JANOG 14 **〜激濃ぃ〜@宮崎市民プラザ** *2004.07.23*

オーバーレイネットワークの可能性とそのインパクト

NTTサービスインテグレーション基盤研究所
情報流通トラヒックサービス品質プロジェクト
亀井聡
kamei.satoshi@lab.ntt.co.jp

発表の概要

- 技術背景
- オーバーレイネットワークとは
  - 概説
  - オーバーレイネットワークとPeer-to-Peer
- 有効な適用先は？
  - インターネット上でのトラヒックエンジニアリングとその限界
  - エンドホストオーバーレイによるトラヒックエンジニアリング
  - ビジネスモデル
- ISP間測定によるインターネットの現状
  - 測定条件
  - P2P技術を用いた分散測定法
  - 品質値の分布
  - 有効領域, AS分析
  - 考察
- まとめと今後の課題

チュートリアル的なとこ

有効な利用法の提案

希望する議論のポイント

提案手法の技術的妥当性

一般的オーバーレイ及び提案手法のビジネス的妥当性

測定データの中身 / 利用法

本日のメイン

- **技術背景**

- **オーバーレイネットワークとは**
  - **概説**
  - **オーバーレイネットワーク**とPeer-to-Peer

- **有効な適用先は？**
  - **インターネット上でのトラヒックエンジニアリングとその限界**
  - **エンドホストオーバーレイによるトラヒックエンジニアリング**
  - **ビジネスモデル**

- ISP**間測定によるインターネットの現状**
  - **測定条件**
  - P2P**技術を用いた分散測定法**
  - **品質値の分布**
  - **有効領域**, AS**分析**
  - **考察**

- **まとめと今後の課題**

➢ IPネットワークでの高品質，高信頼性の要求

- インフラとしてのIPネットワークの普及
- VoIPをはじめとしたリアルタイムアプリケーションの増加
- ISP跨ぎでもある程度以上の品質が要求されることも

➢ End-to-End 環境の複雑化

- アクセス環境の広帯域化，多様化
- バックボーンでのボトルネックも発生
- End-to-End のどこがボトルネックになるか予測困難

➢ エンドホストオーバーレイの実用化

- エンドホストの回線環境，処理能力の著しい向上
- P2Pファイル共有では数100万ノードのオーバーレイネットワークが実現されている

- 技術背景
- オーバーレイネットワークとは
  - 概説
  - オーバーレイネットワークとPeer-to-Peer
- 有効な適用先は？
  - インターネット上でのトラヒックエンジニアリングとその限界
  - エンドホストオーバーレイによるトラヒックエンジニアリング
  - ビジネスモデル
- ISP間測定によるインターネットの現状
  - 測定条件
  - P2P技術を用いた分散測定法
  - 品質値の分布
  - 有効領域, AS分析
  - 考察
- まとめと今後の課題

オーバーレイネットワークとは　～　概説

オーバーレイネットワーク：
既存のネットワークリンクを用いて，その上位層において
目的に応じた仮想的なリンクを形成し，構成するネットワーク．

- **初期のインターネットは電話網のオーバーレイネットワーク？**
- **固定ノードと動的ノード**
  - **かつては固定ノードによる提供が流行** (mbone etc…)**．**
  - **最近はエンドホスト(動的ノード)によって構成されるエンドホストベースのオーバーレイネットワーク流行の兆し ⇒** Peer-to-Peer **もオーバーレイの一種**
- **ビジネス的には？ オーバーレイに課金できるのか？**

➢ 様々なオーバーレイネットワーク

- mbone, 6bone, qbone
  - IP層に multicast や IPv6, DiffServ といった機能を追加
- Akamai / Internap
  - 主に WWW サービス向けの CDN ネットワークを提供 (物理網も込み)
- QRON, SON
  - 固定的に設置された QoS ノード間で品質を管理, 利用者の品質を向上させる枠組み.
- Skype
  - VoIP に特化した P2P ネットワークとソフト. NAT 越えが可能

P2P

- Groove / ArielAirOne
  - P2P ネットワークによって維持されるグループウェア
- Gnutella / Winny
  - P2P ファイル共有ネットワークとソフト

➢ オーバーレイネットワークの中での P2P 技術の位置

- エンドホストオーバーレイネットワークの一種
- 自律性を備えていることがほとんど
- ノードの信頼性は低いが，数の力でそれを補う手法が主流
- スケーラビリティを保ったまま，多数のユーザーノード間でリソースを効率的に交換する機能に優れる．

  - ファイル / ストレージ共有
  - 掲示板 / チャット / 作業内容の共有 (グループウェア)
  - ネットワーク接続環境の融通 (NAT越え)

- 固定ノードベースのオーバーレイネットワークとの大きな違いは，フラッディングや分散ハッシュテーブルといった技術

➢ P2P技術の例

– 分散コンピューティング

- SETI@Home
- United Device の各種プロジェクト

– コミュニケーション／コラボレーション

- Skype
- Groove
- ArielAirOne

– 汎用プラットフォーム/ツールキット

- JXTA
- SIONet
- SOBA Project
- Microsoft Windows XP Peer-to-Peer SDK

etc…

- **フラッディングと分散ハッシュテーブル** (DHT)
  - **フラッディング**
    - **ファイル共有ソフトで広く使われている方式．アプリケーションレイヤーでのフラッディング(広報)によりノードの生死を始めとした様々な情報を伝播する．**
  - **分散ハッシュテーブル**
    - Chord **や** CAN，Tapestry **等が広く知られる．** structured-P2P **とも呼ばれる．**
    - **ハッシュテーブル**：hash(“contents name”)=> contents **等としたテーブル．**
    - DHT: **ハッシュテーブルを分割して持つ(ノード毎に持つ**ID**範囲が決まる)ことにより高速** O(logN) **な検索を実現．**
    - ID**による検索に限定されるため，** DNS**的機能の代替等によく用いられる．ユビキタス方面とか．**

- 技術背景
- オーバーレイネットワークとは
  - 概説
  - オーバーレイネットワークとPeer-to-Peer
- 有効な適用先は？
  - インターネット上でのトラヒックエンジニアリングとその限界
  - エンドホストオーバーレイによるトラヒックエンジニアリング
  - ビジネスモデル
- ISP間測定によるインターネットの現状
  - 測定条件
  - P2P技術を用いた分散測定法
  - 品質値の分布
  - 有効領域, AS分析
  - 考察
- まとめと今後の課題

有効な適用先は？

- **利点**
  - 下位ネットワーク非依存
  - 高スケーラビリティ
  - 高い自律性，対故障性能
  - リソース配分

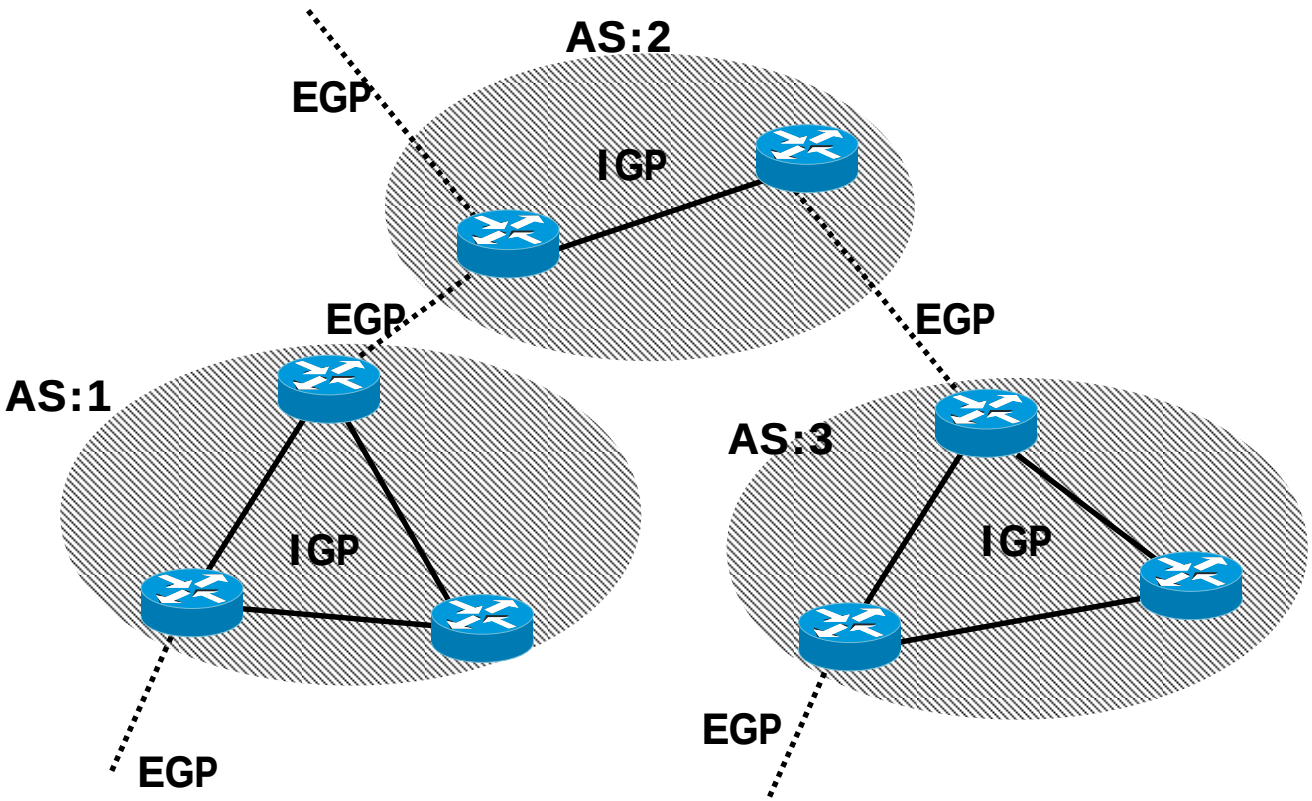
メンテナンスいらずのインフラとして有

- **適用先は…**
  - AS またぎで品質向上をはかる機能をオーバーレイで付加．
  - IPルーティングは宛先ベースなので，場所によって見え方が違う．
  - 他ノードを経由することによってルーティングを変えるインフラは作れそう．

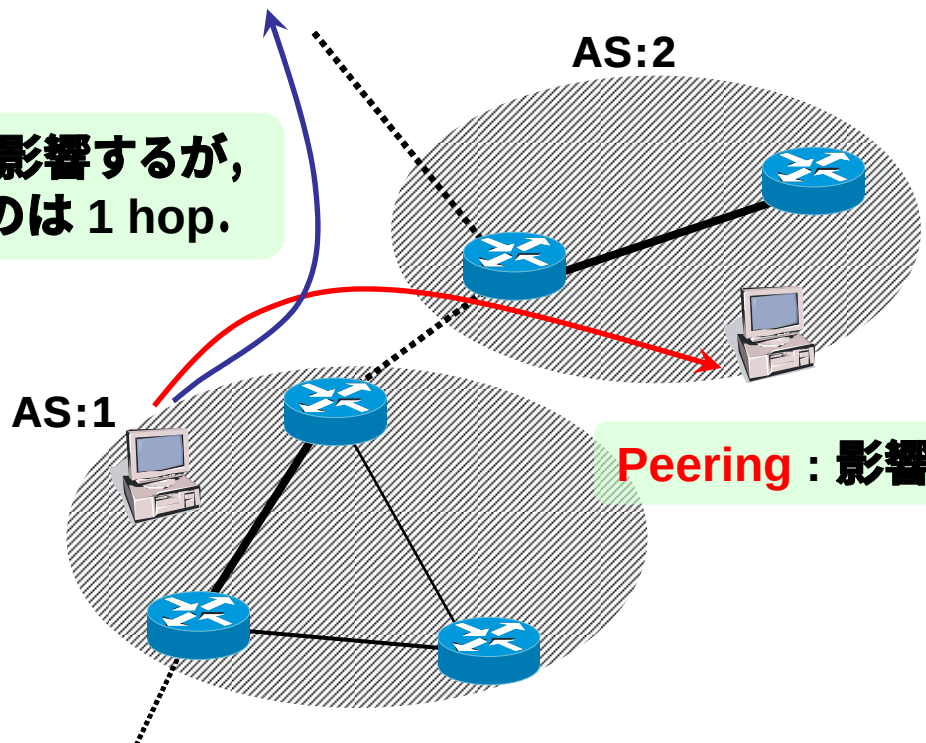
エンドホストオーバーレイネットワークによる
トラヒックエンジニアリング
って有効では？

- **現在のインターネットトラヒックエンジニアリングの枠組**
  - ISP内はIGP(OSPF等), ISP間はEGP (BGP4)

- **ISP内では高度な品質制御も可能**
  - DiffServ, MPLS, etc…



- **ISP間では…**
  - DiffServ / MPLS **の適用はルータ更改とISP間ポリシ調整が必要.**
  - BGP**の枠内でもある程度は対応可能だが, 実現は困難. また,** Transit **と** Peering **の比率は国内では** 1:5 **程度であり, 有効範囲も限定的.**
  - **そもそもAS内で品質が一定ではない.**



Transit：**複数 hop に影響するが, 制御可能なのは 1 hop.**

Peering：**影響範囲は 1 hop**

有効な適用先は？　～　エンドホストオーバーレイによるトラヒックエンジニアリング

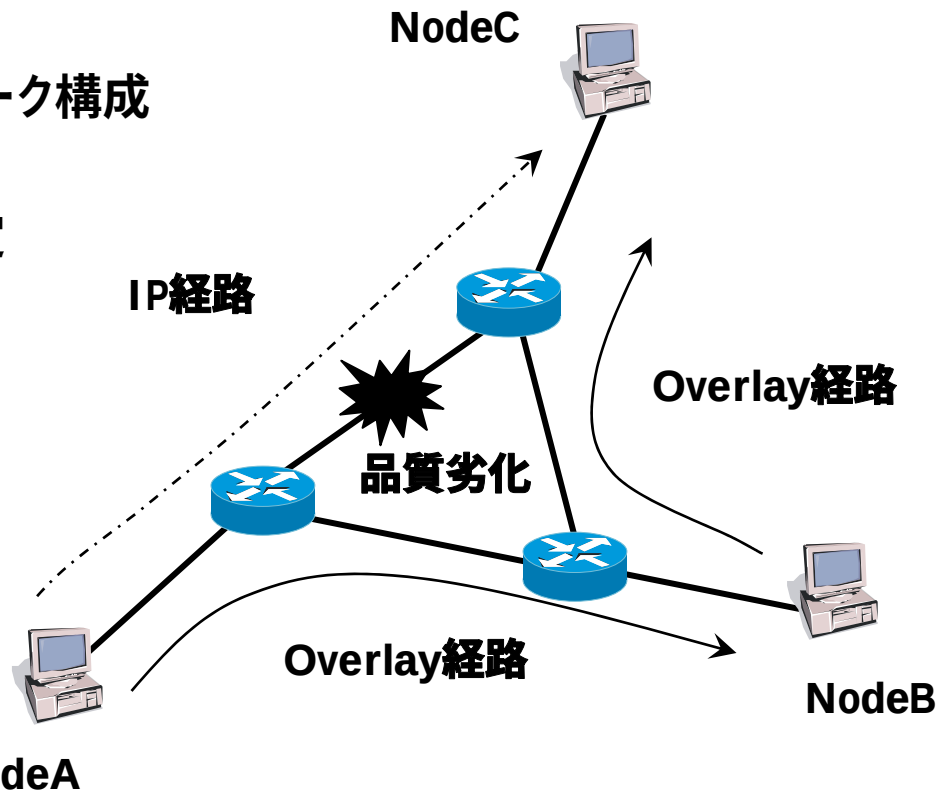
- MPLSやDiffServ が多数の AS に導入された未来は現実的か？
- BGP 拡張についてもポリシやビジネスモデル等, 越えるべき多くの壁.
- AS内にボトルネックが存在した時点でBGPの枠では無理そう.

- IPネットワークへの機能追加手段としてのオーバーレイネットワークが有力な選択肢として存在.
- 特定アプリケーションを指向したエンドホストオーバーレイの実用化が進み, P2Pファイル共有では数100万ノードの規模が実現されている.

エンドホストオーバーレイネットワークによる
トラヒックエンジニアリングが現実的な選択肢となりつつある

有効な適用先は？　～　エンドホストオーバーレイによるトラヒックエンジニアリング

- **配備機能**
  - 自律的ネットワーク構成
  - 情報の伝播
  - 品質測定／推定
  - 経路制御



- **特徴**
  - オーバーレイ層でのIPネットワーク拡張によるEnd-to-End QoS の向上
  - IP測定技術とトラヒック予測技術に基づく測定対地，測定間隔の限定による，スケーラブルなIPレイヤ変動へのアダプティブな追従
  - P2P技術を用いた自律分散化 による高スケーラビリティ，高信頼性の実現

有効な適用先は？　～　ビジネスモデル

- **どうやって課金するんだろう**
  - **そもそも**
    - **オーバーレイネットワークがどんどん普及して，高機能化した時に** ISP **は儲からないでトラヒックだけ増えることも．**
    - **インターネット** vs. **電話会社 ふたたび？**
    - **とりあえず** P2P **ファイル共有についてはトラヒック量課金の方向性も**
  - **では，トラヒックエンジニアリングインフラ(提案手法)では？**
    - BGP**ポリシーの裏をかくのってそもそも許されるんだろうか？**
    - **ユーザーに課金，貢献度に応じて割引く**
    - **ユーザーには無料でばらまき，構築したインフラ利用権を** xSP **に卸す**
    - ISP**からSLA回線を購入したり，データセンターにノード設置**
    - ISP**単位で突っこめば**BGP**オペレーションいらなくなるかも**

etc…

- 技術背景
- オーバーレイネットワークとは
  - 概説
  - オーバーレイネットワークとPeer-to-Peer
- 有効な適用先は？
  - インターネット上でのトラヒックエンジニアリングとその限界
  - エンドホストオーバーレイによるトラヒックエンジニアリング
  - ビジネスモデル
- ISP間測定によるインターネットの現状
  - 測定条件
  - P2P技術を用いた分散測定法
  - 品質値の分布
  - 有効領域, AS分析
  - 考察
- まとめと今後の課題

ISP間測定によるインターネットの現状　～　測定条件

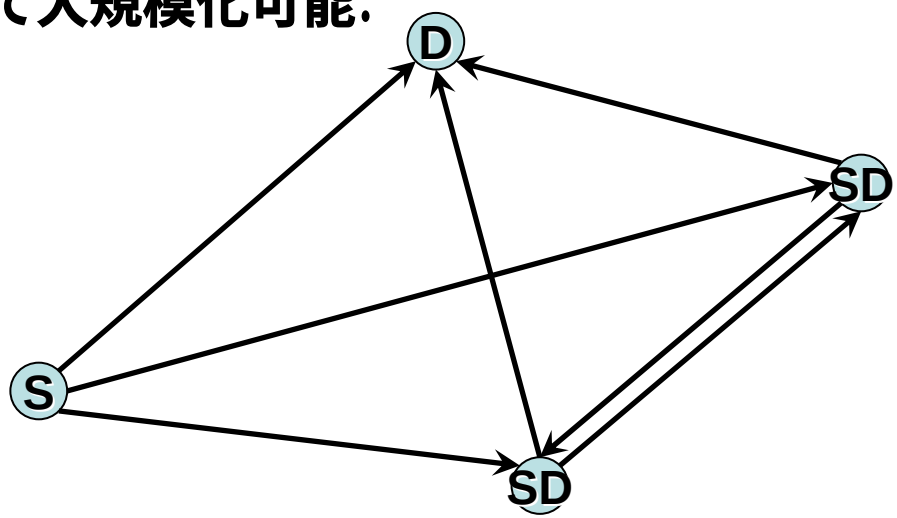
- **本手法の有効性検証のためのISP間品質の実情把握を目的とし，以下の条件でISP間品質測定を実施．**
  - **測定期間：約3週間 (有効データ95%以上)**
  - **測定対地：18 ISP**
    - **東京x6, 大阪x4, 札幌x4, 九州x4，うち10 ISPがメジャー(?)ISP**
  - **測定品質：遅延/損失率**
    - **各対地の組に対し，1時間に1回3分間連続で icmp パケットを1秒間隔で送出し，平均遅延，最大遅延，最小遅延，損失率を計算**
    - **同時に traceroute も測定．**
  - **サンプル数：約20万**
  - JXTA**によって実装したP2P技術を用いた分散測定システムを利用**

ISP間測定によるインターネットの現状　～　P2P技術を用いた分散測定法

- **広様な回線種別と多数ノードからなる広域IPネットワークにおいて，変動する多ノード間での品質/トラヒックを効率的に測定する広域分散測定法.**

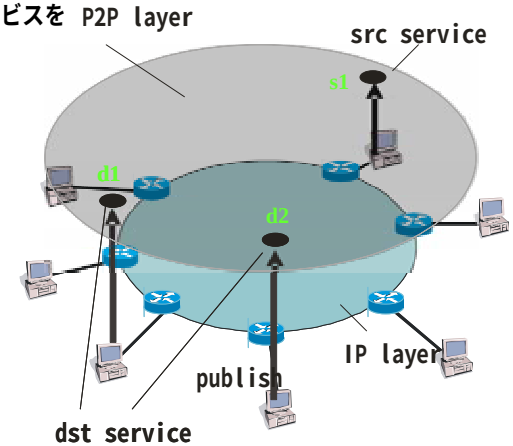
- **各測定ノードではIPアドレスのコンフィグレーションの他には，測定種別（遅延，スループットetc）毎にsrc/dstの宣言をするのみ.**

- **P2P技術を用いた，スケーラブルな自動測定スケジュール設定機能により大規模化，柔軟化に対応. 通常 $O(n^2)$ 必要なスケジュールコストを $O(n)$ に保って大規模化可能.**

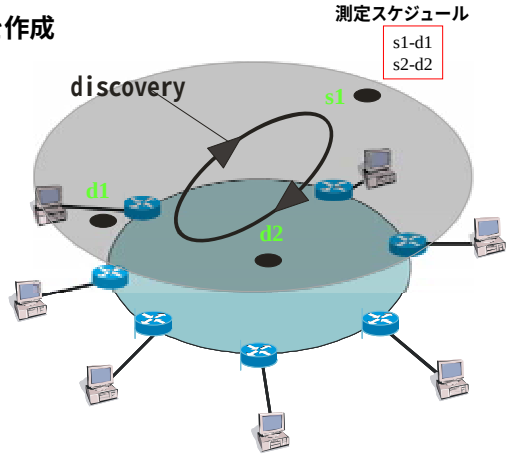


ISP間測定によるインターネットの現状　～　P2P技術を用いた分散測定法

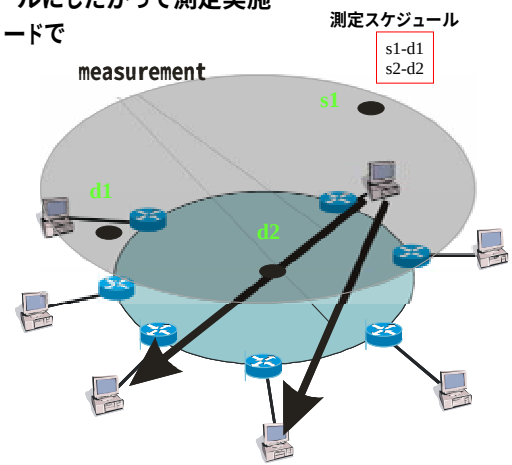
- **準備フェーズ（サービス起動）**
  - P2P層で測定元(src)，測定先(dst)のサービスを起動



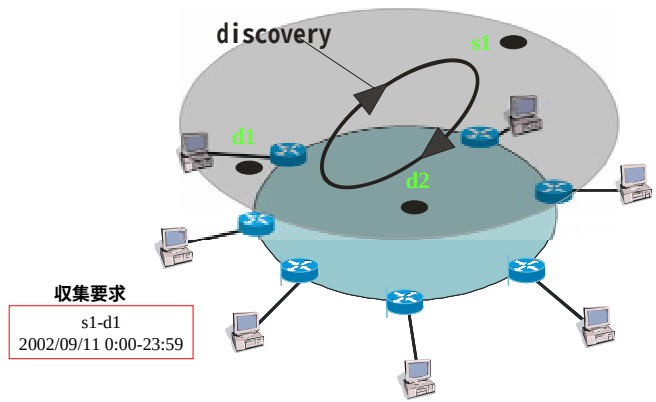
- **計画フェーズ（測定のスケジューリング）**
  - dstをdiscovery．
  - 各srcでsrc-dstの組を作成
  - 測定スケジュール化



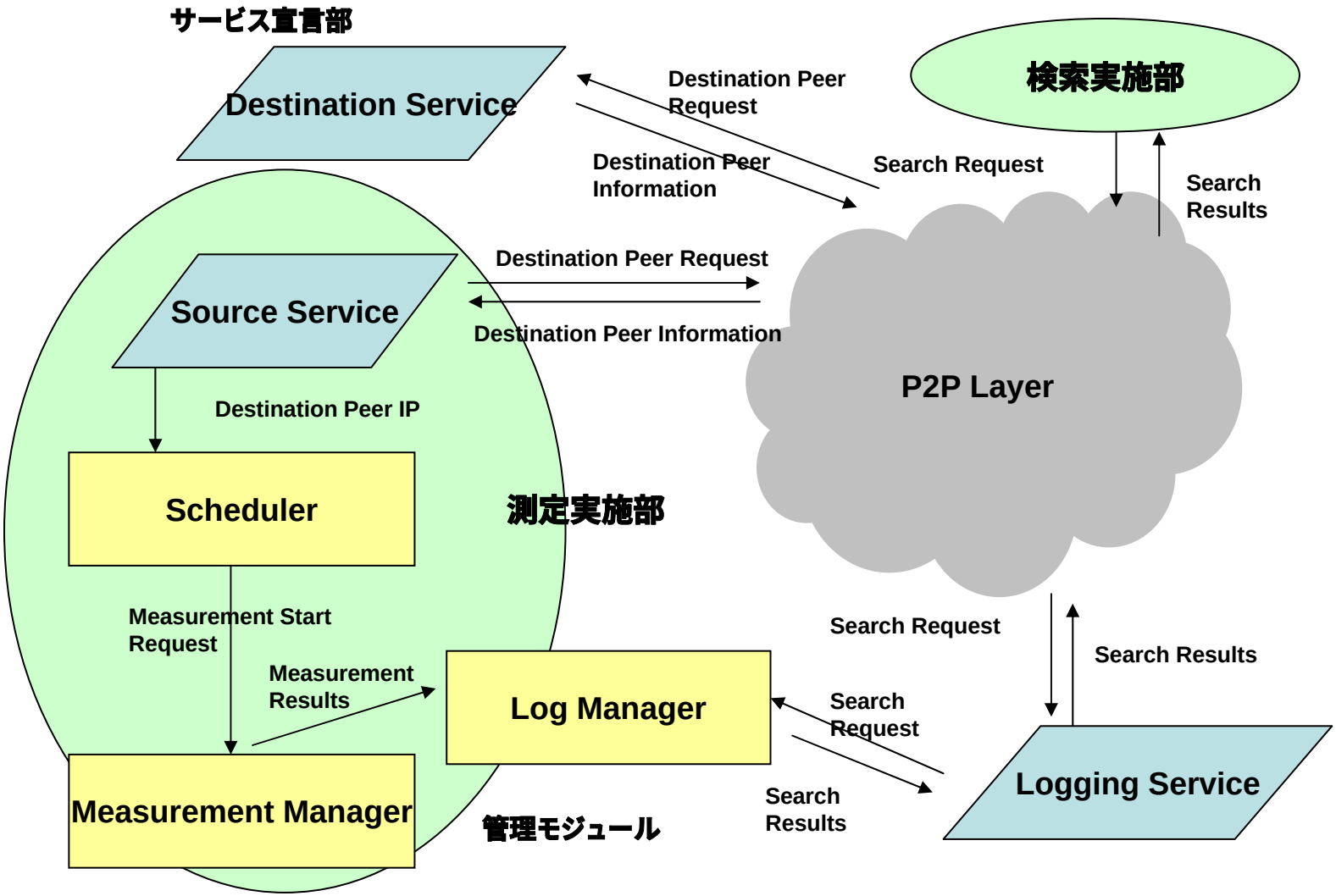
- **測定フェーズ（測定実施）**
  - IP層で測定スケジュールにしたがって測定実施
  - データはsrc or dst ノードで蓄積



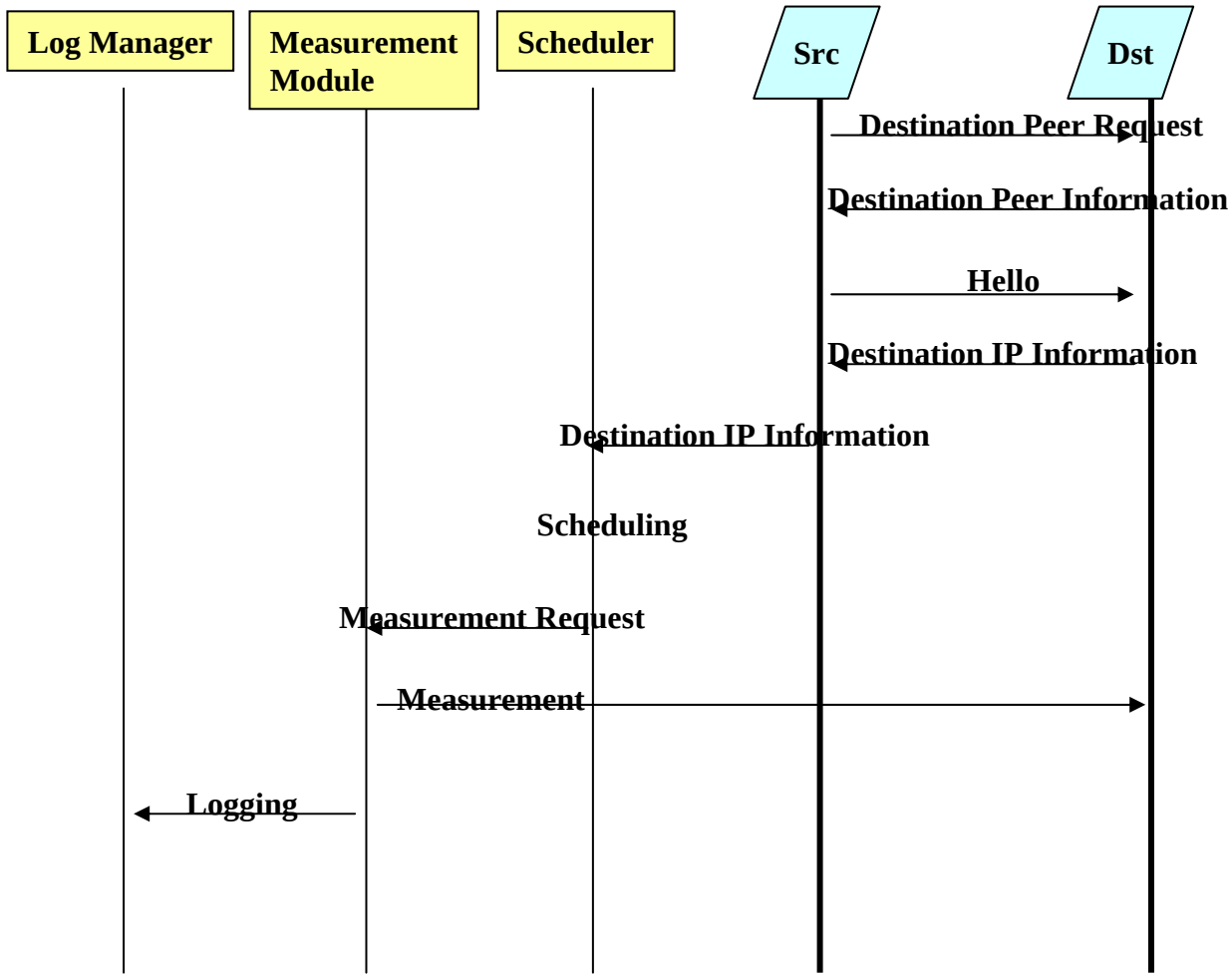
- **収集フェーズ（測定結果抽出）**
  - ノードに蓄積された測定データを条件，範囲を指定して抽出



ISP間測定によるインターネットの現状　～　P2P技術を用いた分散測定法



ISP間測定によるインターネットの現状　～　P2P技術を用いた分散測定法



ISP間測定によるインターネットの現状 ～ 品質値の分布

- **特定対地ペア** ex.) a→b **について** 1 **時間に**1**回**3**分連続で**180**発** icmp **パケットを送出.**
  - **サンプル数**610**時間,** 20**万サンプル**
  - **遅延** min, max, avr **と損失率の分布を計算**

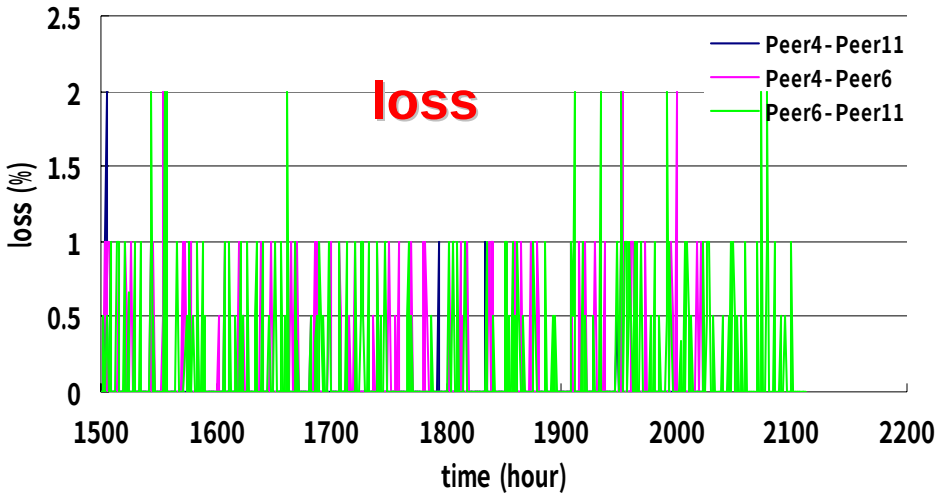
	min (ms)	max (ms)	avr (ms)	loss (%)
平均値	**26.49**	**97.69**	**30.32**	**0.29**
中央値	**25**	**70**	**29**	**0**
95%値	**54**	**201**	**58**	**1.1**

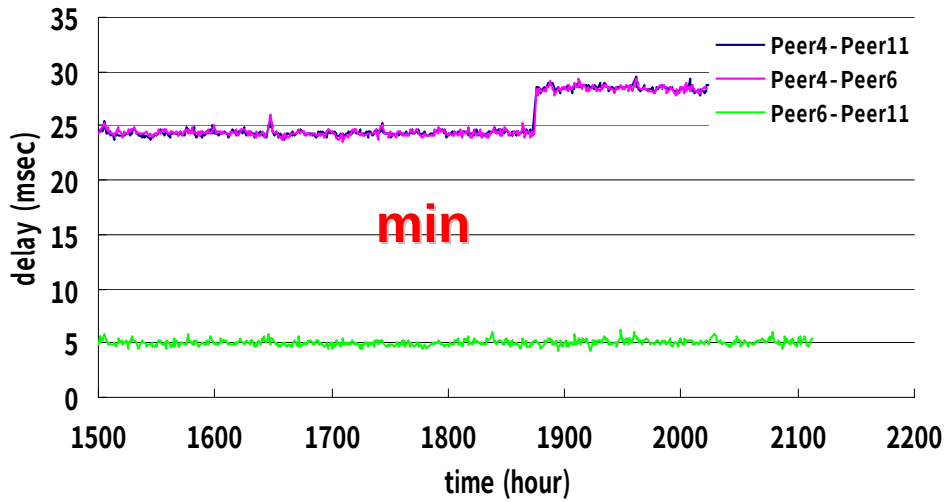
  - **平均値で見ると意外に**(?)**高品質**
  - ex) VoIP

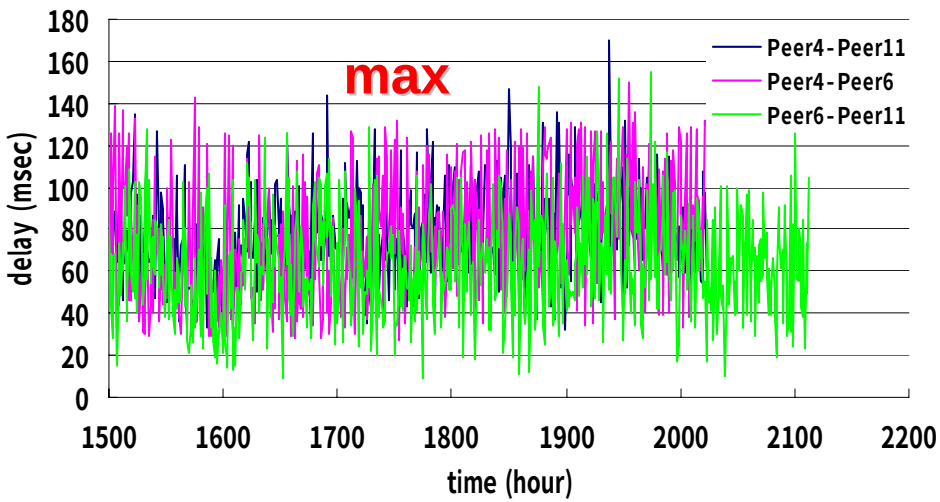
    クラスC：**片道**400ms→**クラス**B：**片道**150ms (-**端末遅延**)

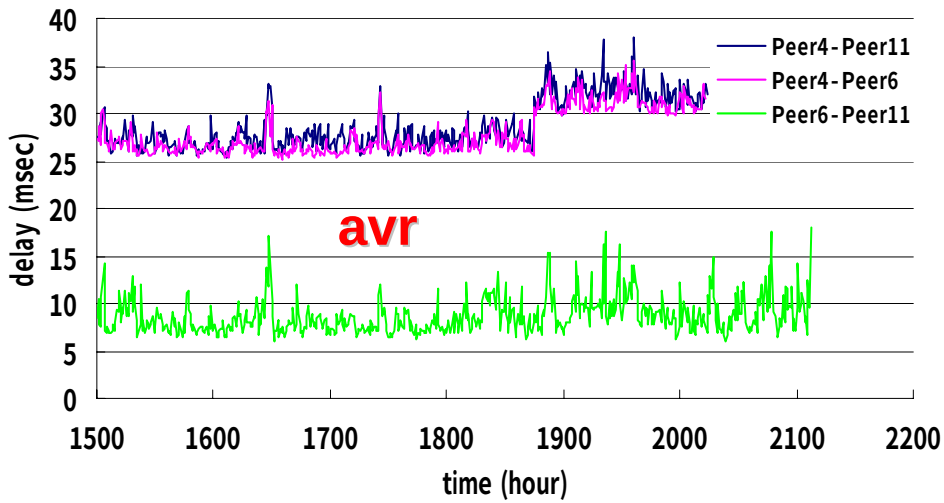
ISP間測定によるインターネットの現状 ～ 品質値の分布

- **個別に見てみると…**



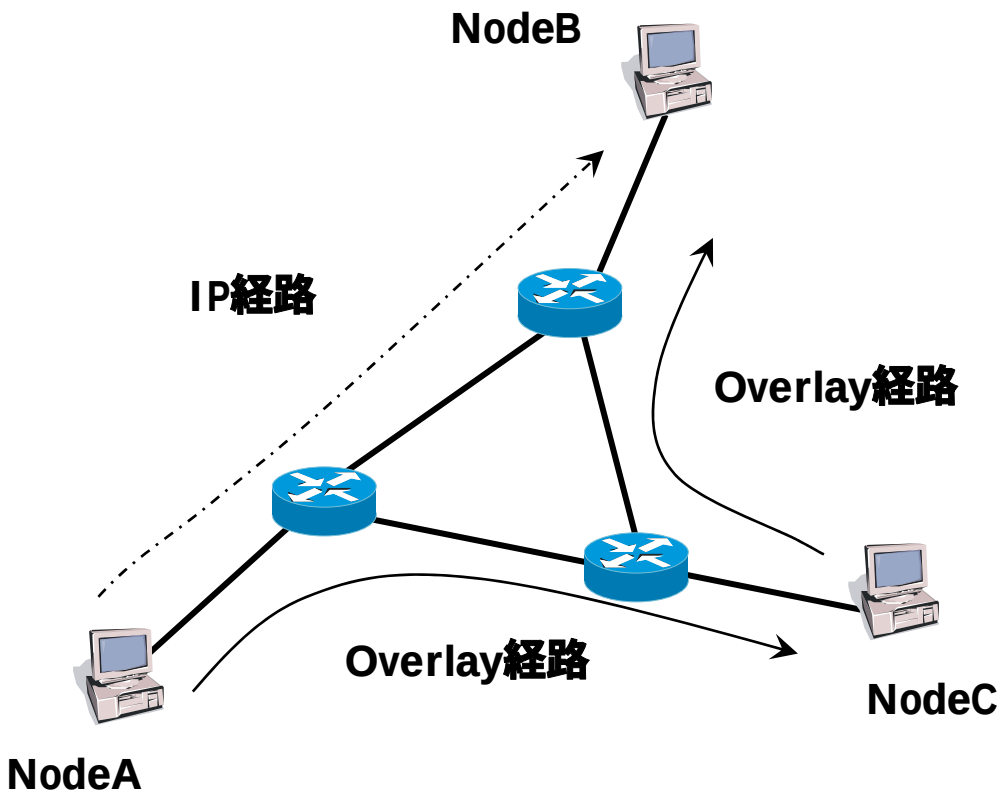




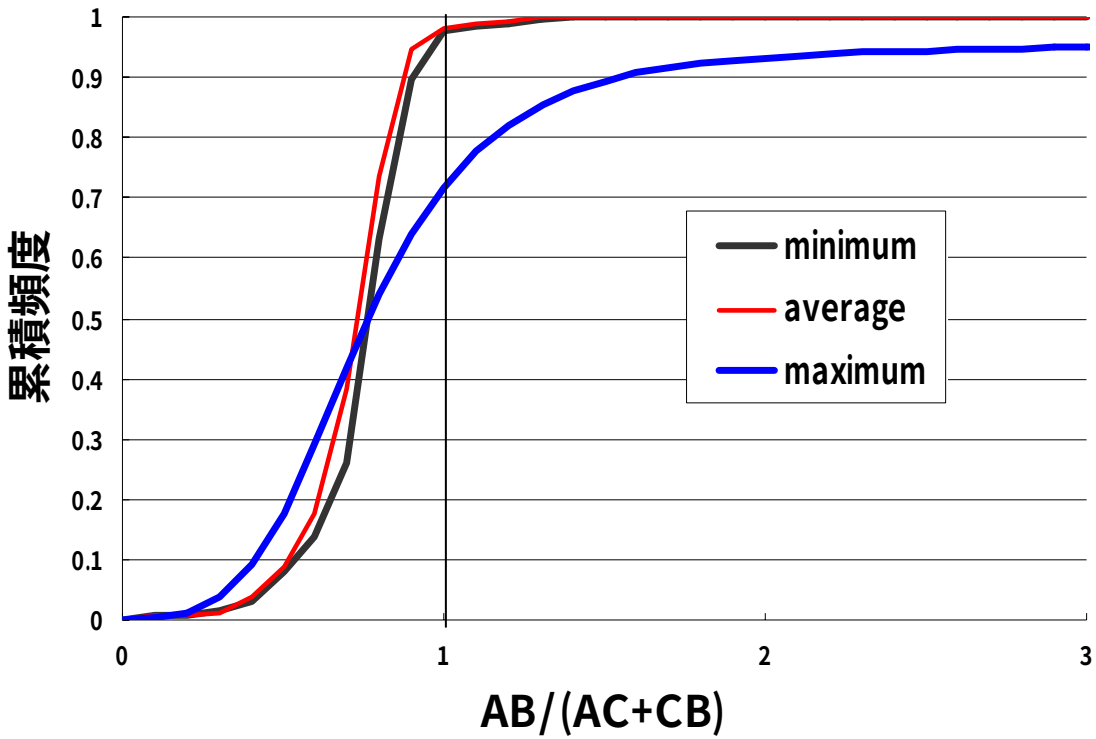


➢ 遅延に関する有効領域の評価

– ping_max(t, a, b) > ping_max(t, a, c) + ping_max(t, c, b) **となる経路が存在すれば有効と考えられる.**



ISP間測定によるインターネットの現状　～　有効領域



遅延に関する有効領域

– max **は最大値同士の加算であるため，控え目(有効領域が少な目)の評価になっている点に注意．**

ISP間測定によるインターネットの現状　～　有効領域

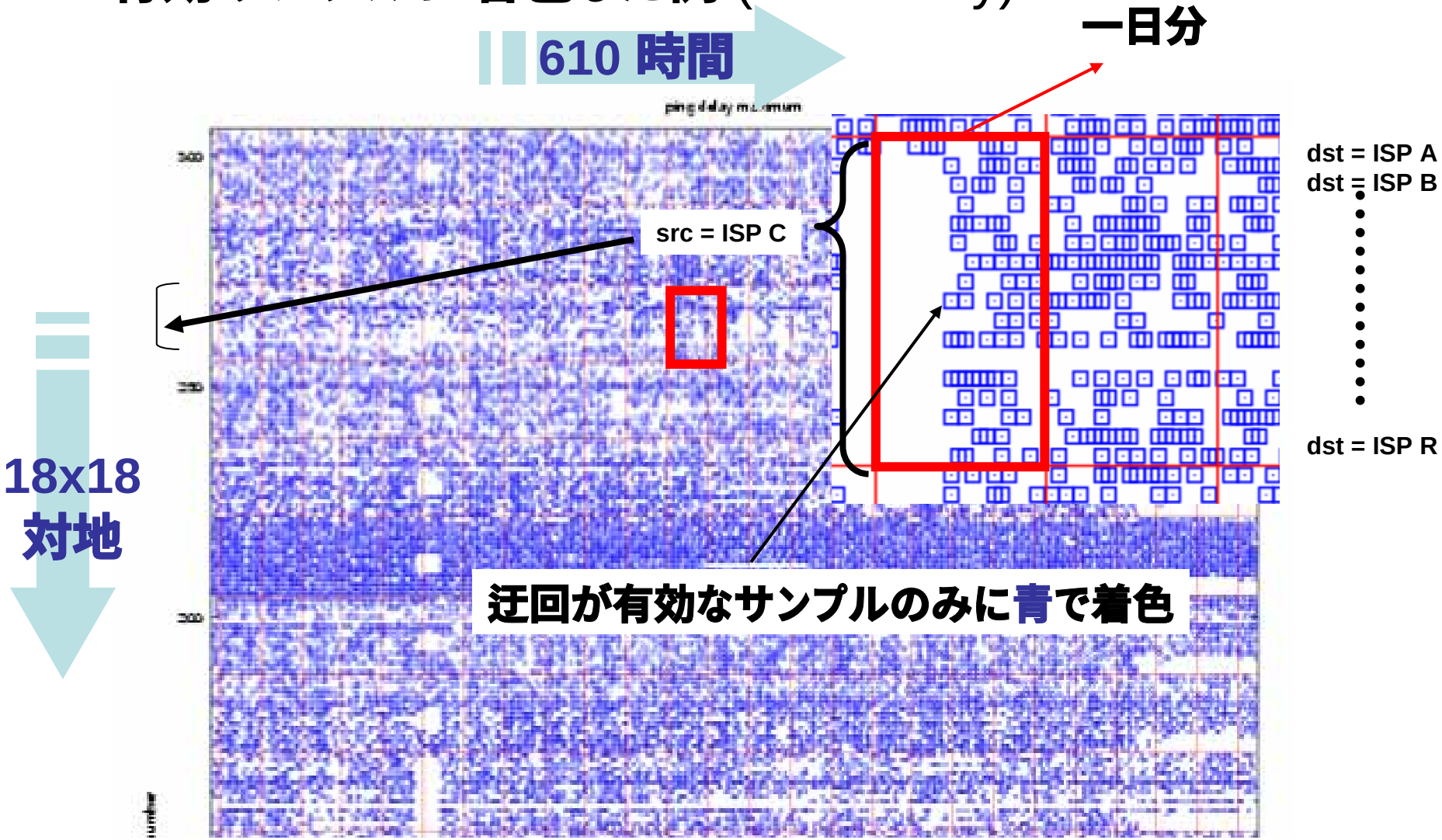
- **有効領域のまとめ**
  - 全時間スロット(1時間毎)における各対地の組を分母とし，それらの中で他のノードを経由(遅延を加算)した場合に値が改善する対地の組の比率を有効範囲と定義．

品質値	有効範囲
最小遅延	3%
最大遅延	**28%**
平均遅延	2%
損失率	**24%**

- **遅延の最大値や損失率については多くの領域での改善が期待できる．**
- **端末遅延や切換に伴うオーバーヘッドは未評価**

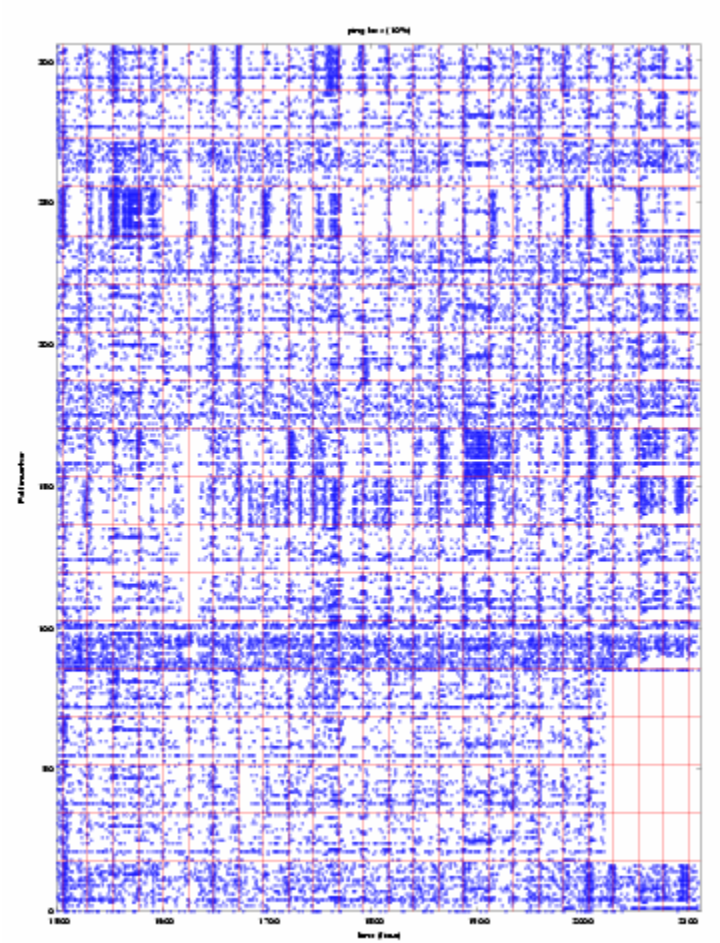
ISP間測定によるインターネットの現状 ～ 有効領域

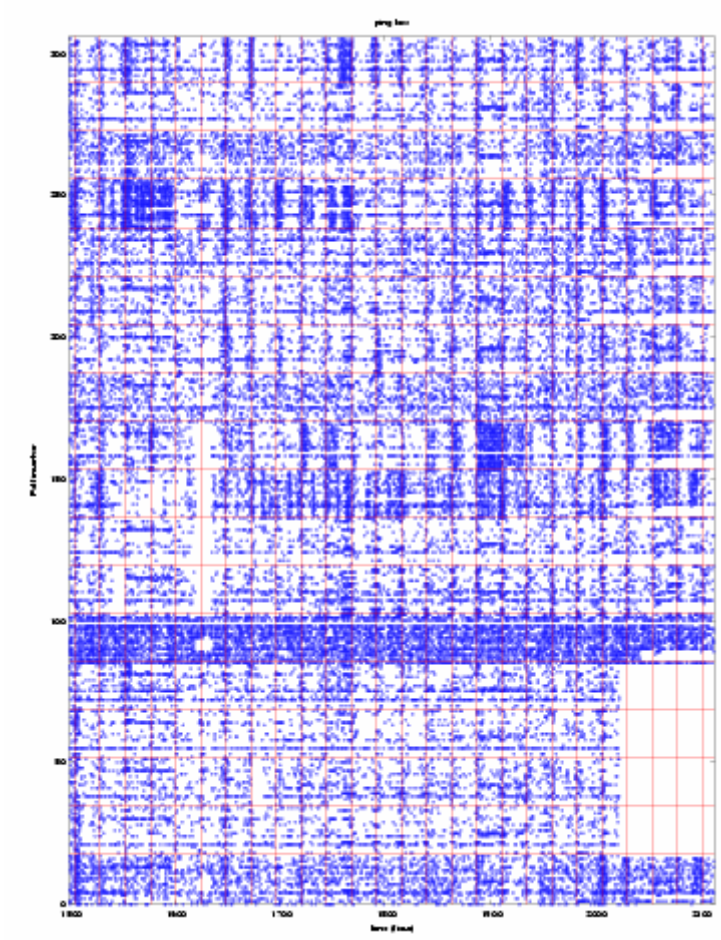
- **有効サンプルに着色した例** (max delay)



ISP間測定によるインターネットの現状　～　有効領域

- **平均値からの乖離で評価**
  - **損失率 左：平均から10%悪化**　　**右：有効領域(迂回により改善)**





ISP間測定によるインターネットの現状　～　有効領域

- **縦の相関が全体としてなんとなく見られる**
  - **同一箇所の劣化に巻き込まれている？ → ASで分析**
  - **全て国内ISPなので時間変動があるのかもしれない**
  - **でも迂回が有効な部分も結構あるのは ISP 体力の差か？**

- **単純な品質劣化と迂回が有効な領域との相関**
  - **品質劣化→迂回が有効 だから当然？**
  - **今回の評価目的は迂回の有効性確認だが，国内ISP間の品質劣化に関する評価も可能そう．**
    - **トラヒックの一極集中 / 障害連鎖 といった総務省次世代インフラ的課題**
    - **地域IXの配置や，有効な Transit や Peer の選択方法とか**

- **劣化箇所(迂回時に品質が向上する箇所)の分析**
  - **劣化箇所における特定ASの出現頻度をカウント**
  - **劣化箇所以外も含めた全経路における当該ASの出現頻度で割って正規化**

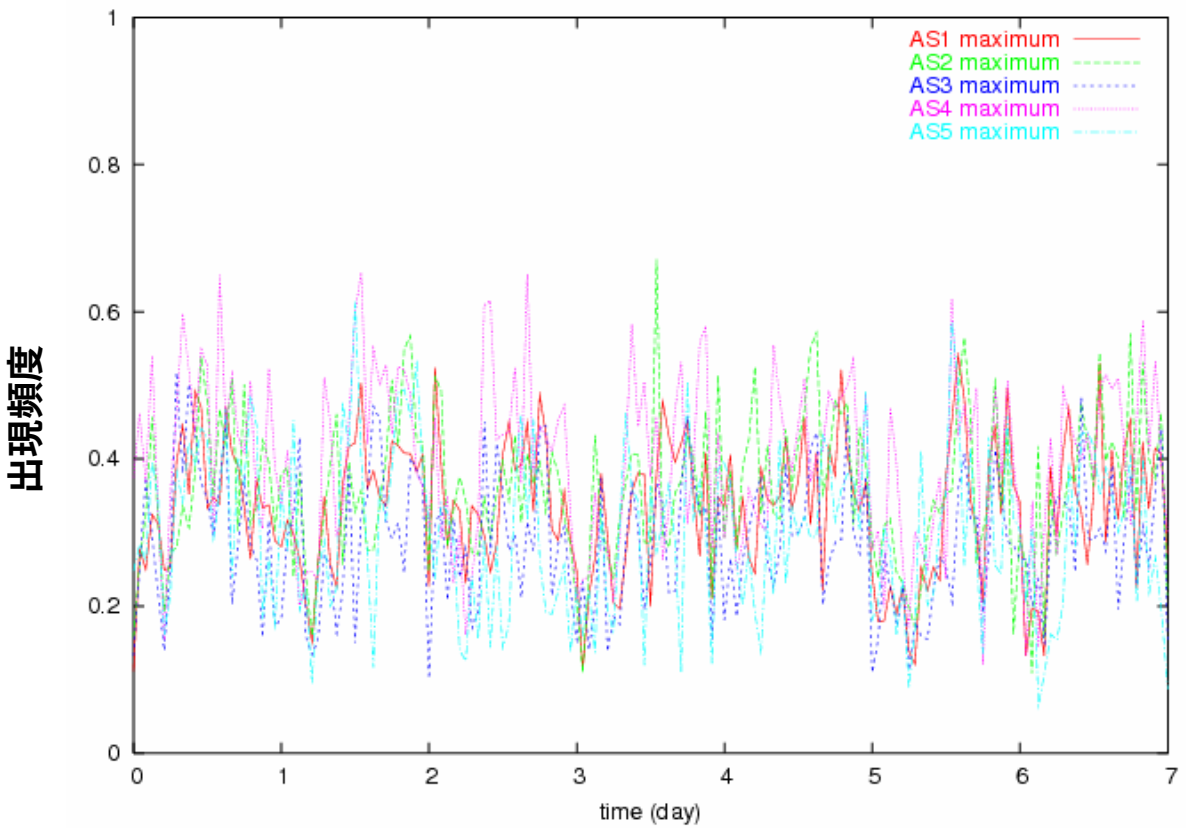
注) AS間の相関は評価できていないため，高頻度で劣化区間に登場する AS が必ずしも劣化の原因とは限らない．

ex)

- **劣化箇所**1：AS1 → AS2 → AS2 → AS3 → AS4 → AS4 → AS5
- **全経路での** AS 2 **の出現頻度** : 20**回**
- AS2**の劣化箇所**1**での出現頻度**：2/20

ISP間測定によるインターネットの現状　～　AS分析

- **全体での出現頻度の高い5 AS についての分析結果**

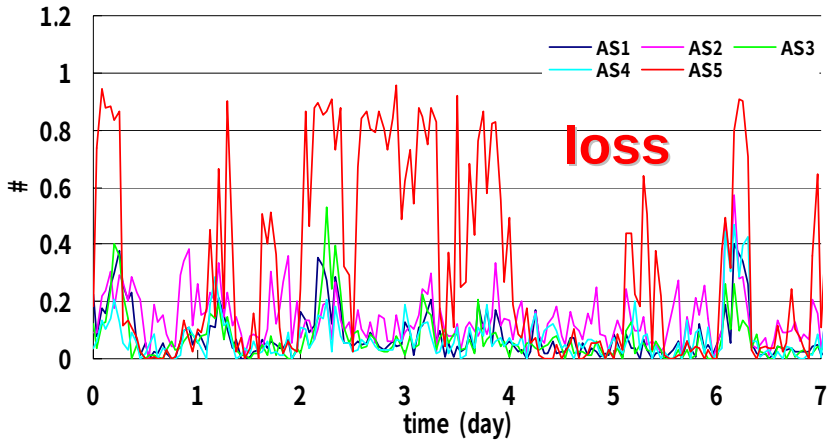


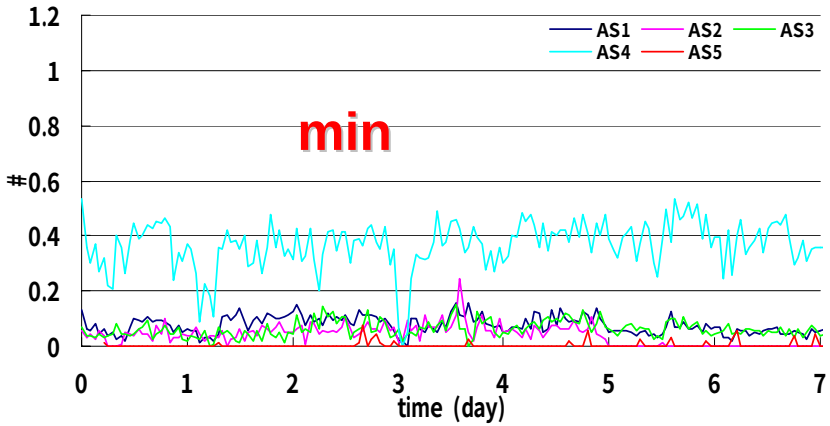
最大遅延劣化箇所におけるAS出現頻度

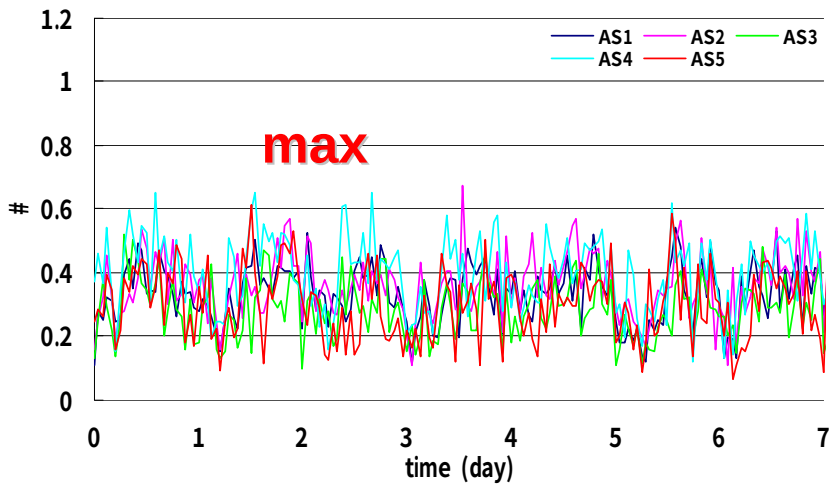
- **品質劣化箇所とAS番号との相関はあまり見られない.**

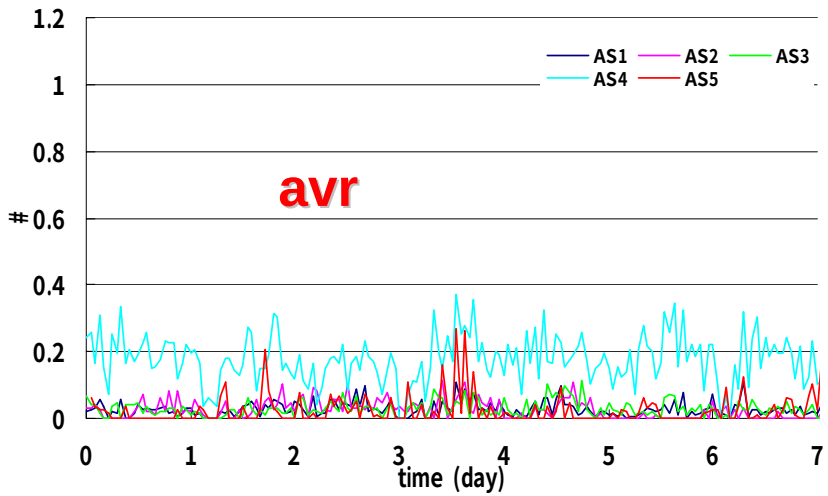
ISP間測定によるインターネットの現状 ～ AS分析

と思ったのだが…



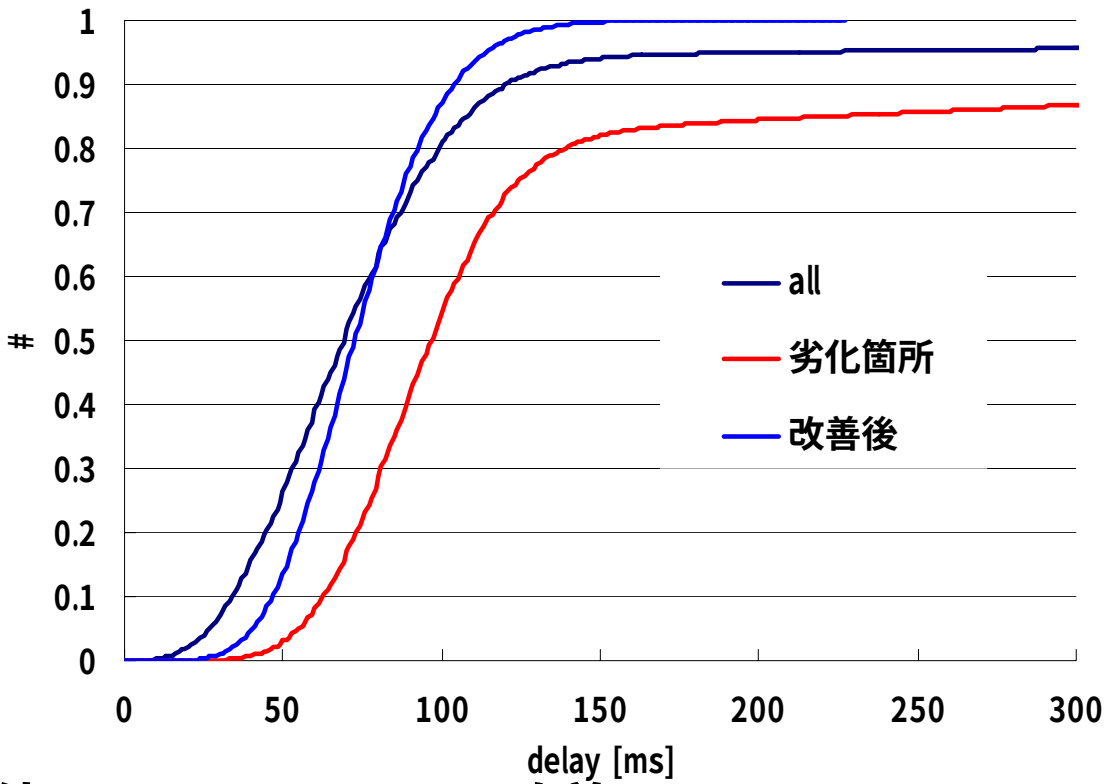






ISP間測定によるインターネットの現状　～　改善効果

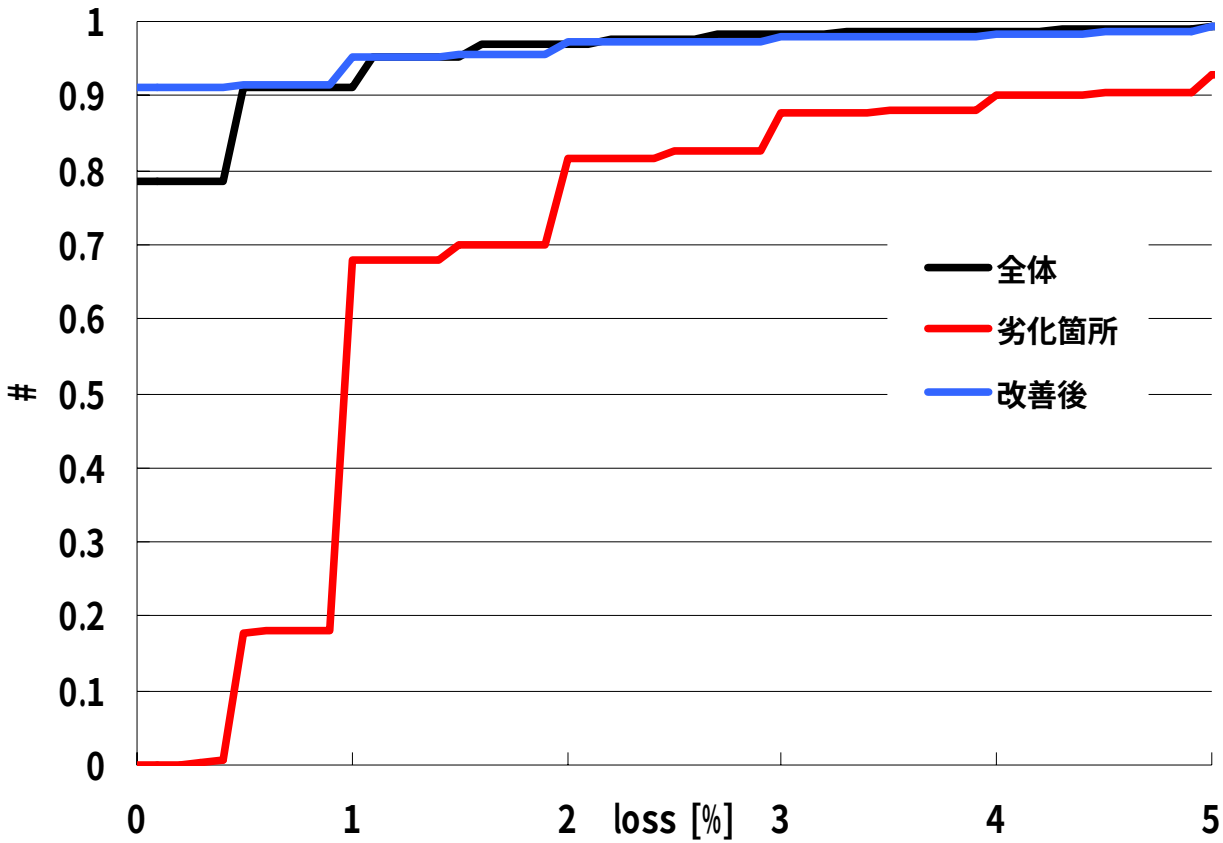
➢ 最大遅延の改善効果



- 95%**値で**683ms→114ms**と改善**
- **例えば** VoIP **では,**
  - **クラス**C：**片道**400ms→**クラス**B：**片道**150ms**へと改善**

ISP間測定によるインターネットの現状　～　改善効果

➢ 損失率の改善効果



– 95%**値での改善は**7%→1%

ISP間測定によるインターネットの現状　～　考察

- **3割弱の領域で遅延の最大値や損失率改善が期待できる．**
  - 突発的な輻輳や故障を回避できる可能性．
  - 最小遅延や平均遅延の改善領域が小さい(2～3%)のはドメイン間ルーティングに起因する根本的なねじれが少ないためか？
- **品質劣化の生じる箇所は特定ASに偏らない場合もある．**
  - 品質劣化箇所はAS間よりAS内に多く分布している可能性も
  - ドメイン間ルーティングの問題ではない？

ドメイン間ルーティング等の既存の枠組みで解決できない問題が
エンドホストオーバーレイによって改善できる可能性

- 技術背景
- オーバーレイネットワークとは
  - 概説
  - オーバーレイネットワークとPeer-to-Peer
- 有効な適用先は？
  - インターネット上でのトラヒックエンジニアリングとその限界
  - エンドホストオーバーレイによるトラヒックエンジニアリング
  - ビジネスモデル
- ISP間測定によるインターネットの現状
  - 測定条件
  - P2P技術を用いた分散測定法
  - 品質値の分布
  - 有効領域, AS分析
  - 考察
- まとめと今後の課題

まとめと今後の課題

- エンドホストベースのオーバーレイネットワークによるトラヒックエンジニアリングという枠組みを提案，
- 実際のISP間トラヒックの現状から，本手法の有効可能性を示した

議論して欲しいポイント

- **提案手法の技術的妥当性**
  - 提案手法はニッチな解決方法を狙っているが，オーソドックスにはどうすればいいか
- **一般的オーバーレイ及び提案手法のビジネス的妥当性**
  - このようなオーバーレイによる中抜きビジネスに対する対抗策は
- **測定データの中身 / 利用法**
  - ISP間の測定結果は実感と合っているかどうか
  - データの有効な分析法，利用法について

等々議論頂ければ幸いです．

参考文献

- Distributed Hash Table (DHT)
  - S.Ratnasamy et al, “A Scalable Content-Addressable Network”, Proc. of ACM Sigcomm, Aug, 2001.
  - I.Stoica et al, “Chord: A scalable peer-to-peer lookup service for Internet applications”, Proc. of ACM Sigcomm, Aug, 2001.
- Overlay Network
  - M.R. Macedonia et al, “MBone Provides Audio and Video Across the Internet”, IEEE Computer, Apr, 1994.
  - L.Zhi and P.Mohapatra, QRON: QoS-aware routing in overlay networks, IEEE J. Select. Areas Commun, Jan, 2004.
  - Y.T.Hou, Z.Duan and Z.Zhang, “Service overlay networks: SLA, QoS and bandwidth provisioning”，Proc. Of IEEE ICNP'02, Nov, 2002.
- Peer-to-Peer
  - **松本**, **横田**, **亀井**, **田山**, **中原**，“**広がる**P2P**サービスとインターネットインフラへの影響**”，JANOG13，Jan，2004.
- QoS Overlay Network & Peer-to-Peer Distributed Measurement
  - **亀井**，**川原**，“**エンドホストオーバーレイネットワークによるトラヒックエンジニアリングとその有効性**”，**信学会**IN**研究会**, Jul, 2004.
  - **亀井**，**川原**，**阿部**，“**国内**ISP**間の測定データを用いた迂回による品質向上効果の評価**”，**地域ネットワーク連携ワークショップ** in **京都**, Jul, 2004.
  - **亀井**，**木村**，” P2P**技術を用いた広域分散測定法の提案**”, **信学会**IN**研究会**, Mar, 2003.

