

広域環境における RTT を用いたネットワークトポロジー推定

水野 悠[†] 柴田 剛志[‡] 田浦 健次朗[‡] 近山 隆^{††}

[†] 東京大学 工学部 電子情報工学科 [‡] 東京大学大学院 情報理工学系研究科 ^{††} 東京大学大学院 工学系研究科

1 はじめに

分散環境における並列アプリケーションの性能向上には、ネットワークの連結の形状 (ネットワークトポロジー) を考慮した最適化を行う必要がある。並列アプリケーションの最適化に関する研究では、ネットワークトポロジーが既知のものであるという前提のもとで行われることが多い。

しかし、実際の環境ではネットワークトポロジーを知るためには、例えば複数のスイッチから情報を収集する必要があると、大きな手間と時間がかかる。また、分散環境を用いるユーザーが皆、スイッチから情報を得る権限を持っているとは限らず、そもそも測定が出来ないという場合も多い。

これらの問題を解決するため、各ホスト間の Round Trip Time(RTT) を用いたトポロジーの推定手法が提案されている [5]。特別なプロトコルを用いないのでヘテロな環境で動作することの他、高速であること、ネットワークへの負荷が低いといった利点を持つ。しかし、この手法ではネットワークをツリー構造として仮定しており、それ以外の構造では用いることが出来ない。本論文ではこの既存研究をより一般化し、実際の WAN 環境などに見られる環状のリンクがある場合など、より一般的なネットワークトポロジーを推定する手法を提案する。

2 関連研究

ネットワークトポロジーの推定には目的・前提の異なる様々な手法が存在する。

レイヤ 2 のトポロジーを推定する研究として、traceroute を用いた手法 [2] や、SNMP などのプロトコルを用いる手法 [1] がある。これらの研究の主な目的はネットワークの管理やトラブルシュートであり、物理的なトポロジーを推定する手法である。決定的な推定を行うことが可能であることや、長い時間をかけて推定を行うことが特徴として上げられる。

パケットロス率やパケットの遅延など、エンドホスト間の測定を用いる研究 [3][4] の多くは、測定結果を統計的に用いて推定を行う。測定に用いたホストの連結に関与しないリンクやスイッチについては情報を得ることが

出来ないで、推定結果は論理的なトポロジーとなる。論理的なトポロジーは物理的なトポロジーの近似であるが、各通信路が同じリンクを利用しているかといった、並列アプリケーションの性能向上に必要な情報を含んでいる。

白井ら [5] はトポロジーとしてツリー構造を仮定し、各ホスト間の RTT のみを用いて推定を行う手法を提案している。ホストやスイッチでは遅延が起これず各リンクは一定の遅延を持つとして、トポロジーをノードと長さを持ったエッジからなるグラフとしてモデル化し、リンクの繋がりごと各リンクの持つ遅延の推定を行う。

任意に選んだ 3 ノード間それぞれの RTT を知ることが出来れば、式 (1) により、図 1 に示すように 3 ノードのツリー構造を一意に求めることができる。AB, BC, CA は各ノード間の RTT の値である。

$$\begin{aligned} x &= (AC + AB - BC)/4, \\ y &= (AB + BC - AC)/4, \\ z &= (BC + AC - AB)/4, \end{aligned} \quad (1)$$

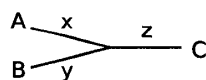


図 1 3 ノードのツリー構造 図 2 ループを含む構造

この方法によって、初めに 3 ノードからなるツリーを作り、そのツリーに 1 つずつノードを加えていくことにより全体のトポロジーを一意に推定する手法を提案している。高速でネットワークへの負荷が軽く、ヘテロな環境でも動作するという利点があるが、トポロジーがツリー構造の場合にしか用いることが出来ないという欠点がある。

3 提案手法

3.1 目的

本論文では前述の白井らによる手法 [5] を先行研究として、ツリー構造にエッジを 1 本加えたネットワークにおいて適用が可能なトポロジー推定手法を提案する。先行研究と同様にネットワークをグラフとするモデル化を行い、リンクの繋がりごと各リンクにおける遅延の推定を低負荷・高速で行うことを目的とする。

3.2 アルゴリズム

一般のトポロジーにおいては任意の 3 ノードが図 1 で表されるツリー構造を取るとは限らず、図 2 で表される

Network Topology Inference Using Round Trip Time on in Environment
by Yu Mizuno[†], Takeshi Shibata[‡], Kenjiro Taura[‡], and Takashi Chikayama^{††}(The University of Tokyo)

構造を取る場合がある。このため、ツリーに順番にノードを加えていく手法を使うことが出来ない。

そこで、以下に述べるように、まずトポロジーの部分木を構成し、それらを繋ぎ合わせることでトポロジーの推定を行う。

1. ノードのグループへの分割

まず初めに、ホストをいくつかの「グループ」に分ける。この「グループ」の定義は「トポロジー全体からループになっているエッジを全て外したとき、連結されているノード群」である。図 3 にグループの一例を示す。グループ内にはループが含まれないので、各グループのトポロジーは必ずツリー構造となる。

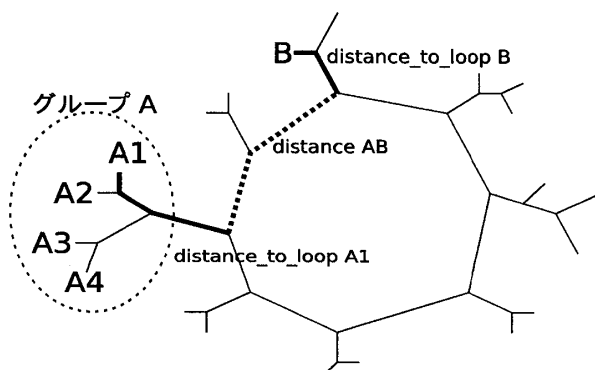


図 3 グループ

「2 ノード A_1, A_2 を選んだとき、あらゆる 2 ノード B, C の組み合わせに対して、 A_1, A_2, B, C の 4 ノード間それぞれの RTT を説明できるツリー構造を作ることが出来れば A_1, A_2 は同じグループに属する」という命題が成り立つので、これを利用してノードのグループ分けを行う。

2. 各グループのトポロジーの構成

各グループのトポロジーを推定する。グループ内はツリー構造になっているので先行研究の手法で用いることができる。

なお、全ノードが同じグループに分類された場合、全体のトポロジーはツリー構造になっていることが分かるので、この段階で全体のトポロジーの推定が行われ、以下の手順 3, 4 を行う必要はない。

3. グループからループへの長さ

各グループからノードを一つずつ選び、そのノードからループまでの経路の長さを求める。

ノード A を選んだとき、あらゆる 2 グループ B, C の組み合わせに対して、それぞれの要素 $B \in B, C \in C$ と A の 3 ノード間のそれぞれの RTT を用いて式 (1) から x を計算すると、その中で最小の x がノード A からループへの経路の長さ $distance_to_loop_A$ である。

4. 各グループの接続

各グループのトポロジーを繋ぎ合わせる。

2 つのグループ A, B のループ上での経路の長さ $distance_{AB}$ は、 $distance_{AB} = AB - distance_to_loop_A - distance_to_loop_B$ で表される。 $distance_to_loop_A, distance_to_loop_B$ は、 $A \in A, B \in B$ のそれぞれからループまでの経路の長さ。 AB は A と B 間の RTT である。

これにより各グループ間のループ上での経路の長さを求めることが出来るので、各グループを組み合わせることで全体のトポロジーを推定することができる。

4 おわりに

本論文ではツリー構造にエッジを 1 本加えた構造のネットワークにおいて、RTT のみを用いてトポロジーを推定する手法を示した。

今後は、より一般的なネットワークにおいて適用可能となるように手法の改良を進める。

また、本手法のアルゴリズムは、各ホスト間の RTT が正確に分かっているということを前提にしているが、実際の環境で動作させる場合には、推定に必要な RTT を効率的な順序で測定し推定の高速化を行うことや、RTT 測定で生じる誤差にも対応できることが必要となってくる。これらも合わせて今後の課題とする。

参考文献

- [1] Y. Breitbart, M. Garofalakis, B. Jai, C. Martin, R. Rastogi, and A. Silberschatz. Topology discovery in heterogeneous IP networks: the NetInventory system. *IEEE/ACM Transactions on Networking (TON)*, Vol. 12, No. 3, pp. 401–414, 2004.
- [2] M. Den Burger, T. Kielmann, and H. E. Bal. TOPOMON: A monitoring tool for grid network topology. In *International Conference on Computational Science (2)*, pp. 558–567. Springer, 2002.
- [3] M. Coates, R. Castro, R. Nowak, M. Gadhiok, R. King, and Y. Tsang. Maximum likelihood network topology identification from edge-based unicast measurements. *ACM SIGMETRICS Performance Evaluation Review*, Vol. 30, No. 1, p. 20, 2002.
- [4] N. G. Duffield, J. Horowitz, F. L. Presti, and D. Towsley. Multicast topology inference from measured end-to-end loss. *IEEE Transactions on Information Theory*, Vol. 48, No. 1, pp. 26–45, 2002.
- [5] 白井達也, 斎藤秀樹, 田浦健次朗. 高速なトポロジー推定—ネットワークを考慮した並列計算のための基盤として. *情報処理学会論文誌*, Vol. 47, No. 4, apr 2006.