

QoS 保証マルチキャストプロトコル SRSVP における階層化 PathQoS

関口 隆昭[†] 藤川 賢治^{††}
岡部 寿男^{††} 岩間 一雄[†]

[†]京都大学大学院情報学研究科通信情報システム専攻
^{††}京都大学大学院情報学研究科知能情報学専攻

あらまし 階層化されたネットワークにおいて各マルチキャストフロー毎の情報を集める方法について論ずる。QoS 保証マルチキャストプロトコル SRSVP は、PQ と呼ばれるフロー毎の情報をシグナリングメッセージに含めて収集し経路計算を行うように設計されている。また、QoS 保証ユニキャストプロトコル HQLIP は、スケーラブルな QoS ルーティングのためにネットワークを階層化して扱う。本研究では、SRSVP がより良い QoS 経路を計算できるようにするために、階層化されたネットワークにおける各エリア間の集約されたリンク情報に対応する PQ (階層化 PQ) を計算するアルゴリズムを設計し実装した。また、各ルータの動作を検証することによりアルゴリズムの効率化を試みた。

Hierarchical Path QoS on a QoS-based Multicast Protocol SRSVP

TAKA AKI SEKIGUCHI[†], KENJI FUJIKAWA^{††}, YASUO OKABE^{††}
and KAZUO IWAMA[†]

[†]Department of Communications and Computer Engineering
^{††}Department of Intelligence Science and Technology,
Graduate School of Informatics, Kyoto University

Abstract In this paper, we argue a method to collect information of each existing multicast flow on hierarchical networks. SRSVP, a QoS-based multicast routing protocol, is designed as it collects flow-specific information, called PQ, by putting it into signaling messages, so that the derived QoS path becomes more efficient. HQLIP, an underlying QoS-based unicast routing protocol, handles a network as a hierarchical structure for scalable QoS-based routing. We have designed and implemented an algorithm to compute PQ (hierarchical PQ) corresponding to aggregated link information on hierarchical networks for SRSVP to compute better QoS paths. We have attempted to make the algorithm more efficient by examining behaviors of routers.

1. はじめに

IP マルチキャストとは、マルチキャストグループのメンバとして設定された複数の受信者へパケットを配送する方式である¹⁾。これを行うためのマルチキャストルーティングプロトコルとして PIM-SM²⁾ など様々なものが開発されている。しかし、これらの既存のプロトコルは best-effort 転送に依存しており QoS (Quality of Service) の保証が考慮されていない。次世代インターネットでは全インターネットにわたって行われるビデオ会議・インターネット放送などの実用化が望まれるが、

そのためには大規模ネットワークにおいて QoS 保証を行うマルチキャストルーティング技術が要求される。

QoS にもとづくルーティングでは、OSPF³⁾ 等の従来のルーティングプロトコルで用いられる単一のメトリックに加えて、利用可能な帯域、遅延などが広告される。これらの値は比較的動的に変化するため、これらの広告によってネットワーク資源が大量に消費されることになり、大規模ネットワークではスケーラビリティの問題が発生する。これを防ぐためには、インターネットを階層的な構造として扱い、広告される経路情報の局所的な集約を行う必要がある⁴⁾。

現在、大規模ネットワークにおいても利用可能な QoS 保証マルチキャストのために、Real Internet Consortium (RIC, <http://www.real-internet.org/>) により QoS 保証マルチキャストルーティングプロトコル SRSVP⁵⁾ および QoS 保証ユニキャストルーティングプロトコル HQLIP⁶⁾ が提案されている。SRSVP は受信者をマルチキャスト配送木に参加させるための QoS 経路を計算するために PQC (Path QoS Collection)⁷⁾ と呼ばれる各マルチキャストフロー毎の情報を収集する方式を採用している。また HQLIP はネットワークを階層化して扱い、スケーラブルな QoS ルーティングを実現している。

本研究では、HQLIP が扱う階層化されたネットワークにおいても SRSVP が効果的な QoS 経路計算を行えるようにするために、階層化されたネットワークにおいて PQC を収集するアルゴリズムの設計および実装について論ずる。

2. 前提となる技術

2.1 PQC

PQC は、QoS ルーティングのために各フローの状態を収集する方式の 1 つである。

QoS 保証マルチキャストルーティングにおいては、各フロー毎の情報をどのように集めるかが重要な問題の 1 つである。つまり、自分が参加しようとしているマルチキャスト配送木が現在どのような状態なのかを受信者がどれだけ知っているかで、ルーティングの戦略が異なってくる。

例えば、PNNI⁴⁾ のシグナリングプロトコルでは、各フローの状態を全く把握せずに、QoS を満たす経路を計算する。このような方式をマルチキャストルーティングに適用する場合、現在の配送木を考慮した効率的なルーティングは行えない。また、各フローによって消費されているリソースに関する情報の欠如のために QoS を満たす経路が存在しないように見えて、経路計算が失敗することもある。

逆に、各フローの状態を完全に把握しようとする方式としては、例えば QOSPF⁸⁾ がある。QOSPF では、各ルータがリンク上のフローの状態を RRA と呼ばれるメッセージで広告する。ネットワーク上の各ルータはそれによって各フローの状態を完全に把握することができ、効率的なルーティングを行える可能性がある^{9),10)}。しかし RRA のようなメッセージはフロー数が増加するにつれてメッセージ数が多くなり、大規模なネットワークにおける利用の場合にはスケーラビリティの問題があり実現が難しい。

PQC は、シグナリングプロトコルの Path メッセージに PQ (Path QoS) と呼ばれるリンクにおける各フローの情報を含めて各ルータに集め、それによって QoS ルーティングプロトコルによって広告されている情報を修正し経路計算を行う方式である。PQ には、あるフローの観点から見た空き帯域および遅延が含まれる。例えば、あるリンク上を 3Mbps のフローが流れていて、QoS ルーティングプロトコルがそのリンクは空き帯域が 6Mbps だという広告をしていた場合、PQ はそのリンクには 9Mbps の空きがあることを示すことになる。

以降で PQC の例を示す、図 1 に示すようなネットワークがあるとす。各リンクの近くの数値はそのリンクの空き帯域を示している (ここでは簡単のため双方向の帯域は考えない)。

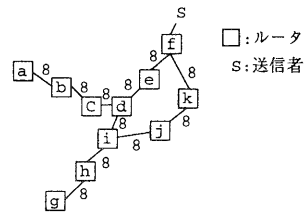


図 1 ネットワーク

最初に、受信者 R1 が送信者 S からのフローに対して 5Mbps の帯域を予約できる経路を確立しようとする。その Resv メッセージは QoS ルーティングプロトコルによって a-b-c-d-e-f のように経路付けされる。次に、別の受信者 R2 が同一フローに対して 4Mbps の帯域を予約できる経路を確立しようとする。その Resv メッセージは、おそらく、g-h-i-j-k-f のように経路付けされる。何故なら他の経路は帯域要求を満たさないからである。ネットワークの状態は図 2(a) のようになる。

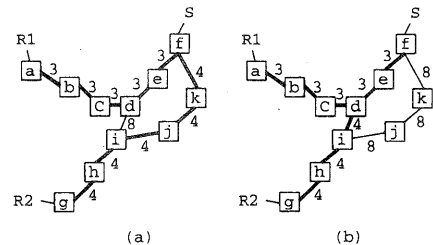


図 2 (a) PQC を使わない予約 / (b) PQC を使った予約

PQC を用いると以下ようになる。Path メッセージが f-e-d-i-h-g のように流れていく際に、Path メッ

セージを受信した各ルータは、リンク上のフローの状態を調べてPQを生成し、Pathメッセージに追加していく。これによって、Pathメッセージが受信者R2に到達したとき、R2はルータd、e、f間のリンク上をR2が受信しようとしているフローと同一のものが流れていることを知る事ができる。R2は、Pathメッセージ内のPQを用いて広告されているリンク情報を修正し、経路計算を行い、ResvメッセージをSへ送信する。その経路は、おそらく、g-h-i-d-e-fようになる(図2(b))。図2(a)と比較して、ネットワーク帯域を効率的に使用している。

2.2 SRSVP

SRSVPはQoS保証マルチキャストルーティングプロトコルであり、RSVP¹¹⁾に似た資源予約機構とPIM-SMに似たマルチキャストルーティング機構を合わせ持つ。

DVMRP^{11),12)}やMOSPF¹³⁾のような従来のマルチキャストルーティングプロトコルは受信者発見を定期的なブロードキャストに頼るため、スケーラビリティの問題がある。SRSVPはPIM-SMと同様にRendezvous Point(以下RVP)の概念を採用し、スケーラブルなマルチキャストルーティングを実現している。SRSVPでは送信者からRVPへはユニキャストによる配送が行われ、RVPから各受信者へはマルチキャストによる配送が行われる。したがって送信者とRVP間およびRVPと各受信者間では異なる資源予約が行われる。

SRSVPでのマルチキャスト配送木・資源予約の成立過程は以下のとおりである。マルチキャストに参加しようとする受信者は、まずResv0と呼ばれるメッセージをマルチキャストの送信者(あるいはRVP)へ向けて送信し、経路上の各ルータはbest-effort経路に沿ってメッセージを転送する。送信者はResv0を受け取ると、その応答としてPathメッセージを送信する。経路上の各ルータはResv0が送信されてきた経路の逆パスに沿ってPathメッセージを転送していく。このとき、SRSVPはPathメッセージを用いて前述のPQを収集する。受信者および経路上の各ルータは、収集したPQを利用してQoS経路計算を行い、求めた経路に沿って今度はResv1と呼ばれるメッセージを送信者へ向けて送る。送信者はその応答としてPathメッセージを送信する。経路上の各ルータはResv1が送信されてきた経路の逆パスに沿ってPathメッセージを転送していく。これに伴ってマルチキャスト経路および資源予約が確立する。

2.3 HQLIP

HQLIPはQoS保証ユニキャストルーティングプロ

トコルであり、OSPFの拡張である。HQLIPは多段の階層を持ち、QoS情報の集約および迅速な伝達を可能とする。HQLIPは、PNNIにおけるクランクバックのような処理が必要無いように設計されている。

HQLIPでは、個々のルータの物理インターフェースをレベル0のエリアとし、ルータを介して繋がっているレベル*i*以下のエリア(サブエリア)をまとめて、レベル*i*+1のエリア(親エリア)とする。エリアの境界はルータ上にあり、親子関係に無い複数のエリアに属するルータをボーダールータと呼ぶ。また、それぞれのエリアにはQoS情報集約のための中心が存在する。

HQLIPによって広告されるリンク情報には、内部リンク情報および外部リンク情報の2種類がある。レベル*i*のエリア内でフラッドされる、そのサブエリア(レベル*j*と*k*のエリア($j, k < i$))とする)同士がどのようにつながっているかの情報が、レベル(*j, k*)の内部リンク情報である。レベル*i*のエリア内でフラッドされる、そのサブエリア(レベル*j*($j < i$))がレベル*i*のエリアに隣接するエリア(レベル*k*)とどのようにつながっているかの情報が、レベル(*j, k*)の外部リンク情報である。

以降で、HQLIPによるリンク情報の生成について説明する。なお、本稿ではリンク情報の表記として($x \leftarrow y, bw=u, dly=v$)を用いる。これは、エリア*y*からエリア*x*へ至るまでのQoSが空き帯域*u*・遅延*v*だということを意味する。

それぞれのルータは、自分のインターフェースに直結したリンクに対して、自分からパケットを出す向きについてのレベル(0,0)の内部リンク情報を生成し、そのインターフェースの親エリア内にフラッドする(図3)。

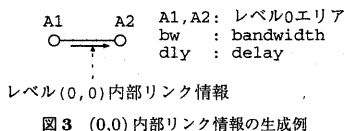


図3 (0,0)内部リンク情報の生成例

レベル*i*(≥ 1)のエリアの境界を内部にもつルータは、レベル*i*のエリアの中心から自分にパケットを受ける際のQoSを計算し、そのエリアに属さず、そのエリアの親エリアに属するインターフェースのリンク情報(レベル(0,*i*)の外部リンク情報)とし、そのインターフェースの親エリア内にフラッドする(図4)。

レベル*j*(≥ 1)のエリアの中心を持つルータは、そのエリア内にフラッドされる内部リンク情報と外部リンク情報をもとに、外部リンク情報でのレベル*i*のエリアから自分にパケットを受ける際のQoSを計算し、レベル(*j, i*)のリンク情報として親エリアにフラッドする。こ

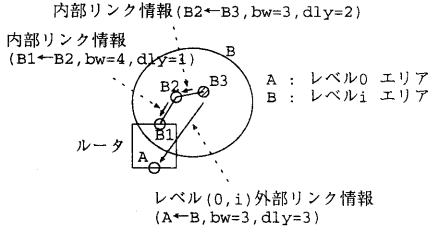


図4 (0, i) 外部リンク情報の生成例

のリンク情報は、親エリアがレベル i のエリアの親エリアでもある場合には内部リンク情報、レベル i のエリアが親エリアの外にある場合には外部リンク情報となる。

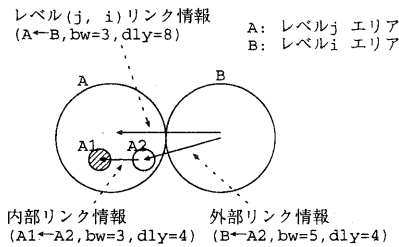


図5 (j, i) リンク情報の生成例

3. 階層化 PQ

3.1 階層化 PQ とは

HQLIP による階層化されたネットワーク上で、SRSVP による PQ を用いる方式の経路計算を行うためには、各ルータ間のリンク情報に対応する従来の PQ に加えて、各エリア間のリンク情報に対応する PQ を生成する必要がある。これを階層化 PQ と呼ぶ。

3.2 階層化 PQ の生成処理の分類

送信者から受信者までの Path メッセージの流れから、階層化 PQ の生成処理は5つのケースに分類されることがわかる。

Path メッセージがあるエリア B の境界を越える時にエリア B が送信者を含まない場合には、ルータは、エリア B の前段のエリアの中でレベルが最大のエリアを A とすると、以下の情報を用いて、エリア A からエリア B への PQ を計算し、Path メッセージに追加する。

Case 1 エリア B がレベル 0 かつエリア A がレベル 0 の場合、 A から B へのリンク情報、リンク上でのそのフローの状態、Path メッセージ自身によって消費されたリソース等を用いて $PQ(B \leftarrow A)$ を計算する（従来の PQ と同様）。

Case 2 エリア B がレベル 0 かつエリア A がレベ

ル 0 より大きい場合、 A から B に至るまでの外部リンク情報と内部リンク情報および前段のルータから受信した Path メッセージに含まれている PQ を用いて $PQ(B \leftarrow A)$ を計算する。

Case 3 エリア B のレベルが 0 より大きい場合、 A から B のサブエリアへの外部リンク情報、 B のサブエリア間の内部リンク情報、前段のルータから受信した Path メッセージに含まれている PQ を用いて $PQ(B \leftarrow A)$ を計算する。

また、Path メッセージが送信者を含むエリアの境界を越えるときには、以下の情報を用いて、FAQ(First Aggregated QoS) と呼ばれる情報が生成される。

Case 4 Path メッセージが送信者を含むレベル 1 以上のエリア A を越える場合、エリア A 内のサブエリアの内部リンク情報、前段のルータから受信した Path メッセージに含まれている PQ、前段のルータから受信した Path メッセージに含まれている送信者から送信者を含む A のサブエリア A' までの FAQ を用いて送信者から A の中心までの QoS 情報を計算し、それを送信者からエリア A への $FAQ(A \leftarrow S)$ とする。

PQ と違って、FAQ に対応するリンク情報は存在しない。すなわち、FAQ は QoS 経路計算時に使用されるのではなく、計算された経路が送信者近辺においても要求された QoS を満たすかどうかを判断するために使用される。

最後に、Path メッセージが受信者に到達したときには以下の処理が行われる。

Case 5 Path メッセージが受信者 R に到達したとき、受信者 R を完全に含む (R 上に境界が無い) エリアに対してそれらのエリアを越えるものとみなし、低いレベルのエリアから順に Case 3 と同様の処理を行う。

以降で Case 2, Case 3, Case 4 の例を示す。図 6 にネットワークの例を示す。図中で S は送信者を、 R は受信者を、 T はルータを示し、それ以外の記号はエリアを示している。斜線がかかったエリアはその親エリアの中心を含むことを示しているが、以降の説明が必要が無い中心に関しては特に示していない。ここで、Path メッセージはルータを $T1 \rightarrow T2 \rightarrow T3 \rightarrow T4 \rightarrow T5$ の順に通過していくとする。

図 7 に、ルータ $T5$ が $PQ(C1a \leftarrow B)$ を生成する様子を示す (Case 2)。 $PQ(C1a \leftarrow B)$ を計算するために以下の情報が用いられる。

- 外部リンク情報 ($B2a \leftarrow B3$, $bw=4$, $dly=2$)
- 内部リンク情報 ($B2b \leftarrow B2a$, $bw=1$, $dly=1$)
- $PQ(B2b \leftarrow B2a$, $bw=3$, $dly=3$)

$PQ(B2b \leftarrow B2a)$ によって内部リンク情報 ($B2b \leftarrow B2a$)

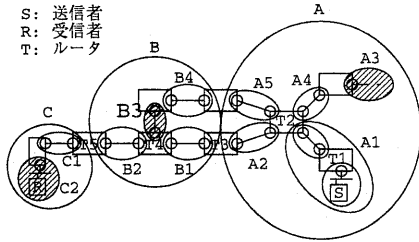


図6 階層化されたネットワーク

は置き換えられ、結果として $PQ(C1a \leftarrow B)$ は以下の情報から求められる。

- 外部リンク情報 ($B2a \leftarrow B3$, $bw=4$, $dly=2$)
- $PQ(B2b \leftarrow B2a$, $bw=3$, $dly=3$)

帯域は上記の値の最小値、遅延は上記の値の和をとることにより、求まる PQ は、 $PQ(C1a \leftarrow B$, $bw=3$, $dly=5$) となる。

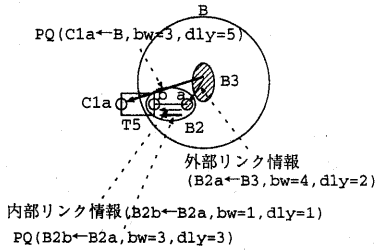


図7 $PQ(C1a \leftarrow B)$ の生成 (Case 2)

図8に、ルータ T5 が $PQ(B \leftarrow A)$ を生成する様子を示す (Case 3) . リンク情報 ($B \leftarrow A$) が外部リンク情報 ($B1 \leftarrow A$) と内部リンク情報 ($B3 \leftarrow B1$) から計算されているとすると、 $PQ(B \leftarrow A)$ を計算するためには以下の情報が用いられる。

- 外部リンク情報 ($B1 \leftarrow A$, $bw=1$, $dly=2$)
- 内部リンク情報 ($B3 \leftarrow B1$, $bw=3$, $dly=1$)
- $PQ(B1 \leftarrow A$, $bw=2$, $dly=2$)

$PQ(B1 \leftarrow A)$ によって外部リンク情報 ($B1 \leftarrow A$) は置き換えられ、結果として $PQ(C1a \leftarrow B)$ は、 $PQ(B \leftarrow A$, $bw=2$, $dly=3$) となる。

図9に、ルータ T3 が $FAQ(A \leftarrow S)$ を生成する様子を示す (Case 4) . $FAQ(A \leftarrow S)$ は以下の情報から求められる。

- 内部リンク情報 ($A3 \leftarrow A4$, $bw=1$, $dly=2$)
- 内部リンク情報 ($A4 \leftarrow A1$, $bw=3$, $dly=2$)
- $FAQ(A1 \leftarrow S$, $bw=2$, $dly=1$)

求まる FAQ は、 $FAQ(A \leftarrow S$, $bw=2$, $dly=5$) となる。

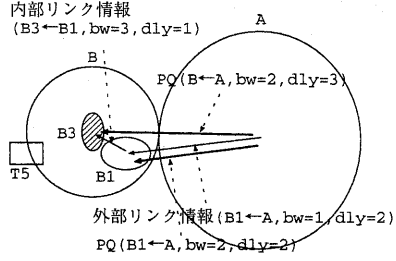


図8 $PQ(B \leftarrow A)$ の生成 (Case 3)

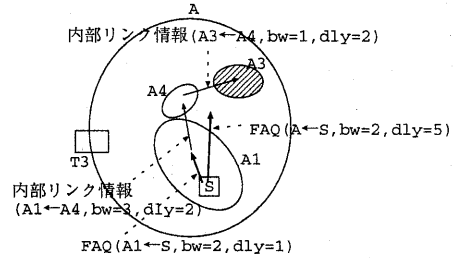


図9 $FAQ(A \leftarrow S)$ の生成 (Case 4)

3.3 各ルータでの動作

Case 1 から Case 5 までのそれぞれの処理に必要な情報等を検証すると、階層化 PQ を生成するために各ルータが行う動作の流れは以下にならない。

```
procedure hierarchical_pathqos
begin
```

```
  a := parent(D);
  while prev(a) ≠ nil and
    level(a) ≤ level(B) do
    Case3(a, prev(a));
    a := parent(a);
  endwhile;
  if C ≠ nil then
    while level(a) ≤ level(B) do
      Case4(a);
      a := parent(a);
    endwhile;
    Case2(C, B);
  else
    while prev(a) ≠ nil do
      Case5(a, prev(a));
      a := parent(a);
    endwhile;
  endif
end
```

ここで、 B は越えるエリアのうち最大のレベルのエリア、 C は Path メッセージの出力インタフェースを示すエリア、 D は Path メッセージの入力インタフェースを示すエリアを示す。また、 $level(x)$ はエリア x のレ

ベルを返す関数, $\text{parent}(x)$ は x の親エリアを返す関数, $\text{prev}(x)$ は x の前段のエリアを返す関数であり, $\text{parent}(x) \cdot \text{prev}(x)$ は, 処理に必要なとなる情報をルータが持っていないければ nil を返すとす。Case2(x, y) は $\text{PQ}(x \leftarrow y)$ を生成する処理であり, Case3(x, y) と Case5(x, y) も同様とする。Case4(x) は $\text{FAQ}(x \leftarrow S)$ を生成する処理とする。

3.4 前段のエリアの求め方

上記のアルゴリズムにおける関数 $\text{prev}(x)$ は, エリア x の前段のエリアを返すものである。しかし, 前段のエリアを求めるためには Path メッセージが流れてきた経路を知る必要があり, そのためには Path メッセージに含まれる PQ を走査しなければならない。すなわち, 各ルータにおいて, 越えるエリアの数だけ Path メッセージを走査することになる。

各ルータが上記のアルゴリズムを実行し, 新たに生成した PQ を, 生成した順番に Path メッセージの最後に付け加えると仮定すると, Path メッセージの走査を後部から行うことにより, この走査は全体として各ルータにおいて 1 回とすることができる。

```
function prev(x:area): area
begin
  while pq ≠ nil do
    a := begging area of pq;
    b := parent(a);
    c := parent(ending area of pq);
    if b ≠ c and c = x then return a;
    pq := previous element of pq;
  endwhile;
  return nil;
end
```

前段のルータから受信した Path メッセージに含まれている PQ はリスト構造で格納されているとし, 変数 pq は Path メッセージ受信時にリスト構造の最後の要素がセットされているとする。

4. おわりに

本研究では, SRSVP がより効果的な QoS 経路計算を行えるようにするために, HQLIP によって階層化されたネットワークにおいて階層化 PQ を計算するアルゴリズムについて設計した。また, 新たに生成した PQ を Path メッセージの最後尾に付け加えて各ルータが収集することにより得られる効率化の方法を示した。

SRSVP と HQLIP の既存の実装の 1 つとして, Real Internet Consortium による RICD がある。今回は RICD に上記のアルゴリズムにもとづいて階層化 PQ を収集する機構の実装を行った。現在, SRSVP と

HQLIP は新しい形での実装が進められており, 今後はそちらをベースに階層化 PQ を収集する機構の実装および評価を行う予定である。

参考文献

- 1) C. Semeria, T. Maufer: Introduction to IP Multicast Routing, <http://www.3com.com/>, 3Com Corporation (1997).
- 2) D. Estrin, et. al.: Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification, RFC 2362, Network Working Group (1997).
- 3) J. Moy: OSPF Version 2, RFC 1583, Network Working Group (1994).
- 4) ATM Forum PNNI subworking group: Private Network-Network Interface Spec. v1.0 (PNNI 1.0), afpnni-0055.00 (1996).
- 5) K. Fujikawa, I. Sheng, Y. Goto: Simple Resource ReSerVation Protocol (SRSVP), Real Internet Consortium (1999).
- 6) M. Ohta: Hierarchical QoS Link Information Protocol (HQLIP), Real Internet Consortium (1999).
- 7) Y. Goto, M. Ohta, K. Araki: Path QoS Collection for Stable Hop-by-Hop QoS Routing, inet97, proceedings (1997).
- 8) Z. Wang, C. Sanchez, B. Salkewicz, E. Crawley: Quality of Service Extensions to OSPF or Quality of Service Path First Routing (QOSPF), Internet Draft (1996).
- 9) B. M. Waxman: Routing of Multipoint Connections, IEEE JSAC 6, pp. 1617-1622 (1988).
- 10) M. Imase, B. Waxman: Dynamic Steiner Tree Problem, SIAM J. Disc, Vol. 4, No. 3, pp. 369-384 (1991).
- 11) R. Braden, et. al.: Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification, RFC 2205, Network Working Group (1997).
- 12) D. Waitzman, C. Partridge, S. Deering: Distance Vector Multicast Routing Protocol, RFC 1075, Network Working Group (1988).
- 13) J. Moy: Multicast Extensions to OSPF, RFC 1584, Network Working Group (1994).