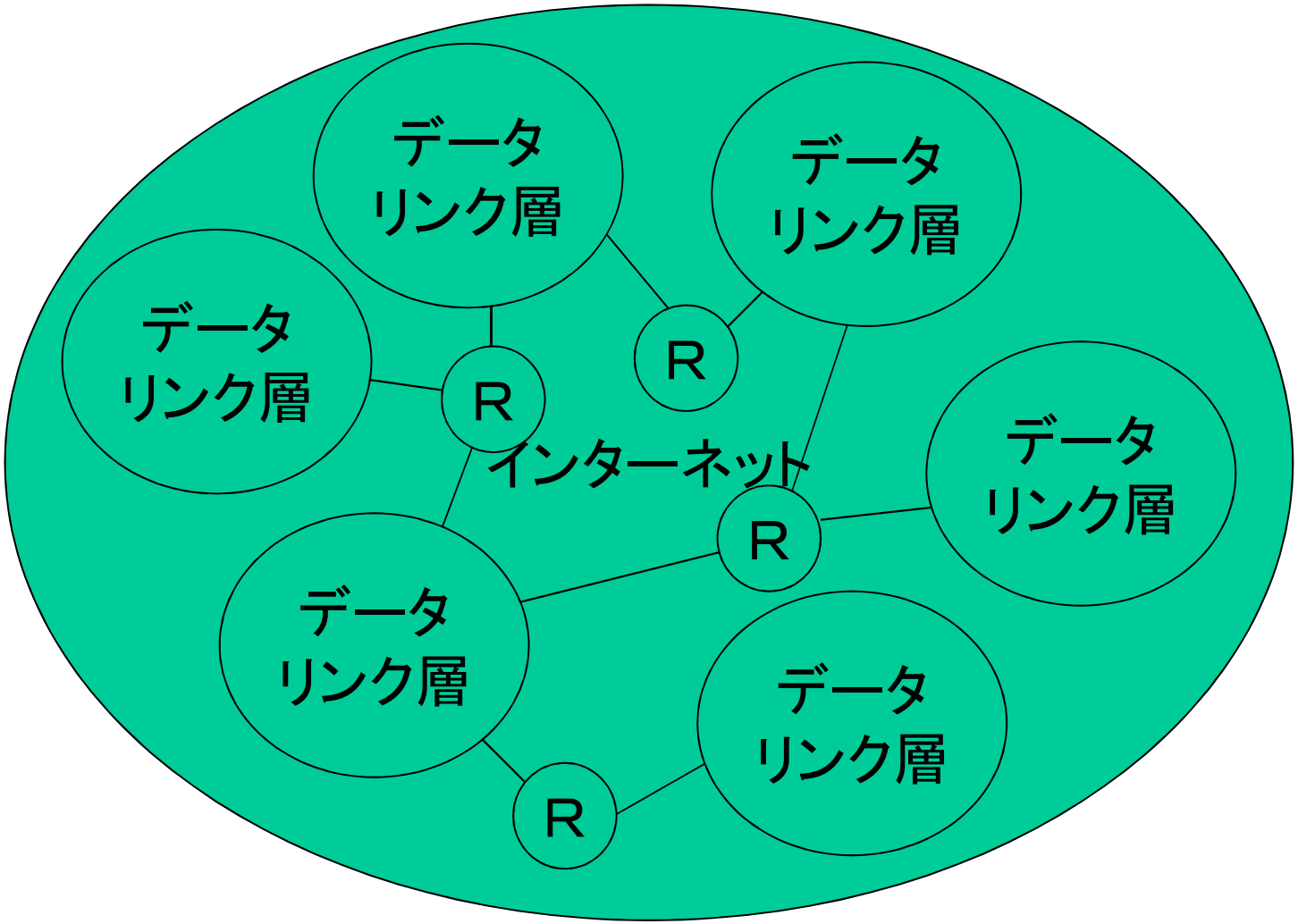
太田昌孝

mohta@necom830.hpcl.titech.ac.jp

ftp://chacha.hpcl.titech.ac.jp/infra9.ppt

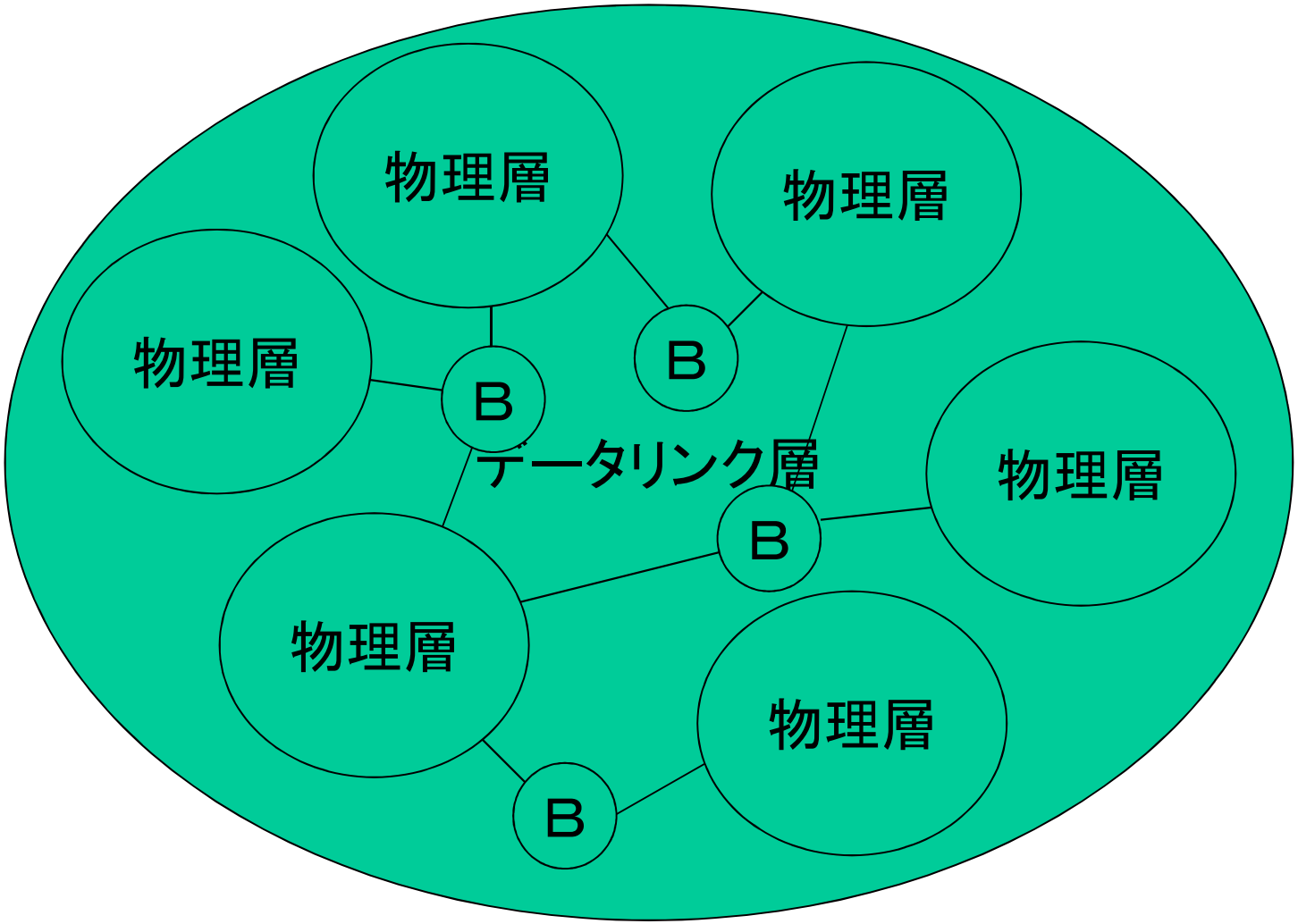
データリンク層でのルーティング

- 複雑なデータリンク層では必要
  - ブロードキャストは非効率、ループに弱い
- データリンク（MAC）アドレスを見て行う
  - IPアドレスとは無関係
- とにかく届けばいい場合は
  - ラーニングブリッジ
  - スパニングツリー
- 最短（最適）経路の設定も可能



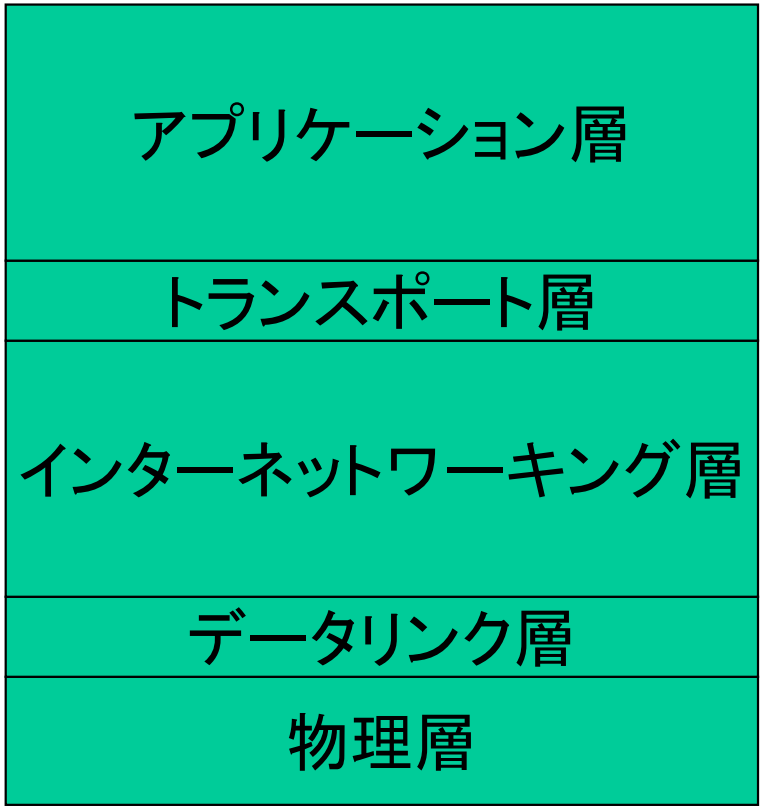
R　：ルータ

CATENETモデル



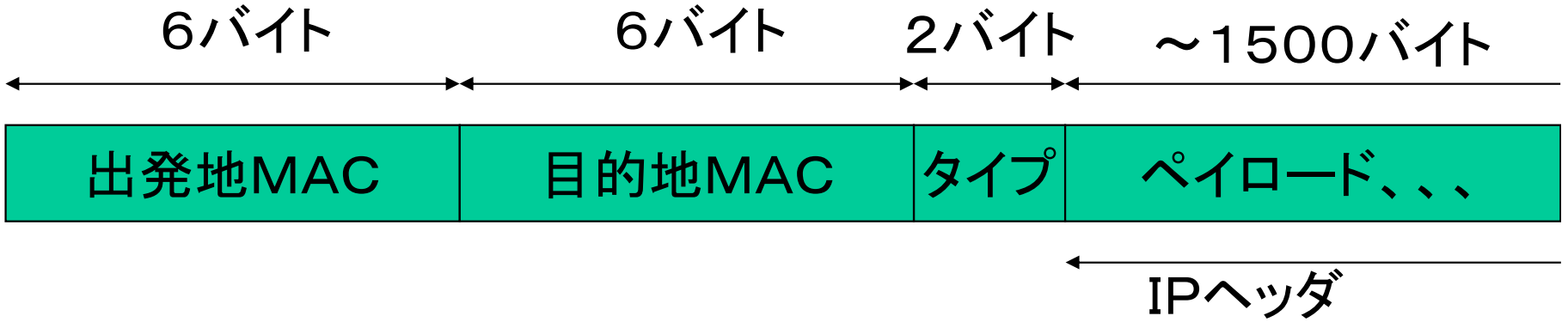
B　:ブリッジ

複雑なデータリンク層

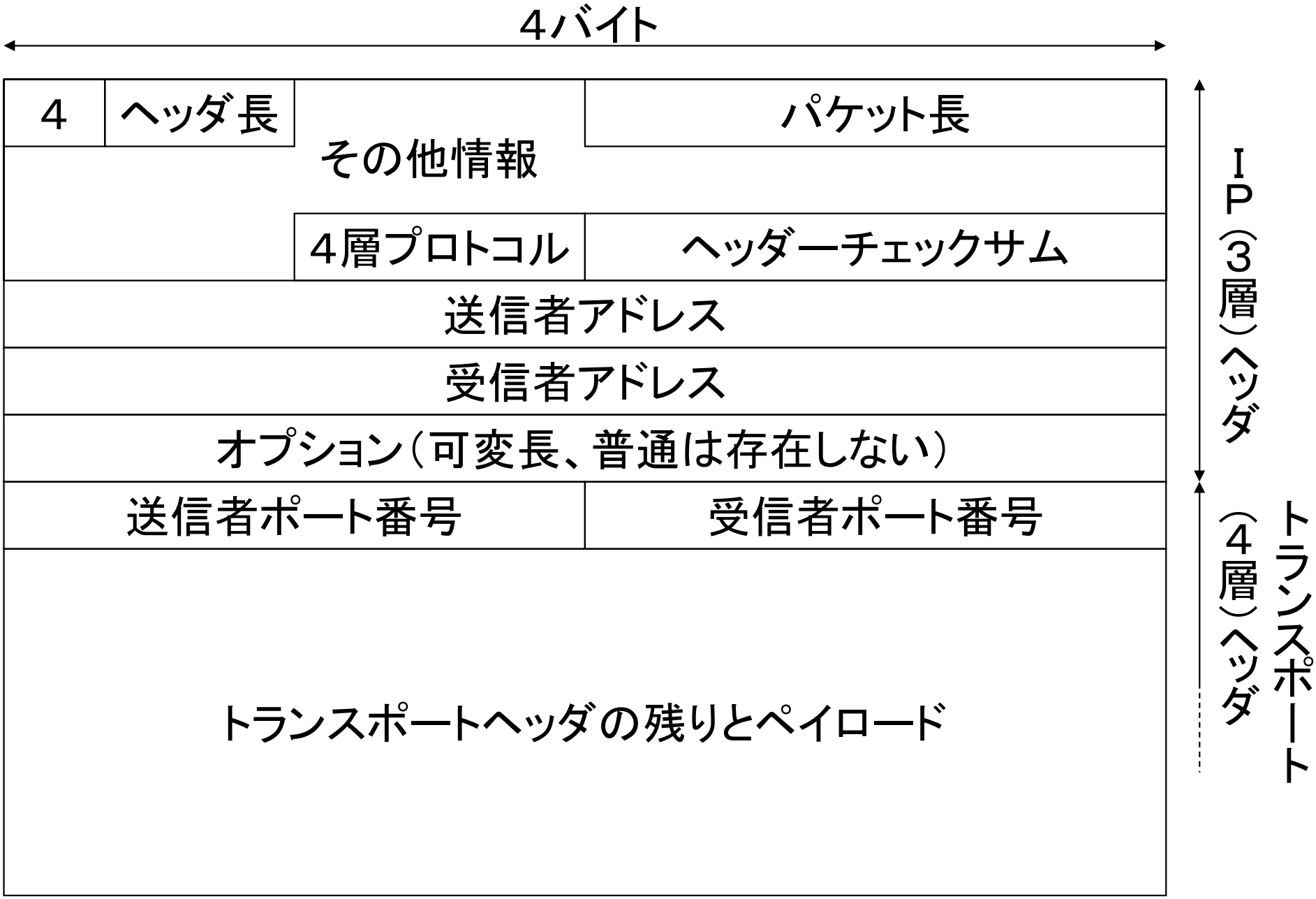


こここそインターネット

インターネットのレイヤリング構造



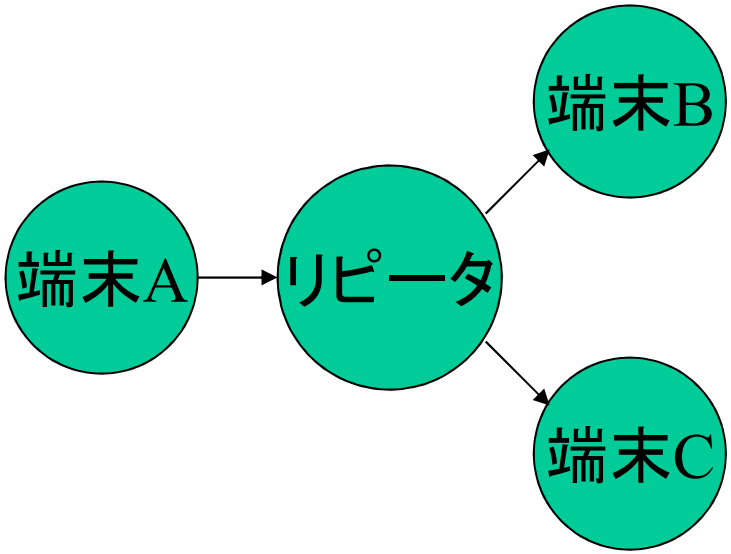
イーサネットのフレーム



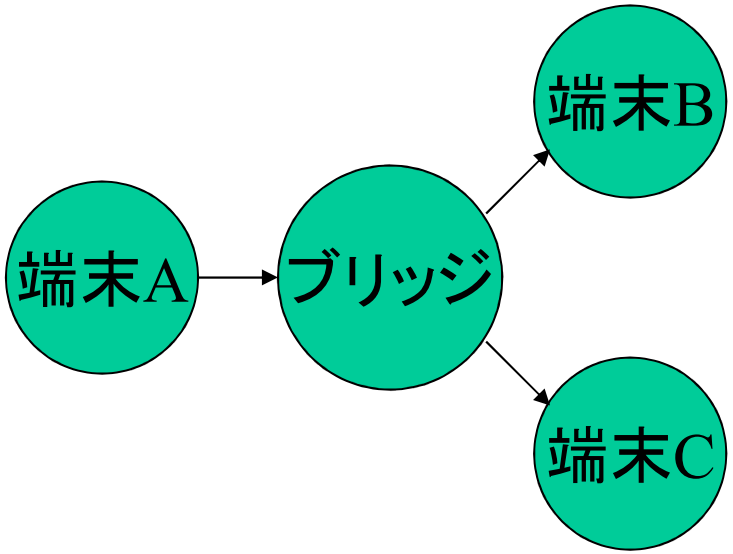
IPv4パケットフォーマット

単純なルーティング

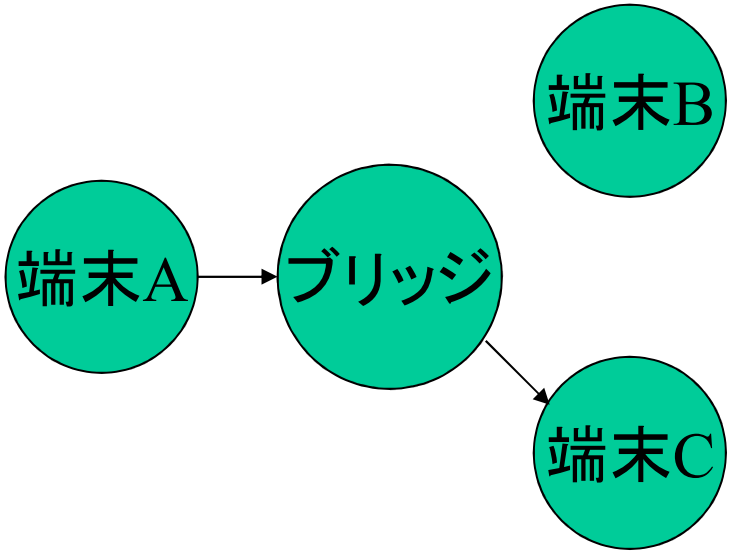
- パケットを全機器にブロードキャスト
  - 帯域が非効率
  - 経路にループがあると無限ループ
- ラーニングブリッジ
  - パケット入力時に出発地MACを学習して
    - 目的地MACの経路として利用
  - ブロードキャストパケットのループは防げず
- スパニングツリー
  - MSTを自動的に計算



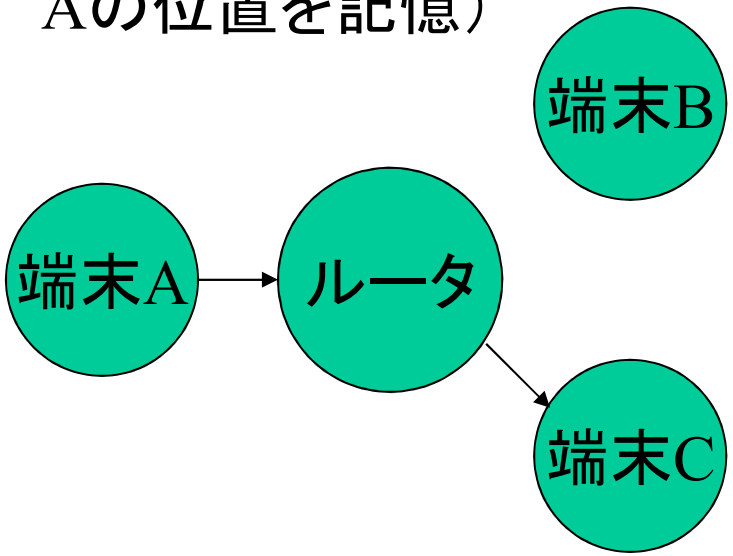
(a)　1層での統合



(b)　2層での統合（最初、Aの位置を記憶）

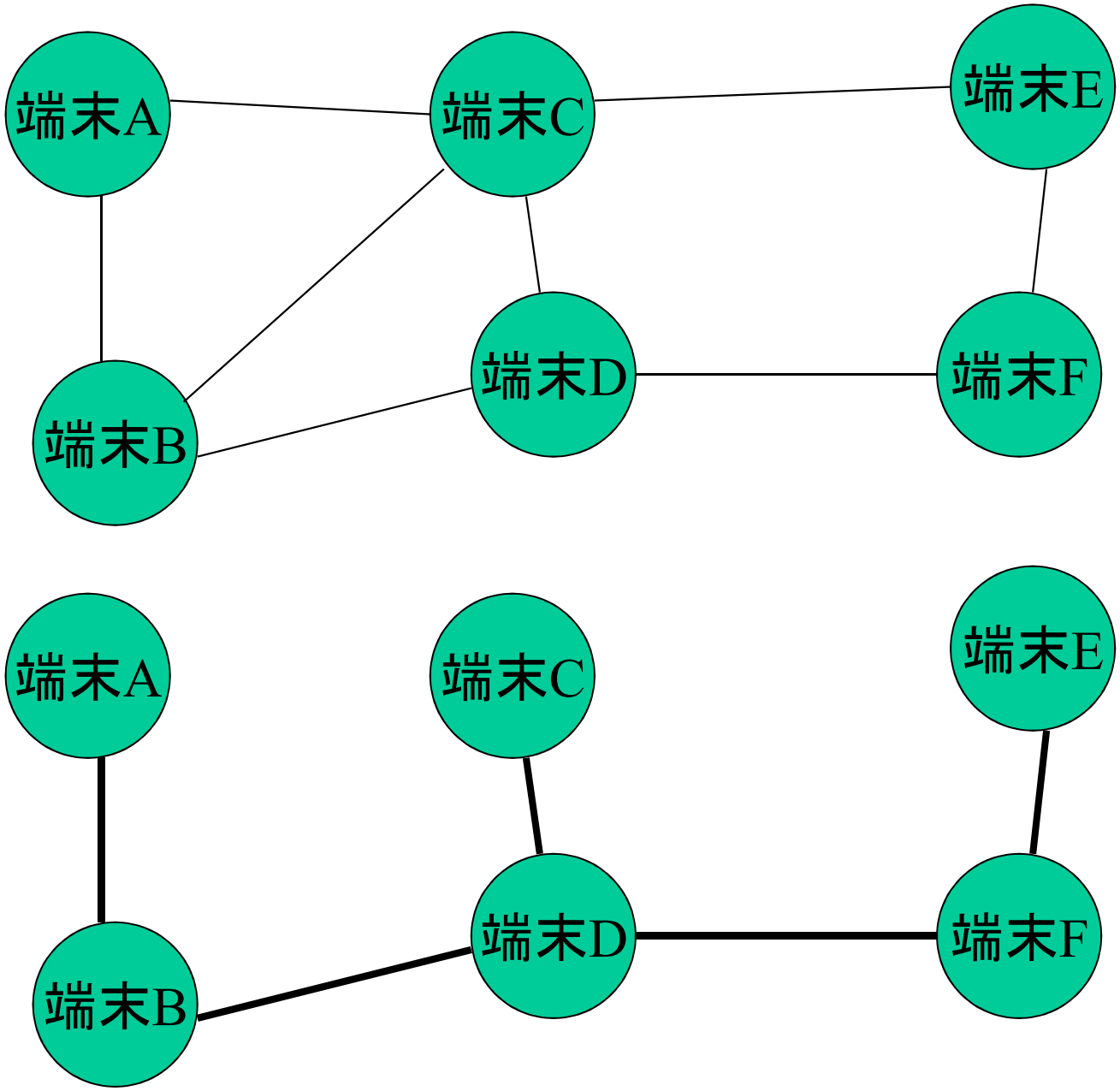


(c)　2層での統合（Cの位置を記憶後）



(d)　3層での統合

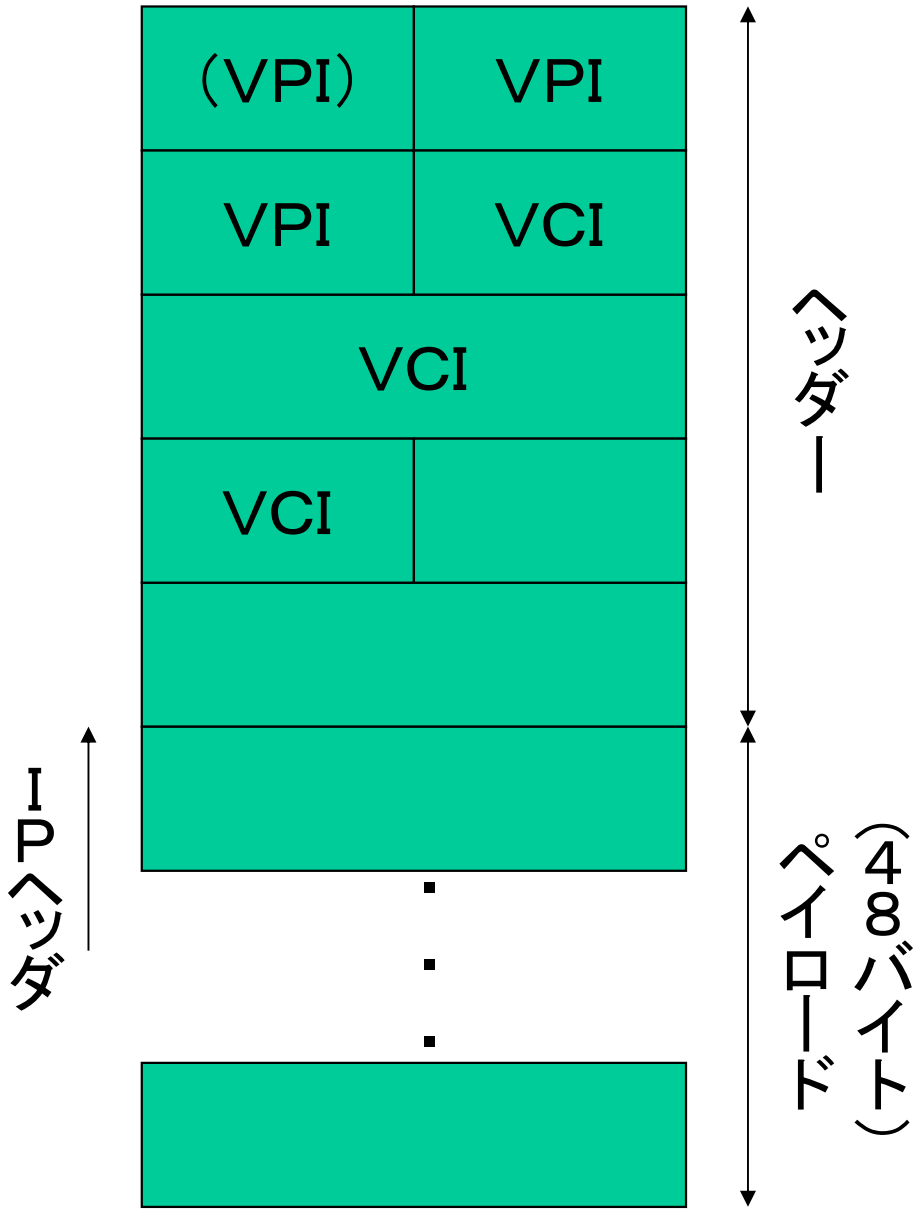
層ごとの中継機器の振る舞い（AからCへのパケット）



ミニマムスパニング木の抽出

複雑なルーティング

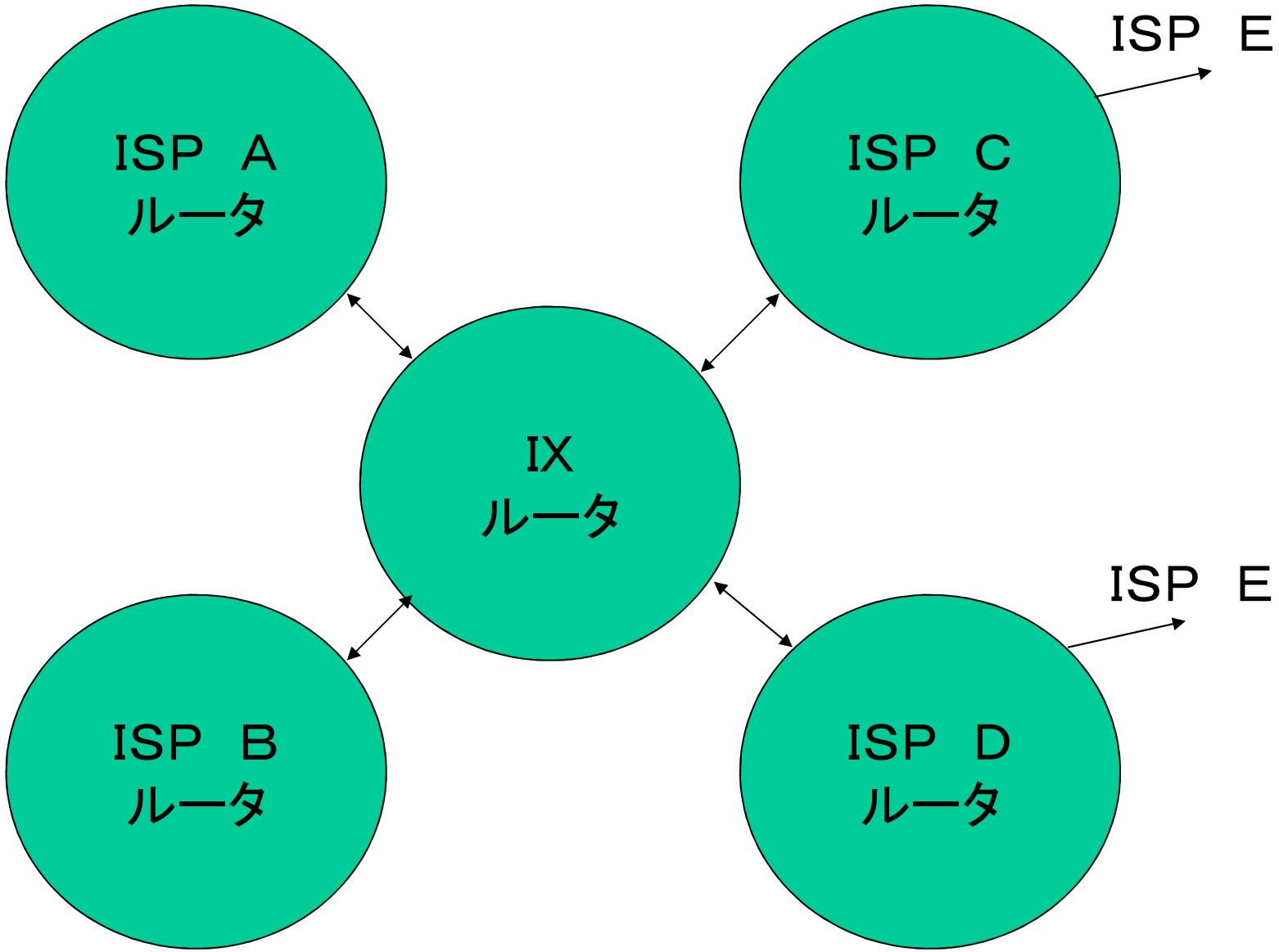
- ATMでは、複雑なスイッチの設定が可能
  - 各スイッチを手動で設定
  - 各スイッチを半自動で設定
  - 各スイッチを完全自動設定（PNNI）
    - 必要QoSに応じたルーティング（QoSルーティング）等



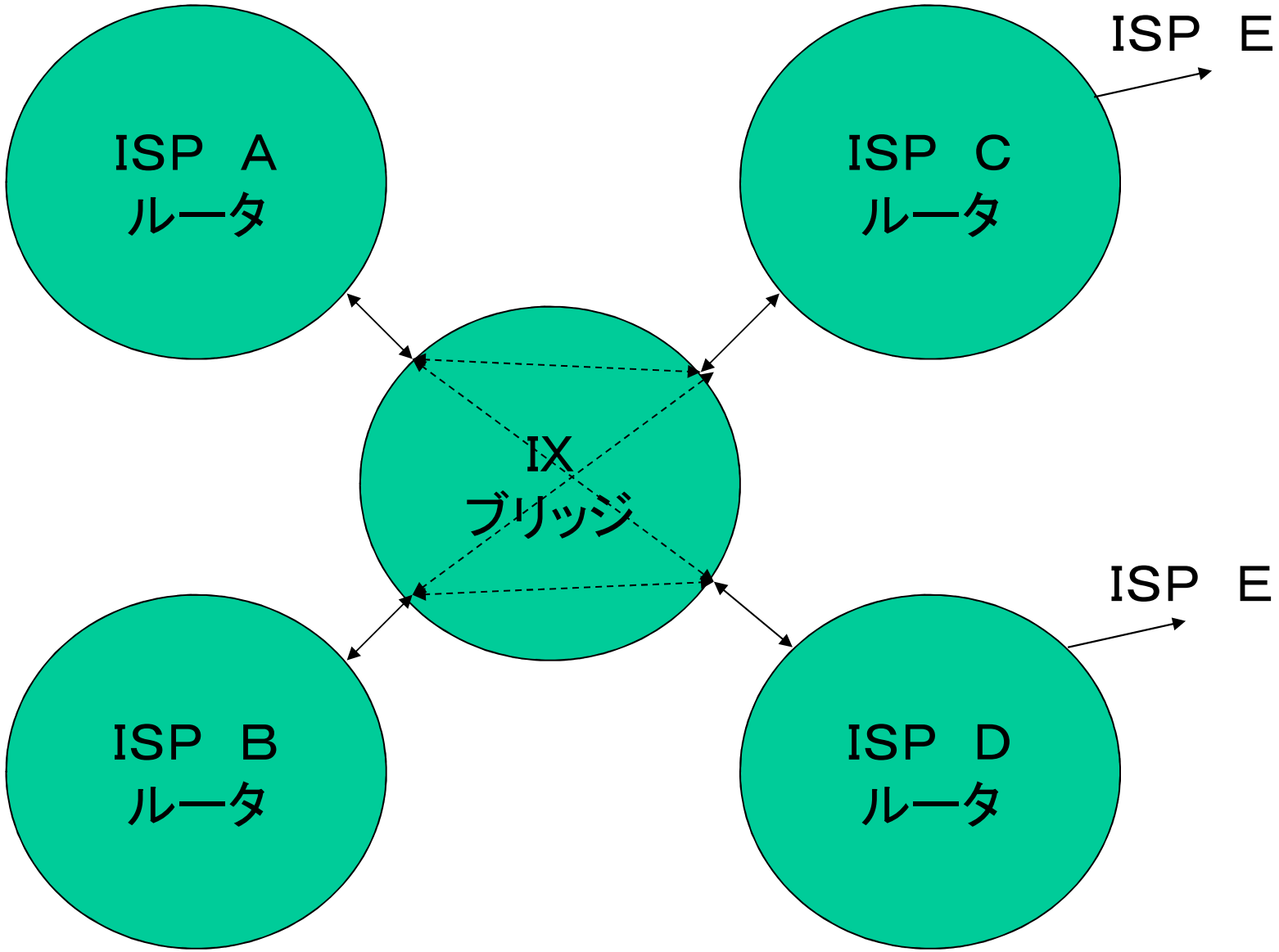
ATMのセル

IX （Internet Exchange）

- ISPがトラフィックを交換する場所
  - ISPが個別にトラフィックを交換するより安い
- L3 IX
  - IXの中央にルータ
  - ISPごとのポリシー制御が不可能
- L2 IX
  - IXの中央にブリッジ（ATMスイッチ）
  - ISPのルータは直接BGP交換



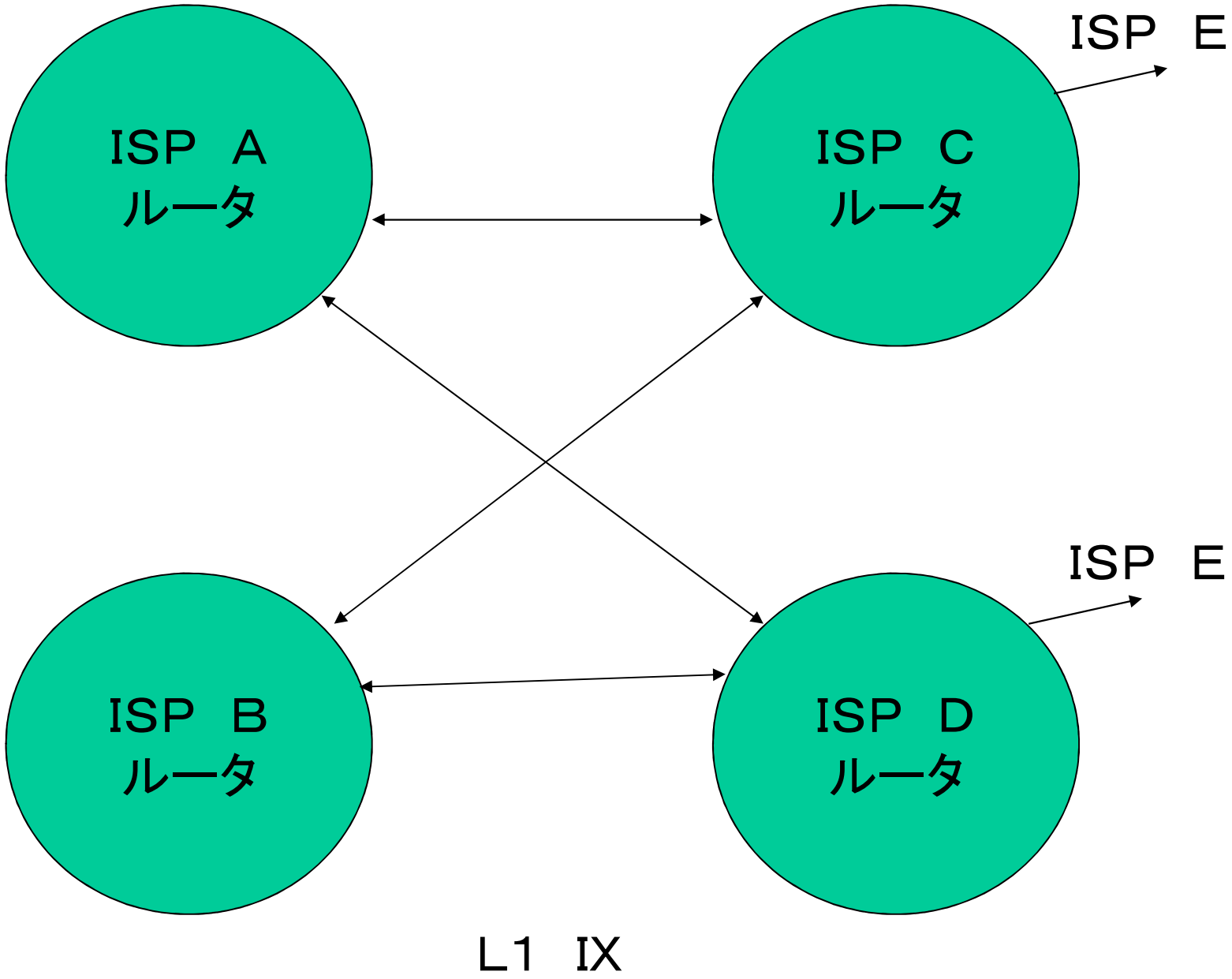
L3 IX

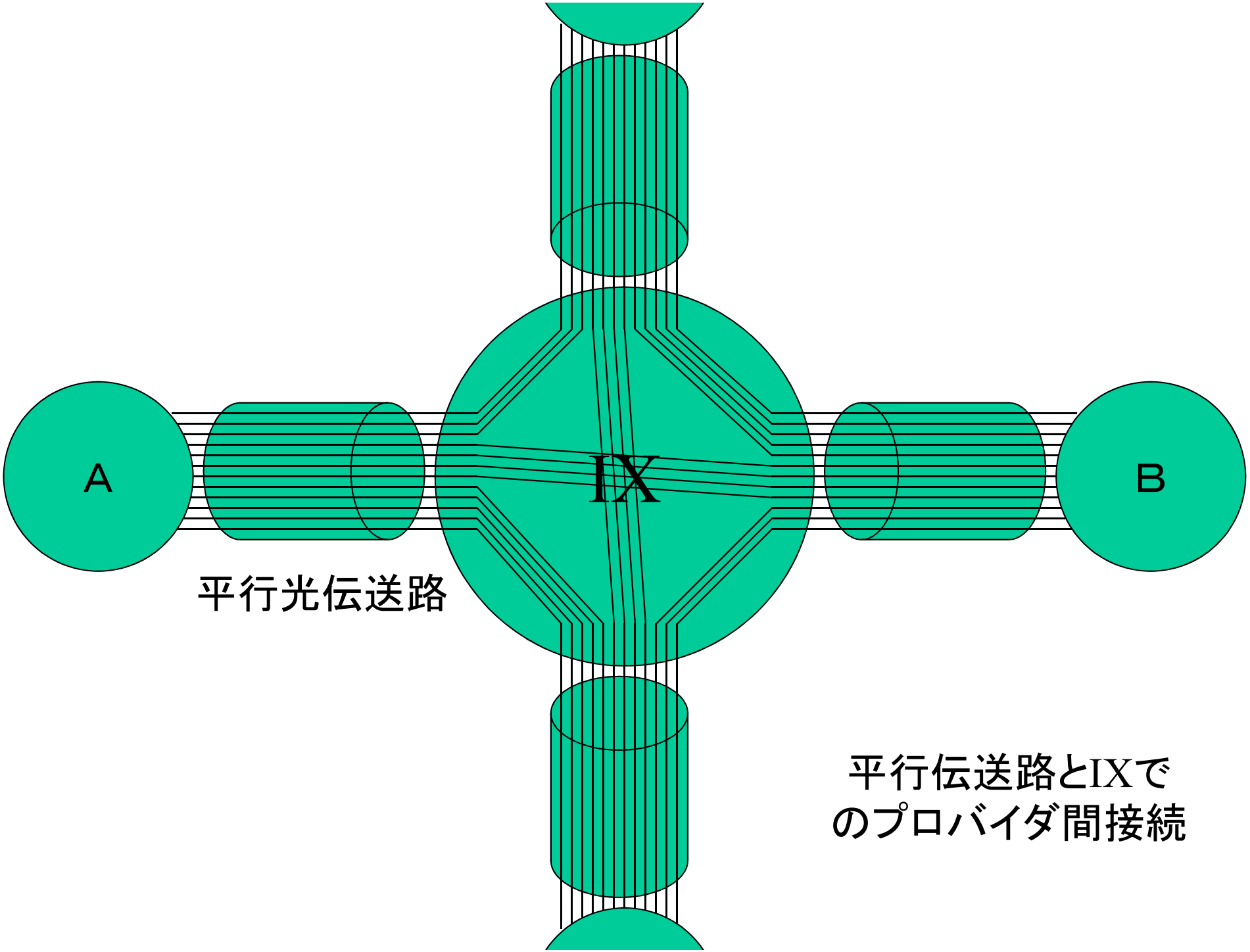


L2 IX

今後のIXは、、、

- L2　IXは
  - 個々のISPのルータはインターフェース1
  - ブリッジで他の多数のISPと接続
    - ブリッジが速度のネックに、、、
- L1　IXは
  - ISP毎にインターフェースを用意
    - ISP間の接続速度がインターフェースの速度を越えたら必然

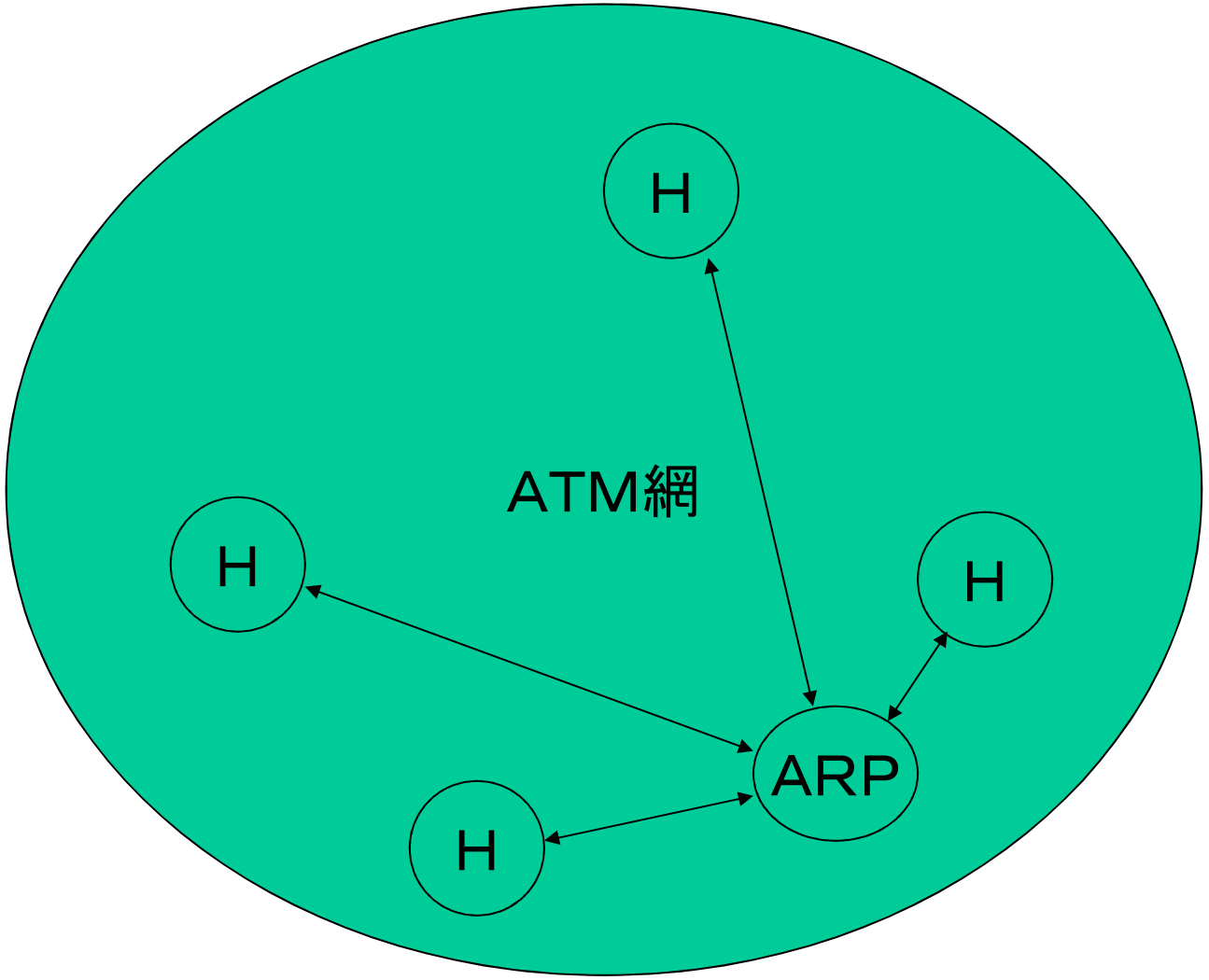

ISP A
ルータ
ISP C
ルータ
ISP E
ISP B
ルータ
ISP D
ルータ
ISP E
L1 IX



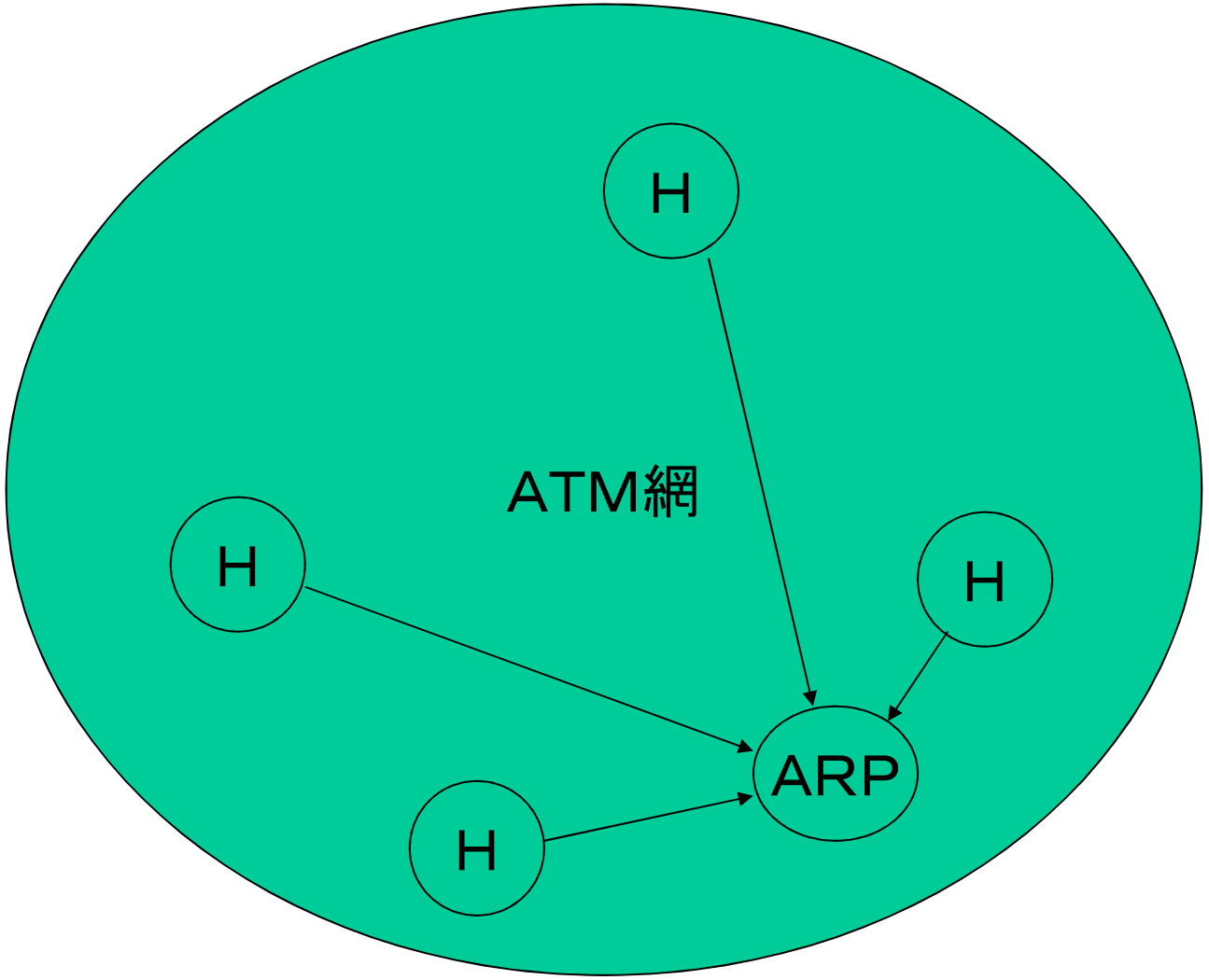
平行伝送路とIXで
のプロバイダ間接続

CLIP（Classical IP and ARP over ATM、 RFC1577）

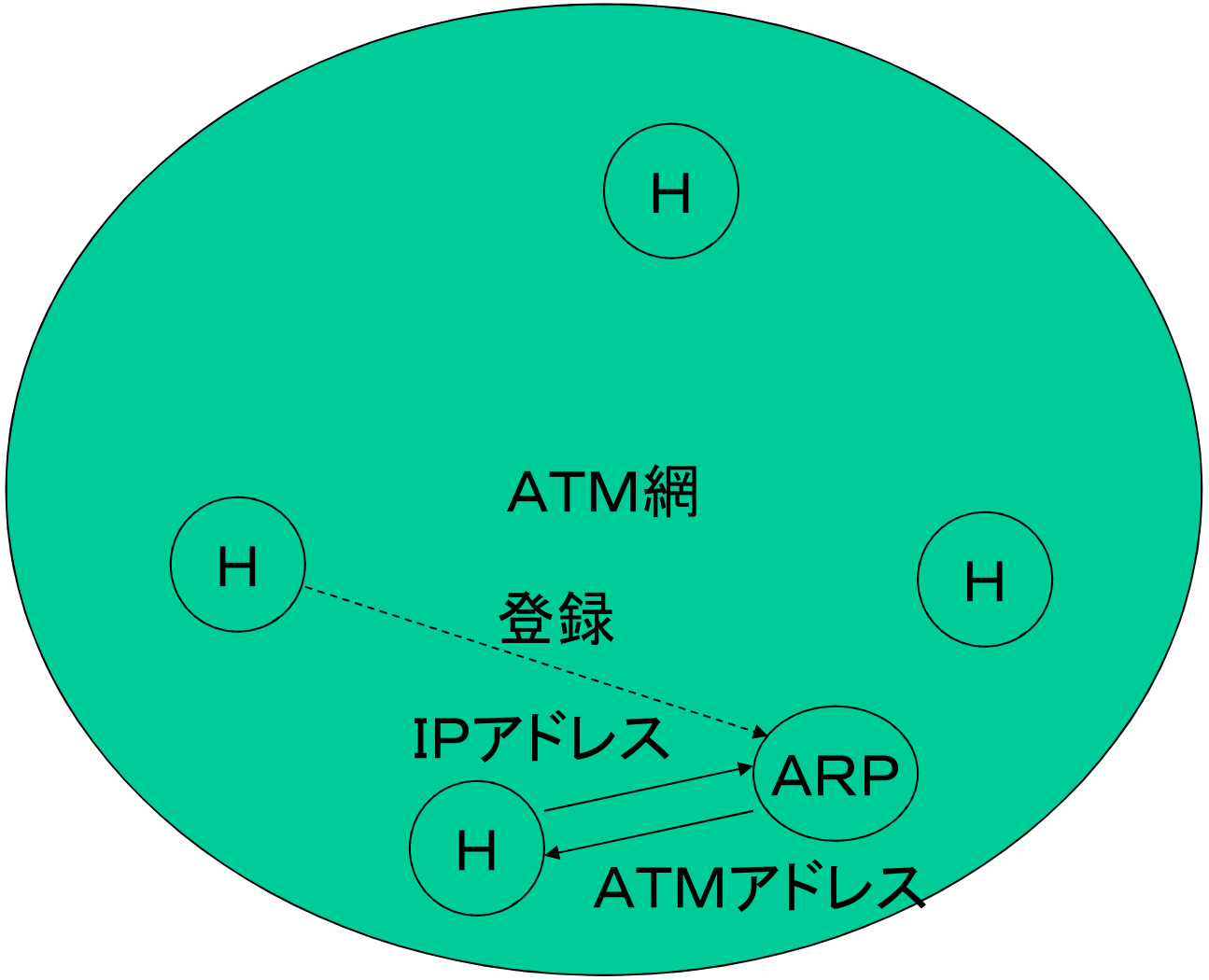
- ATMが世界を制すると思われていた時代があった
- 巨大なATM網が世界を覆う時
  - インターネットはATM網を利用して構築
    - ATM網中の少数のホストが仮想的サブネットを形成
      - サブネットを形成するホストは、ARPサーバを共有
- 仮想サブネット間はルータで中継



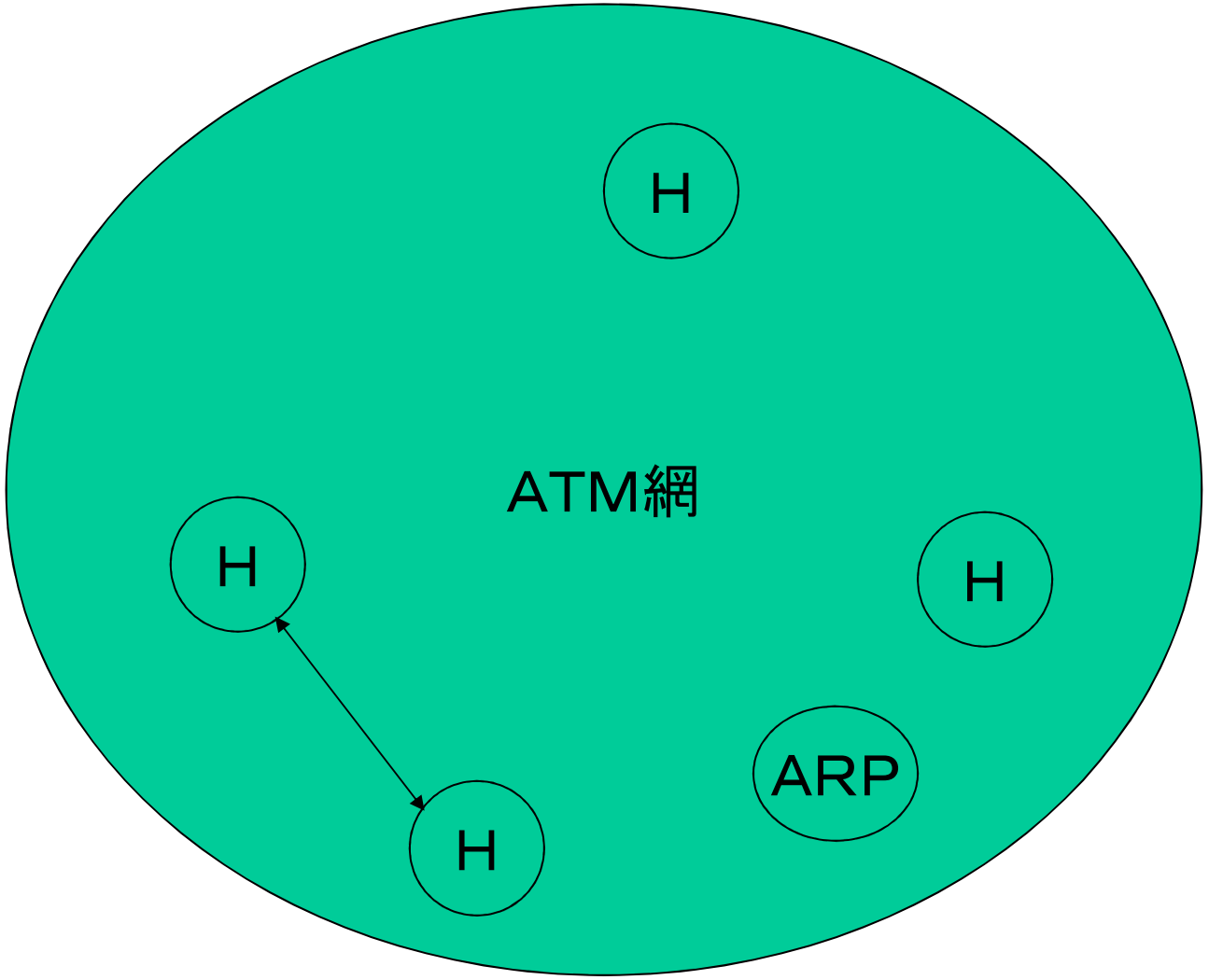
ARPサーバを共有するホストが仮想的サブネットを形成



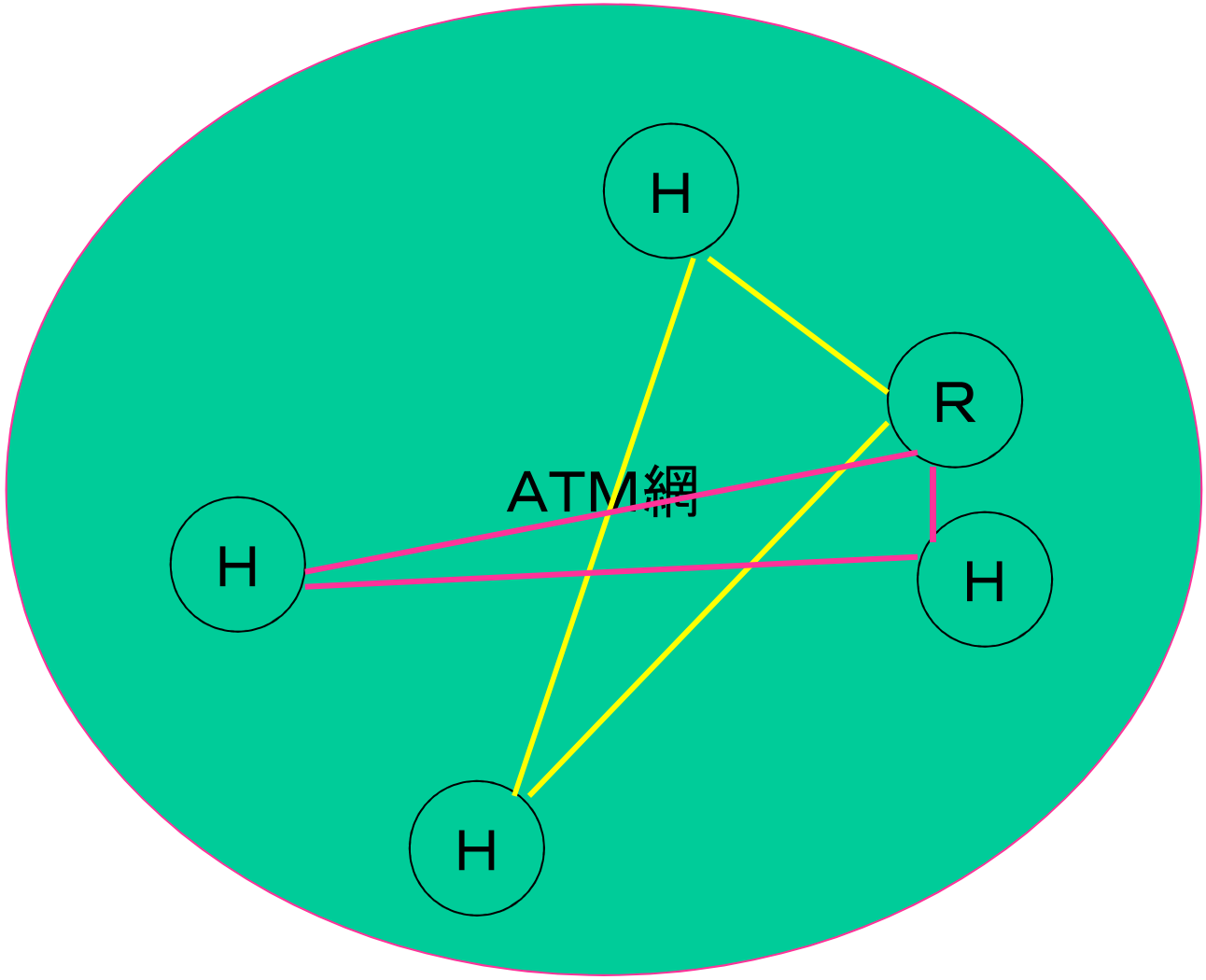
ARPサーバに自分のIPアドレスとATMアドレスを登録



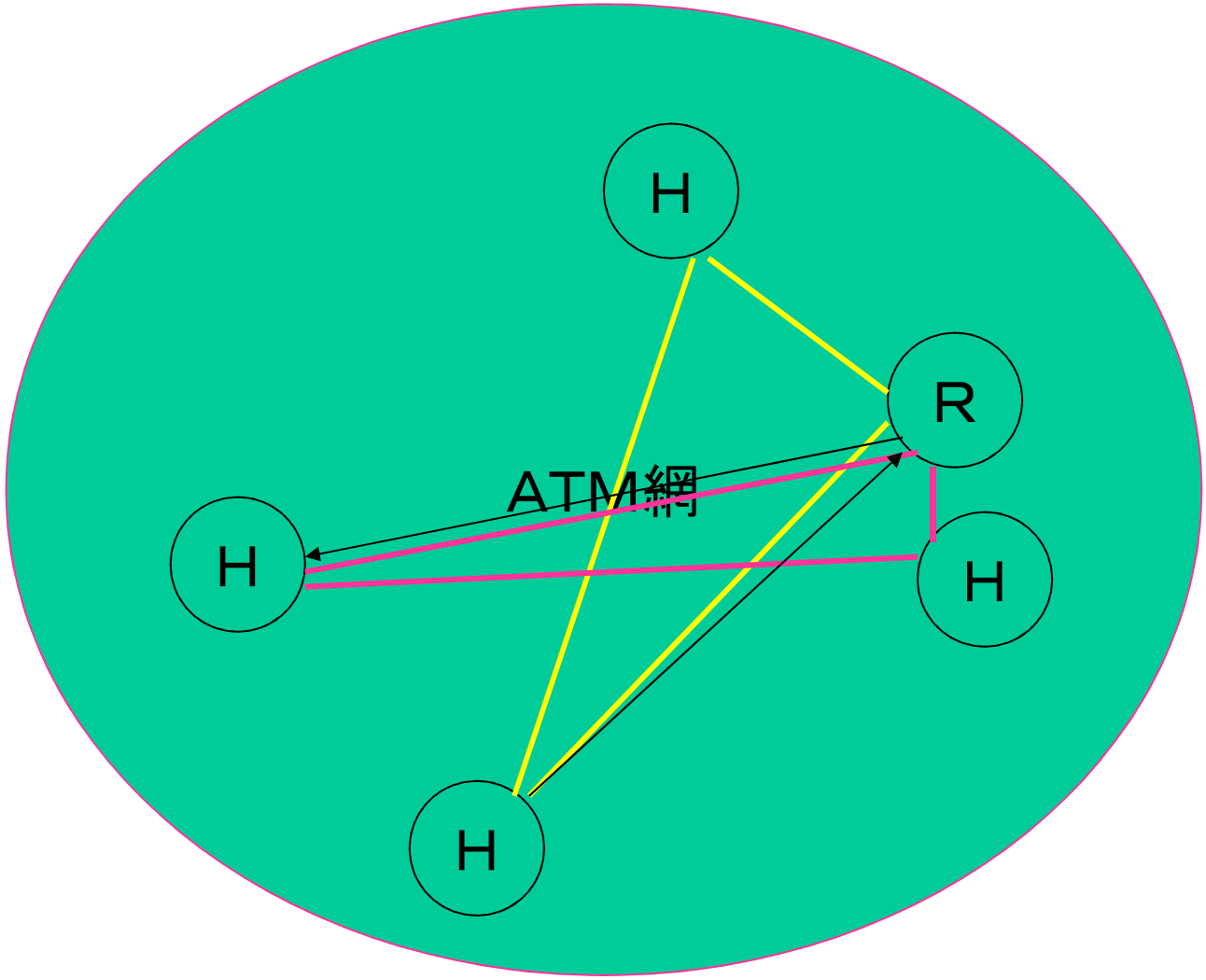
ARPサーバに他のホストのATMアドレスを問い合わせ



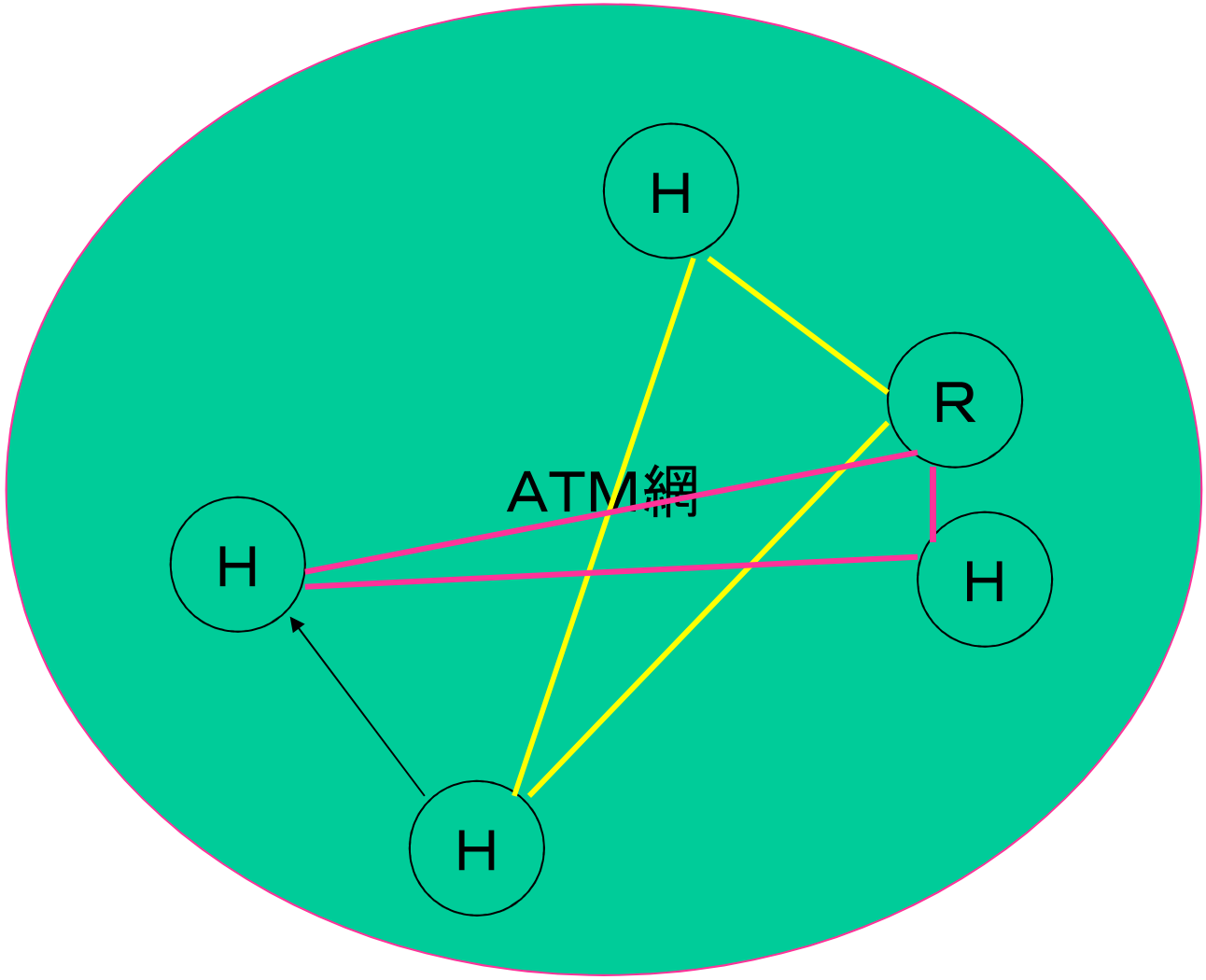
得られたATMアドレスを用いて他のホストと通信



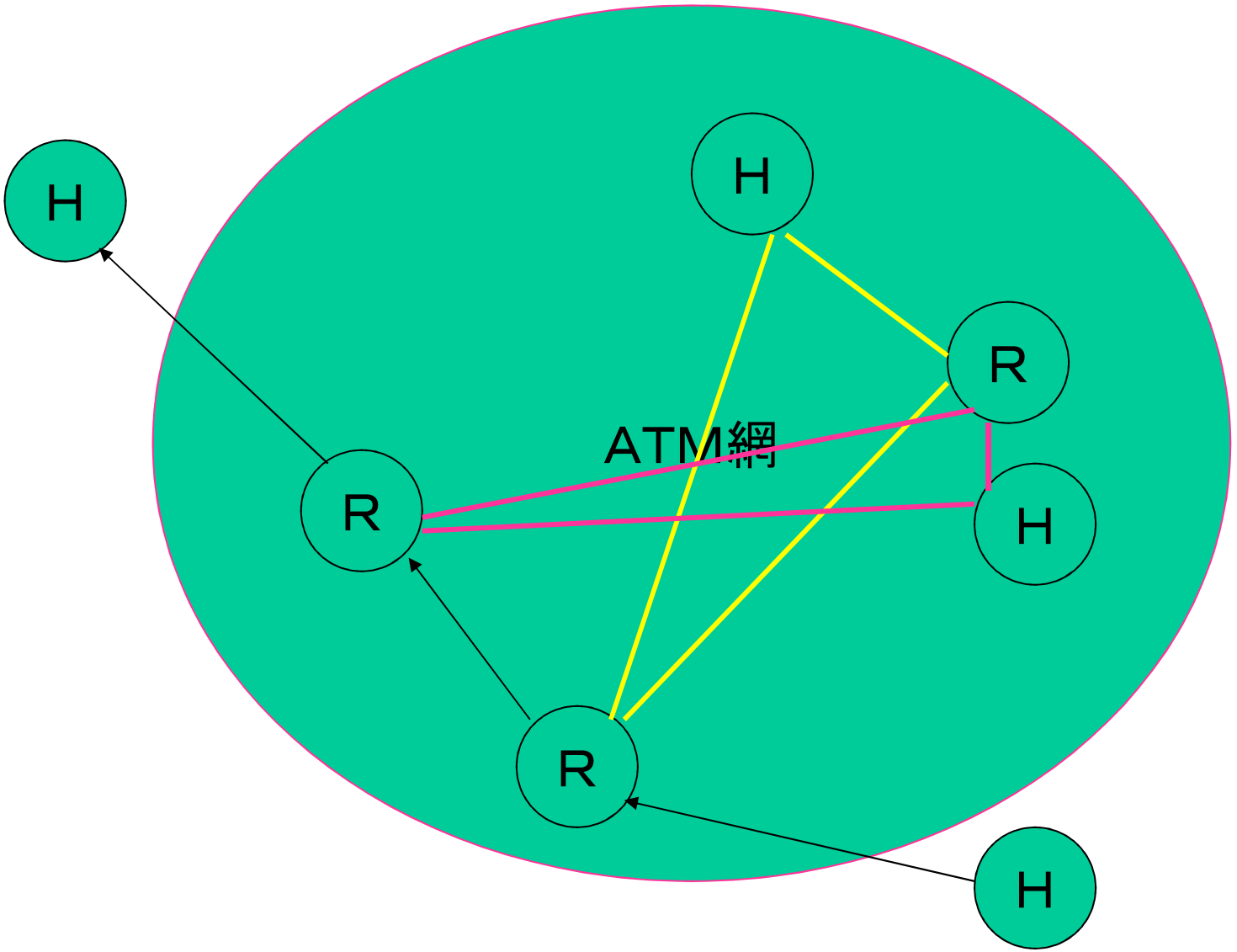
二つの仮想サブネットのルータによる結合



サブネット間の通信



最短距離の通信



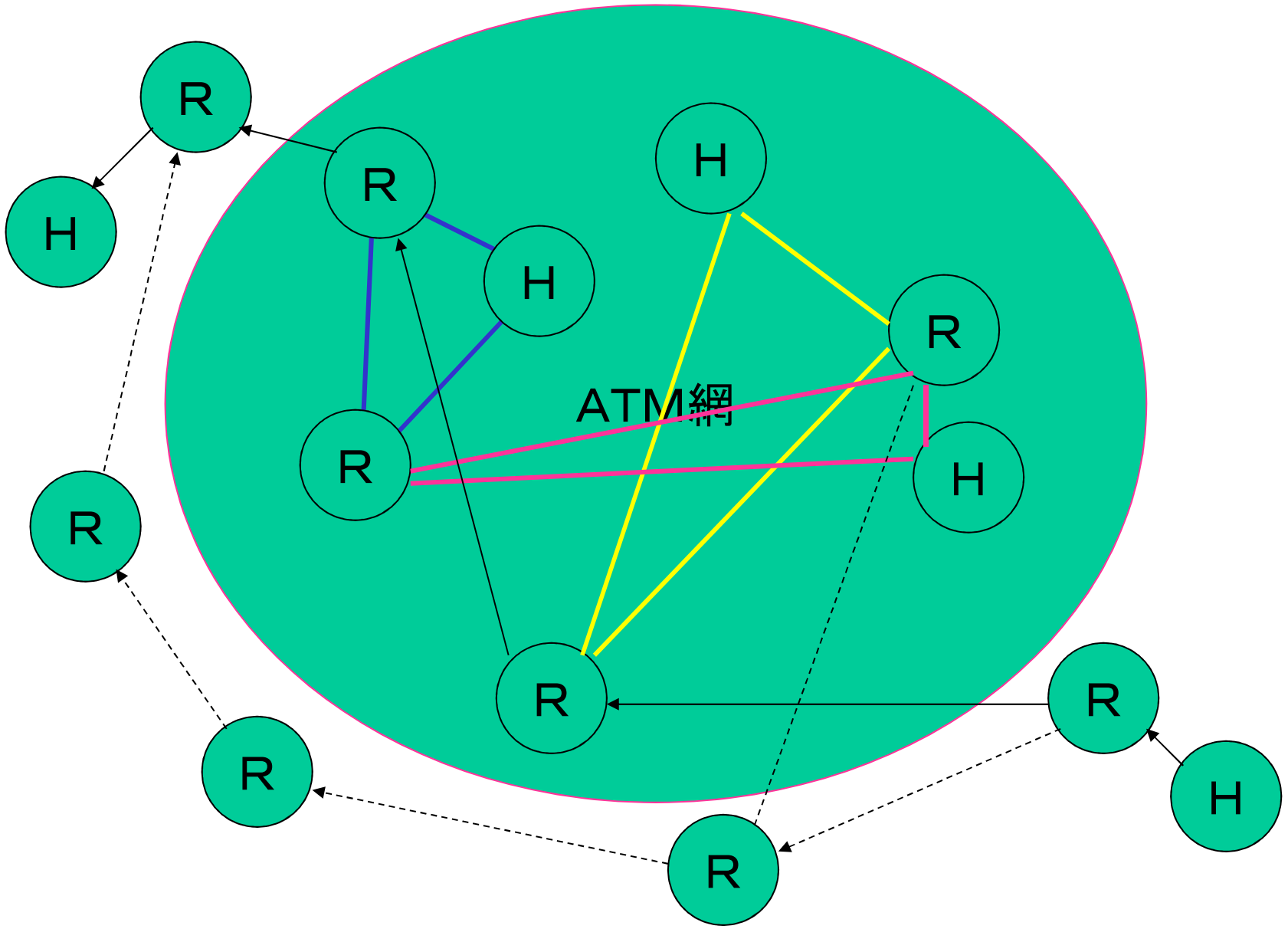
最短距離の通信

ROLC （Routing over Large Cloud）

- CLIPの経路の（ATM網としての）非効率性を取り除く
- ＮＢＭＡ　Ｎｅｘｔ　Ｈｏｐ　Ｒｅｓｏｌｕｔｉｏｎ　Ｐｒｏｔｏｃｏｌ（ＮＨＲＰ）（ＲＦＣ２３３２）
  - 遠くの仮想サブネットにある相手のＡＴＭアドレスを知り、直接通信
  - ルータ間で経路に沿って問い合わせを回す

ROLCの難点

- コネクションが前提
- マルチキャストでは負荷が集中する
- ATM網外部と結合するとかえって効率が悪くなる
  - 最悪の場合ループが発生？



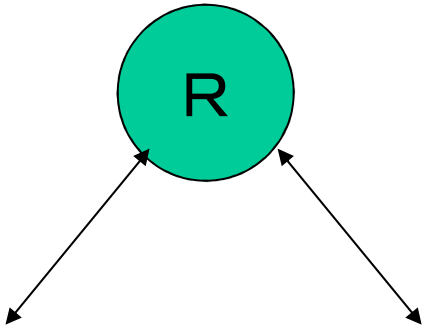
最短距離の通信？

ROLCの本質的困難

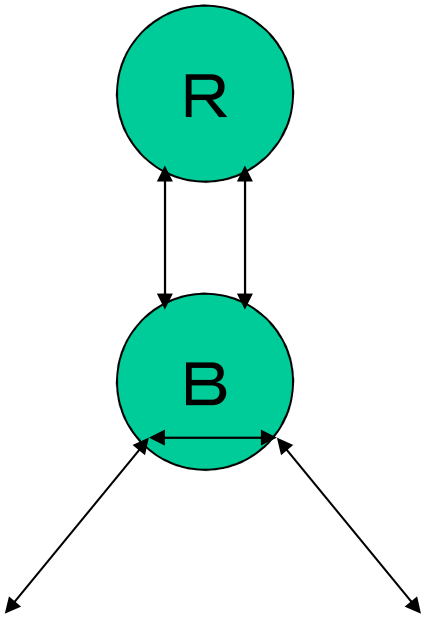
- ATM網上の仮想的IP網がATM網と異なるトポロジーをもつため
  - ATM網的距離とIP網的距離が不整合
    - 両方の距離を最小化することはできない
- ATM網のトポロジーをIP網にあわせてしまえば距離が整合する
  - いっそATMスイッチにIPアドレスをふり、ATM機器の制御にIPパケットを利用すれば、、、

CSR （Cell Switching Router）

- ATMのシグナリングをIPで
  - ATMスイッチは、IPアドレスに沿って配置
- 帯域予約をする通信は、IPパケット（RSVP？）によるATMシグナリング
- あくまで帯域予約の場合
  - BE通信にはBE用のVCをあらかじめ用意
- これで、ROLCは不要



通常のルータ



CSR／IPSILON／MPLSルータ

CSRの発展、 Ipsilon

- BEでも、特定ホストへの通信量が多い場合、専用VCを確保（フロードリブン）
  - ATMの高速性が活かせる？！
  - 大規模ネットワークではVCの数が増えすぎてスケールしない？
- ATM以外でも、データリンク層のラベルがあれば、CSRに利用可能

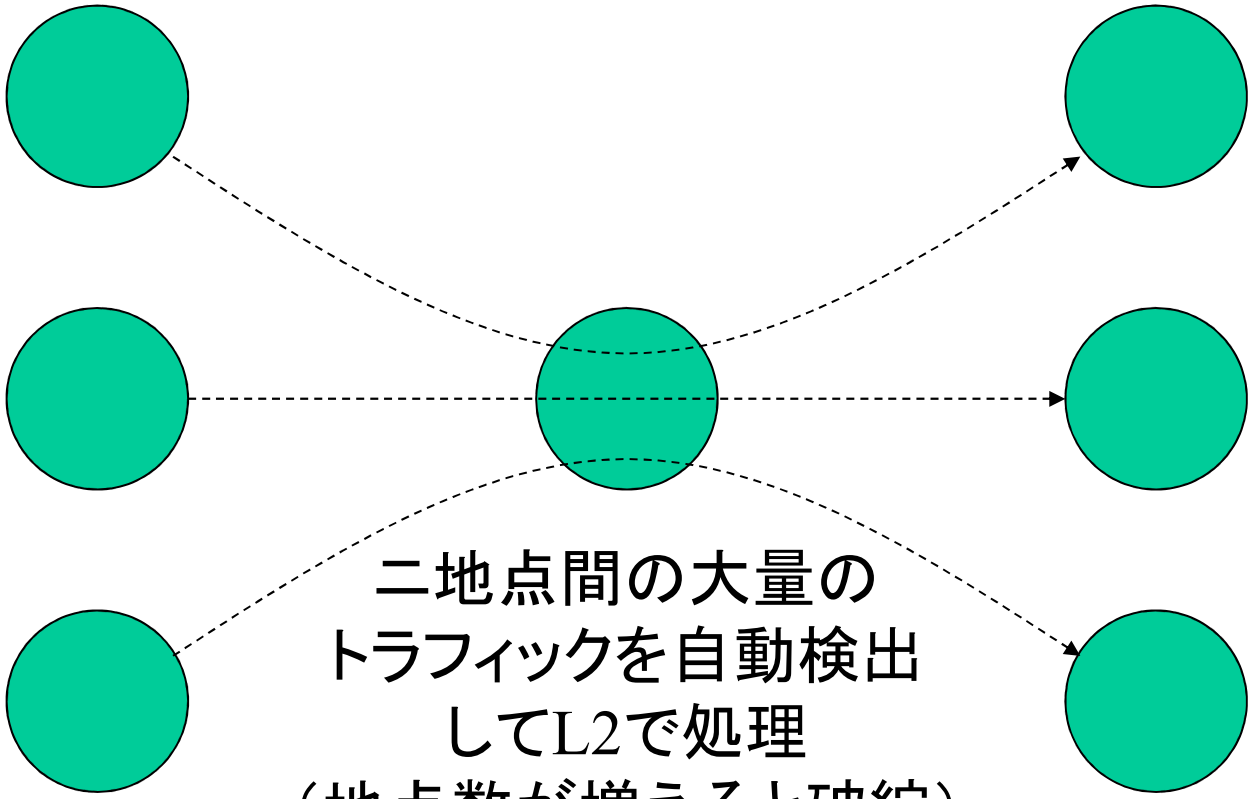
MPLS （Multi Protocol Label Switching）

- フロードリブンはスケールしない
- トポロジードリブン
  - BEのVCをネットワークごとに静的に確保
    - 大規模ネットワークでもスケール
  - ラベルを階層的に持つ
    - 相手ホストのネットワークに到着すると外側のラベルを捨てる
  - 送信ホストは相手ホストのネットワーク内のルート情報を知り、階層的ラベルを付与

MPLSの破綻

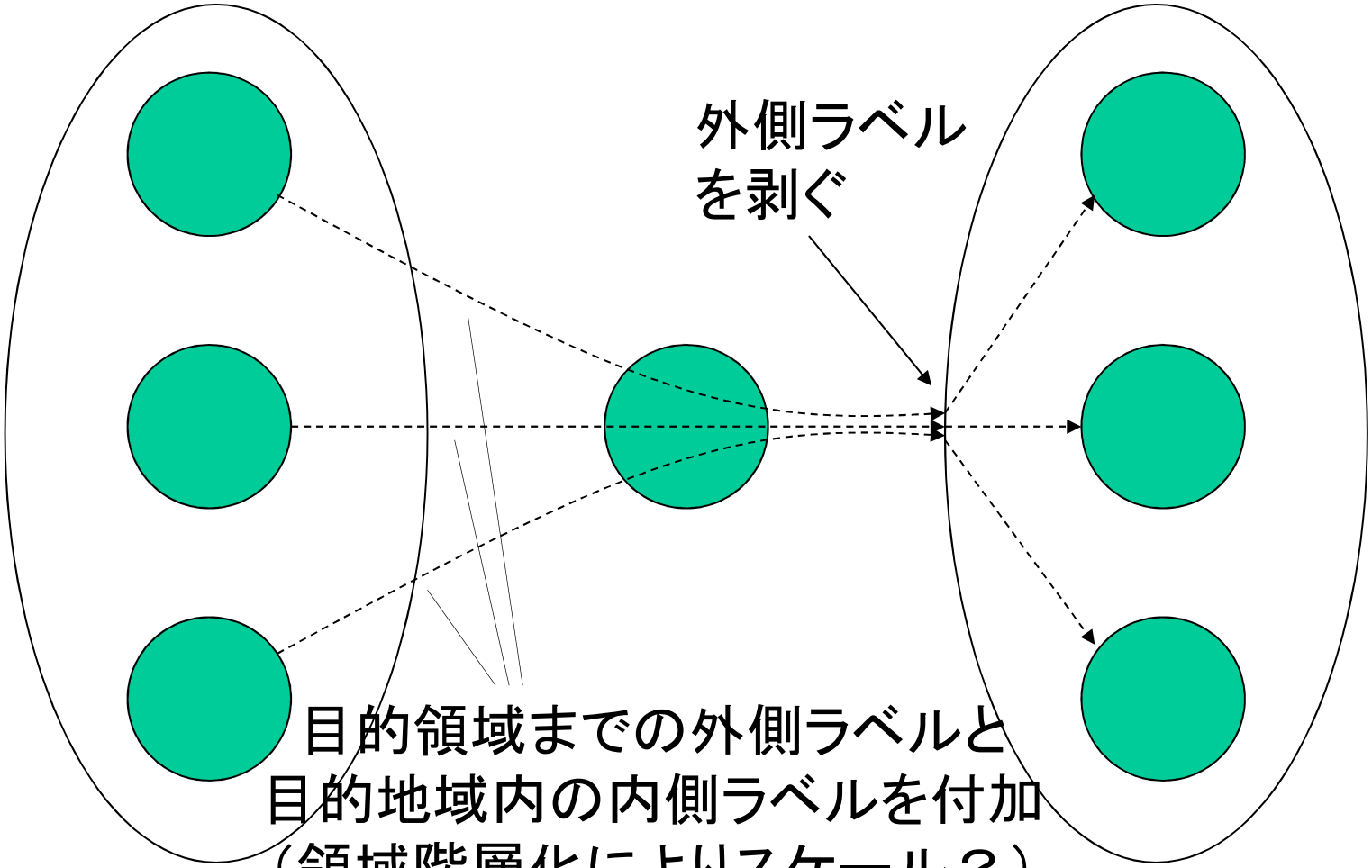
- L2のブリッジは、L3のルータより高速？
  - 実際は、イーサネットでは同程度
  - ATM（CSR）は、パケットルータの10倍遅い
- フロードリブンはスケールしないが、トポロジードリブンはスケールする？
  - トポロジードリブンはフロードリブンにすぎない
- MPLSは使えない？
  - なんとか利用法をみつけて生き残るには、、、

フロードリブン



二地点間の大量の
トラフィックを自動検出
してL2で処理
（地点数が増えると破綻）

トポロジードリブン



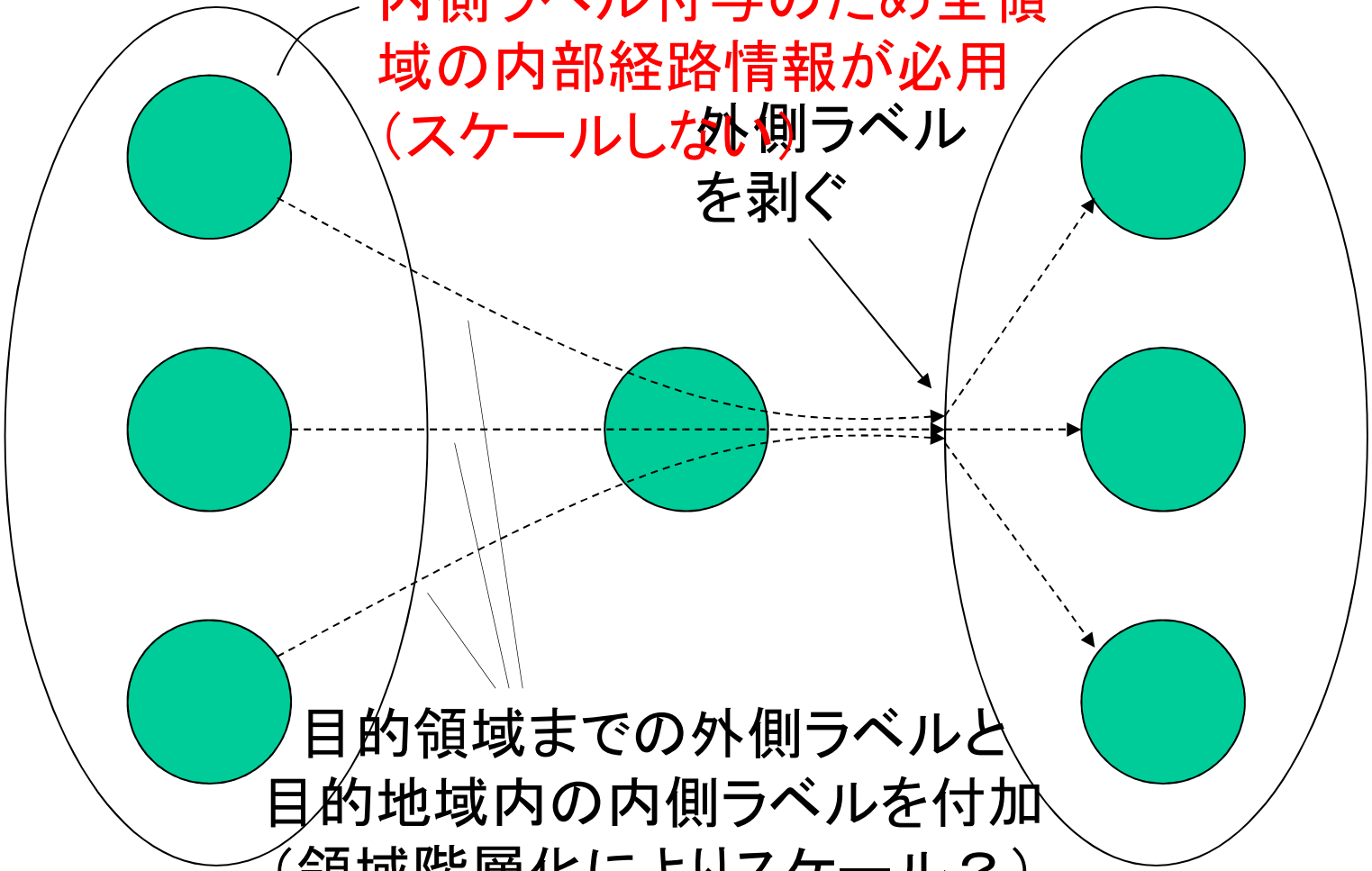
トポロジードリブンの虚構

- 内側のラベルを付加するためには
  - 通信相手付近のトポロジー（経路）情報が必要
  - 大きなネットワークでは相手の付近の経路は分からない
    - 相手と通信する必要のある場合に動的に問い合わせ
    - フロードリブンに他ならない

トポロジードリブン

内側ラベル付与のため全領域の内部経路情報が必用（スケールしない）

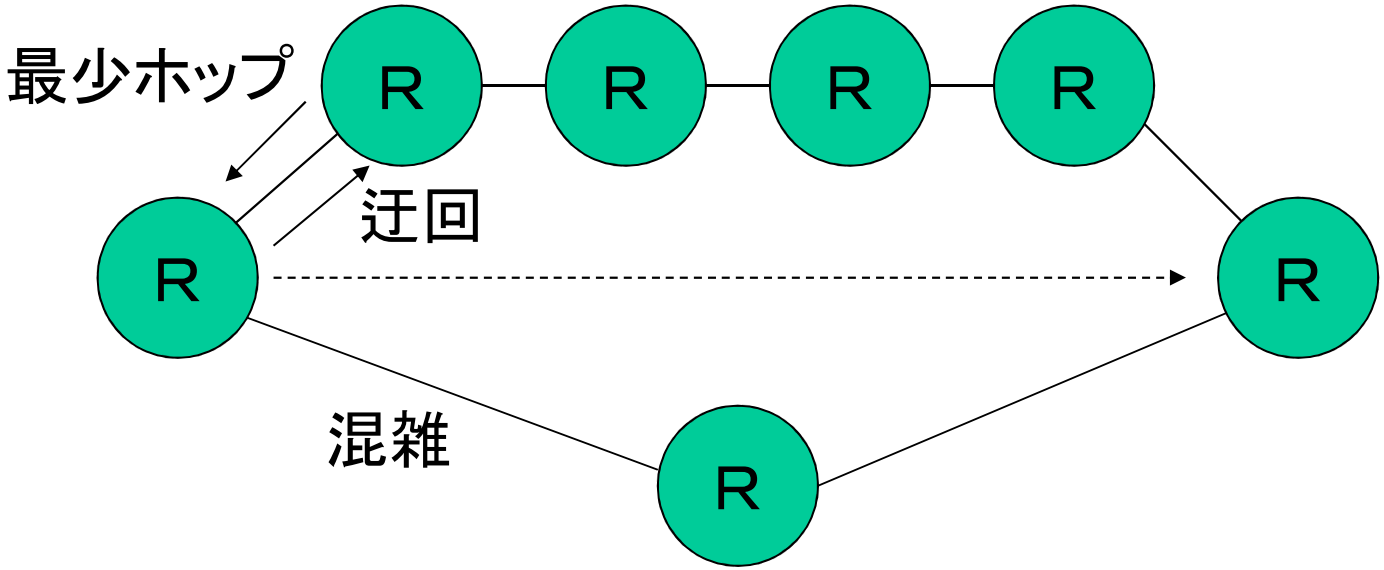
外側ラベルを剥ぐ



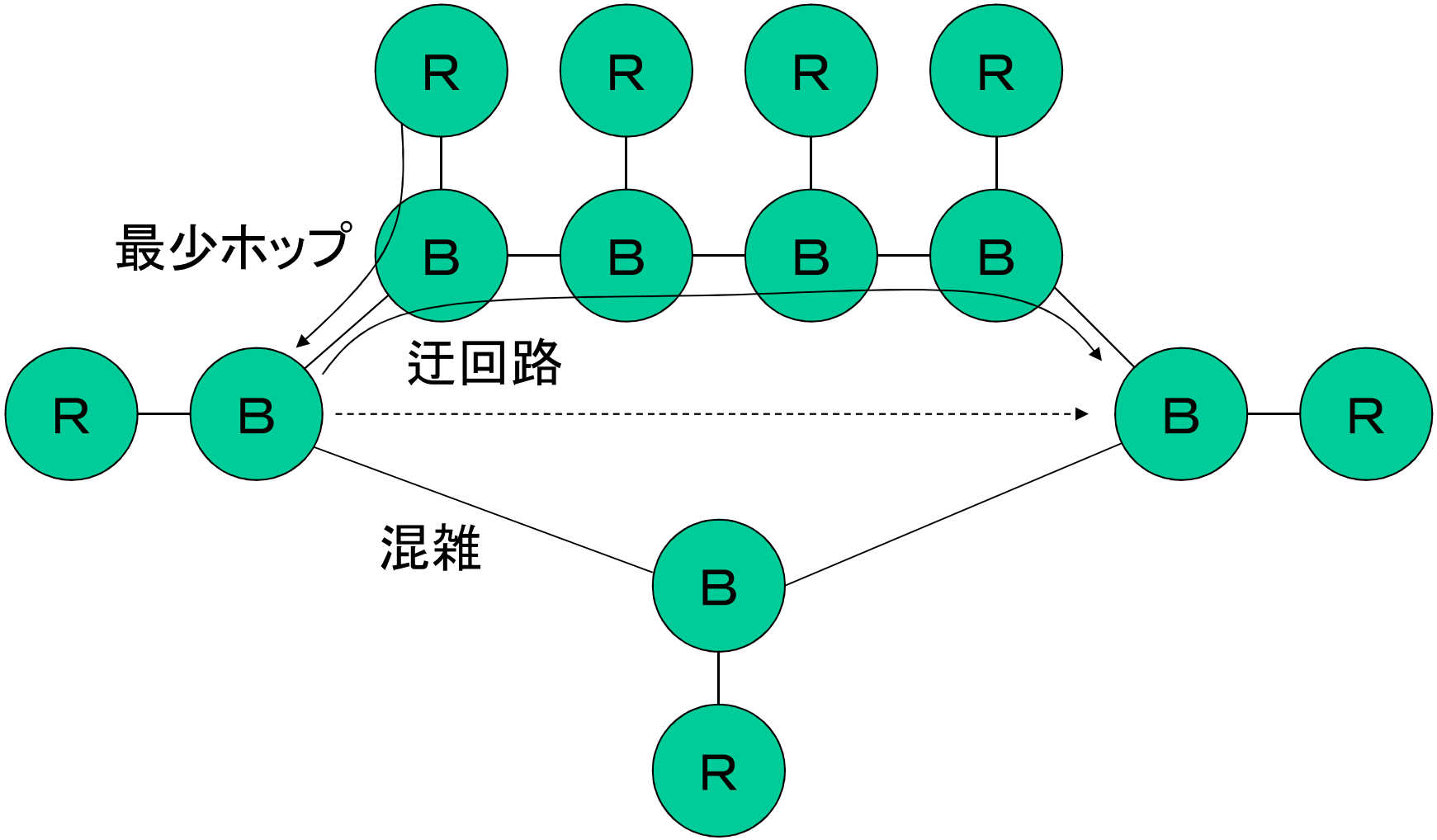
目的領域までの外側ラベルと目的地域内の内側ラベルを付加（領域階層化によりスケール？）

TE （Traffic Engineering）

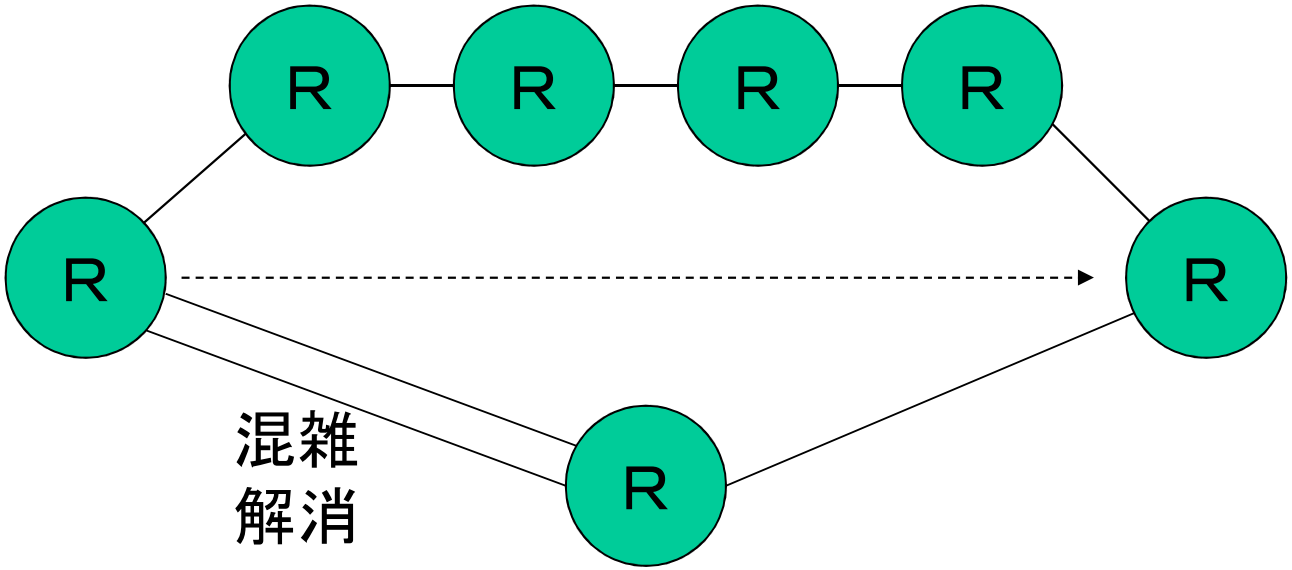
- 本来、トラフィック量を予測して、十分な回線（L1）を敷設しておくこと
  - 電話網ではメインのサービスはL2なので工夫の余地なし
- MPLS生き残りのためには
  - 不十分なL1の上でL3が足りなくなったらL2で迂回路を利用できる
    - L3ルーティングだけでは、迂回ができない
      - 局所的迂回をしようとするとループが発生



L3では局所的な迂回はできない



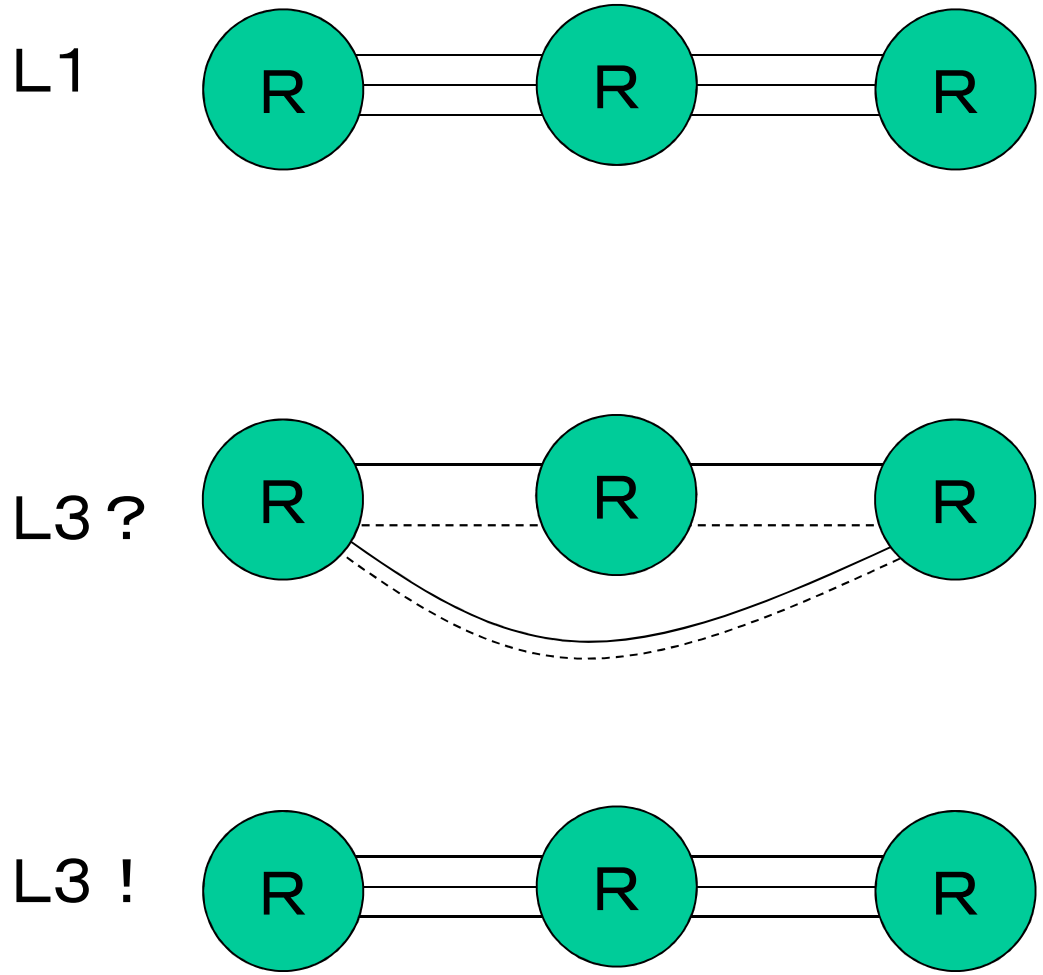
MPLSルータによるL2での迂回



正しい混雑回避（L1＆L3）

MPλS？

- L2スイッチは高価でも、L1スイッチ（鏡）ならそうでもない？
  - 平行リンクが複数あるなら、L3に全て見せたほうがよさげ


L1
R
R
R
L3 ?
R
R
R
L3 !
R
R
R

まとめ

- L2でいろいろやってもL3だけと比べて
  - 機器が複雑化する
  - 管理も複雑化する
  - 速度は遅くなるかせいぜい同程度
- エンドツーエンド原理から
  - L3機器間には余計な機器を入れない
  - 入れても何もやらせない