社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

計測衝突を軽減するための分散型オーバレイネットワーク計測手法

ディンティエン ホアン† 　長谷川　剛†† 　村田　正幸†

† 大阪大学大学院情報科学研究科　〒 565-0871 大阪府吹田市山田丘 1-5
†† 大阪大学サイバーメディアセンター　〒 560-0043 大阪府豊中市待兼山町 1-32
E-mail: †{d-hoang,murata}@ist.osaka-u.ac.jp, ††hasegawa@cmc.osaka-u.ac.jp

あらまし　オーバレイネットワークでは，経路制御効率やアプリケーション品質を向上させるために，アンダーレイネットワークの性能をリアルタイムかつ高精度に計測する必要がある．この時，複数のオーバレイパスの計測を同時に行うことにより，計測にかかる時間を短縮することができるが，計測する複数のパスの経路が重複している場合には，重複箇所において計測衝突が発生し，計測負荷の増大や計測精度の低下が問題となる．本稿では，集中型制御を必要とせず，個々のオーバレイノード自身の挙動，及び周囲のオーバレイノードとの最小限の情報交換により，経路重複の状態を判断し，計測衝突を回避する分散型オーバレイネットワーク計測手法を提案する．提案手法は，オーバレイノードが自身を始点とするパスを逐次的に計測し，衝突を回避する．さらに，計測タイミングをランダムに決定することによって，始点オーバレイノードが異なる経路重複パスにおける計測衝突の確率を小さくする．また，周囲のオーバレイノードと経路情報及び計測結果を交換し，統計処理を行うことにより，計測結果の精度を向上する．
キーワード　オーバレイネットワーク，ネットワーク計測，計測衝突，経路制御，統計処理

A distributed measurement method for reducing measurement conflict in overlay networks

Dinh TIEN HOANG†, Go HASEGAWA††, and Masayuki MURATA†

† Graduate School of Information Science and Technology, Osaka University　1-5, Yamadaoka, Suita,Osaka, 565-0871 Japan
†† Cybermedia Center, Osaka University　1-32, Machikaneyama-cho, Toyonaka, Osaka, 560-0043 Japan
E-mail: †{d-hoang,murata}@ist.osaka-u.ac.jp, ††hasegawa@cmc.osaka-u.ac.jp

Abstract　In overlay networks, in order to obtain accurate measurement results, it is important to take care of the measurement conflict problem. This problem occurs when overlapped paths are measured simultaneously. In this report, we propose a measurement method which reduce the number of measurement conflicts without centralized control to schedule the measurement. In this method, an overlay node uses traceroute to get path information to other overlay nodes and exchanges with nearby overlay nodes to estimate path overlaps. Based on the number of overlapped paths, the overlay node calculates an appropriate measurement frequency and a measurement timing to minimize the probability of measurement conflicts among overlapped paths. Furthermore, the overlay node exchanges measurement results with overlay nodes of overlapped paths to statistically obtain more exact measurement results.
Key words　Overlay networks, network measurement, measurement conflict, routing, statistical method

1. まえがき

オーバレイネットワークはネットワークサービスの迅速な展開を可能にする技術として，近年注目されている．一般にオーバレイネットワークとは下位ネットワーク上に構築された論理的なネットワークのことである．例えば，ダイヤルアップを利用したインターネットは電話網上のオーバレイネットワークといえる．本稿では，IP ネットワーク上に構築されたオーバレイネットワークを対象とする．

一般的なオーバレイネットワークでは，オーバレイノードはエンドホスト上にアプリケーションプログラムの形で実現されている．この場合，エンドホスト間通信が通過するネットワーク内でのルーティングやトラヒック制御は，そのネットワーク内ルータによって行われ，オーバレイノードが行うことができないため，効率的なオーバレイパスの選択ができない．一方，ネットワーク内ルータ上にオーバレイノードを設置することに

よって，ネットワーク内においてもオーバレイルーティングなどが可能となる．本稿では，ルータ上にオーバレイノードを設置することを前提とし，より効率的なオーバレイネットワークを構築することを目指す．

オーバレイネットワークはその性能の維持と向上のために，定期的にアンダーレイネットワークの資源や性能に関する情報（帯域，伝播遅延時間，パケット廃棄率など）を計測によって得る必要がある．オーバレイネットワークにおける計測手法は今まで数多く提案されている [1]〜[6]. [1] では，すべてのオーバレイノード間のパスを単純な方法によって計測するため，計測オーバヘッドは $O(n^2)$ となる．ここで，n はオーバレイノード数である．そのため，この手法はオーバレイノード数が 50 程度までのネットワークにしか適用できないと指摘されている [7].

そこで，計測オーバヘッドを削減する方法が提案されている [2]〜[6]. [2] では，オーバレイパス重複の状態に基づいて，オーバレイパスをいくつかのセグメントに分割する．次に，そのすべてのセグメントを通過するオーバレイパスの集合の中で，要素数の最小の集合をヒューリスティックアルゴリズムを用いて決定し，計測の対象とする．一般に，オーバレイネットワークにおけるパスの経路は重複することが多いため，すべてのセグメントを通過する要素数の最小のオーバレイパス集合は，その要素数が全体のオーバレイパスの数と比べて少ない．そのため，計測するパス数が $O(nlog(n))$ まで削減される．この手法では要素数の最小のオーバレイパス集合のオーバレイパスを計測し，計測結果を用いてセグメントの性能を近似的に推測する．次に，その推測結果を用いて全体のオーバレイパスを近似的に推測する．そのため，この手法では正確な計測結果が得られない．

[3] では，パスの性能を，そのパスの経路に含まれるリンク性能の線形方程式として表しパス上の一部のリンクの性能を計測することで，すべてのリンクの性能，及びパスの性能を推定する．一般的に，ネットワーク内のリンク数はパス数よりも小さいため，計測の必要なパス数は $O(nlog(n))$ となる．本手法は [2] の手法と比べて，正確な結果が得られるものの，一般的なネットワーク内のリンク数は, [2] におけるセグメント数よりも大きいため，リンク性能の推定の際の計算量が増加すると考えられる．

[4] は，利用可能帯域の計測手法 BRoute を提案している．BRoute は，以下に示すインターネット上のオーバレイネットワークの 2 つの性質に基づいて，計測オーバヘッドを $O(n)$ まで削減する．すなわち，(1) ボトルネックリンクがオーバレイパスの両端からおよそ 4 ホップ以内に存在すること，及び，(2) オーバレイパスはその両端に近い場所で経路重複が発生しやすいこと，である．そのため，各オーバレイパスの両端に近い部分の利用可能帯域を計測することで，パス全体の利用可能帯域が得られ，計測オーバヘッドを大きく削減できる．しかし，この手法は遅延時間とパケット廃棄率の計測に適用できない．

オーバレイネットワークにおいては，オーバレイノードの密度 (ルータ数に対するオーバレイノード数) が高くなるにつれて，経路が重複するパス数が増加する．重複するパス同士で同時に計測を行うと，計測トラヒックが衝突し，計測結果に誤差が生じる．しかし，上述した各手法 [1]〜[4] は衝突を回避する仕組みを提供していない．

計測衝突を回避するために, [5] では，アンダーレイネットワークの経路情報を一箇所（スーパーノードと呼ぶ）に集約し，経路重複の状態を基に計測タイミングをスケジューリングすることで，計測衝突を回避する手法を提案している．この手法では，ひとつのパスの計測を計測タスクと定義し，各計測タスクの実行時間，及び間隔に基づいて，同じ実行時間を持つ部分タスクに分割し，同時実行可能な部分タスク数を最大化するスケジューリングを行う．これにより，衝突を回避しながら，全体の計測時間を短縮することができる．しかし，計測タスクは分割できるとは限らない．例えば，PathChirp [8] のような利用可能帯域計測ツールでは，計測用パケットの送信タイミングを調整し，受信パケットの振る舞いによって利用可能帯域を推測し，次回の計測パケットの送信間隔を調整するため，計測タスクを分割することができない．

また，上述した各手法 [2], [3], [5] に共通する特徴として，アンダーレイ経路の情報をスーパーノードに集約し，スーパーノードが計測タイミングや計測パスを決定し，個々のオーバレイノードへ指示を与える，ということが挙げられる．そのため，アンダーレイ経路情報の収集にかかる時間及びトラヒック量が大きいこと，また，アンダーレイネットワークの変化やオーバレイネットワーク構成の変更に伴い，スケジューリングなどを再度行う必要があるため，その間計測性能が低下するという問題がある．

[6] は，スーパーノードを使用せず，計測衝突を回避できる利用可能帯域計測システム ImSystemPlus を提案している．ImSystemPlus では，オーバレイノード間で，重複する経路に対して，重複の度合及びトラヒック量に基づいて，経路重複が発生しているパスの計測タイミングをずらし，衝突を回避する．しかし，そのために各オーバレイノードは IP ネットワークの完全なトポロジを把握できるという前提が必要である．

そこで本稿では，スーパーノードを使用せず，かつ，IP ネットワークの完全なトポロジ情報を必要としない，オーバレイネットワーク計測手法を提案する．具体的には，個々のオーバレイノードが自身を始点とするオーバレイパスの計測タイミングを決定し，計測衝突を回避する．提案手法は，個々のオーバレイノードが他のオーバレイノードまでのアンダーレイ経路情報を取得し，他のオーバレイノードと経路情報を交換することにより，自身を始点とする経路と，他のオーバレイノードを始点とするパスの経路重複の状態を推定する．1 つのオーバレイノードを始点とする複数のパスは，逐次的に計測を行うことで，計測衝突を回避する．一方，始点が異なる経路は，始点オーバレイノードがランダムに計測タイミングを決定することで，衝突を確率的に回避する．

以下，本稿は次のように構成される．2. 章では，オーバレイパスの経路重複を検出するための方法を説明する．3. 章では，計測衝突を回避するオーバレイパス計測手法について述べる．4. 章では，提案手法を適用した計測システムの構成を説明する．5. 章では，提案手法の評価を行う．最後に 6. 章で本稿のまとめと今後の課題を述べる．

2. 経路重複の検出方法

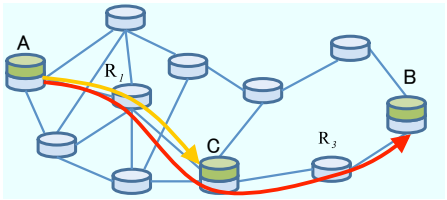
図 1 は，2 つのオーバレイパスのアンダーレイ経路が重複，すなわち，オーバレイパスの一部が同じルータ及びルータ間リンクを通過している状態を分類したものである．本稿では，経路重複の状態に従って以下の 3 通りに分類する．

- 完全重複：1 つのパスが他方のパスを完全に含む．
- 片側重複：2 つのパスが始点オーバレイノードからオーバレイノードでない一部のルータまでの経路を共有する．
- 部分重複：2 つのパスが経路上のオーバレイノードを含まない一部の経路を共有する．

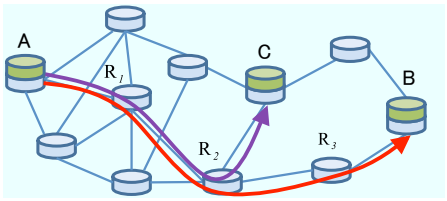
完全重複及び片側重複は，オーバレイパスの始点オーバレイノードが他のオーバレイノードへ traceroute を発行し，その結果を処理することによって検出することができる [9]．例えば，図 1(a) では，A から B,C へ traceroute を発行すると，パス AB とパス AC が完全重複していることを検出することができる．同様に，図 1(b) では，A から B,C へ traceroute を発行すると，パス AB とパス AC 途中のルータ R_2 までの経路を共有していることをを検出することができる．一方，部分重複の場合，traceroute を用いるだけでは重複を正しく検出することができない．そこで，本稿は，以下の部分重複の状態の検出方法を提案する．

（1） traceroute の結果に基づく部分重複状態の推定

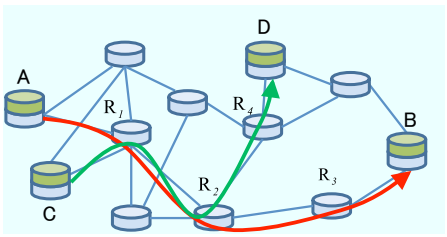
図 1(c) において，A は他のオーバレイノードへ traceroute を発行する．その結果，パス AB とパス AC が片側重複の関係



(a) 完全重複の構成例



(b) 片側重複の構成例



(c) 部分重複の構成例

図 1　オーバレイパスの経路重複の状態

にあり，パス AB とパス AD が片側重複の関係にあることが検出できたとする．この時，パス AB とパス CD が部分重複の関係にある可能性がある，と判断する．

（2） オーバレイノード間の情報交換による部分重複関係の確定

A は，部分重複関係にあると推測されたパスの始点オーバレイノードである C と経路情報の交換を行い，パス AB とパス CD が部分重複の関係にあるか否か判定する．

本手法により，スーパーノードのような 1 箇所に全ての経路情報を集約することなく，オーバレイノード間の最小限の情報交換により，図 1(c) のような部分重複関係を特定することができる．

3. オーバレイネットワークにおける計測手法

本章では，オーバレイネットワークにおいてオーバレイパスの計測を行う際の，計測衝突を回避する方法を提案する．なお，紙面の制約上，任意のオーバレイパス AB についての具体例を示すことにより提案手法を説明する．まず，パス AB が他のオーバレイパスと経路が重複しない場合，計測衝突が発生しないため，衝突回避方法は不要である．そこで，以下ではパス AB が他のオーバレイパスと経路重複する場合を考える．

経路重複の状態により，以下の 2 つの場合を考える．

（1） パス AB があるオーバレイパスを完全に含む場合

この場合，パス AB とその重複パスは完全重複の関係にあるため，3.1.1 節において説明する手法により，オーバレイパス AB は計測の対象とならない．

（2） パス AB が他のオーバレイパスを含まない場合

この場合，パス AB の片側重複パス数と部分重複パス数に応じて，計測頻度と計測タイミングを決定することにより，衝突回避の確率を小さくする．詳細は 3.1.2 節に示す．

また，オーバレイノード間で計測結果を交換し，統計処理を行うことにより，計測結果の精度を向上する．具体的な手法については 3.2 節で説明する．

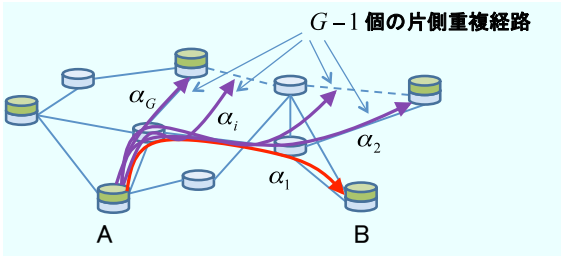


図 2　パス AB の片側重複パス

3.1 計測衝突回避方法

3.1.1 完全重複の場合

この場合は，ホップ数の多いオーバレイパスを直接計測せず，ホップ数の少ないオーバレイパスをまず計測し，その計測結果を用いてホップ数が多いオーバレイパスの性能を推定する．

図 1(a) の例を用いて説明する．図 1(a) ではパス AB がパス AC を含んでいる．ここで，A がパス AB を計測する場合を考える．A が B に対して traceroute を発行すると，traceroute の通信パケットが C を通過するため，C は自身がパス AB 上に存在することを知ることができる．そこで，C はパス CB を計測し，その結果を A に送信する．また，A は traceroute の実行結果より，C がパス AB 上に存在することを知ることができる．そのため，A はパス AB を直接計測せず，パス AC のみを計測する．そして，A はパス AB の計測結果と，C から受信したパス CB の計測結果よりパス AB の計測結果を推定する．本手法の詳細については [9] を参照されたい．

3.1.2 片側重複及び部分重複の場合

パス AB は一般に複数の片側重複パスと部分重複パスを持つ．ここで，図 2 が示しているように，AB は $G-1$ ($G \geqq 1$) 個の片側重複パスを持つとする．さらに，A は 2. 章に述べた推測方法を用いて，パス AB と $G-1$ 個の片側重複パスを合わせた G 本のパスが，それぞれ $K_i - 1$ ($1 \leqq i \leqq G$) 個の部分重複パスを持つと推測されたとする．

オーバレイノード A はパス AB と $G-1$ 個の片側重複パスの計測を逐次的に行うことにより，片側重複パスとの計測衝突を回避することができる．一方，パス AB 及びその $G-1$ 個の片側重複パスの部分重複パスは，始点が A ではないため，A 以外のオーバレイノードが計測する．そのため，計測衝突を完全に回避することはできない．そこで，片側重複パスに対する逐次的な計測と，部分重複パスに対するランダム計測を合わせた手法を提案する．

A は，ある時間間隔 T ごとに，パス AB と $G-1$ 個の片側重複パスの計測結果を集計するとする．一般的に，計測衝突の発生や，パスの性能自身が時間的に変動することが原因となり，パスの計測結果は計測の度に変動する．そのため，その変動の大きさに従い，計測する頻度を調整する必要がある．具体的には，計測結果が大きく変動するパスでは，計測の頻度を大きくする必要があり，計測結果が安定しているパスでは，計測の頻度を小さくしても良いといえる．

計測頻度の大きさを表す指標として，計測時間割合と呼ばれる指標を導入する．計測時間割合は次のように定義する．まず，簡単のため，パスの一回の計測時間はすべて τ であると仮定し，オーバレイパスが時間 t ($t \geqq \tau$) 毎に 1 回の計測を行うなら，そのパスの計測時間割合は τ/t であると定義する．

計測結果の変動の大きさに応じて計測時間割合を決定した結果，パス AB の計測時間割合が β_1，$G-1$ 個の片側重複パスの計測時間割合が β_i ($2 \leqq i \leqq G$) と決定されたとする．これらのパスの計測時間が重ならないためには，計測時間割合の合計が 1 以下（つまり，$\sum_{i=1}^{G} \beta_i \leqq 1$）でなければならない．したがって，計測時間割合の合計が 1 より大きい場合（つまり，$\sum_{i=1}^{G} \beta_i > 1$

の場合)，計測時間割合を削減しなければならない．その場合，以下の式に従い，計測時間割合 x_i を求める

$$\begin{array}{l}\frac{\beta_1-x_1}{x_1}=\frac{\beta_2-x_2}{x_2}=...=\frac{\beta_G-x_G}{x_G}\\ x_1 \leqq \beta_1, x_2 \leqq \beta_2, ..., x_G \leqq \beta_G\\ \sum_{i=1}^{G} x_i = 1\end{array} \quad (1)$$

問題 (1) の解 x_i^* は式 (2) で与えられることが確認することができる．

$$x_i^* = \frac{\beta_i}{\sum_{i=1}^{G}\beta_i}, \quad i=1,...,G \quad (2)$$

提案手法においては，式 (2) で導出される値を用いる．なお，$\sum_{i=1}^{G}\beta_i \leqq 1$ の場合は，$x_i^*=\beta_i$ とする．

ここで，パス AB と $G-1$ 個の片側重複パスの計測時間割合を x_i^* に設定した計測を行うことができることを証明する．

定理 1. パス AB と $G-1$ 個の片側重複パスの計測時間割合が x_i^* となるような計測方法が存在する．

証明. $L_i=1/x_i^*$ とする．ここで，$L_1 \leqq L_2 \leqq ... \leqq L_G$ と仮定しても一般性を失わない．$\sum_{i=1}^{G} 1/L_i = \sum_{i=1}^{G} x_i^* = 1$ であるため，ある $1 \leqq l \leqq G$ が存在し，$L_1 \leqq ... \leqq L_l \leqq G \leqq L_{l+1} \leqq ... \leqq L_G$ を満たす．

1 つのパスを計測するために割り当てられる時間を計測タイムスロットと呼び，パス AB と $G-1$ 個の片側重複パスを計測するのにかかる時間を計測周期と呼ぶ．簡単のため，パス AB と $G-1$ 個の片側重複パスの順番をつけ，パス AB をパス 1 と呼び，$G-1$ 個の片側重複パスをそれぞれパス $i(2 \leqq i \leqq G)$ と呼ぶ．以下の計測方針を考える．

（1） 1 つの計測周期において，パス AB と $G-1$ 個の片側重複パスの計測順番をランダムに決定し，各パスに計測タイムスロットを割り当てる．

（2） $i>l$ を満たすパス i は，自身に割り当てられた計測タイムスロットにおいて，確率 G/L_i で計測を行う．つまり，確率 $1-G/L_i$ でこのパスは計測されない．パス $i(i>l)$ が計測されない場合，その計測タイムスロットをパス $j(j \leqq l)$ の計測に使用する．

（3） $j \leqq l$ を満たすパス j は，自身に割り当てられた計測タイムスロットにおいて，常に計測を行う．これにより，パス j の計測時間割合は $1/G$ $(<1/L_j)$ となり，設定したい計測時間割合 $1/L_j$ と比べて，$1/L_j-1/G$ だけ不足することになる．

上述した計測方法により，すべてのパスの計測時間割合を $1/L_i(i=1,...,G)$ にすることができることを以下に証明する．

- パス $i(i>l)$ の場合

パス i は割り当てられた計測タイムスロットにおいて確率 G/L_i で計測されるが，1 つの計測周期は G 個の計測タイムスロットを持つため，計測時間割合は $1/L_i$ となる．

- パス $j(j \leqq l)$ の場合

パス $j(j \leqq l)$ の計測タイムスロットで計測されるため，設定したい計測時間割合 $1/L_j$ と比べて，計測時間割合が $1/L_j-1/G$ だけ不足している．したがって，すべてのパス $j(j \leqq l)$ の計測時間割合を合わせると，$\sum_{j=1}^{l}(1/L_j-1/G)$ だけ不足している．

一方，パス $i(i>l)$ では，計測タイムスロットで確率 $1-G/L_i$ でこのパスが計測されないため，計測周期では $(1-G/L_i)*1/G=1/G-1/L_i$ の確率で計測されない．

したがって，すべてのパス $i(i>l)$ がある計測周期において計測されない確率を足し合わせると，$\sum_{i=l+1}^{G}(1/G-1/L_i)$ の確率で，パス $j(j \leqq l)$ の計測に使用することができる．

ここで，$\sum_{i=1}^{G} 1/L_i = \sum_{i=1}^{G} x_i^* \leqq 1$ であるため，$\sum_{j=1}^{l}(1/L_j-1/G) \leqq \sum_{i=l+1}^{G}(1/G-1/L_i)$ である．したがって，パス $i(i \leqq l)$ の計測時間割合を $1/L_i$ にすることができる．

□

次に，パス AB と K_1-1 個の部分重複パスとの計測衝突を確率的に回避する方法を説明する．上述したように，パス AB と部分重複関係にあるパスは，A 以外のオーバレイノードが計測するため，AB とその部分重複パスの衝突を完全には回避できない．そこで，パス AB と部分重複パスとの計測衝突の確率を小さくする方法を以下に提案する．

パス AB が K_1 個の部分重複パスを持つ場合，パス AB の計測時間割合を $1/K_1$ 以下に抑える．つまり，パス AB の計測時間割合はは以下の式により決定する．

$$y_1^* = \min\{x_1^*, 1/K_1\} \quad (3)$$

以上の手法により，パス AB の計測時間割合 y_1^* を決定する．パス AB は，決定された計測時間割合に従い，1 回の計測周期中のランダムなタイミングで計測を行う．

3.2 情報交換および統計処理による精度向上

部分重複関係にあるパスは，3.1 節に示した方法により計測衝突を確率的に回避するが，完全に回避できないため，計測衝突による精度の低下が避けられない．そこで提案手法においては，パスの経路が重複している箇所の計測をより綿密に行い，計測結果をオーバレイノード間で交換し，統計処理を行うことによって，精度の補完を行う．本節では，経路重複部分の計測，情報交換，および統計処理手法を説明する．

3.2.1 部分重複箇所の計測方法

この節では，部分重複パスを持つパス AB の計測手順を説明する．ここでは，遅延時間を計測する場合について述べる．

パス AB 上の経路において，部分重複関係にある他のパスと経路が重複している箇所は一般的に複数ある．それら重複箇所の端点を A 側から順に R_1，R_2，...，R_l とする．提案手法においては，AB 間の計測のみを行うのではなく，これらの端点間 R_1R_2，...，$R_{l-1}R_l$ のネットワーク性能を個別に計測する．図 3 はパス AB とルータ R_1，R_2，...，R_l の状態を示している．

以下では，A がパス AB の遅延時間を計測する手順を説明する．

A は，B へ traceroute を発行し，パス AB 上のルータを検出する．次に，A はこれらのルータに対して，ping を発行し，その結果を基づいて，AR_1，R_1R_2，...，$R_{l-1}R_l$，R_lB の遅延時間を取得する．取得方法は以下の通りである．AR_1，AR_2，...，AR_l，AB の遅延時間はそれぞれ t_{A,R_1}，t_{A,R_2}，...，t_{A,R_l}，$t_{A,B}$ で表し，A から各ルータへの ping によって取得する．その結果を用いて，パス AB 上の各ルータ間の AR_1，R_1R_2，...，$R_{l-1}R_l$，R_lB の遅延時間 t_{A,R_1}，t_{R_1,R_2}，...，t_{R_{l-1},R_l}，$t_{R_l,B}$ を以下のようにして算出する．

$$\begin{array}{ll}t_{R_i,R_{i+1}} = t_{A,R_{i+1}}-t_{A,R_i} & ,i=1,...,l-1\\ t_{R_l,B} = t_{A,B}-t_{A,R_l} & \end{array} \quad (4)$$

3.2.2 計測結果の情報交換

上述の計測結果，および AB 間の経路情報を，部分重複パスの始点オーバレイノードに送信する．その際，相手のオーバレイノードから同様の情報を受信する．

3.2.3 統計処理手法

最後に，情報交換によって得られた情報を用いた統計処理手法，およびパス AB の計測結果の決定手法を説明する．

A はパス AB に対する自身の計測結果と，ほかのオーバレイノードとの情報交換によって得られた計測結果から，経路上のルータ間の遅延時間を計算する．以下では，パス AB の経路上

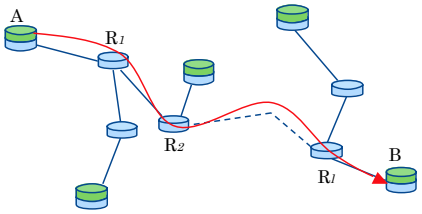


図 3　パス AB と AB 上のルータの状態

のルータ R_1R_2 間での遅延時間の計算方法を説明する．

A は他の $m-1$ 個のオーバレイノードと，R_1R_2 間の計測結果を交換するとする．A とその他の $m-1$ 個のオーバレイノードが R_1R_2 間を計測するとき，衝突が発生する可能性があり，計測衝突の状態によって，異なる計測結果が得られると考えられる．つまり，A とその他の $m-1$ 個の，合計 m 個のオーバレイノードの中の i 番目オーバレイノードの計測結果を X_i $(1 \leqq i \leqq m)$ で表すと，X_i は確率変数と見なすことができる．

以下の X_1，X_2，...，X_m の平均値 X を R_1R_2 間の遅延時間とする．

$$X = \left(\sum_{i=1}^{m} X_i\right)/m \tag{5}$$

以降，X の期待値は，各オーバレイノードから得られた遅延時間の計測結果の期待値の平均値となり，さらにその分散は，情報交換相手が自身の計測結果のみを用いて算出した分散の平均値より小さくなることを証明する．

ここで，確率変数 Z の期待値を $E(Z)$ で，分散を $V(Z)$ で表す．まず，式 (5) より，式 (6) が成り立つ．

$$E(X) = \left(\sum_{i=1}^{m} E(X_i)\right)/m \tag{6}$$

また，分散 $V(X)$ は式 (7) により評価することができる．

定理 2.

$$V(X) \leqq \left(\sum_{i=1}^{m} V(X_i)\right)/m \tag{7}$$

証明. 任意の確率変数 Y，Z と任意の定数 a に対して，以下の各式が成り立つ．

$$E(Y+Z) = E(Y)+E(Z) \tag{8}$$

$$E(aY) = aE(Y) \tag{9}$$

$$V(Y+Z) = V(Y)+V(Z)+2cov(Y,Z) \tag{10}$$

$$V(aY) = a^2V(Y) \tag{11}$$

ただし，$cov(Y,Z)$ は Y と Z の共分散であり，式 (12) を満たす．

$$cov(Y,Z) = E(YZ)-E(Y)E(Z) \tag{12}$$

また，式 (12) より，以下が成り立つことを容易に確かめられる．

$$cov(Y,-Z) = -cov(Y,Z) \tag{13}$$

さらに任意の確率変数 W に対して，式 (12) より，以下が成り立つ．

$$cov(Y+W,Z) = cov(Y,Z)+cov(W,Z) \tag{14}$$

したがって，式 (10)，(11) と (13) より以下の式が成り立つ．

$$\begin{aligned} V(Y-Z) &= V(Y)+V(-Z)+2cov(Y,-Z) \\ &= V(Y)+V(Z)-2cov(Y,Z) \end{aligned} \tag{15}$$

また，$V(Y-Z) \geqq 0$ であるため，式 (15) より，

$$V(Y)+V(Z) \geqq 2cov(Y,Z) \tag{16}$$

が成り立つ．

さらに，式 (10) と (14) より，式 (17) のように展開することができる．

$$V(\sum_{i=1}^{m} X_i) = \sum_{i=1}^{m} V(X_i)+2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} cov(X_i,X_j) \tag{17}$$

したがって，式 (16) より，

$$\begin{aligned} V(\sum_{i=1}^{m} X_i) &\leqq \sum_{i=1}^{m} V(X_i)+\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}(V(X_i)+V(X_j)) \\ &= m\sum_{i=1}^{m} V(X_i) \end{aligned} \tag{18}$$

最後に，式 (9) より，$V\left((\sum_{i=1}^{m} X_i)/m\right) = \left(V(\sum_{i=1}^{m} X_i)\right)/m^2$ となり，これを式 (18) に適用すると，式 (7) が得られる． □

式 (6) は，R_1R_2 間の遅延時間の期待値が情報交換によって得られた遅延時間の計測結果の期待値の平均値になることを示している．また，式 (7) は，各オーバレイノードの計測結果を集めて，統計的に処理することによって，計測結果の分散はそれぞれのオーバレイノードで処理した計測結果の分散の平均値よりも小さくなることを示している．分散が小さくなれば，信頼区間幅も狭くなり，精度が高くなるため，オーバレイノード間の計測結果の交換は有効である．

上述のようにして得られた，パス AB 上の各ルータ間の AR_1，R_1R_2，...，$R_{l-1}R_l$，R_lB の遅延時間の分散をそれぞれ v_{A,R_1}，v_{R_1,R_2}，...，v_{R_{l-1},R_l}，$v_{R_l,B}$ とする．なお，v_{R_1,R_2}，...，v_{R_{l-1},R_l} は，式 (7) を用いて評価することができる．

また，AR_1，R_1R_2，...，$R_{l-1}R_l$，R_lB は重複しないため，それぞれの遅延時間を表す確率変数は互いに独立であると考えられる．さらに，それぞれの遅延時間を表す確率変数は正規分布に従うと仮定すると，正規分布の合成の性質により，パス AB 全体の遅延時間も正規分布に従う．

そのため，パス AB 全体の遅延時間の平均値 T と分散 V と 95% の信頼区間はそれぞれ式 (19)，(20)，(21) により計算することができる．

$$T = t_{A,R_1}+\left(\sum_{i=1}^{l-1} t_{R_i,R_{i+1}}\right)+t_{R_l,B} \tag{19}$$

$$V = v_{A,R_1}+\left(\sum_{i=1}^{l-1} v_{R_i,R_{i+1}}\right)+v_{R_l,B} \tag{20}$$

$$\text{信頼区間：}\left(T-1.96\sqrt{\frac{V}{n}},T+1.96\sqrt{\frac{V}{n}}\right) \tag{21}$$

4. 計測システム

本章では，前章に述べた計測手法を適用した計測システム構成について説明する．紙面の制約上，以下では，任意のオーバレイノード A が行う計測の手順について説明する．

- Job1：traceroute による経路重複状態の調査

A はまず，他のすべてのオーバレイノードへ traceroute を実行し，A を始端としたパスを調査する．そして，実行結果に基づいて，それぞれのパスの重複状態を判断し，片側重複関係にあるパス群を各グループにまとめる．

- Job2：パス計測，情報交換，及び統計処理

以下の 3 つの手順は同時に実行可能である．

− Job2.1：パスの計測

A は 3.1 節で提案した手法に基づいて各パスの計測を行う．

− Job2.2：他のオーバレイノードとの情報交換

A は他のオーバレイノードとの間でパス情報や計測結果を交換する．

− Job2.3：パスの計測結果の計算

A は 3.2 節で説明した手法に基づき，自身の計測結果と情報交換によって得られた他ノードによる計測結果を用いて統計処理を行い，計測結果を確定する．

アンダーレイネットワークの経路は変化する可能性があるため，Job1 は定期的に実行する．一般的に，アンダーレイ経路の変化の頻度は，パスの性能の変化の頻度よりも小さいため，Job1 の実行頻度は Job2 の実行頻度より小さくする．

5. 性能評価

提案手法では各オーバレイノードが自身を始点とするパスの計測頻度及び計測タイミングを決定する．これにより，スーパーノードを必要とする従来手法と違い，計測トラヒックや計算オーバヘッドが一ヶ所に集中せず，分散的に実行することができる．

既存手法 [2], [3], [5] では，それぞれのオーバレイノードが自身を始点とする経路情報を取得し，その計測結果をスーパーノードに送信する必要があるが，オーバレイノード数を n とすると，使われる可能性のあるオーバレイパス数は $O(n^2)$ であるため，スーパーノードにおける計測結果収集のためのトラヒック量は $O(n^2)$ になる．一方，提案手法では，経路情報の 1 箇所への集約は行わず，それぞれのオーバレイノードが自身を始点とするパスの重複部分の計測結果をやり取りするため，オーバヘッドは $O(n)$ となる．

計測パス数に関しては [2] の手法では，$O(nlogn)$ となる．一方，提案手法では，完全重複関係にあるパスは計測する必要がないため，計測パス数を削減することができる．以下では，[10] における議論に従い，提案手法とフルメッシュ計測の計測パス数を比較し，提案手法の有効性を確かめる．まず，L_{max} をオーバレイパスの最大経路長 (通過する IP ルータ間のリンク数)，$P(r)$ を全経路に対する経路長が r であるパスの割合，N をルータ数，d をオーバレイノード数の割合 ($d = n/N$) とする．[10] によると，提案手法の計測パス数は式 (22) より，与えられる．

$$M = \sum_{r=1}^{L_{max}} \left(P(r)Nd(Nd-1) \prod_{i=1}^{r-1} \left(1 - \tfrac{Nd-2}{N-1-i}\right) \right) \tag{22}$$

一方，フルメッシュ計測の計測パス数は式 (23) より，与えられる．

$$M_{full} = Nd(Nd-1) \tag{23}$$

従って，提案手法の計測パス数とフルメッシュ計測の計測パス数の比は式 (24) より，計算できる．

$$\mu = \tfrac{M}{M_{full}} = \sum_{r=1}^{L_{max}} \left(P(r) \prod_{i=1}^{r-1} \left(1 - \tfrac{Nd-2}{N-1-i}\right) \right) \tag{24}$$

$d \leqq \frac{2}{i+1}$ を満たす最小の整数 i を i_0 とする．

- $i \leqq i_0$ の場合，

この場合，N が十分大きいとき，式 (25) のように近似できる．

$$\tfrac{Nd-2}{N-1-i} \approx d \tag{25}$$

- $i > i_0$ の場合

この場合，式 (26) が成立する．

$$\tfrac{Nd-2}{N-1-i} > d \tag{26}$$

したがって，N が十分大きい場合，μ の値は式 (27) より評価できる．

$$\begin{aligned} \mu &\leqq \sum_{r=1}^{L_{max}} \left(P(r) \prod_{i=1}^{r-1} (1-d) \right) \\ &= \sum_{r=1}^{L_{max}} \left(P(r)(1-d)^{r-1} \right) \end{aligned} \tag{27}$$

式 (27) では，$\sum_{r=1}^{L_{max}} P(r) = 1$，$(1-d)^{r-1} < 1, \forall r > 1$ であるため，式の右辺が 1 より小さく，d が 1 に近づくほど小さくなる．

したがって，d が大きい場合，提案手法はフルメッシュ計測手法と比べて，計測数が大きく削減される．これは [10] において，シミュレーション評価の結果より検証されており，4 つのネットワークモデルに対して，計測パス数が最大で約 1/4000 に削減されることが確認されている．

6. おわりに

本稿では，オーバレイノードが自身を始点とするパスと他のパスとの重複状態を推測し，パスの計測頻度及び計測タイミングを適切に調整することにより，計測衝突の確率を小さくし，計測精度を向上することができる，分散型オーバレイネットワーク計測手法を提案した．

今後の課題として，提案手法の計測精度や計測オーバヘッドの定量的評価，および既存手法との比較評価が挙げられる．

謝辞 本研究の一部は，情報通信研究機構からの委託研究「ダイナミックネットワーク技術の研究開発 課題カ」によっている．ここに記して謝意を示す．

文　献

[1] D. Andersen, H. Balakrishnan, M.Kaashoek, and R. Morris, “Resilient overlay networks,” in *Proceedings of SOSP 2001*, Oct. 2001.

[2] C. Tang and P. McKinley, “On the cost-quality tradeoff in topology-aware overlay path probing,” in *Proceedings of ICNP 2003*, Nov. 2003.

[3] Y. Chen, D. Bindel, H. Song, and R. Katz, “An algebraic approach to practical and scalable overlay network monitoring,” in *Proceedings of SIGCOMM 2004*, Aug. 2004.

[4] N. Hu and P. Steenkiste, “Exploiting internet route sharing for large scale available bandwidth estimation,” in *Proceedings of the IMC’05*, Oct. 2005.

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

[6] C. L. T. Man, G. Hasegawa, and M. Murata, “Monitoring overlay path bandwidth using an inline measurement technique,” *IARIA International Journal on Advances in Systems and Measurements*, vol. 1, no. 1, pp. 50–60, 2008.

[7] A. Nakao, L. Peterson, and A. Bavier, “Scalable routing overlay networks,” *ACM SIGOPS Operating Systems Review*, vol. 40, pp. 49–61, Jan. 2006.

[8] V. J.Ribeiro, R. H. Riedi, R. G. Baraniuk, J. Navratil, and L. Cottrell, “pathchirp: Efficient available bandwidth estimation for network paths,” in *Proceedings of Passive and Active Measurement Workshop 2003*, Apr. 2003.

[9] G. Hasegawa and M. Murata, “Scalable and density-aware measurement strategies for overlay networks,” in *Proceedings of ICIMP 2009*, pp. 21–26, May 2009.

[10] 森弘樹, 長谷川剛, 村田正幸, “オーバレイネットワークにおける経路重複を利用した計測手法,” **電子情報通信学会技術研究報告** *(ICM2008-67)*, vol. 108, pp. 53–58, Mar. 2009.