

マルチドメイン P2P 型 CDN における経路制御アーキテクチャの提案と評価

浅井 大史*

江崎 浩*

概要

近年、デジタルコンテンツの配布手段として P2P 型 CDN システムが利用されるようになってきている。しかし、一般的に、IP ネットワークにおける経路制御の階層化構造と P2P 型 CDN システムにより構成されるオーバーレイネットワークにおける階層化構造とが異なるため、マルチドメイン P2P 型 CDN システムにおいて、ネットワーク資源を必要以上に消費してしまう状況が発生しているのが現状である。本論文では、マルチドメイン P2P 型のコンテンツ配信システムにおいて、ネットワークトポロジー情報及び ISP 間の相互接続情報を利用することで、ネットワーク資源の効率的利用を実現する配信経路制御手法を提案し、シミュレーションによる評価を行うことで提案手法の有効性を示す。

A Content Delivery Path Control Architecture for the Multi-domain P2P CDN System

Hirochika Asai* Hiroshi Esaki*

Abstract

In order to achieve effective and high quality contents delivery over the Internet, the CDN system adopting the P2P (Peer-to-Peer) technology has been developed and deployed. These P2P CDN systems are generally so called the overlay networks, which construct their own network topology. Since the network topology of P2P CDN system is generally different from the layer 3 network topology and it never take into account the inter-domain routing policy, the P2P CDN system frequently consumes larger network resource and larger network cost from the layer 3 operators' point of view. In this paper, we propose and evaluate a novel content delivery path control architecture for the efficient network resources utilization, using the network topology and the relationship among ISPs.

1. はじめに

1960 年代の研究者間の高価な計算機を共有して利用することを目的として発明・開発されたインターネットは、いまや我々の社会生活や経済および産業活動における重要な情報通信基盤となってきた。近年では、高速 DSL, FTTH といった安価なブロードバンド環境が一般家庭に普及しつつあり、またモバイルネットワーク環境においても高速な通信網が整備されつつある。

一般家庭におけるブロードバンド化は、動画やソフトウェアファイルイメージなどの大容量データの配布を CD, DVD 等の物理媒体ではなく、インターネットを介したデータ通信により実現可能なものとしつつある。この結果として、インターネットで流通・交換されるデー

タトラヒックの総量は現在でも指数関数状に増大している。

このような大容量データの配布手段として、従来はクライアント-サーバ型のコンテンツ配信ネットワーク (CDN) システムが利用されていた。しかし、クライアント-サーバ型の場合にはサーバへの負荷の増大が指摘されており、近年ではサーバ負荷の分散を目的として、BitTorrent [1] に代表されるピア・ツー・ピア (P2P) 型の CDN システムが利用されるようになってきている。P2P 型の CDN システムを利用した場合には、IP ネットワークにおける経路制御の階層化構造と P2P 型 CDN システムにより構成されるオーバーレイネットワークにおける階層化構造とが異なるのが一般的である。結果として、同一インターネットサービスプロバイダ (ISP) 内の閉じた通信でコンテンツ配信が可能な場合でも、別の ISP を経由したデータ配信を行う場合が少なくなく、

* 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology,
The University of Tokyo

ネットワーク資源を必要以上に消費してしまう状況が発生しているのが現状である。Huangらは、ISP間の接続情報を考慮せずにP2Pネットワークを構築した場合、ISP間を通過するトラフィックが増大してしまうことを示している [5]。

通信品質を向上させるために、Round Trip Time (RTT) や通過ルータ数 (ホップ数) をメトリックとして、P2Pシステムなどのオーバーレイネットワーク上での経路制御を行う手法 [3,9] やISP内トポロジー情報を考慮した配信経路制御手法 [8] が数多く研究されてきたが、既存の経路制御手法においてISP間の相互接続情報を経路制御メトリックの要素として利用するものはなかった。

本論文では、ISP間の相互接続情報をマルチドメインP2P型CDNにおける配信経路制御の経路評価マトリックスとして導入することで、より適切なデータ配信経路制御を実現することができる手法を提案する。すなわち、オーバーレイネットワークのトポロジー情報に、ISP内のトポロジー情報及びISP間の相互接続情報を付加した「リンクコスト」を新たに定義した配信経路制御手法を提案し、その性能をシミュレーションにより評価する。

2. IP ネットワークにおける経路制御と階層化構造

IPネットワークにおけるパケットの階層的配送経路制御は、単一の運用主体 (AS:Autonomous System) により運用されるネットワーク内配送制御及び、AS間ネットワーク配送制御により構成されている [4]。

AS間及びAS内で交換される経路情報は、一般にはそれぞれ異なるルーティングプロトコルを用いて交換・共有・構築される。相互接続されるネットワークは、それぞれの運用ポリシーを持ち、それに従ってネットワークの相互接続が行われる [7]。AS間の接続形態としては、以下の3つに分類できる。

1. シブリング (Sibling)

同一組織に属するネットワークプロバイダ (ISP等) によって運用されるネットワーク同士の接続形態。

2. ピアリング (Peering)

同規模のネットワークを運用するネットワークプロバイダ (ISP等) 同士の対等な関係での接続形態。

3. トランジット (Transit)

大規模なネットワークを運用するネットワークプロバイダ (ISP等) とそれより小規模なネットワークを運用するネットワークプロバイダ (ISP等) の接続形態。

シブリングは単一組織内部での接続であり、また、ピアリングはトラフィックの交換量等から対等な立場となるため、これらの間ではトラフィックの交換を自由に行うことができる (無償での相互接続 [7])。しかし、トランジットの場合は、小規模なISPは大規模なISPからグローバルなインターネットへの接続性を購入することになり、プロバイダと顧客の関係が存在する。つまり、シブリング・ピアリングと比較し、トランジットはより貴重な (コストのかかる) 回線と言える。

3. オーバーレイネットワーク上での経路制御

一般にP2Pシステムは、個々のシステムに依存した構築アルゴリズムに従って、IPネットワークとは異なる独自のオーバーレイネットワークを構築する。既存の構築アルゴリズムにおいては、ISP間の接続関係が考慮されることはなく、コンテンツ配信の際にISP間のネットワーク資源を効率利用できていないのが現状である。つまり、P2Pシステムを利用した際、同一ISP内の閉じた通信でコンテンツ配信が可能な場合でも別のISPを経由したデータ配信を行う場合が少なくなく、結果としてISP間のネットワーク資源を浪費してしまう場合が発生している。

3.1 物理ネットワークトポロジーを考慮した経路制御

物理ネットワークトポロジーを考慮したオーバーレイネットワーク構築、配送経路制御手法として、物理的なネットワーク距離推定の際指標となるRound Trip Time (RTT) を利用する手法 [6] (RTT極小化選択手法) や、ネットワークトポロジー推定の際指標となるホップ数 (通過ルータ数) を利用する手法 [3,9] (ホップ数極小化選択手法) などが多く研究されている。しかしながら、これらの既存手法においてはIPネットワークを均質なものと仮定しており、現実のインターネットで存在するISP間接続における関係を数値化した情報 (メトリック) として定義、利用した手法は存在しない。

3.2 リンクコストによる経路制御

本論文では、物理ネットワークを考慮したオーバーレイネットワークを構築する際に必要となる、個々の ISP ネットワーク内部トポロジー情報及び ISP 間の相互接続情報を、1次元に集約したパラメータ「リンクコスト」として定義する。

リンクコストは、ネットワーク層のノードにあたる各ルータ間のリンクに設定するパラメータであり、ISP 間(特にトランジット)を通過するような貴重なリンクには、より高いリンクコストを設定するものとする。

ある隣接するルータ r_i, r_j 間のリンクコストを lc_{ij} とすると、オーバーレイネットワーク上のノード $m \rightarrow n$ のリンクコストの和 $LC(m, n)$ は、

$$LC(m, n) = \sum_{i,j:r_i,r_j \in m \rightarrow n \text{ の経路}} lc_{ij}$$

と表すことができ、このリンクコストの和 $LC(m, n)$ を各ノード間の経路制御のメトリックとして用いる経路制御手法を提案手法とする。

ここで、最適経路を取得・選択するためのアーキテクチャとしてプローブ方式とリゾルバ方式の2種類を提案する。

3.2.1 プローブ方式

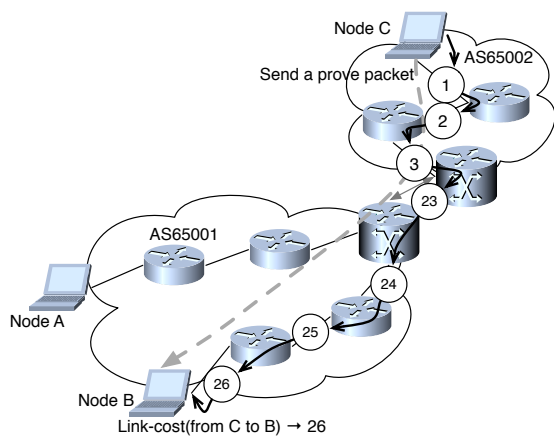


図1 プローブ方式

「プローブ方式」は、図1に示したように、一方のノード(図1における Node C)からプローブパケットを送出し、中継するルータがリンクコストを加算しながら他方のノード(図1における Node B)に配送することで、他方のノードにリンクコストの和を通知する方式である。プローブパケットを受け取ったノードは、各ノードから受け取ったリンクコストの和を比較し、経路を決定

する。

しかし、この方式はノード間のすべてのルータにプローブパケットに対して適切なリンクコストを加算する機能を実装する必要があり、実現が困難であると考えられる。

3.2.2 リゾルバ方式

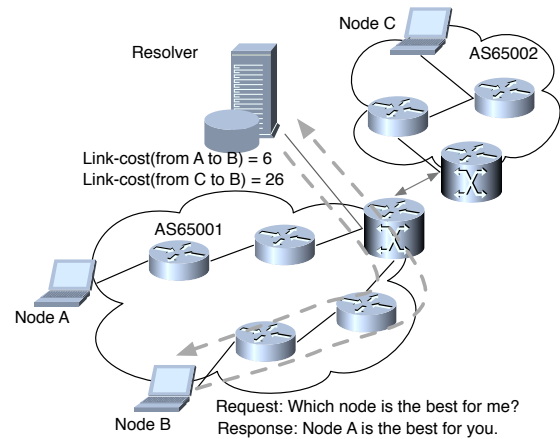


図2 リゾルバ方式

「リゾルバ方式」は、図2に示したように、ノードの問い合わせに対して、リゾルバ上のデータベースにあらかじめ登録された各リンクのリンクコストから、リゾルバが最適経路計算を行い、最適経路をノードに返答する方式である。

この方式では、リゾルバが最適経路計算を行うため、各ノードが最適経路を計算する必要がなく、また、エンドユーザには最適経路のみを提供するため、プローブ方式と比較して、エンドユーザに公開する情報が減り、ISP が内部のトポロジー情報及び ISP 間の相互接続情報を提供しやすいという利点がある。一方、経路情報をリゾルバで集中管理することになるため、リゾルバの規模対応性を検討する必要がある。

4. 評価

シミュレーションにより従来手法と提案手法の比較をすることで、提案手法の有効性を評価した。従来手法として、ランダム選択手法、RTT 極小化選択手法、ホップ数極小化選択手法を比較対象とした。

4.1 評価モデル

4.1.1 ネットワークトポロジー

本論文では、ISP A~H の8つの ISP から構成されるネットワークを評価に用いた。各 ISP の相互接続関係は図3の通りである。また、各 ISP はそれぞれ次のネッ

トワークを構成するものと定義した。

- ISP A
 - サービス地域: 日本, USA
 - コアルータのトポロジー: 図 4
- ISP B, C, E, F
 - サービス地域: 日本
 - コアルータのトポロジー: 図 4 の日本の部分と同一のトポロジー
- ISP D, G
 - サービス地域: USA
 - コアルータのトポロジー: 図 4 の USA の部分と同一のトポロジー
- ISP H
 - サービス地域: 日本
 - コアルータのトポロジー: 図 4 の日本の相互接続ポイントのみにコアルータを有する

なお、相互接続ポイントにおける ISP 間の通信の遅延は、 $1[ms]$ とした。

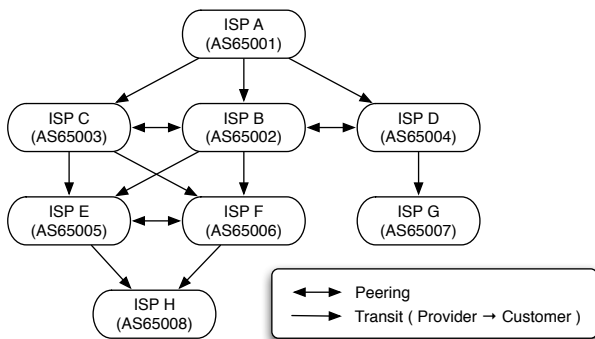


図 3 ISP 間の相互接続関係

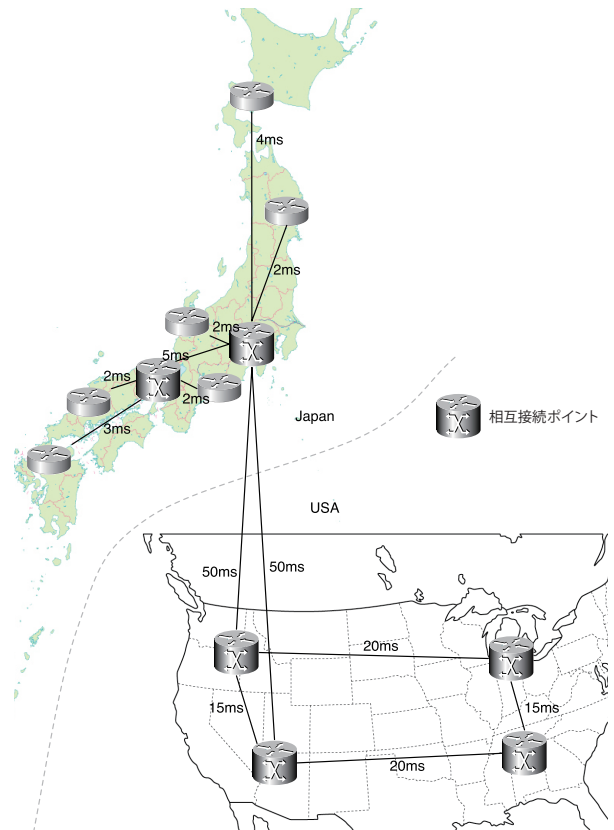


図 4 ISP A のコアルータのトポロジー (数値は各リンクにおける片道遅延を示す)

また、各 ISP に属するオーバーレイネットワークのノードのトポロジーは次に従うものとした。

- ISP A (日本), B, C, E, F
 - コアルータからのトポロジー
 - ホップ数 = $random(2, 3)^{*1}$
 - 遅延 = ホップ数 $[ms]$
- ISP A (USA), D, G
 - コアルータからのトポロジー
 - ホップ数 = $random(2, 7)^{*1}$
 - 遅延 = $\frac{(ホップ数-2) \times 10}{5} + 2[ms]$
- ISP H
 - コアルータからのトポロジー
 - ホップ数 = $random(2, 4)^{*1}$
 - 遅延 = $\frac{(ホップ数-2) \times 5}{2} + 2[ms]$

なお、各ノードは各コアルータに一樣に分布 (接続) するものとした。

*1 $random(min, max)$ は min 以上 max 以下の整数を得る一様分布ランダム関数である

4.1.2 ノードの配置

P2P 型 CDN システムにおけるオーバーレイネットワーク上の経路選択は、コンテンツを提供するノード (Seeder) の数によってその特徴が異なるため、本論文では、Seeder の数を 1~100 まで変化させ、シミュレーションによる評価を行った。Seeder の配置は、各 ISP 上に、一様分布に従うランダム関数にて配置した。また、Seeder からコンテンツを取得するノードも同様に、各 ISP 上に、一様分布に従うランダム関数にて 1 ノード配置し、各 Seeder 数につき 100,000 パターンのノード配置を試行した。なお、各手法の比較においては、同一のトポロジー及びノード配置を用いた。

4.1.3 IP ネットワークにおける経路交換ポリシー及び経路選択

IP ネットワークにおける ISP 間の経路交換ポリシーには次のポリシーを適用した。

- ピアリング
互いの顧客の経路を交換する。
- トランジット
プロバイダから顧客へはフルルート (全経路) の広告を行う。顧客からプロバイダへは、顧客及びその顧客から受け取った経路を広告する。

また、ISP 間の経路制御における経路 (ベストパス) 選択は次の手順に従うものとした。

1. Next hop AS がピアリング又はトランジット (プロバイダ→顧客) である経路
2. 通過 AS 数が最小となる経路
3. Next hop AS の AS 番号が小さいもの

なお、ISP 内の経路はホップ数を最小とする経路制御に従うものとした。

4.1.4 リンクコスト

表1 リンクコスト

接続関係	リンクコスト
ISP 内のリンク	1
ピアリング	20
トランジット (プロバイダ → 顧客)	30
トランジット (顧客 → プロバイダ)	50

各ルータ間のリンクコストとして表1の値を用いた。

4.2 シミュレーションによる評価結果

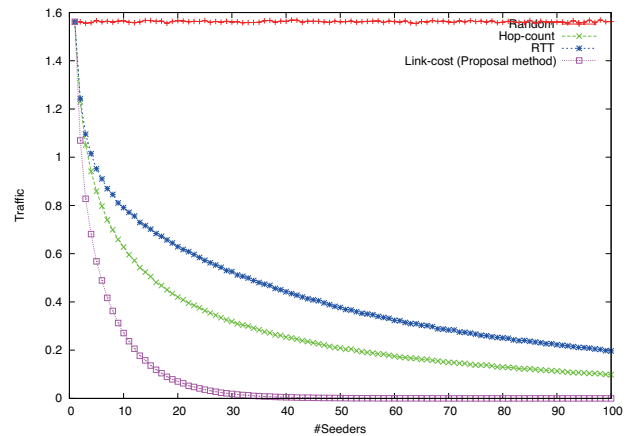


図5 ISP 間リンク通過トラフィック

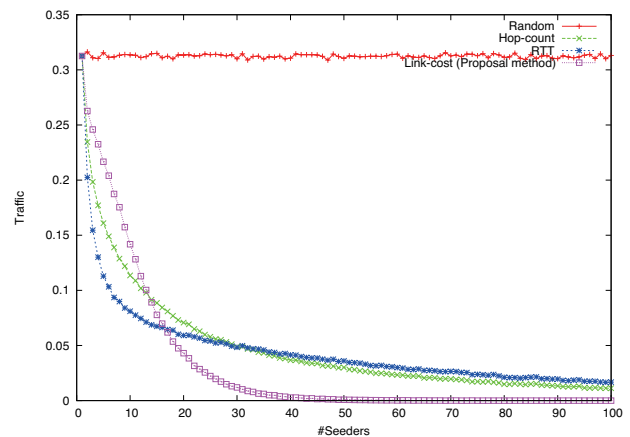


図6 ピアリング通過トラフィック

表2 各経路制御手法におけるトラフィックの評価結果 (#Seeders = 10)

評価マトリクス 経路制御手法	ピアリング		トランジット (プロバイダ→顧客)		トランジット (顧客→プロバイダ)	
	ISP間リンク	ピアリング	トランジット	トランジット	トランジット	トランジット
ランダム選択	1.566	0.314	0.628	0.628	0.625	0.625
ホップ数極小化	0.628	0.114	0.262	0.262	0.252	0.252
RTT極小化	0.791	0.081	0.343	0.343	0.367	0.367
提案手法	0.271	0.142	0.094	0.094	0.035	0.035

図5, 6, 7, 8に各経路制御手法における Seeder の数と各種 ISP 間リンクを通過するトラフィックの関係を示した。なお、トラフィックは、ノードが Seeder からコンテンツを取得する1処理あたりの各種 ISP 間リンク通過数の平均で表している。また、図9に各経路制御手法における Seeder の数と平均 RTT の関係を示し、図10に Seeder の数が10のときの遅延特性を RTT の累積度数分布として示した。

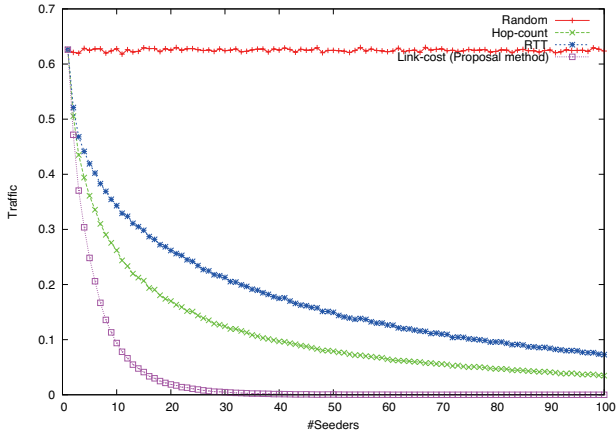


図7 トランジット (プロバイダ→顧客) 通過トラフィック

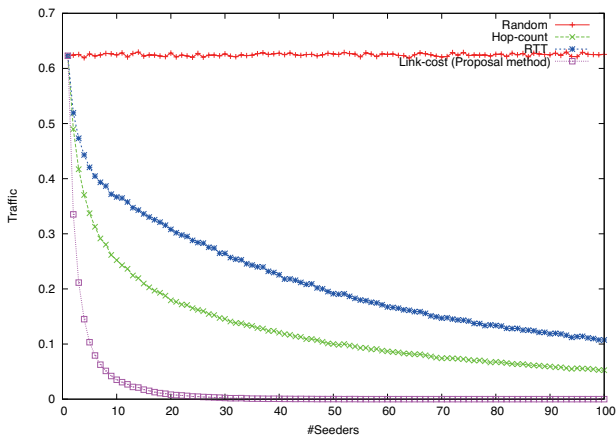


図8 トランジット (顧客→プロバイダ) 通過トラフィック

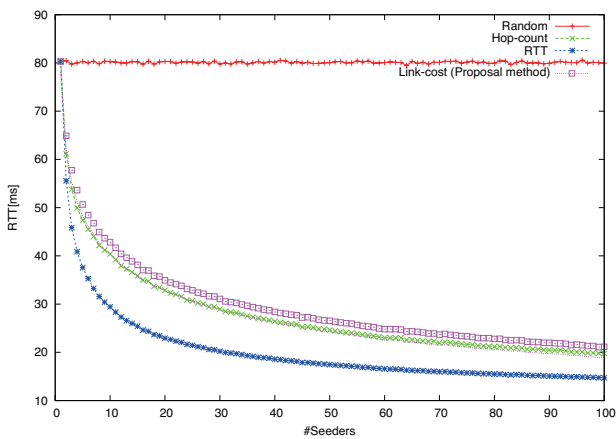


図9 平均 RTT

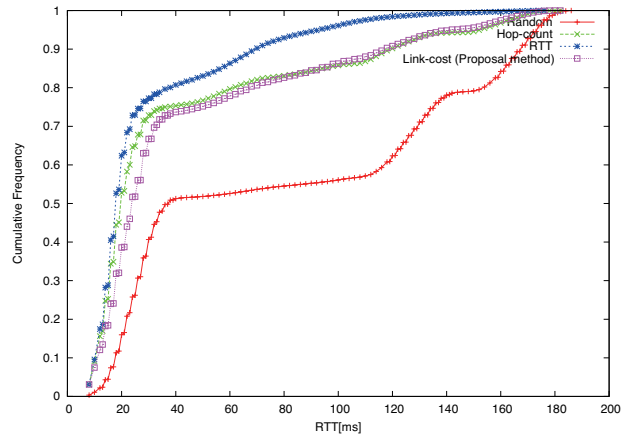


図10 RTT の累積度数分布 (#Seeders = 10)

4.2.1 ISP 間リンク通過トラフィックの削減

図5より、ランダム選択手法では、ISP間リンクを通過するトラフィックを Seeder の数に依らず一定量発生させており、ネットワーク資源を浪費していることがわかる。また、提案手法により、今回評価した全ての Seeder 数について、ランダム選択手法・従来手法と比べ、ISP間リンクを通過するトラフィックを削減できることが示された。

図6, 7, 8より、提案手法では、従来手法と比べ、より高いリンクコストを設定したトランジット (顧客→プロバイダ) よりもピアリングを優先して選択することで、トランジット等の貴重なリンクを通過するトラフィックをより削減できることを示すことができた。

Seeder の数が10のときの評価結果 (表2) を比較すると、提案手法によりISP間リンクを通過するトラフィックをランダム選択手法に比べ約83%削減できており、従来手法のホップ数極小化選択手法と比較しても約67%削減できた。

しかし、図9は、提案手法により、遅延の大きい経路でも、リンクコストが小さければ優先的に選択してしまうというを示している。図10より、RTTの累積度数分布の95%値は、RTT極小化選択手法で92[ms]、提案手法では142[ms]となっており、通信品質が低下してしまっていることがわかる。今後、リンクコストに通信品質を含む経路制御を検討する必要がある。

5. まとめ

本論文では、マルチドメイン P2P 型 CDN における配信経路制御において、ネットワークトポロジー情報及びISP間の相互接続情報を経路制御の評価マトリックスと

して含むようなリンクコストを導入することで、ISP 間通過トラフィックを削減できることを示した。

5.1 今後の課題

本論文では、ネットワークトポロジー情報及び ISP 間の相互接続情報を用いた経路制御により、全体として ISP 間 (特にトランジット) を通過するトラフィックを削減できることを示したが、各 ISP ごとの評価は行わなかった。例えば、上位 ISP としては、経済的な理由から、より多くのトラフィックを顧客のトランジットへ流したいという場合がある。今後は、各 ISP における評価も行う必要がある。また、このような視点において、適切なリンクコストの値を考察する必要がある。

また、本論文では、8つの ISP から構成されるネットワークにおいてシミュレーションを行ったが、今後は、現実のネットワークにおいての有効性を示すために、CAIDA [2] により提供されている “Inferred AS Relationships Dataset” を用いたより現実に近いネットワークでの大規模なシミュレーションや、提案手法の実装及び実証実験を行う必要がある。

さらに、今後は遅延特性、利用可能帯域やリンクの安定性などの通信品質をコストとして含むような経路制御の検討と評価を行う必要がある。

参考文献

- [1] BitTorrent, <http://www.bittorrent.com/>.
- [2] Cooperative Association for Internet Data Analysis (CAIDA), <http://www.caida.org/>.
- [3] Frank Dabek, Russ Cox, Frans Kaashoek and Robert Morris, “Vivaldi: A Decentralized Network Coordinate System”, *ACM SIGCOMM*, 2004.
- [4] J. Hawkinson, BBN Planet and T. Bates, “Guidelines for creation, selection, and registration of an Autonomous System (AS)”, RFC 1930, March 1996.
- [5] Cheng Huang, Jin Li and Keith W. Ross, “Can Internet Video-on-Demand be Profitable?”, *ACM SIGCOMM*, 2007.
- [6] T. S. Eugene Ng and Hui Zhang, “Predicting Internet Network Distance with Coordinates-Based Approaches”, *IEEE INFOCOM*, 2002.
- [7] Iljstch van Beijnum, “BGP TCP/IP ルーティングとオペレーションの実際”, オーム社, 2004.
- [8] Haiyong Xie, Y. Richard Yang, Arvind Krishnamurthy, Yanbin Liu and Avi Silberschatz, “P4P: Provider Portal for Applications”, *ACM SIGCOMM*, 2008.
- [9] 後藤嘉宏, 阿多信吾, 村田正幸, “P2P サービスにおける物理ネットワークを考慮した論理トポロジー構築手法”, 映像情報メディア学会技術報告, 2002.