品質確保のためのネットワーク運用管理技術（継続）

代表研究者　田　中　良　明　早稲田大学大学院国際情報通信研究科教授
共同研究者　ザニケエフ・マラット　早稲田大学国際教養学部助教

1 まえがき

インターネットにおいてできるだけ品質を確保してコンテンツ配信を行うことを考える．根となるノードからユニキャストで配信するのは簡単であるが，ネットワークに余計な負荷が掛かり，品質が劣化する．マルチキャスト木を構築して配信を行えば負荷は少なく，品質への影響も少ない．しかし，インターネットのトポロジーは，ネットワーク管理者が自ネットワークについて知っているだけで，他ネットワークについては分からない．ましてや，コンテンツ配信を行うユーザは，トポロジーをまったく知らない．そこで，トポロジー不明網においてマルチキャスト木を構築する必要が生じる．本稿では，ネットワークにおける測定遅延時間をノード間の距離と考え，マルチキャスト木の構築を行う方法を検討する．

ノードは地理的に一様に分布しているのではなく，かたまって分布していることが多い．そこで，本稿では，遅延時間を手掛かりとするノードのクラスタリングを行って，ノード群を地理的にクラスタリングすることでマルチキャスト木を構築し，更にそのマルチキャスト木の特性を評価する．これによって，ユーザは，遅延時間に基づく他のノードの分布の情報を知ることができ，マルチキャスト木構築に利用できる．ユーザが自身のもつ遅延時間情報のみによってマルチキャスト木を構築できるのが本検討の特徴である．

ただし，ここでいうマルチキャストとは特定のプロトコルを指すのではなく，ネットワーク内で木状に分岐しながらコンテンツを配信する広義の意味とする．分岐ノードによる蓄積配信も含むものとする．

インターネット内のノードの位置を推定する既存の検討としては，Meridian [1]がある．Meridian では，ネットワーク内での位置に基づいてノードの部分集合を形成する際，ノードに関する情報を繰り返し通信で交換して位置を特定していく．それに対し，本手法では，より統計的な手法として，遅延時間の分布に対して階層的な解析を行う．

2 インターネットにおける遅延時間の特徴

検討対象とするネットワークの遅延時間分布データとして，DS² (Delay Space Synthesizer) [2]によって生成したデータを利用する．DS² は，実際のインターネットにおける遅延時間分布に近い遅延時間を生成するソフトウェアである．ここでは，ネットワーク規模を全世界とし，その中からランダムにノードを 3997 個選んで，そのノード間の遅延を生成させる．最大遅延は 1980ms である．DS² は，遅延時間測定を模擬しているので，生成したデータには測定に失敗したケースも含まれる．遅延時間の測定の結果が不明または失敗したペアは全体の 14.8%である．

まず，ネットワークの遅延時間分布の特徴をとらえる必要がある．そこで，ネットワークの中で，ある一つのノードから残りの 3996 個のノードへの測定遅延時間の密度分布を求める．図 1 は，1 区間を 30ms にして密度分布を求めた結果である．この区間が小さ過ぎると，ノードのかたまりがグラフ上で判別できない．また，逆に大き過ぎると全体が一つのノードのかたまりになる．したがって，適切な区間の大きさを用いる必要がある．図 1 を見ると分かるように，インターネットにおける遅延は，3〜4 段階程度に分かれている．ユーザが他のノードの分布が分からない状況でネットワーク内にマルチキャスト木を構築する手法として，遅延時間のこのような特徴を利用して密度分布をかたまりごとに分けることでクラスタリングを行い，マルチキャスト木を構築していく方法が考えられる．

具体的な方法としては，まず，マルチキャストグループの根となるノードからマルチキャストグループのノード全部への遅延時間を測定する．最初は，1 区間を小さくした密度分布を求める．次に 1 区間の大きさを徐々に大きくしていって密度分布図を平滑化し，ノードのかたまりがグラフ上の山となって現れるようにする．このグラフの極小点が遅延時間分布のかたまりの境界である．図 1 では，4，9，16（×30ms）に境界がある．

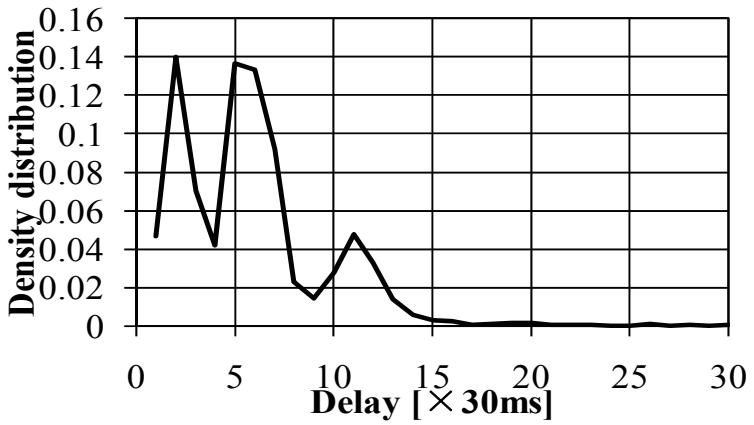


図 1　一つのノードからの測定遅延時間の密度分布（1 区間 30ms）

3 クラスタリング手法

前章で述べた方法で作られるクラスタは，地理的に一つのノードを中心とした同心円を境界とするドーナッツ状である．したがって，同じクラスタに入っているからといって，必ずしも近いわけではない．

そこで，一つのドーナッツ状のクラスタに再度同じ手法でクラスタリングを行う．これにより 2 段階目のクラスタリングができる．これを更にもう一度行うと 3 段階のクラスタリングになる．このクラスタリング手法によって，木構造の階層的なトポロジーが構築され，マルチキャスト木となる．

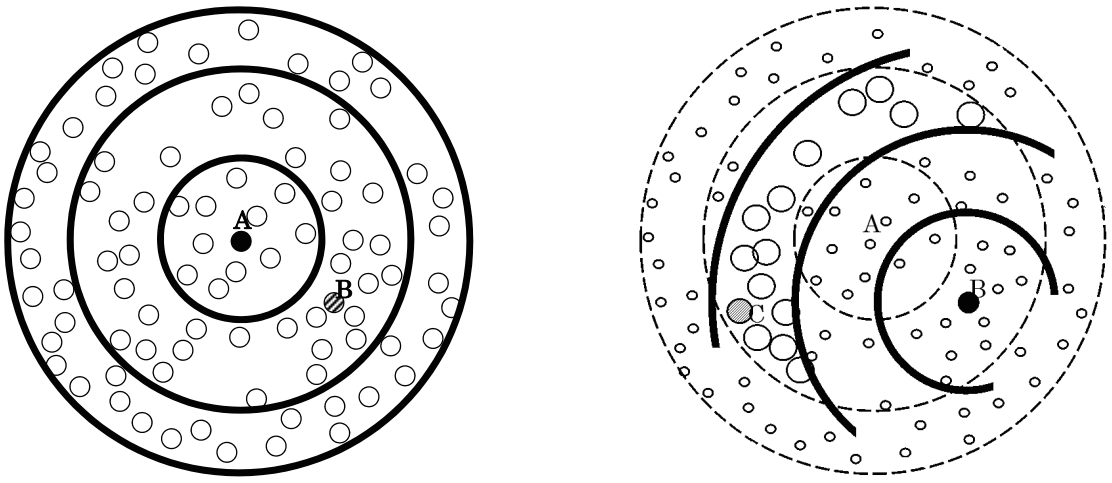
具体的には，以下の手順になる．

(1) 図 2(a)の根となるノード A からマルチキャストグループの他の全ノードへの遅延時間を測定する．遅延時間の密度分布を求め，平滑化を行って 3〜4 個の山が現れるようにする．密度分布の山の境界で切ってクラスタリングを行う．地理的には，ネットワーク内のノードが図 2(a)のノード A を中心とするドーナッツ状のクラスタに分けられる．

(2) ドーナッツ状の各クラスタの中で，それぞれ代表ノードを決定し，(1)と同じことを行う．すると，例えば，図 2(b)のノード B を中心とする同心円によって分けられたクラスタを作ることができる．

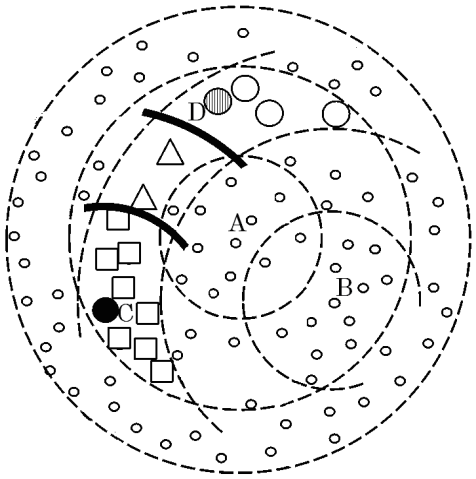
(3) (2)で作った各クラスタの中で，それぞれ代表ノードを決定し，(1)と同じことを行う．すると，例えば，図 2(c)のノード C を中心とする同心円によって分けられたクラスタを作ることができる．

(4) 最後に，分けられたクラスタ内で代表ノードを決定する．これは，図 2(c)におけるノード D である．



(a) Step 1: A を中心として一次境界線を引く　　(b) Step 2: B を中心として二次境界線を引く

図 2　クラスタリングのイメージ



(c) Step 3: C を中心として三次境界線を引く

図 2　クラスタリングのイメージ（続き）

4 クラスタリング結果

DS² によって生成した遅延時間データに提案アルゴリズムのクラスタリングを適用した結果を表 1 に示す．第 1 段階クラスタリング及び第 2 段階クラスタリングにおけるクラスタ数はそれぞれ 4，第 3 段階クラスタリングにおけるクラスタ数は 2 である．

表 1　3 段階クラスタリング適用結果

クラスタ名	ノード数	平均遅延	クラスタ名	ノード数	平均遅延
(0, 0, 0)	**678**	**44**	(2, 0, 0)	131	30
(0, 0, 1)	26	59	(2, 0, 1)	50	214
(0, 1, 0)	95	23	(2, 1, 0)	53	43
(0, 1, 1)	24	159	(2, 1, 1)	154	197
(0, 2, 0)	2	131	(2, 2, 0)	3	494
(0, 2, 1)			(2, 2, 1)	2	1050
(0, 3, 0)	2	446	(2, 3, 0)	2	1080
(0, 3, 1)	2	126	(2, 3, 1)	3	1113
(1, 0, 0)	**978**	**58**	(3, 0, 0)	3	494
(1, 0, 1)	7	83	(3, 0, 1)	2	
(1, 1, 0)	21	51	(3, 1, 0)		
(1, 1, 1)	118	132	(3, 1, 1)		
(1, 2, 0)	2	80	(3, 2, 0)		
(1, 2, 1)	23	222	(3, 2, 1)		
(1, 3, 0)	3	155	(3, 3, 0)		
(1, 3, 1)	6	299	(3, 3, 1)		

表 1 において，「クラスタ名」の列に書かれている座標は，左側から順番にクラスタリングの段階を示しており，数字はクラスタ番号を示している．この数字の集合を座標形式で表現し，各クラスタの固有名としている．また，ノード数は各クラスタに所属するノードの数を表す．この結果から，今回対象としたマルチキャストグループでは，大きなものとして 1000 前後のノード群と，800 前後のノード群が存在し，その他は 0～100 前後の規模のノード群が複数存在していることが分かる．

本手法のクラスタリングによって生成されたクラスタ内のノード同士が，実際に遅延時間の小さいノード

同士の集合となっているか調べ，クラスタリングの良し悪しを視覚的に確認してみる．そこで，各クラスタ内の全ノードペアの遅延時間を測定し，その平均を算出する．

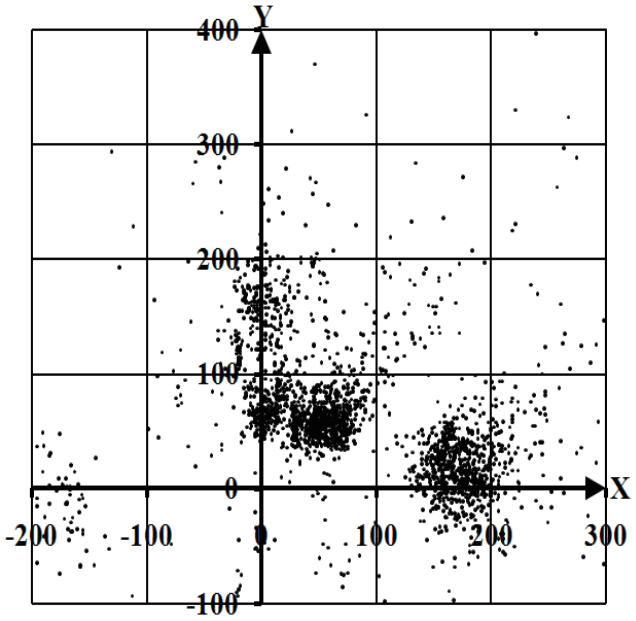
表 1 の平均遅延時間は，クラスタ内における全ノードペアの遅延時間の平均値を示している．また，マルチキャストグループ内における全ノードペアの遅延時間の平均値は 144ms である．この値と表 1 に示す平均遅延時間を比べてみると，本手法のクラスタリングにより，近いノードをクラスタとして抽出するのに成功している場合もあるが，そうでない場合もあることが分かる．

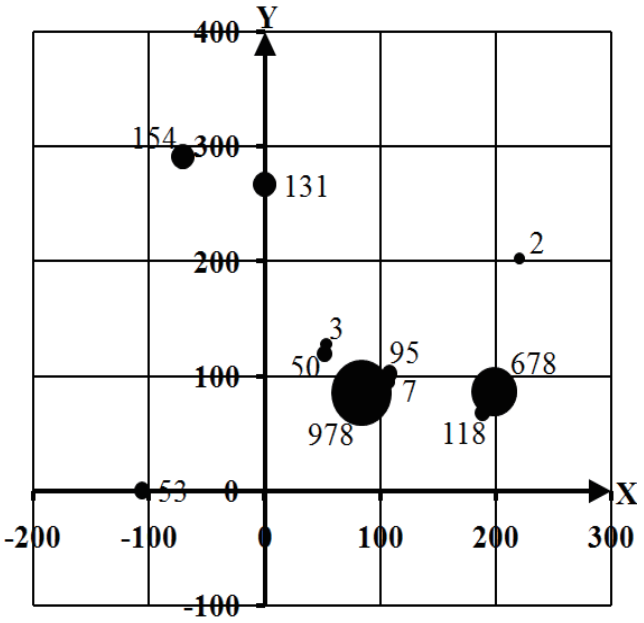
例えば，表 1 の太字のクラスタは，所属ノード数も多く，クラスタ内全ノードペアの遅延時間の平均値も小さいため，クラスタ抽出に成功しているといえる．しかし，各段階のクラスタリングにおいて代表ノードよりも外側に位置するクラスタほど，クラスタ内全ノードペアの遅延時間の平均値が大きな値になっている．原因としては，代表ノードとの通信のルーチングが悪く，実際には近くにあっても遠くのクラスタに配置されていることが考えられる．また，端の方のクラスタほど遅延時間の平均値が大きい．これは，端の方はノード数が少なく極小点が出現しにくいためであると考えられる．また，各段階におけるクラスタリングの一番外側のクラスタは，クラスタ内全ノードペアの遅延時間の平均値も大きくならざるを得ない．全体の 14.8%が遅延時間不明となっているため，3 段階の通信遅延時間測定を通して，全体の 38%前後のノードが途中で失われているのも問題の一つである．

また，完成した各クラスタ内の代表ノードを利用して，クラスタの 2 次元座標上への配置を試みる．これは XY 軸上に基準となる 3 点を設定し，三角測量によって配置するものとする．比較対象として，同じ基準点によってできた三角形の頂点との遅延時間を距離と見ることで，全ノードを 2 次元座標上にプロットした図 3(a)を用意する．XY 軸の単位は ms である．

上記の方法で表 1 のクラスタをプロットしたものが，図 3(b)である．図 3(b)において，プロットされた点の近くの数字は，そのクラスタ内の所属ノード数を示しており，そのクラスタの規模に応じて，黒点を大きく表示している．

図 3(a)と(b)を比較してみると，提案手法による図 3(b)は，理想図である図 3(a)のノード群を圧縮して表示したものとなっており，近いノードを抽出できているクラスタに関しては，望んだ結果が得られているといえる．しかし，所属ノードの平均遅延時間が大きいクラスタに関しては，正確に図示されていないものがある．また，ノード群の広がりを考えると，クラスタの重心により近いサンプルの抽出ができるような改善も必要である．





(a) 個々のすべてのノードのプロット図　　(b) クラスタのプロット図

図 3　三角測量を利用した 2 次元座標へのプロット図比較

5 マルチキャスト木評価方法

本クラスタリングを行うことで構築されたマルチキャスト木について説明する．図 2 において A に当たるノードがクラスタリングを始める中心ノードであり，これが木構造の根に当たる．その次の層には B を置く．更にその次には C を置き，代表ノード D をそのもう一つ下の層に置き，最後の末端に当たる部分に，各クラスタの所属ノードすべてをそれぞれ置くことにする．この様子を図 4 に示す．

このマルチキャスト木を用いたコンテンツ配信と，ユニキャストでのコンテンツ配信を特性比較する．具体的には，コンテンツ配信の待ち時間の分布をユニキャスト配信とマルチキャスト配信で比較し，階層的に構築したマルチキャスト木の特性評価を行う．

ここでいう待ち時間とは，コンテンツ配信が根のノードから開始されてから，個々のノードが受信し終わるまでの時間のことであり，待ち行列遅延時間と伝送遅延時間の和である．また，最下層のクラスタには多くのノードが所属しているため，ノードのリソースに負荷がかかりすぎないように，一つのノードから同時配信可能なノードの数を，ソケット数で制限する．

例えば，ノード P からノード Q1 とノード Q2 にコンテンツを送信することを考える．根のノードから配信が開始されてノード P が受信し終えるまでの時間を W_P とし，ノード P からノード Q1 への通信時間を T_{PQ1} とすると，ノード Q1 の待ち時間は $W_{Q1}=W_P+T_{PQ1}$ である．ノード P のソケットが一つとすると，ノード P からノード Q2 への送信はノード Q1 への送信が終了した後に開始する．ノード P から Q2 への通信時間を T_{PQ2} とすると，ノード Q2 の待ち時間は $W_{Q2}=W_P+T_{PQ1}+T_{PQ2}$ となる．ノード P のソケットが二つとすると，ノード Q1 とノード Q2 に同時に送信できるので $W_{Q2}=W_P+T_{PQ2}$ となる．

ソケットの数が大きいほど同時に配信を行うノードが多いため，一つの通信に割り当てられる帯域が小さくなりリソースもより消費されるが，一方で個々のノードの中の待ち時間の差は小さくなる．

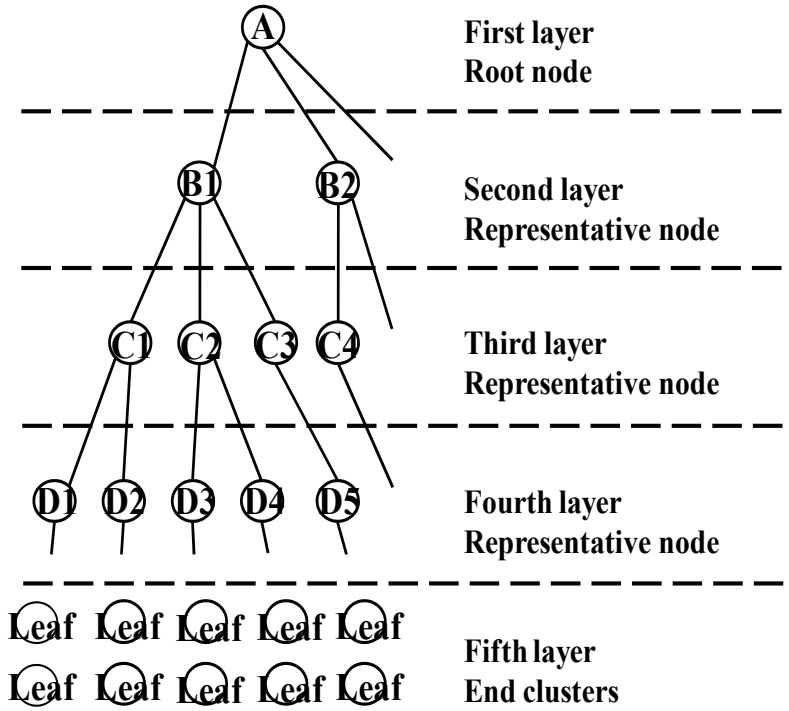


図 4　構築されたマルチキャスト木

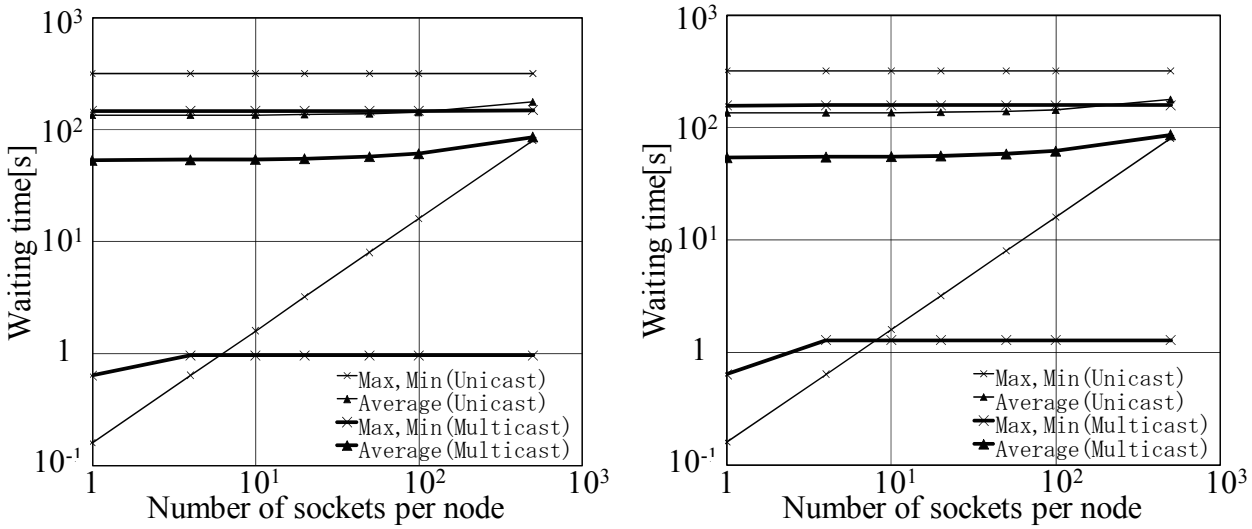
6 マルチキャスト木評価結果

第 2 章ではグループのノード数を 3997 個にしたが，クラスタリングの過程で全体の 38%前後のノードが失われているため，評価対象のノードの数を 2001 個とし，一つのノードから 2000 個のノードへのコンテンツ配信を考えて木を評価する．リンク速度はすべて 100Mbps とし，配信するコンテンツの大きさは 2MByte とする．第 1 段階クラスタリング及び第 2 段階クラスタリングにおけるクラスタ数はそれぞれ 4，第 3 段階クラスタリングにおけるクラスタ数は 2 である．更に，生成されたクラスタの中から代表ノードをそれぞれ選ぶことで，5 階層の木構造を生成する．また，DS^2 で生成したデータが同じでも，クラスタリングにおける根のノードの選び方によって結果は変化するので，異なる根のノードをもつ 3 パターンについて結果を求める．

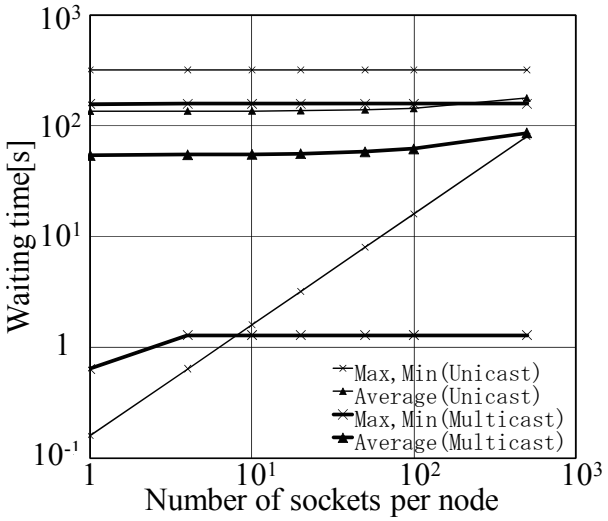
図 5 に，階層的にマルチキャスト木を構築してマルチキャストで配信を行う場合と，ユニキャストで配信を行う場合の待ち時間を示す．待ち時間はノードごとに異なるので，最大値，最小値，平均値の三つの値を示してある．

図 5 を見ると，ユニキャストよりもマルチキャストの方が，最大値と最小値に幅があることが分かる．これは，クラスタリングにおいて所属ノード数が少ないクラスタが多数存在する．そういうノードに対する通信では，ソケット数を増やしても空のままであり，使用している帯域が大きいため，待ち時間が小さくなる．そのため，最小値が大きくならない．

しかし，待ち時間の最大値や平均値は，マルチキャストの方がユニキャストに比べて大きく短縮されていることが分かる．また，ソケット数が増えると帯域が細くなり，ノードがコンテンツを受信するまでの待ち時間は大きくなるので，ソケット数に比例して待ち時間の平均値そのものは増加している．平均値のソケット数に応じた短縮率を表 2 に示す．表 2 より，本手法で構築したマルチキャスト木は待ち時間が大きく改善されており，マルチキャストに適しているといえる．



(a) パターン 1　　(b) パターン 2



(c) パターン 3

図 5　ユニキャスト通信時とマルチキャスト通信時のソケット数に応じた待ち時間

表 2　待ち時間平均値のソケット数に応じた短縮率

	ソケット数						
	1	4	10	20	50	100	500
パターン 1	0.393	0.397	0.399	0.402	0.412	0.426	0.484
パターン 2	0.399	0.404	0.406	0.409	0.418	0.431	0.483
パターン 3	0.402	0.410	0.412	0.416	0.425	0.439	0.489
平均値	0.398	0.404	0.406	0.409	0.418	0.432	0.485

7 むすび

本稿では，インターネットにおいてユーザがマルチキャスト木を構築する方法を提案した．ユーザはネットワークのトポロジーを知らないので，測定によってトポロジーに関する情報を得なければならない．本稿では，遅延時間によってマルチキャストグループのノード群をクラスタリングし，それを階層的に行ってマ

ルチキャスト木を構築した.

このクラスタリングによって形成された木構造をマルチキャスト木として見ることにし，マルチキャスト通信の待ち時間を求めて木を評価した.本手法の適用結果として,マルチキャスト通信における待ち時間は，ソケット数にかかわらず大きく改善されることが分かった．したがって，クラスタリングによって構築されたマルチキャスト木が有効であると結論できる.

本手法の問題点として,クラスタ内の所属ノード数が多いものと少ないものが存在することが挙げられる．所属ノードが多いクラスタでは，親ノードがボトルネックとなり，待ち時間が大きくなる．一方，所属ノードが少ないクラスタでは，帯域を大きく使うことができるため待ち時間が小さくなる．そのため，マルチキャストにおいて，ソケット数を増やしても待ち時間の幅の大きさが改善されていない．また，クラスタ内の全ノードペアの遅延時間の平均値が大きいクラスタが存在する．これは，その部分のクラスタリングが適切でないことを意味している.

今後の課題として，より平均遅延時間が小さく，所属ノード数が偏っていないクラスタリングが実現できるような検討が必要である．クラスタリングの段階数や，各段階のクラスタ数を増やすだけでは，細かいクラスタが無数にできる．そこで解決策として，所属ノードが多いクラスタについて，更に細かいクラスタにするようなアルゴリズムを加えることが考えられる．同時に，細かいクラスタ同士で近いものをまとめるアルゴリズムや，平均遅延時間の大きなクラスタを再クラスタリングするアルゴリズムなども必要になってくる．また，最適な段階数やクラスタ数の設定方法の検討も必要である.

【参考文献】

[1] B. Wong, A. Slivkins, E. G. Sirer, “Meridian: A lightweight network location service without virtual coordinates,” Proc. ACM SIGCOMM 2005, pp.85-96, Philadelphia, USA, Aug. 2005.

[2] DS2 Delay Space Synthesizer. Available at: http://www.cs.rice.edu/~bozhang/ds2/

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
測定遅延時間の多次元クラスタリングによるトポロジー推定	電子情報通信学会通信ソサイエティ大会	2008 年 9 月
トポロジー不明網におけるマルチキャスト木構築法	電子情報通信学会技術研究報告	2009 年 5 月