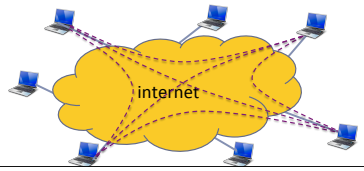
# Distributed Solution Approaches for Large-scale Network Measurement Exploiting Local Information Exchange

大阪大学 大学院情報科学研究科 情報ネットワーク学専攻
先進ネットワークアーキテクチャ講座
ディン ティエン ホアン

## 研究の背景
## エンドホストによるネットワークサービス

- エンドホストによって構成されるネットワークの技術とそれらを利用したサービスの普及
  - P2P（KaZaA, BitTorrent, ...）, CDN（Akamai, CoDeeN, ...）, ...
- エンドホスト間の全パスの性能情報に基づく経路制御
  - 性能情報：遅延時間（RTT）、パケットロス率、利用可能帯域など
  - 性能が最も良いパスの選択、障害の迂回パスの探索

全てのエンドホスト間パスの性能情報を取得する必要がある






## 研究の目的

- エンドホストによって構成されるネットワークにおける性能情報の計測手法を提案
- 性能情報の性質に基づいて、三つの手法を提案
  - 遅延時間とパケットロス率の計測手法
  - 利用可能帯域の計測手法
  - リンク障害の検出方法



## 博士論文の構成

- Chapter 1 Introduction
- Chapter 2 Measurement method for end-to-end additive quality metrics
  - 遅延時間とパケットロス率の計測手法
- Chapter 3 Measurement method for end-to-end available bandwidth
  - 利用可能帯域の計測手法
- Chapter 4 Measurement method for link fault diagnosis
  - リンク障害の検出手法
- Chapter 5 Conclusion



## Chapter 2
### Measurement method for end-to-end additive quality metrics

- Dinh Tien Hoang, Go Hasegawa and Masayuki Murata, ``A low-cost, distributed and conflict-aware measurement method for overlay network services utilizing local information exchange,'' *IEICE Transactions on Communications*, vol.E96-B, no.2, pp.459--469, February 2013
- Dinh Tien Hoang, Go Hasegawa and Masayuki Murata, ``A distributed measurement method for reducing measurement conflict frequency in overlay networks,'' in Proceedings of *the 2011 International Communications Quality and Reliability Workshop (IEEE CQR 2011)*, pp.1--6, May 2011.
- Dinh Tien Hoang, Go Hasegawa and Masayuki Murata, ``A distributed and conflict-aware measurement method based on local information exchange in overlay networks,'' in Proceedings of *the 2012 Australasian Telecommunication Networks and Applications Conference (ATNAC 2012)*, pp.1--6, November 2012.
- Dinh Tien Hoang, Go Hasegawa and Masayuki Murata, ``A distributed measurement method for reducing measurement conflict in overlay networks,'' *Technical Report of IEICE(CQ2010-57)*, vol. 110, no.287, pp.49--54, November 2010 (in Japanese).
- Dinh Tien Hoang, Go Hasegawa and Masayuki Murata, ``A distributed and conflict-aware measurement method using local information exchange in overlay networks,'' *Technical Report of IEICE(IN2012-35)*, vol.112, no.134, pp.13--18, July 2012.

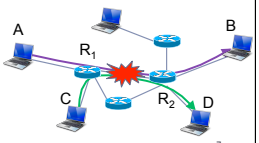


## Chapter 3
### Measurement method for end-to-end available bandwidth

- Dinh Tien Hoang, Go Hasegawa and Masayuki Murata, ``A distributed mechanism for probing overlay path bandwidth using local information exchange,'' submitted to *IEICE Transactions on Communications* [accepted], January 2014.
- Dinh Tien Hoang, Go Hasegawa and Masayuki Murata, ``A distributed measurement method exploiting path overlapping in large scale network systems,'' in *the 1st Workshop on Large Scale Network Measurements (31st NMRG meeting)*, Zurich, Switzerland, October 2013.
- Dinh Tien Hoang, Go Hasegawa and Masayuki Murata, ``Monitoring available bandwidth in overlay networks using local information exchange,'' in Proceedings of *the 2013 Australasian Telecommunication Networks and Applications Conference (ATNAC 2013)*, November 2013.
- Dinh Tien Hoang, Go Hasegawa and Masayuki Murata, ``A distributed method for measuring available bandwidth in overlay networks exploiting path overlap,'' *Technical Report of IEICE(NS2013-71)*, vol.113, no.205, pp.1--6, September 2013.

## 研究の背景：性能情報の計測

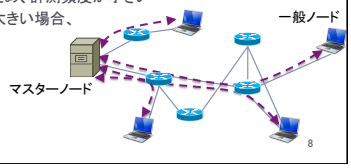
- 性能情報はエンドホスト間の計測により取得する
  - 遅延時間とパケットロス率の計測ツール：ping, tracerouteなど
  - 利用可能帯域の計測ツール：Pathload, pathChirpなど
- 性能情報が時間とともに変動
  - 一般に正確に取得するために複数回の計測結果の平均値をとる
  - 変動が大きい場合、高頻度に計測する必要がある
- 経路が重複しているパス（以降、重複パス）が多く存在
  - 重複パス：(A,$R_1$,$R_2$,B)と(C,$R_1$,$R_2$,D)
- 重複パスを同時に計測するとき、計測の競合が発生
  - 計測精度が低下






## 研究の背景：既存の計測手法

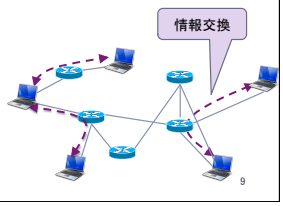
- 集中制御により、計測競合を回避
  - マスターノードにネットワーク全体の経路情報を収集し、重複パスを検出
  - 重複パスを同時に計測しないようにスケジューリングする
  - 計測競合を完全に回避可能
- 大規模ネットワークに適用するときの問題点
  - マスターノードにトラヒック量と計算量が集中
  - ➡ 故障しやすく、ネットワーク全体が停止
  - 重複パスの数が多いため、計測頻度が小さい
  - ➡ 性能情報の変動が大きい場合、計測精度が低い






## 研究の目的とアプローチ

- **研究の目的**：大規模ネットワークにおける分散型計測手法を提案
- **研究のアプローチ**：計測精度を向上するために、重複パスの計測結果の関連性に着目し、エンドホスト間の情報交換を利用
  - 計測結果のサンプル数を増大
  - 計測トラヒック量と計測競合を軽減






## 提案手法の概要

(1) 重複パスの検出
- 経路情報の交換を利用

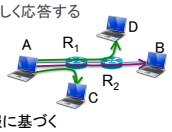
(2) 計測タイミングの決定
- 計測頻度をパスの重複の度合により調整する
- 計測タイミングをランダムに決定する

(3) 計測の実行
- 性能情報の性質に基づいて、計測方法を提案
  - 遅延時間とパケットロス率の計測方法
  - 利用可能帯域の計測方法
- 計測結果の交換により、計測精度を向上



## 重複パスの検出方法（1/3）

- 検出の方針
  - エンドホストは他のエンドホストまでのパスの経路情報をtracerouteを用いて、取得する
  - エンドホスト間で、経路情報を交換し、重複パスを検出する
- 前提
  - エンドホストは他の全てのエンドホストのIPアドレスを保持する
  - エンドホスト間パス上のルータはtracerouteに正しく応答する
- 検出の手順



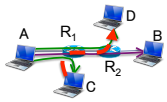

(1) 始点ノードが同じ重複パスの検出
  - 始点ノードから他のエンドホストへの経路情報に基づく
  - 例：AはB,CとDまでの経路情報を取得し、ACとADがABの重複パスであることを検出



## 重複パスの検出方法（2/3）

- 検出の手順

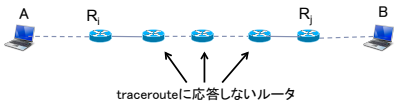
(2) 始点ノードが異なる重複パスの推定
  - 始点ノードが同じ重複パスの重複状況に基づく
  - 例：
    - ABとACは、Aと$R_1$の間で重複
    - ABとADは、Aと$R_2$の間で重複
    - $R_1$Dは、$R_2$を通過する
    - → AはCDがABの重複パスであると推定




(3) 始点ノードが異なる重複パスの確定
  - 経路情報の交換により、重複パスであるかどうかを確定
  - 例：
    - AとCはABとCDの経路情報を交換することにより、ABとCDが重複していることを検出

## 重複パスの検出方法(3/3)

- 一部のルータが正しくtracerouteに応答しない場合の対策
  - それらのルータを含むパスの部分を仮想的なリンクと見なす
  - 例:パスABは$R_i$と$R_j$の間のルータがtracerouteに応答しない
  - → パスABの経路情報は(A,...,$R_i$,Rj,...,B)であると見なす




＊この対策は博士論文の第2章に追記した

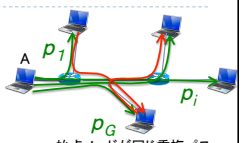


## 計測タイミングの決定方法

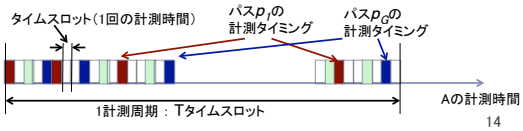
(1)パスの重複状況に基づいて計測頻度を算出

$$f_i = \min\{\frac{\beta_i}{\sum_{j=1}^{G}\beta_j}, \frac{1}{K_i+1}\}$$

$K_i$:パス$p_i$の始点ノードが異なる重複パスの数




(2)計測タイミングが均等に分布するようにランダムに割り当てる

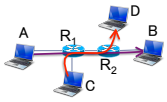





## 遅延時間とパケットロス率の計測精度向上方法

計測ツール:pingやtracerouteなど

計測ツールの性質
- 始点ノードからパス上のルータまでの性能を計測できる
  - 前提:ルータが正しく応答する
- 部分パスの計測結果からパス全体の計測結果を推定できる
  - 例: $RTT_{AB} = RTT_{AR_i} + RTT_{R_iR_2} + RTT_{R_2B}$



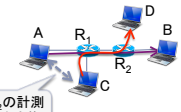

計測精度向上方法:重複部分の計測結果を交換し、重複部分の精度を向上することにより、パス全体の計測精度を向上



## 遅延時間とパケットロス率の計測手順

(1) 他パスとの重複関係に基づいてパスを分割し、部分パスを計測する
- AはA$R_1$, $R_1R_2$, $R_2$Bの遅延時間を計測する
- CはC$R_1$, $R_1R_2$, $R_2$Dの遅延時間を計測する

(2) 重複部分の計測結果を交換する
- AとCは$R_1R_2$の計測結果を交換する




(3) 統計処理により、重複部分の計測結果の精度を向上する
- 情報交換により$R_1R_2$の計測結果のサンプル数が多くなるため、精度が向上される

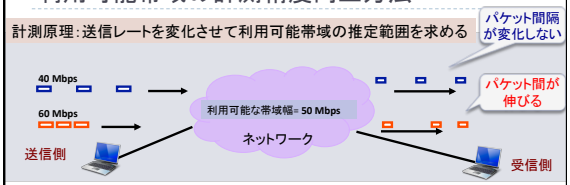
(4) 部分パスの計測結果から全体の計測結果を推定

$$RTT_{AB} = RTT_{AR_i} + RTT_{R_iR_2} + RTT_{R_2B}$$
$$RTT_{CD} = RTT_{CR_i} + RTT_{R_iR_2} + RTT_{R_2D}$$

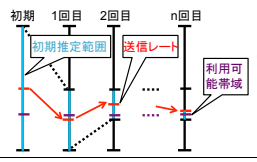


## 利用可能帯域の計測精度向上方法

計測原理:送信レートを変化させて利用可能帯域の推定範囲を求める




**[推定範囲の求め方:二分探索]**




計測トラヒック量は初期推定範囲に依存する

利用可能帯域に近い初期推定範囲を設定することにより、計測トラヒックを削減し、計測の競合を軽減し、計測精度を向上する

## 利用可能帯域の計測手順

- 初期推定範囲の計算方法の基本アイデア

(1) 利用可能帯域が連続的に変動する
➡ 直近の計測結果を初期推定範囲の計算に利用できる

(2) ボトルネックリンクが重複部分に存在する場合、重複パスの計測結果が同じになる
➡ 重複パスの直近の計測結果を初期推定範囲の計算に利用できる

- 利用可能帯域の計測手順

(1) 自パスとその重複パスの計測結果を用いて、初期推定範囲を計算し、計測を行う
(2) 重複パスの計測結果を交換する

## シミュレーション評価

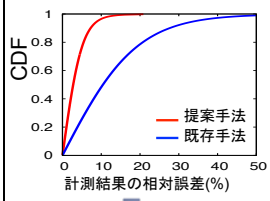
- 評価方法
  - 既存手法[12]と比較
  - 評価指標
    - 計測精度：計測結果の相対誤差
    - 計測トラヒック量
- シミュレーション設定
  - アンダーレイネットワークトポロジ：AT&T、BA、ランダム
    - ノード数：523、リンク数：1304
    - 全てのリンクの物理帯域：100Mbps
  - エンドホスト数：ノード数の20%
  - 計測の競合による誤差は確率統計理論により計算する

[12] M. Fraiwan and G. Manimaran, "Scheduling algorithms for conducting conflict-free measurements in overlay networks", *Computer Networks*, vol 52, pp. 2819-2830, Oct. 2008

19

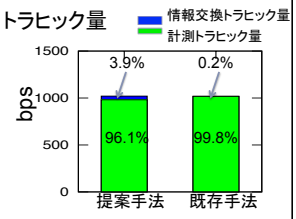
## 評価結果（遅延時間の計測）

計測精度




既存手法と比べて、提案手法の相対誤差が大きく下回る

トラヒック量



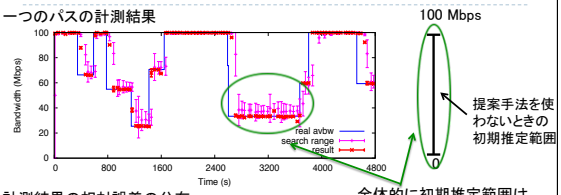

提案手法は情報交換トラヒック量が大きいが、計測トラヒックが小さい

計測に費していたオーバヘッドのごく一部を情報交換に費やすことで計測精度が大きく改善

20

## 評価結果（利用可能帯域の計測）

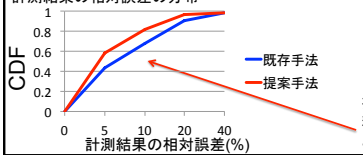
一つのパスの計測結果




100 Mbps

提案手法を使わないときの初期推定範囲

計測結果の相対誤差の分布




全体的に初期推定範囲は提案手法を使わないときと比べて、小さく利用可能帯域に近い

提案手法の計測精度は既存手法より大きく上回る

21

## 2章と3章のまとめ

- ネットワーク内のすべてのエンドホスト間パスの遅延時間、パケットロス率と利用可能帯域の計測手法を提案
  - 経路情報の交換により、重複パスを検出
  - パスの経路重複の状況に基づき、計測タイミングを決定
    - 計測競合を軽減
  - 計測結果の交換による計測精度の改善
    - 重複部分の計測結果の交換により、重複部分およびパス全体の計測精度を向上
    - 一回の計測に必要な計測時間の削減により、計測競合を軽減
  - シミュレーションによる性能評価
    - 提案手法の計測精度が既存手法と比べて上回ることを確認した
    - 一回の計測に必要な計測トラヒック量を削減できることを確認した

22

## Chapter 4

Measurement method for link fault diagnosis

- Dinh Tien Hoang, Go Hasegawa and Masayuki Murata, ``Spatial and temporal solutions for fault diagnosis in large-scale network systems," submitted to *IEEE/ACM Transactions on Networking*, November 2013.

23

## 本章の背景

- エンドホスト間パス上のリンクには、様々な原因により障害が発生する
  - ハードウェアの障害、ソフトウェアのバグ、設定ミス、…
- 一定の期間毎にすべてのリンクに障害が発生するかどうかを監視する必要がある
  - 以降、この期間を監視期間と呼ぶ
  - 監視の方法：監視期間中にすべてのリンクに対して、1回以上計測する
- 計測トラヒックがネットワークに負荷を与えるため、少ない計測トラヒック量で早期に障害を検出することが求められる

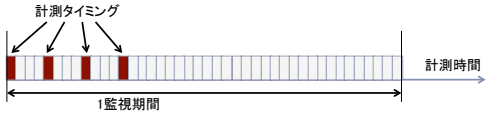
24

## 既存の障害検出方法[57]

- 一つの監視期間において、等間隔に一定数のパスを選択し、計測を行う
  - 多くのリンクを網羅できるように、パスを選択する
- 計測タイミングが監視期間の冒頭に集中する

⇒

- 障害が監視期間の末期に発生する場合、障害を検出するための時間が長くなる




[57] P. Barford, N. Duffield, A. Ron and J. Sommers, "Network performance anomaly detection and localization", in *Proc. INFOCOM 2009*, pp. 1377-1385, 2009

25

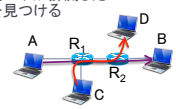
## 本章の目的とアプローチ

- **研究の目的**：早期にリンク障害を検出する分散型手法を提案
- **研究のアプローチ**：検出時間を短縮するために、重複パスの計測結果の交換を利用
  - 交換した計測結果に基づいて、監視期間中に計測するパスの数を削減
  - 計測パス数の削減に従って、計測タイミングの間隔を大きくすることにより、計測タイミングを監視期間中に広く分布させる

26

## 提案の障害検出方法(1/2)

- 重複パスの計測結果を交換し、計測パス数を動的に減らす
  - エンドホストはパスを一つずつ計測する
  - 計測の後に、すぐに計測結果を重複パスの始点ノードに送信
  - エンドホストは自分が計測したパスと他のエンドホストが計測したパスの経路情報から、計測する必要がなくなるパスを見つける




- 例：
  - Aが計測するパス：{AB,AC,AD}，Cが計測するパス：{CA,CB,CD}
  - AはABを計測し、AB上のリンクが障害なしの結果をCに送信
  - CはCDを計測し、CD上のリンクが障害なしの結果をAに送信
  - Aは自身とCの計測結果から、AD上のリンクが障害ないと判断し、ADの計測を省略する

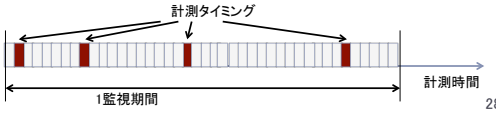
27

## 提案の障害検出方法(2/2)

- 計測パス数に基づいて、動的に計測タイミングを決定
  - 監視期間を計測パス数分だけの等間隔タイムスロットに分割
  - タイムスロットの中に、ランダムに計測タイミングを決定する
- 情報交換により、計測パス数が減少するため、タイムスロットの幅が大きくなり、計測タイミングが監視期間の中に広く分布するようになる

⇒

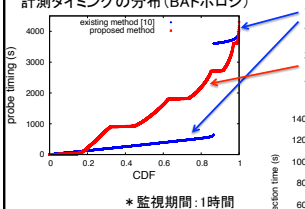
障害を早期に検出できる




28

## シミュレーション評価

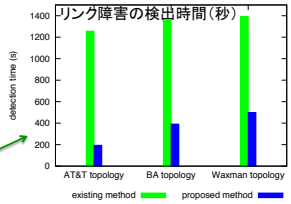
計測タイミングの分布(BAトポロジ)




＊監視期間：1時間

既存手法の計測タイミングが監視機関の冒頭に集中

提案手法の計測タイミングが監視機関の全体に広く分布

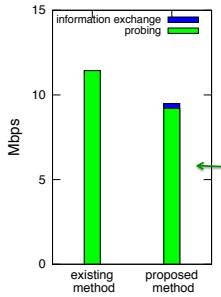



提案手法の障害検出時間は既存手法より大きく下回る

29

## シミュレーション評価

計測トラヒック量と情報交換トラヒック量(AT&Tトポロジ)




設定：
(1)パケットロス率の計測を用いて、リンク故障を検出する
(2)一回計測は10000パケットを用いる
(3)一回の計測結果の交換は1パケットを用いる
(4)パケットサイズは50バイト

提案手法は情報交換トラヒック量が大きいが、計測トラヒックが小さい

計測に費していたオーバヘッドのごく一部を情報交換に費やすことで障害検出時間を大幅に削減できる

＊この評価結果は博士論文に追記した

30

## 本章のまとめ

- 分散型リンク障害検出手法の提案
  - 計測結果の交換に基づき、動的に計測タイミングを決定
  - 計測パス数を減らすとともに計測間隔を増加させ、計測タイミングを監視期間に広くに分布させる
- シミュレーションによる性能評価
  - 提案手法の障害検出時間が既存手法と比べて大幅に短縮できることを確認した

31

## 本研究のまとめ

- エンドホストによって構成されるネットワークアプリケーションのための分散型計測手法を提案
  - 遅延時間とパケットロス率の計測手法
    - 重複部分の計測結果を交換することにより、計測精度を改善できることを確認した
  - 利用可能帯域の計測手法
    - 重複パスの計測結果を交換し、計測における初期推定範囲を推定することにより、計測時間を短くし、計測競合を軽減することにより、計測精度を向上できることを確認した
  - リンク障害の検出手法
    - 重複パスの計測結果を交換することにより、パスの計測時間を広く分布でき、リンクの障害を早期に検出できることを確認した
- 従来手法が計測に費やしていた時間やオーバヘッドのごく一部を情報交換に費やすことで、計測精度の向上や障害検出時間の短縮に役立つことを確認した

32

## 今後の課題

- 提案手法を実環境で評価する
  - 大規模ネットワークに提案手法を適用し、有効性を確認する
- 提案手法を他のネットワーク分野に適用する
  - 特にネットワーク資源の競合が存在する環境において、競合を解決する方法

33