オーバレイルーティングと
ネットワークただ乗り問題

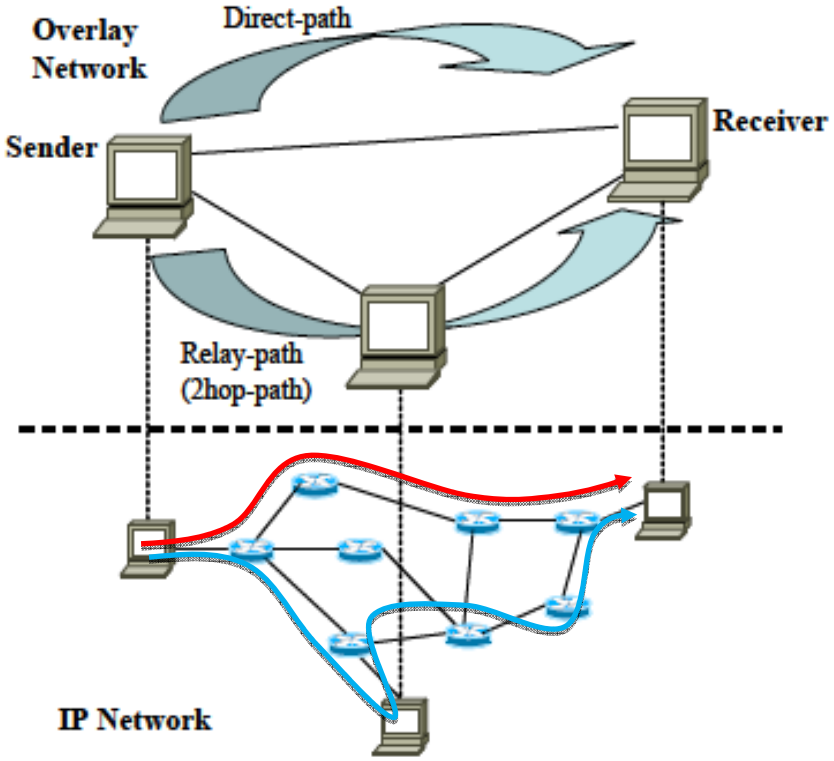
長谷川 剛
大阪大学サイバーメディアセンター

目次

- 背景
- オーバレイルーティング
  - 概要
  - 有効性の評価
- ネットワークただ乗り問題
  - 問題定義
  - 評価指標
  - 評価結果
- まとめと今後の課題

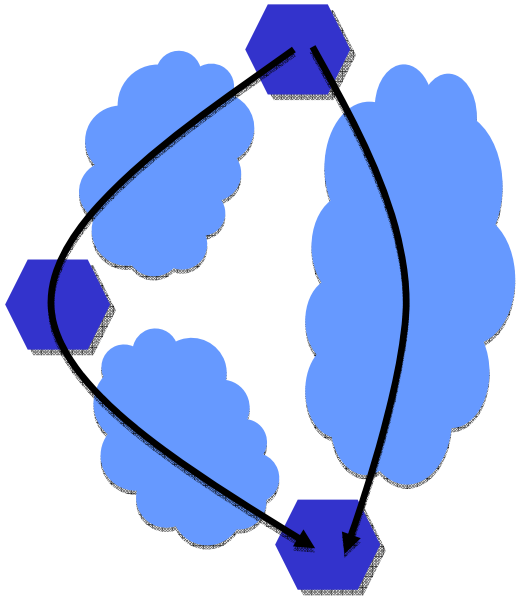
オーバレイルーティング

- オーバレイネットワークで行われるアプケーションレベルの経路制御
  - オーバレイノード間でネットワーク性能の計測
  - エンド間（オーバレイノード間）の性能が向上するように経路を選択
    - 直接経路：送受信ノード間をIPに任せて配送
    - 迂回経路：途中、（単一、または複数の）別のノードを経由させて配送



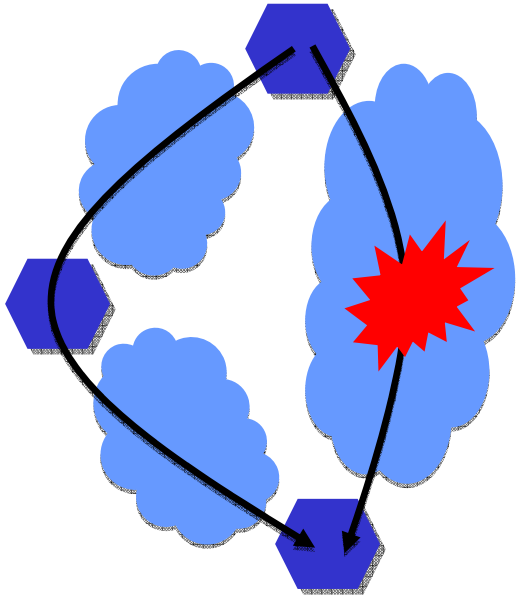
オーバレイルーティングの効果

- エンド間性能の向上
  - 遅延時間の短縮
    - 全体の30-40%の転送において改善可能
    - 最悪値が大きく改善
- 「遠回りした方が近い」
  - 現状のISP間の経路制御が原因
    - ISP間リンクの課金体系
      - トランジット／ピアリング
    - BGPによる経路制御
      - ユーザ性能ではなく、ISPのコスト構造を第一とした設定
  - アプリケーションレベルで行われるオーバレイルーティング
    - 単純にエンド間性能を指標として経路制御を行う
    - トランジット・ピアリングの区別なくリンクを利用



オーバレイルーティングの効果（2）

- ネットワーク障害への対応
  - フルメッシュオーバレイネットワークを構築
  - 定期的にノードの生死確認
    - 当然、オーバヘッドは大きい
  - 障害発生時には、障害地点からの距離（ホップ数）に関係なく、固定時間で障害検出・代替経路を発見
    - BGPの場合、ホップ数に比例した時間が必要



発表内容

- ネットワーク帯域に関する指標を用いたオーバレイルーティングの評価
  - 利用可能帯域、TCPスループット
    - 転送データサイズの大きなアプリケーションにとって重要な指標
- オーバレイルーティングの負の側面の評価
  - ネットワークただ乗り問題

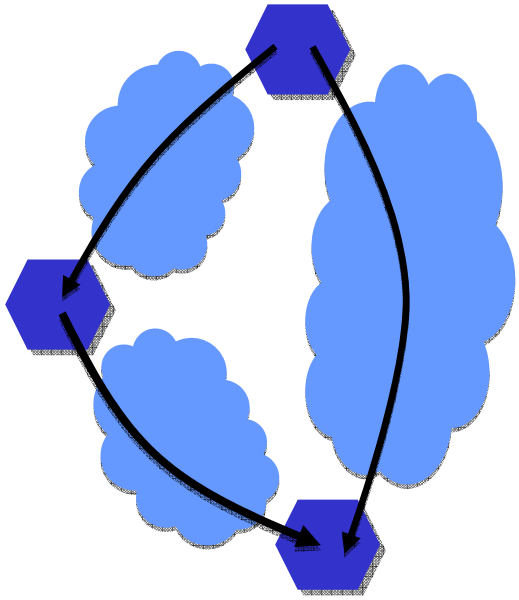
ネットワーク帯域に関する指標を用いた オーバレイルーティングの評価

使用データ

- データ元：Scalable Sensing Service（S-cube）
  - ネットワーク環境： PlanetLab
  - 参加ノード間の遅延時間，利用可能帯域，パケット廃棄率を計測
  - 計測日時：2006年10月25日
  - 参加ノード数：588 （AS数 179）
  - 参加ノードを AS 番号ごとにグルーピング
    - AS ごとにオーバレイノードが 1つずつあると仮定
      - 同一ネットワーク内で中継することは無意味
      - AS 間のデータが複数存在するときはその値の平均値をとる

オーバレイルーティングの指標（1）

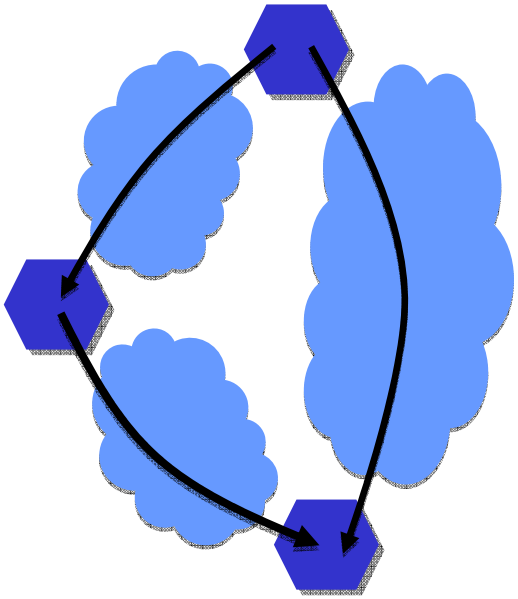
- 迂回経路は3ホップパスまでを利用
- 遅延時間
  - S-cubeの計測データを利用
  - 迂回経路：和を利用
- 利用可能帯域
  - S-cubeの計測データを利用
  - 迂回経路：最小値を利用



オーバレイルーティングの指標(2)

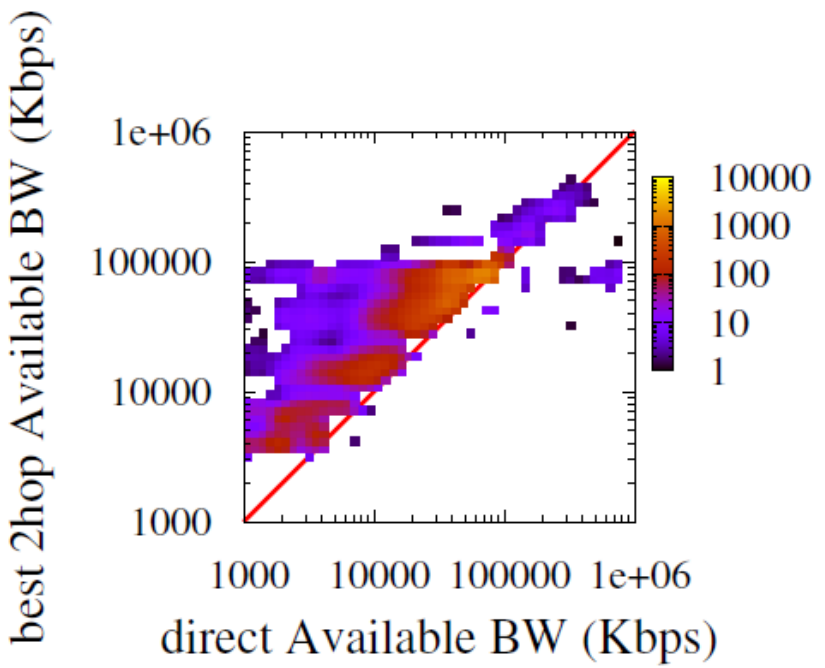
- TCPスループット
  - 遅延時間、利用可能帯域から算出

$$P_{ij}^{1} = \min\left(\frac{(8 \cdot MSS)\sqrt{1.5}}{D_{ij}^{1}\sqrt{n_{ij} \cdot L}} \ , \ B_{ij}^{1} \right) \ \text{(bps)}$$

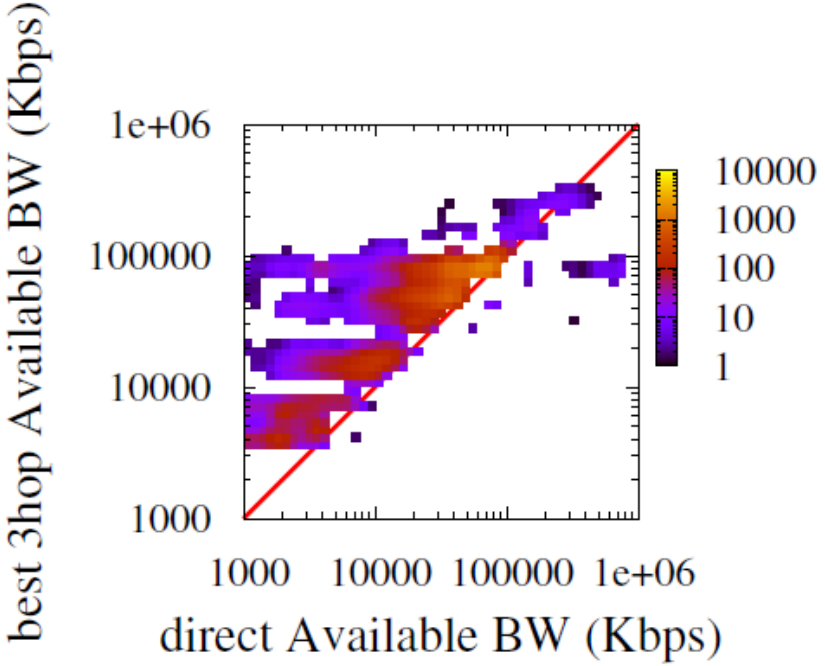


  - ロス率はパラメータとする
    - 経路のASホップに比例
    - オーバレイレベルのホップ数に比例
  - 迂回経路
    - TCPを終端する場合：各途中経路の算出値の最小値
    - TCPを終端しない場合：迂回経路の遅延時間、利用可能帯域から算出

直接パス，迂回パスの性能比較（利用可能帯域）



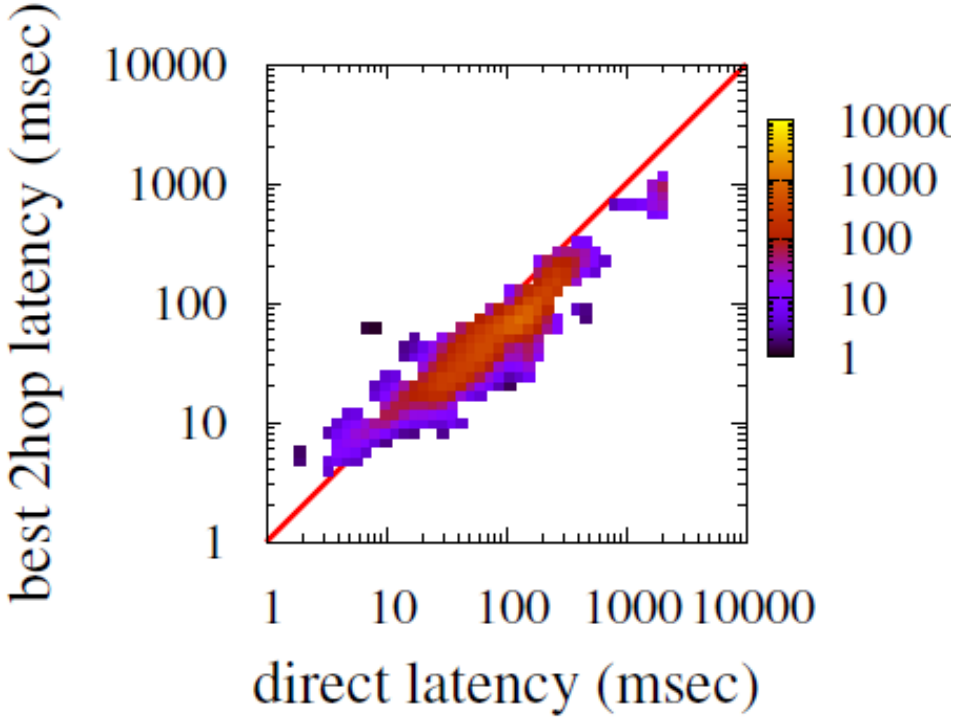
(a) 最適 2 ホップパス



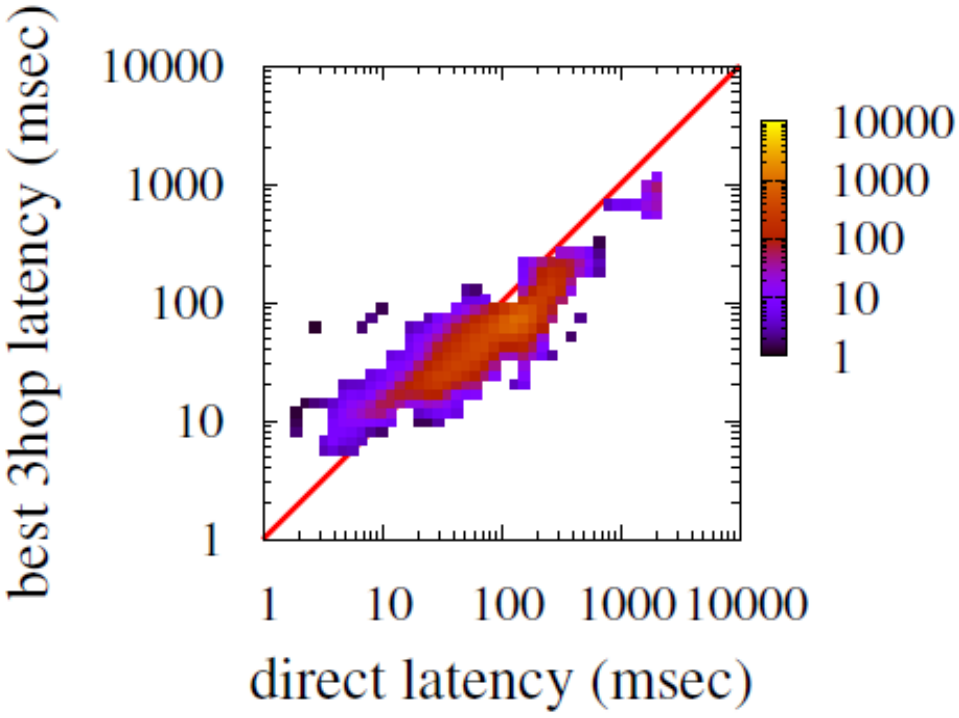
(b) 最適 3 ホップパス

- 迂回パスが直接パスより有効となる割合
  - 2ホップ迂回パス： 96.6%、3ホップ迂回パス： 97.7%
- 3ホップ迂回パスの有効度
  - 2ホップ迂回パスで改善せず，3ホップ迂回パスで改善する割合： 46.9%
  - 2ホップ迂回パスで改善し，3ホップ迂回パスでさらに改善する割合： 51.6%

直接パス，迂回パスの性能比較（遅延時間）



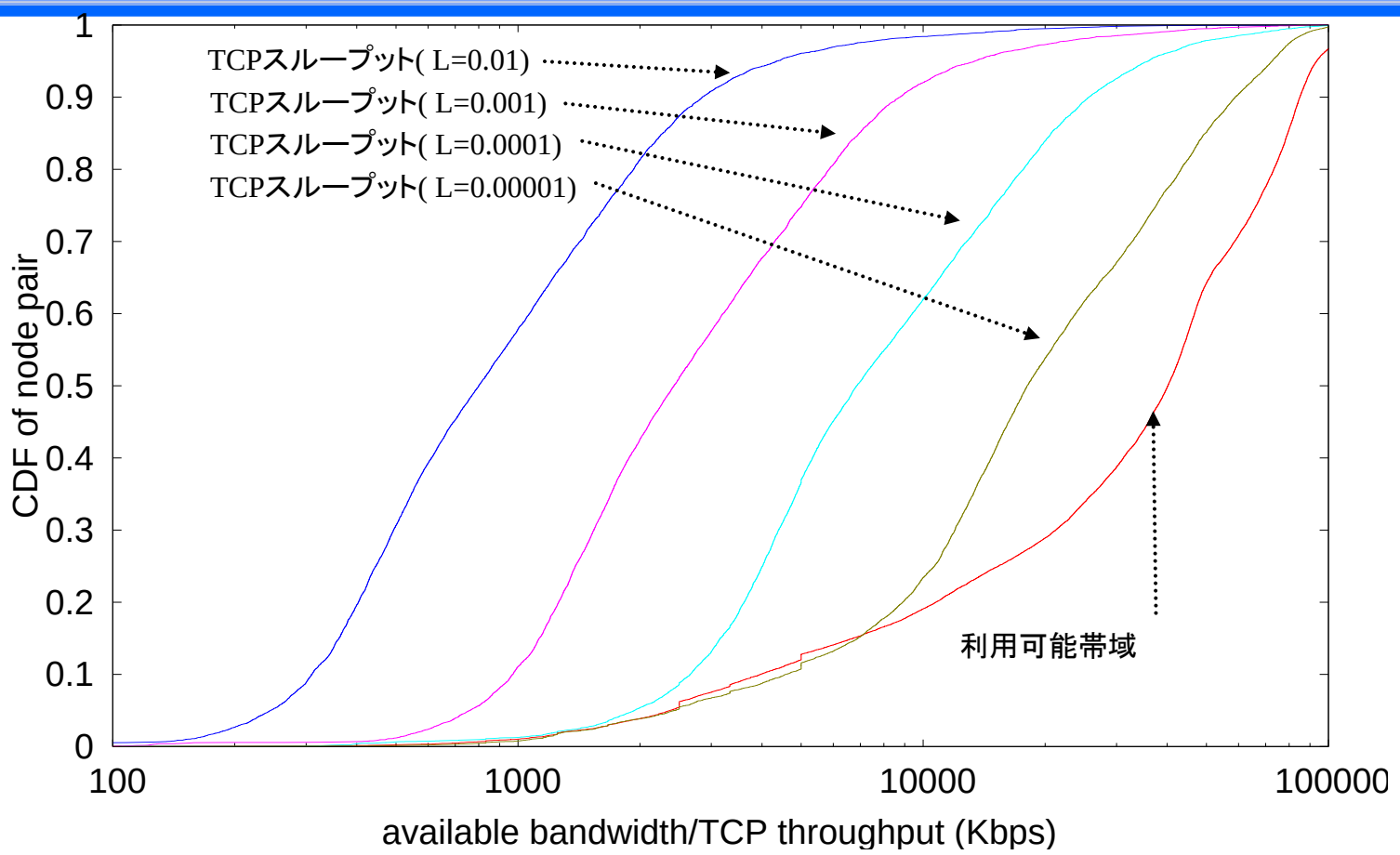
(a) 最適 2 ホップパス



(b) 最適 3 ホップパス

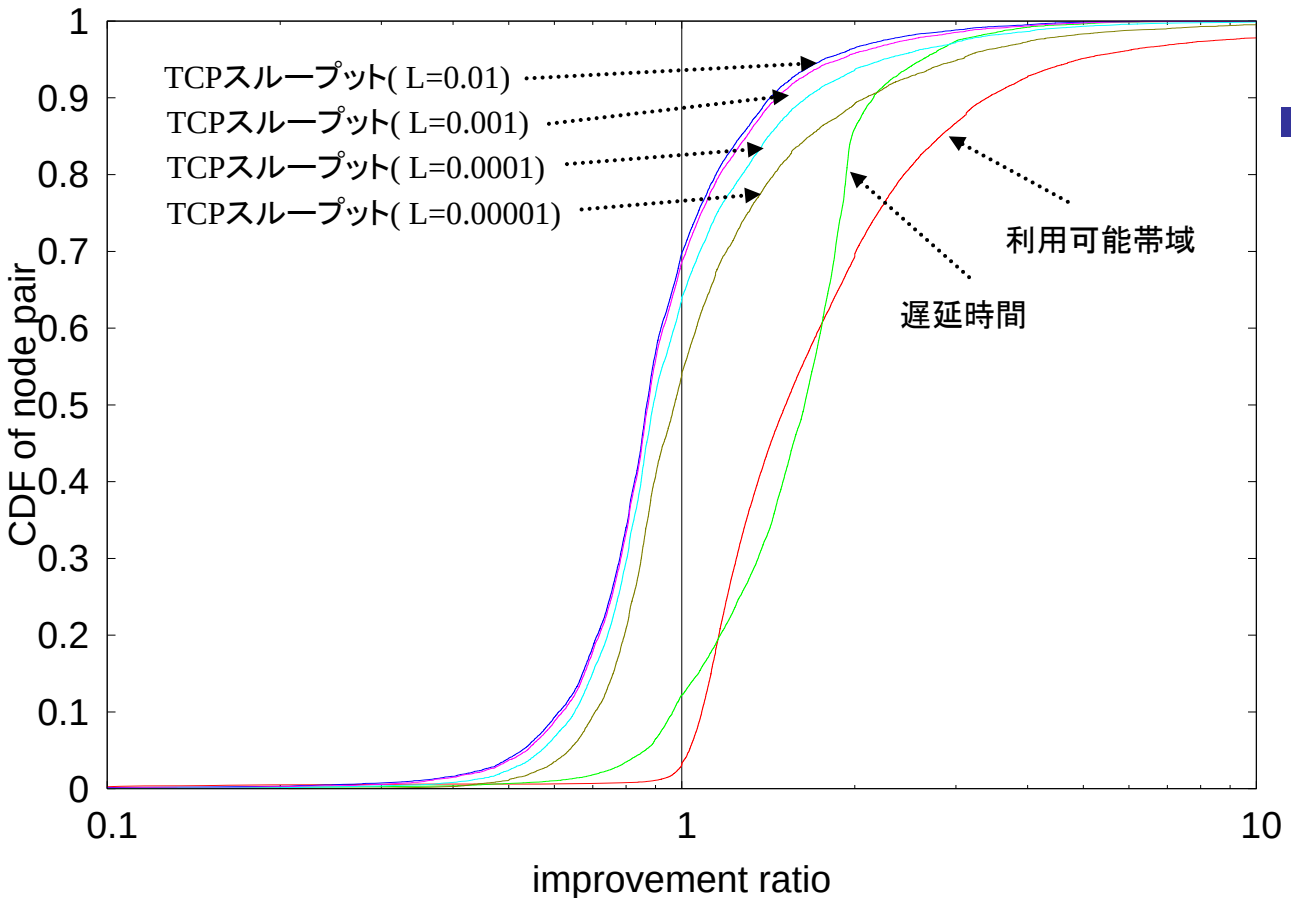
- 迂回パスが直接パスより有効となる割合
  - 2ホップ迂回パス： 87.5%、3ホップ迂回パス： 85.4%
- 遅延時間の改善度が利用可能帯域と比べて小さい
  → IP ルーティングがホップ数を基準としているため

TCPスループットの分布



- L= 0.00001のときでも利用可能帯域と大きな差
  - PlanetLab の環境が，ネットワーク資源が潤沢であるため，TCP スループット値に対し，利用可能帯域が大きいことが原因

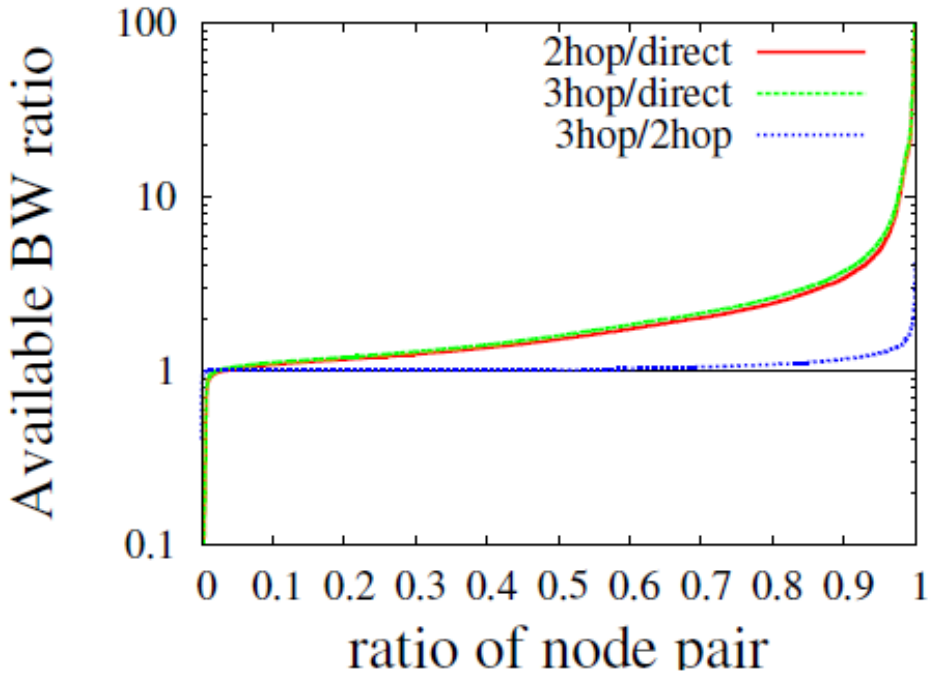
最適迂回パスの性能分布



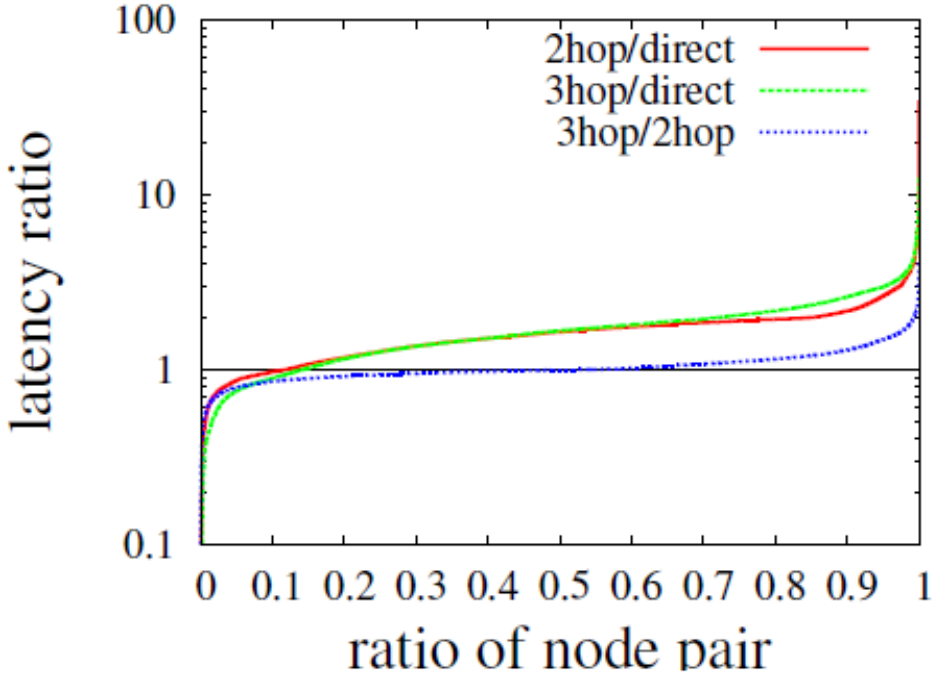
- 性能改善度:
  直接パスの性能を 1 としたときの迂回パスの性能

- **TCP スループットを用いたときは迂回パスの効果が小さい**
  ⇒ 迂回パスは経路長が長くなり、RTT・パケット廃棄率が増加
  - **中継ノードで TCP コネクションを終端することで改善可能**

3ホップパスの効果



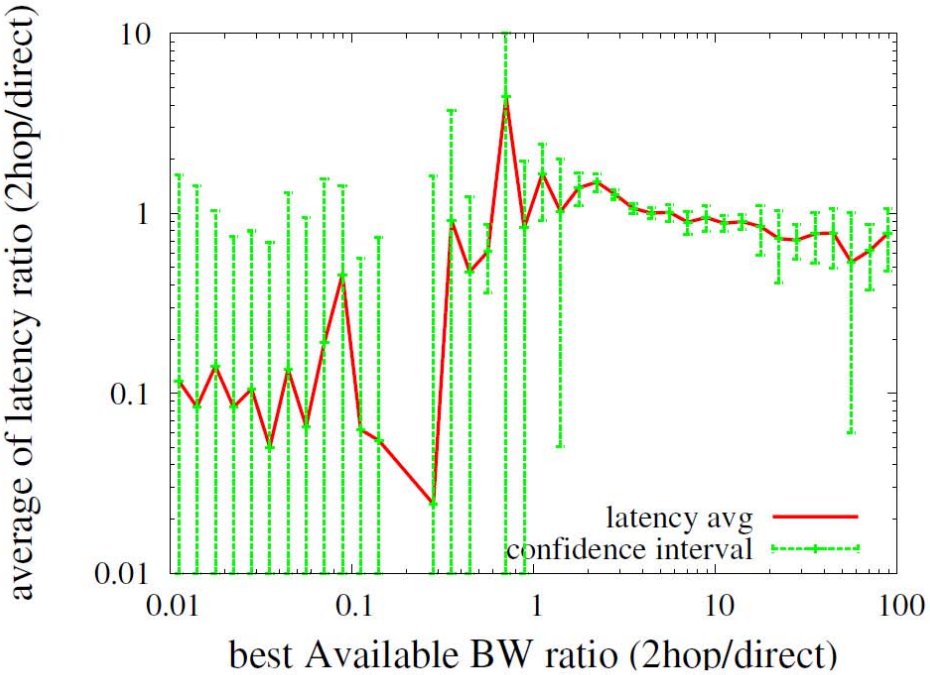
(a) 利用可能帯域



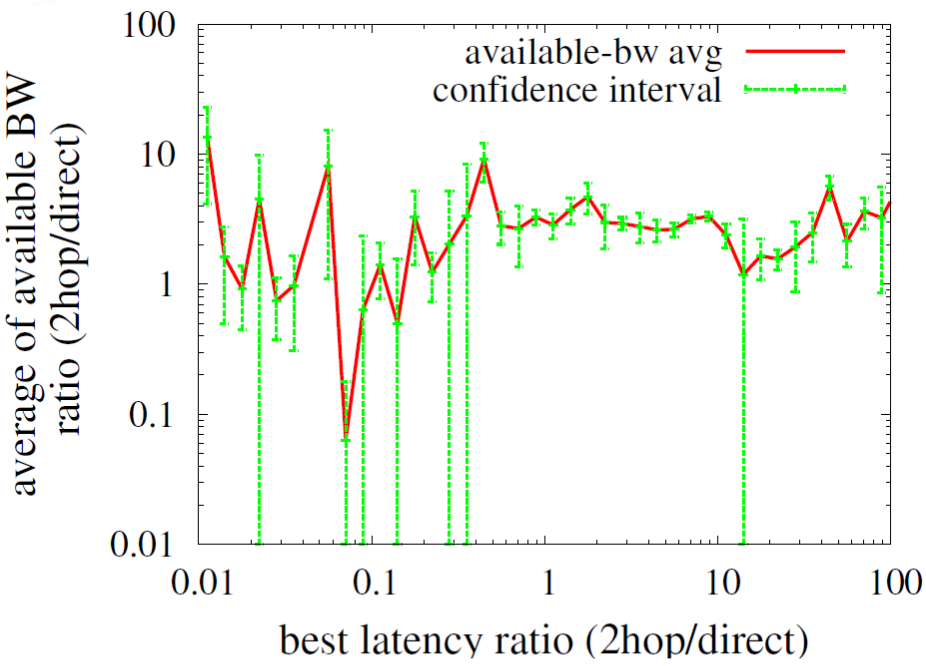
(b) 遅延時間

- 3ホップ迂回パス / 2ホップ迂回パスは、ほとんどが1付近
  - 3ホップ迂回パスを利用することでより良いパスが見つかる場合はあるが，その改善度は低い

帯域と遅延の相関



利用可能帯域がベストのパスの遅延



遅延がベストのパスの利用可能帯域

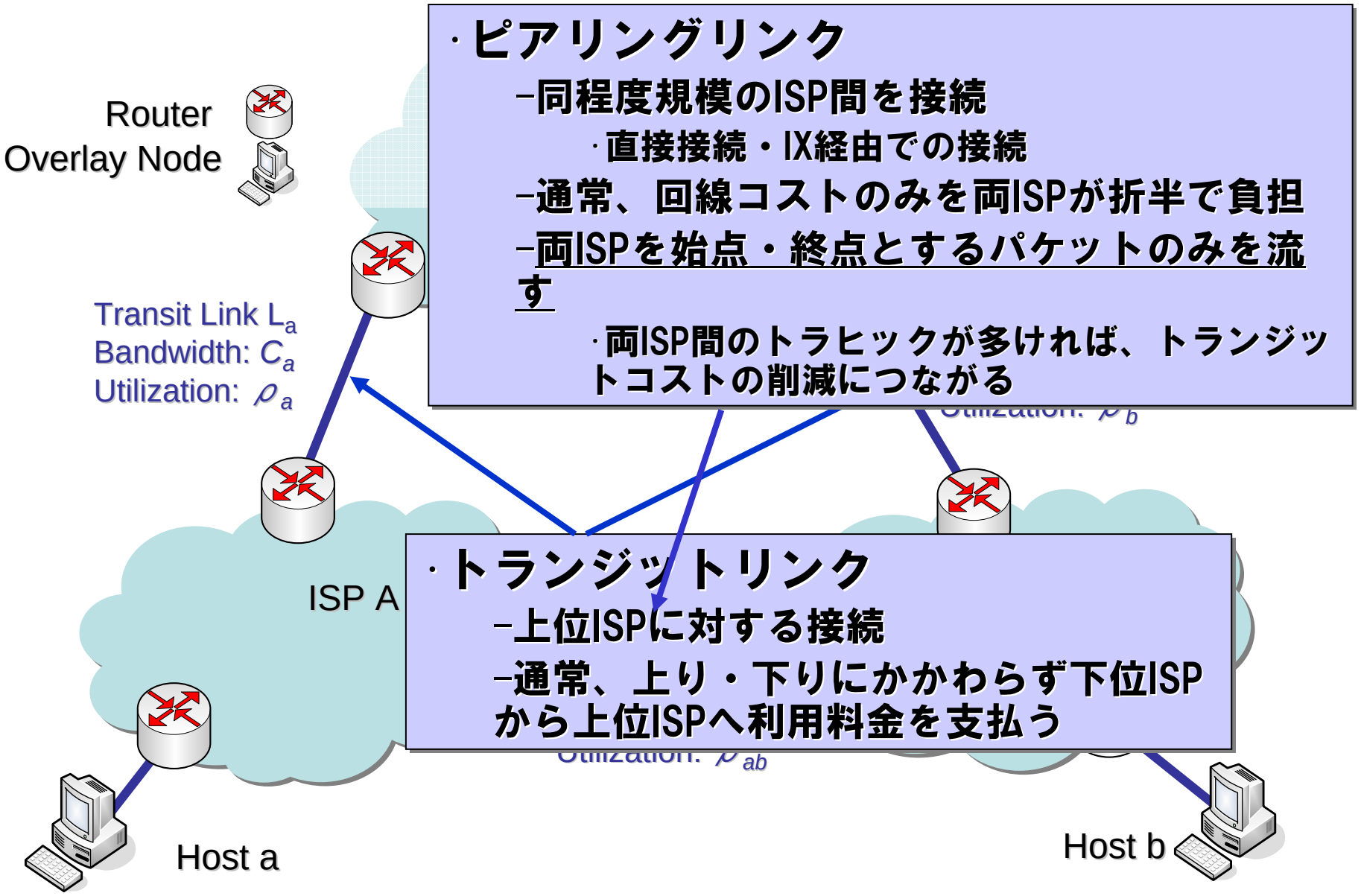
- 利用可能帯域がベストのパスを選択すると、そのパスの遅延は直接パスに比べて悪化
- 遅延がベストのパスを選択すると、そのパスの利用可能帯域は直接パスに比べて良化

まとめ

- 利用可能帯域を指標としたオーバレイルーティングにおいて、迂回した方が良い可能性は高い
  - ほとんど全ての転送で、良い迂回経路が存在
- 3ホップパスは、シングルパス転送においては用いる必要はない
  - マルチパス転送では利用価値がある
- 遅延の良いパスの帯域は良いが、帯域が良いパスの遅延は良くない
  - 遅延で経路を探せば、それなりに帯域の良いパスは発見できるが、本当に帯域が良いパスは帯域で探さないと発見できない

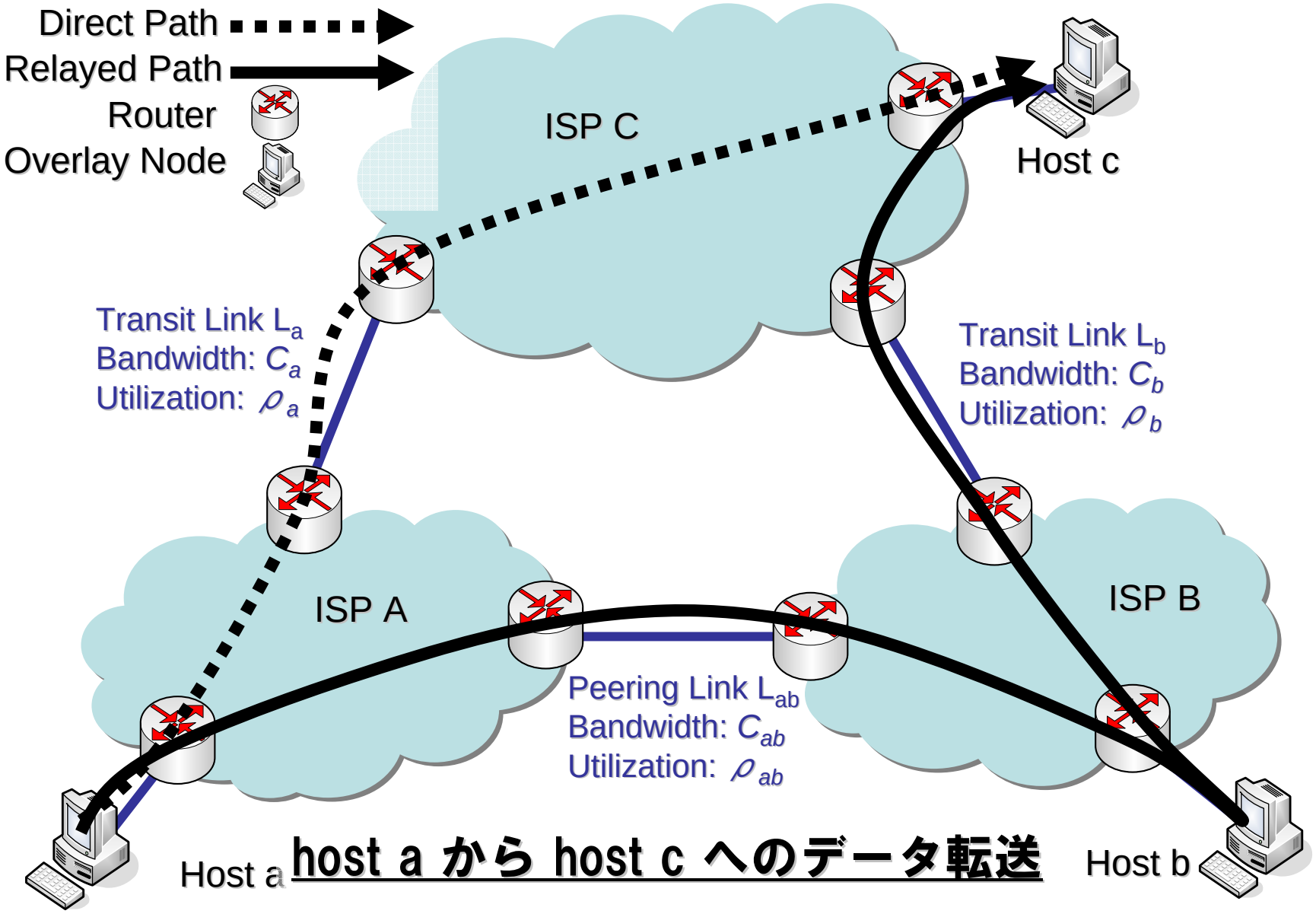
ネットワークただ乗り問題

ネットワークモデル（1）



- ピアリングリンク
  - 同程度規模のISP間を接続
    - 直接接続・IX経由での接続
  - 通常、回線コストのみを両ISPが折半で負担
  - 両ISPを始点・終点とするパケットのみを流す
    - 両ISP間のトラヒックが多ければ、トランジットコストの削減につながる
- トランジットリンク
  - 上位ISPに対する接続
  - 通常、上り・下りにかかわらず下位ISPから上位ISPへ利用料金を支払う

ネットワークモデル（2）

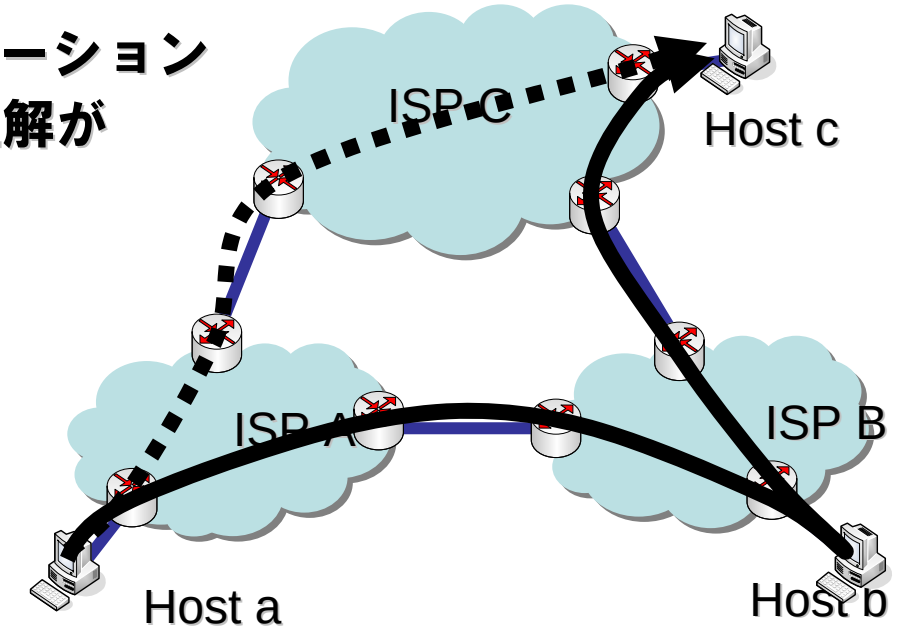

Direct Path
Relayed Path
Router
Overlay Node
ISP C
Host c
Transit Link L_a
Bandwidth: C_a
Utilization: ρ_a
Transit Link L_b
Bandwidth: C_b
Utilization: ρ_b
ISP A
ISP B
Peering Link L_{ab}
Bandwidth: C_{ab}
Utilization: ρ_{ab}
Host a
host a から host c へのデータ転送
Host b

オーバレイトラヒックのコスト負担

- Direct pathを使う場合
  - ISP A 内の host a が、ISP Aのトランジットリンクを使ってトラヒックを運ぶ
    - host a から課金可能
- Relay pathを使う場合
  - ISP A 内の host a が、 ISP A-B 間のピアリングリンクと、ISP B のトランジットリンクを使ってトラヒックを運ぶ
    - ピアリングリンク：ISP A, Bで折半
    - ISP B のトランジットリンク：ISP Bが負担
  - ネットワークただ乗り問題
    - ISP A 内のホストが発生させたトラヒックのトランジットコストを、直接的には関係のない ISP B が負担
  - ピアリングリンク上のトラヒック量に偏りが発生すると、コスト負担を要求されることもある

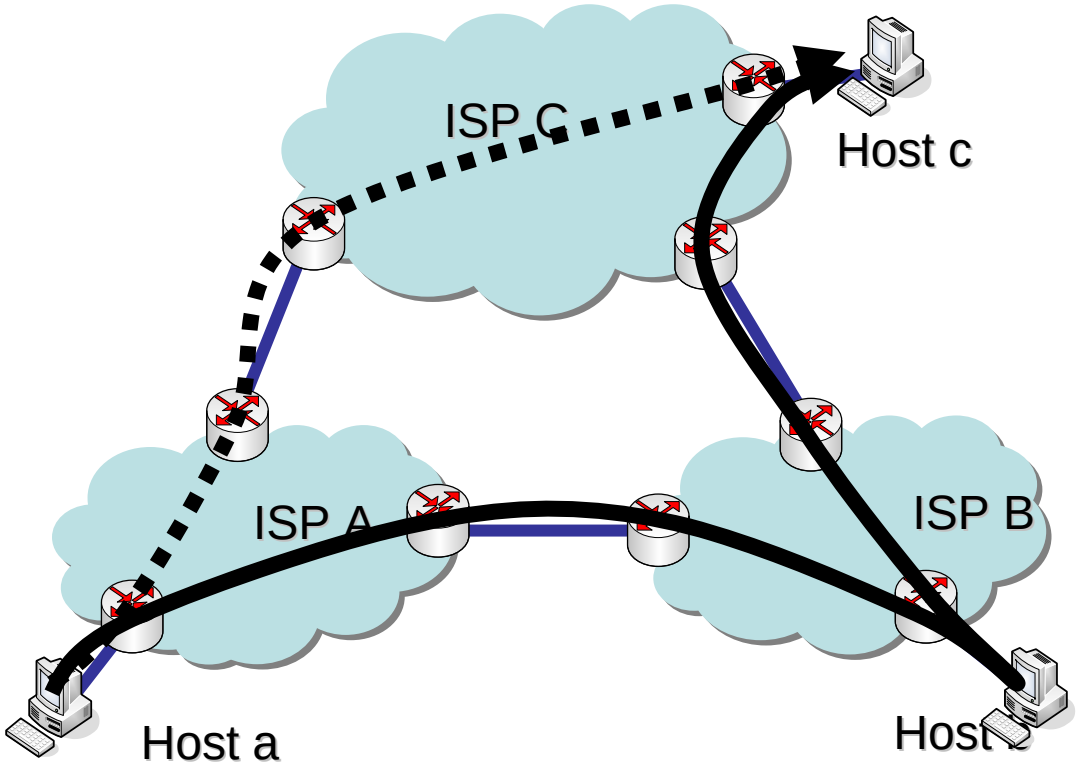
考えられる解決方法とその問題点（1）

- オーバレイネットワークに参加している host b から課金する
  - ユーザは自分のPCが中継に使われていることを自覚してない
- アプリケーションそのものを遮断・制限
  - Winny であれば可能
  - Skype は？
    - 「有用」なアプリケーションの場合、ユーザの理解が得られない可能性



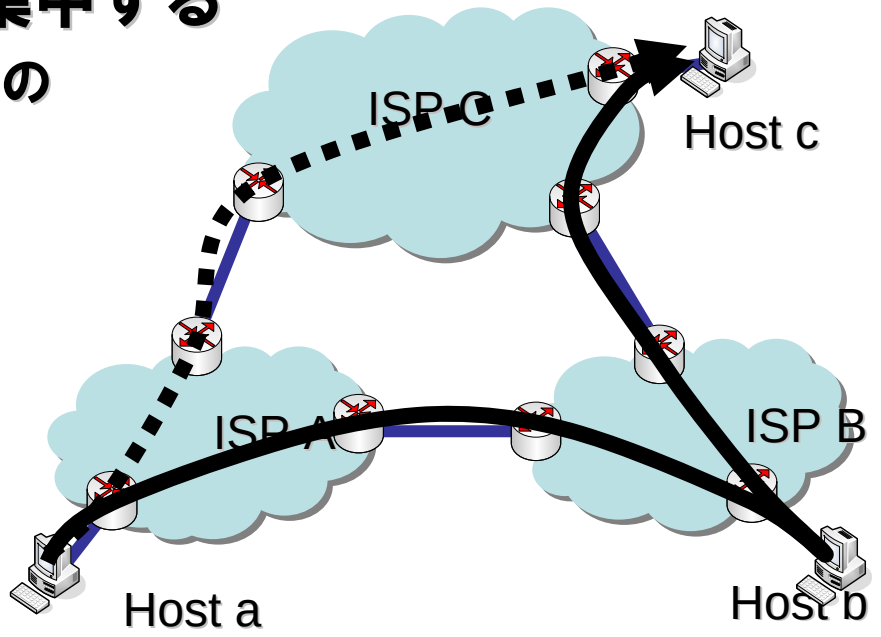
考えられる解決方法とその問題点（2）

- Relayed pathを使うトラヒックを検出し、課金／遮断する
  - 通常のトラヒック（ISP A→ISP B）と区別できない
    - どちらも src: host a、dst: host bのパケット



考えられる解決方法とその問題点（3）

- ISP Bが設備投資を行い、トランジットリンク帯域を増強
  - オーバレイトラヒックのうち、中継経路（Relayed Path）が使われる割合が増える
    - ISP A内のユーザが恩恵を受ける
    - ISP B内のユーザには思ったような効果が得られない
- 品質の良いネットワークを構築した結果、そこへオーバレイトラヒックが集中する
  - そのコストをトラヒックの発生元から回収できない



Relayed pathを使うトラヒック量（1）

- 利用可能帯域の比で経路を選択する場合
  - ISP Bがトランジットリンク帯域を増強することによって、流れ込んでくるオーバレイトラヒック量

$$\Delta x_r = \frac{A_a(C_b' - C_b)}{(A_a + C_b'(1-\rho_b'))(A_a + A_b)}x, \quad \lim_{C_b' \to \infty} \Delta x_r = \frac{A_a}{(A_a + A_b)}x$$

  - 帯域増強が進むと、オーバレイトラヒックのほとんど全てが中継経路へ流れる
- 利用可能帯域の大きい経路を選択する場合

$$\Delta x_r = \frac{1}{2}(C_b' - C_b)$$

  - 増強した帯域の50%を、オーバレイトラヒックの中継に使われる

Relayed pathを使うトラヒック量（2）

- 遅延時間(RTT)の小さい経路を選択する場合
  - 各経路のRTT

$$RTT_{d,r} = \tau_{d,r} + \frac{1}{\mu_{d,r}(1-\rho_{d,r})}$$

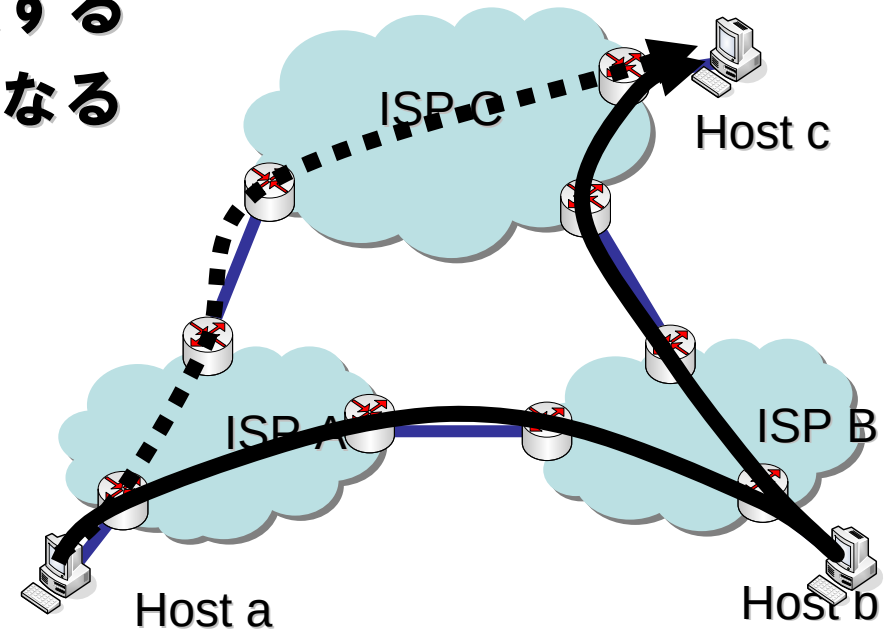
  - 経路を流れるトラヒック量が増えると、RTTが増大する
  - RTTが等しくなる時の、それぞれの経路の帯域利用率の差

$$\Delta\rho = \frac{\Delta\tau \cdot \mu(1-\rho_d)^2}{1-\Delta\tau \cdot \mu(1-\rho_d)}$$

  - ネットワーク全体の負荷が高いと、中継経路がより多く使われる

Relayed pathを使うトラヒック量（3）

- ボトルネックが両経路が通過する場所（例: host cに近い場所）にある場合
  - オーバレイトラヒック量に関係なく、使われる経路は固定的になる
  - RTTで経路を選択する場合、常に全てのトラヒックが中継経路を流れる場合もある
  - 利用可能帯域を選択する場合は、50%ずつになる



ただ乗り問題の評価指標

ただ乗り量

- 迂回パスを構成するリンクの集合の中で直接パスで使用していないトランジットリンクの集合の要素数

$$F_{ikj} = \{x | (x \in T_{ikj}) \& (x \notin T_{ij})\}$$

  - ノードペア N_i, N_j 間のデータ転送
  - 直接パスのトランジットリンクの集合 T_{ij}
  - ノード N_k を中継する迂回パスのトランジットリンクの集合 T_{ikj}
  - 直接パスで使用していない迂回パスのトランジットリンクの集合 F_{ikj}
- F_{ikj} の要素数 $|F_{ikj}|$ をノード N_k を中継する迂回パスのただ乗り量と定義する

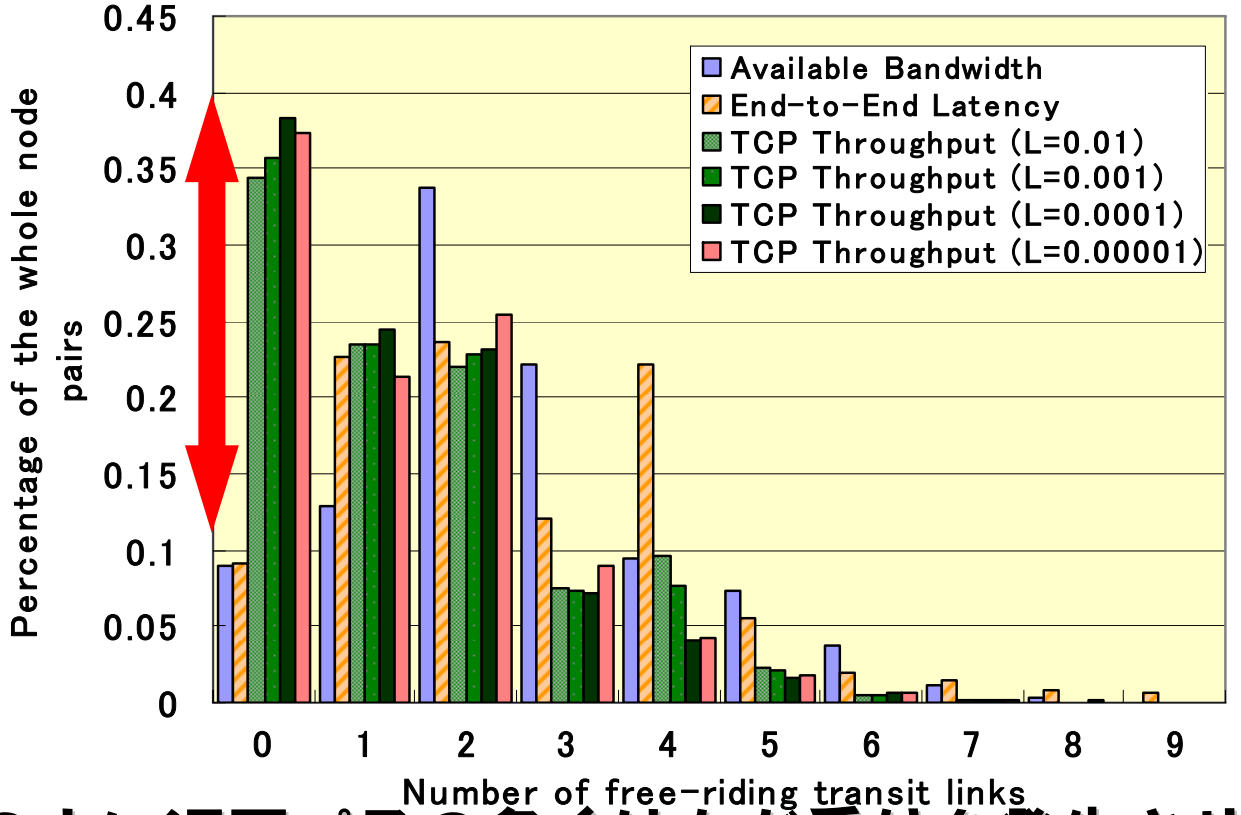
ただ乗り問題評価のためのデータ

- PlanetLab ノード間のネットワーク性能の計測結果
  - S-cube が計測，公開しているデータを使用
    - 物理帯域，利用可能帯域，遅延時間，パケット廃棄率
- PlanetLab ノード間の経路を構成するリンクの種類
  - AS 間リンクの種類
    - CAIDA が調査，公開しているデータを使用
      - BGP テーブルの情報と AS の次数（他 AS とのリンク数）から推測［*］
  - PlanetLab ノード間の AS レベルの経路
    - traceroute コマンドでルータレベルの経路を取得
    - traceroute.org のルートサーバから各ルータのAS番号を取得
  - リンクの種類が判別できない AS 間リンクが多数存在
    - ［*］の調査で用いられた BGP テーブルから得られないリンクが多い

⇒ ピアリングリンクと仮定してただ乗りに寄与しないと仮定

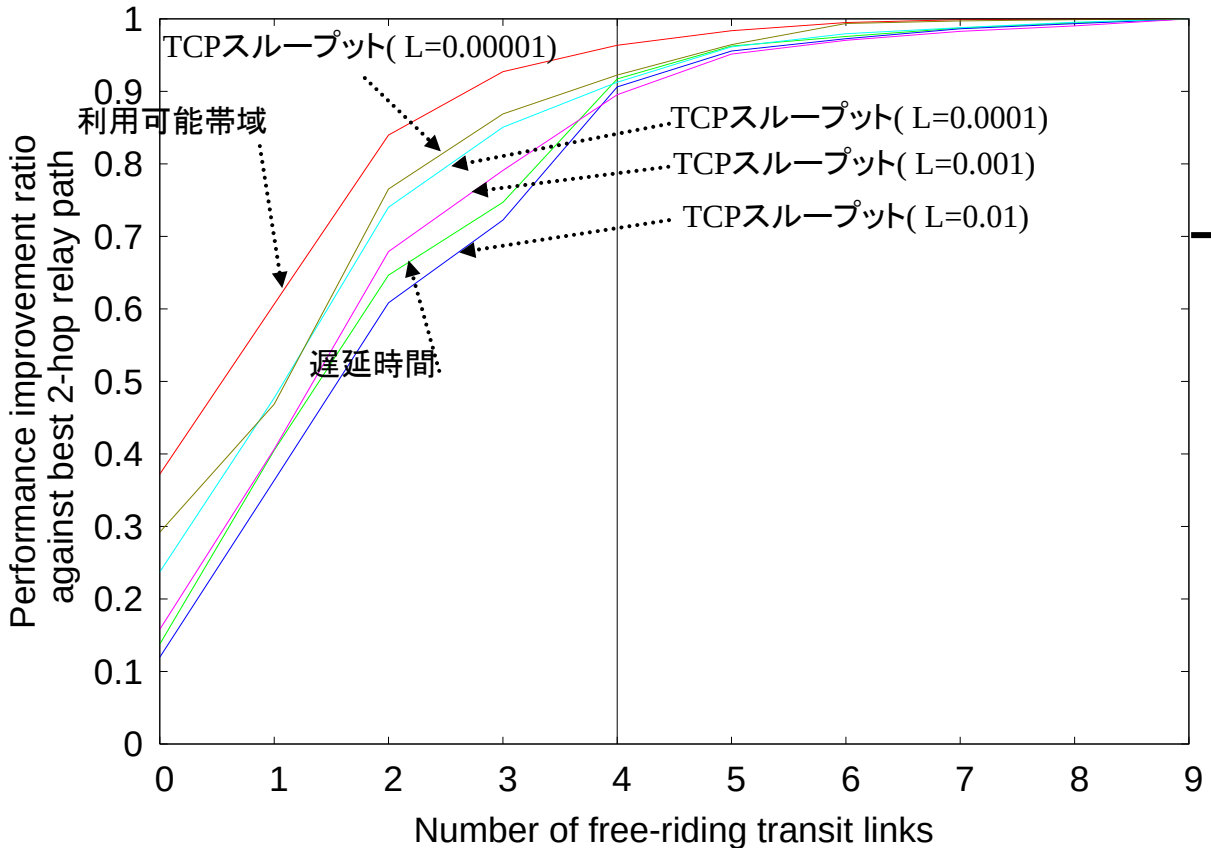
[13] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, K. claffy, and G. Riley, "AS relationships: Inference and validation," *ACM SIGCOMM Computer Communication Review*, vol. 37, No.1, Jan. 2007.

ただ乗り量の分布



- 性能のよい迂回パスの多くはただ乗りを発生させる
- TCP スループットをメトリックとしたときはただ乗り量の少ない　　最適迂回パスの割合が多い
  - AS ホップ数が小さいパスが最適迂回パスとして選ばれやすいから

ただ乗り量の上限と性能改善度の関係



■ Y軸:

（ただ乗り量が上限以下の迂回 パス中で最適な迂回パスの性能）

―――――――――――――――

（ただ乗り量の制限が無いときの 最適迂回パスの性能）

- ただ乗りを許容しない場合、オーバレイルーティングによる性能向上は限定的
- ただ乗り量を 4本まで許容することで、85-90%程度の性能を得られる
- 利用可能帯域を用いる場合、より少ないただ乗り量で高い性能が得られる

まとめ

- ネットワークただ乗り問題の定量的評価
  - 性能が改善する迂回パスの多くでただ乗りが発生
  - 問題の緩和と性能改善の両方を実現するパスの評価
    - 利用可能帯域をメトリックとした場合，他のメトリックに比べただ乗り量の上限数が少なくても性能が改善する割合が高い