

4-4 分散計算環境構築に向けた高機能光パス設定技術

4-4 Advanced Lightpath Establishment for Distributed Computing

橋 拓至 徐 蘇鋼

TACHIBANA Takuji and XU Sugang

要旨

現在、波長分割多重(WDM)と光パススイッチングを導入した光パスネットワーク上で、分散計算環境を構築することが検討されている。本稿では、光パスネットワーク上で分散計算環境を構築するために、広域かつ多地点間でのデータ通信を実現する新たな二つの光パス設定技術を示す。この二つの光パス設定技術を用いることで、複数のドメインで構成された広域 WDM ネットワーク上で効率の良い光パス設定が可能となり、さらには多地点間でデータを通信するための光リングを動的に構築できる。

Currently, it is expected that distributed computing environment is developed in wide-area networks by using wavelength division multiplexing (WDM) and lightpath switching. In this paper, in order to develop the distributed computing environment over lightpath switching networks, we study two new lightpath establishment approaches. These two approaches enable lightpaths to be effectively established in wide-area WDM networks and enable optical ring to be dynamically developed for the data transmission in multiple points.

[キーワード]

光パススイッチング, グリッド, RSVP-TE, 複数ドメイン, 光リング
Lightpath switching, Grid, RSVP-TE, Multiple domains, Optical ring

1 まえがき

近年、インターネットを介した広域の分散計算環境(グリッド)の構築が進められている。しかしながら、インターネットでは伝送遅延や伝送帯域の保証が難しいため、テラバイトからペタバイト級の大容量データを扱う分散計算環境をインターネットで構築することは困難である。

そこで現在、波長分割多重(WDM)と光パススイッチング[1]を導入した光パスネットワーク上で、分散計算環境を構築することが検討されている[2]–[5]。光パスネットワークでは、二つのノード(送信ノードと受信ノード)間にデータ伝送用の波長(光パス)を設定し、設定した光パスを使ってデータを伝送する。この光パスを、分散計算を

行うエンドホスト間に直結するように設定することで、エンドホスト間にギガバイト以上の伝送帯域を保証できる。

このような分散計算環境を広域に展開する場合、エンドホスト間に複数のドメインを経由する光パスを設定する必要がある。各ドメイン内のネットワーク構成は複雑で、さらに各ドメインは異なるポリシーに従ってネットワークを管理している。そのため、ネットワーク内部の波長利用状況を隠すことなく他のドメインに伝達することは期待できない。したがって、ネットワーク内部の波長利用情報を他のドメインにできるだけ隠しながら、複数ドメインを経由する光パスを設定しなければならない。そこで著者らは、限られた数の波長利用情報を有効に利用して光パスを設定するラ

シミュレーションによる性能評価から、提案方式が複数ドメインを経由する光パス設定に有効であることを示した。

さらに、分散計算環境について考えると、光パスネットワーク上で一対一通信のみならず多地点間通信を行う必要がある。そこで著者らは、多地点間通信のための光パス設定技術及び管理方式を提案した[8][9]。本方式により、光通信技術、光ネットワーク技術、分散計算技術を駆使して、エンドホスト間に保証された広域の分散計算環境を構築可能となる。さらに上記提案方式を実装し、提案方式及び実装システムの有効性を検証した[9]。

本稿では、光パスネットワークで用いられる光パス設定技術について説明し、複数ドメインを経由する光パスの効率の良い設定方式について示す。さらに、光パスを使って片方向リングの広域分散計算環境を構築する方法を示す。

2 RSVP-TE を用いた光パス設定技術

WDM ネットワークでは、Generalized Multi-Protocol Label Switching (GMPLS) の Resource Reservation Protocol-Traffic Extension (RSVP-TE) シグナリングを用いた光パス設定が検討されている[10][11]。図1に RSVP-TE シグナリングによる光パス設定例を示す。RSVP-TE では Path メッセージと Resv メッセージを使って光パスを設定する。

図1では、送信ノード i がユーザから受信ノード j までの光パス設定要求を受信した場合を考える。ノード i は Path メッセージを生成し、次リンクで利用可能な波長の情報を Path メッセージに設定する。図1では、 $\{\lambda_1, \lambda_2, \lambda_4, \lambda_5, \lambda_6, \lambda_8\}$ が設定されている。それから送信ノード i は、Path メッセージを受信ノード j に向けて送信する。

中間ノードが Path メッセージを受信すると、Path メッセージ内に設定されている波長情報から、次リンクで利用できない波長の情報を削除する。それから中間ノードは、Path メッセージを受信ノード j へ向けて送信する。

受信ノード j が Path メッセージを受信すると、

Path メッセージ内の波長情報 $\{\lambda_1, \lambda_4, \lambda_6, \lambda_8\}$ から光パス設定に使用する波長を選択し、選択した波長 λ_4 が利用可能であれば予約する。その後、受信ノード j は Resv メッセージを生成し、選択した波長 λ_4 を Resv メッセージに設定する。それから、Resv メッセージを送信ノード i に向けて送信する。

中間ノードが Resv メッセージを受信すると、Resv メッセージ内に設定されている波長 λ_4 の利用状況をチェックする。波長 λ_4 が利用可能であれば、中間ノードは波長を予約し、Resv メッセージを送信ノード i へ向けて送信する。波長 λ_4 が既に他の光パスに利用されている場合は、光パス設定が失敗し、送信ノード i へ向けて ResvErr メッセージを送信する。送信ノード i が Resv メッセージを受信すると光パス設定が成功する。

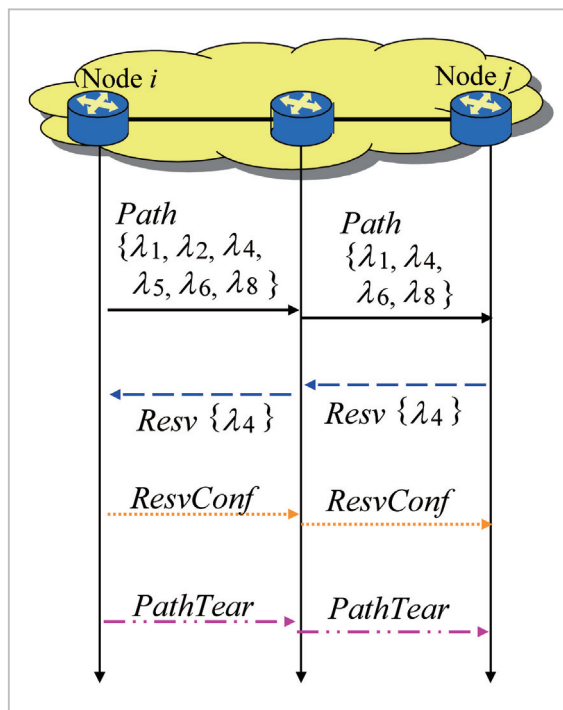


図1 RSVP-TE シグナリングを用いた光パス設定

3 複数ドメイン WDM ネットワークにおける光パス設定技術

図2に、複数ドメインで構成された WDM ネットワークモデルを示す。このような複数ドメイン WDM ネットワークにおいて、前章で説明し

た RSVP-TE プロトコルを用いて光パスを設定する場合、ドメイン内の波長利用情報を他のドメインに通知する必要がある。しかしながら、各ドメインは異なるポリシーでネットワークを運営・管理しているため、ネットワーク内の全波長の利用情報を他ドメインに通知することは期待できない。そこで[6][7]では、各ドメインが限られた数の波長利用情報だけを他のドメインへ通知して光パスを効率よく設定するランク計算型光パス設定方式を提案している。

ランク計算型光パス設定方式では、各ドメインがあらかじめ利用情報を通知する波長数 K に関して契約を行い、契約波長数以下の波長利用情報を隣接ノードに通知する。この契約によって、各ドメインは限られた数の波長利用情報だけを他のドメインへ通知すればよい。また、利用情報を収集する波長を選択する際に、その時点で使用可能な波長を優先的に選択できるように、各波長に対して光パス設定の優先度を示す波長ランクを設定する。送信ノードはこの波長ランクに従って契約波長数分の波長を選択し、限られた数の波長利用情報を効率よく使用して複数ドメインを経由する光パスを設定しやすくする。

図2にランク計算型光パス設定方式のシグナリング手順を示す。図2では契約波長数を $K=3$ としている。光パス設定を行う送信ノードは契約波長数以下の波長 $\{\lambda_1, \lambda_2, \lambda_4\}$ を選択し、Path メッセージと Resv メッセージを使って図2のように光パスを設定する。

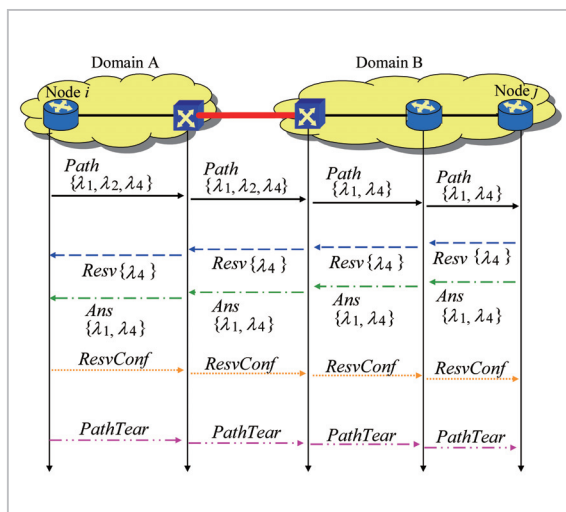


図2 ランク計算型光パス設定方式の光パス設定シグナリング

さらに受信ノードは、Path メッセージが収集した波長利用情報 $\{\lambda_1, \lambda_4\}$ を、Ans メッセージと呼ばれる制御メッセージを使って送信ノードへ返信する。Ans メッセージを受け取った送信ノードは、以下の式(1)に示す指数加重平均などを使って波長ランクを更新する。

$$r_{ij}^n(m) = \begin{cases} (1-\alpha)r_{ij}^n(m-1) + \alpha, & \text{波長} n \text{ は利用可能,} \\ (1-\alpha)r_{ij}^n(m-1), & \text{波長} n \text{ は利用不可能,} \\ r_{ij}^n(m-1), & \text{波長} n \text{ は Ans メッセージに} \\ & \text{含まれていない.} \end{cases} \quad (1)$$

ここで、 $r_{ij}^n(m)$ は送信ノード i から受信ノード j への m 回目の光パス設定終了後の波長 n のランクを示している。また、パラメータ α は $0 \leq \alpha \leq 1$ で与えられる。

式(1)から、光パス設定に利用可能だった波長のランクは増加し、利用不可能だった波長のランクは減少する。それゆえ、波長ランクの高い波長を優先的に使用することで光パスが設定しやすくなる。

図3は、二つのドメインで構成された WDM ネットワークにおけるランク計算型光パス設定方式のシミュレーションによる性能評価結果を示している。各ドメインは 14 ノードの NSFNET で構成されており、各リンクの波長数を 8 とする。このネットワークに対して、二つのドメインを経由する光パスの設定要求が率 1.0 でポアソン到着する。また、各ドメイン内部で設定される光パスの設定要求が率 ξ でポアソン到着する。各光パスの設定保持時間は平均 1 秒の指数分布に従う。また、各ノードでの制御メッセージの処理時間は 1 ミリ秒とする。

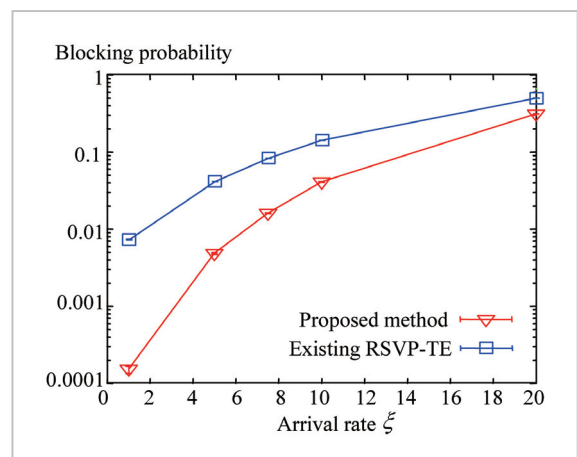


図3 シミュレーション結果

図3では、契約波長数 $K=3$ の場合に関して、ランク計算型光パス設定方式の光パス設定失敗確率を示している。また、従来の RSVP-TE を用いた光パス設定方式の光パス設定失敗確率も示している。図3から ξ の値によらず、ランク計算型光パス設定方式の失敗確率が従来方式の失敗確率よりも常に小さいことが分かる。したがって、ランク計算型光パス設定方式を用いることで、少ない波長利用情報を有効に使用して光パスを設定できる。

4 多地点間通信を実現する光パスを用いた光リング構築技術

分散計算では、一対一通信だけではなく多地点間通信が行われ、複数ホスト間でデータ交換をしなければならない。そこで著者らは、光ネットワークで多地点間通信を実現するために、光パスを動的に設定して片方向光リングを構築する方法を提案している [8] [9]。

図4は片方向光リングを用いた多地点間分散計算環境を示している。光リングでは、光ツリーで必要とされているマルチキャスト機能などが不必要であるためコストを削減でき、さらに光パス設定を片方向に限定することで波長資源の利用率が上が期待できる。

提案方式では、 G 個のホスト $S = \{h_p, h_1, h_2, \dots, h_{G-1}\}$ を含む片方向リングを構築するために以下のアルゴリズムに従って光リングの構成が決定される。

(1) ホスト集合 $S = \{h_p, h_1, h_2, \dots, h_{G-1}\}$ に

対して情報を初期化する。既に光パスが設定されたホスト集合 L と、まだ光パスが設定されていないホスト集合 U を作成する。すなわち、 $L = \{ \}$ 、 $U = S = \{h_p, h_1, h_2, \dots, h_{G-1}\}$ とする。

- (2) 集合 U から親ホスト h_p を削除し L に追加する。すなわち、 $L = \{h_p\}$ 、 $U = \{h_1, h_2, \dots, h_{G-1}\}$ とする。 U に属するホストの中で、ホスト h_p からの経路コスト (ホップ数) が最小になるホスト h を求め、受信ホストとする。ホスト h_p からホスト h に対し光パスを設定する。
- (3) ホスト h を集合 U から外し L に追加する。例えば $h = h_{G-1}$ であれば、 $L = \{h_p, h_{G-1}\}$ 、 $U = \{h_1, h_2, \dots, h_{G-2}\}$ とする。その後、 $U = \{ \}$ でなければ、 U に属するホスト中、 h からの経路コストが最小となるホストを求め、受信ホストとする。ホスト h からそのホストに対し、光パスを設定する。本操作を $U = \{ \}$ となるまで繰り返す。
- (4) L に最後に追加されたホストから親ホスト h_p に対し、光パスを設定する。

ここで、上記アルゴリズムのステップ (2) から (4) では、光パス設定の制御シグナリングが要求される。図5に、片方向光リングを構築するための RSVP-TE シグナリングを拡張した光パス設定シグナリングを示す。図5は三つのノード a, b, c からリング網を構成する例を示している。 $a-b$ 間、 $b-c$ 間及び $c-a$ 間の光パスは、2で説明した RSVP-TE シグナリングに従って設定される。集合 L 及び集合 U の情報は GRID_Group_Msg を

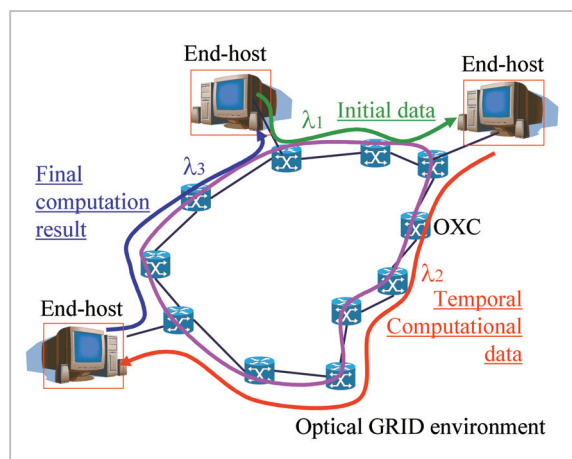


図4 光リングを用いた多地点間分散計算

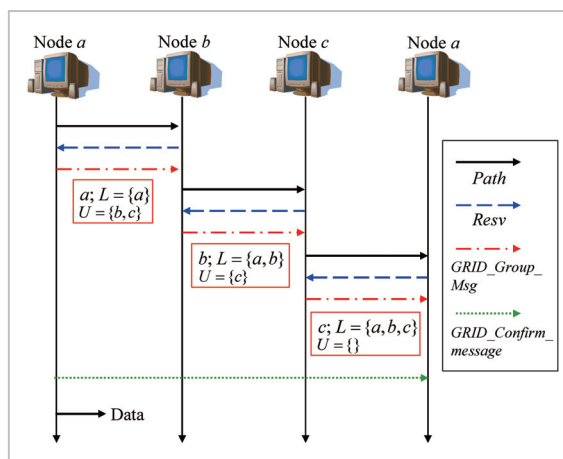


図5 光リング設定シグナリング

使って伝送され、さらに光パス設定終了を確認するために GRID_Confirm_message が使用される。

図6に光リング設定シグナリングの実装実験結果を示す[9]。実験ネットワークは3ノード(a, b, c)で構成され、ノード間のデータ伝送用リンクは点線で示されている。また、各リンクの波長多重数を2とし、制御プレーンはLANで構築されている。実験結果から、本実装を用いることで光リ

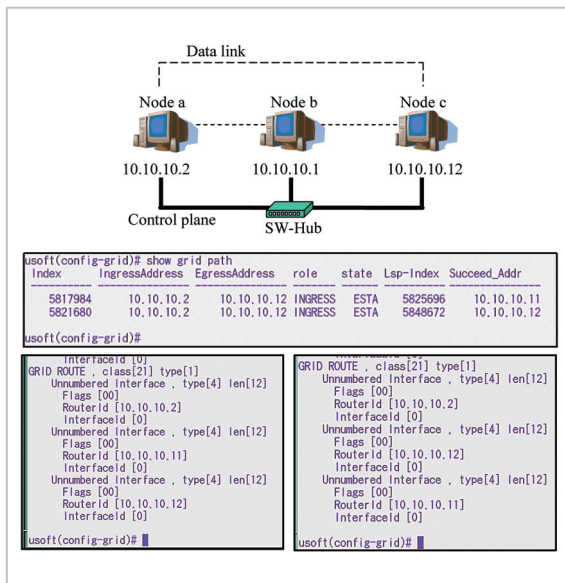


図6 実装実験結果

参考文献

- 1 I.Chlamtac, A.Ganz, and G.Karmi, "Lightpath communications: An approach to high bandwidth optical WAN's", IEEE Transactions on Communications, Vol.40, pp.1171-1182, July 1992.
- 2 "OptIPuter", <http://www.optiputer.net/>.
- 3 "CA*Net 4", <http://www.canarie.ca/>.
- 4 "National Lambda Rail", <http://www.nlr.net/>.
- 5 H.Nakamoto, K.Baba, and M.Murata, "Proposal of a Shared Memory Access Method for Lambda Computing Environment", in Proc. Op-NeTec 2004, Oct. 2004.
- 6 T.Tachibana and H.Harai, "End-to-End Lightpath Establishment Based on Rank Accounting without Wavelength Conversion for Multi-Domain WDM Networks", in Proc., 31st European Conference and Exhibition on Optical Communication Conference (ECOC2005), Sept. 2005, No.We4.P.087, pp.675-676.
- 7 T.Tachibana and H.Harai, "Lightpath Establishment based on Rank Accounting in Multi-Domain WDM Networks", to appear in IEICE Transaction on Communications.
- 8 H.Harai and M.Murata, "Establishing Lightpaths of an Optical Ring for Distributed Computing Environment", in Proc. IEEE GridNets2005 (Second International Workshop on Networks for Grid Applications), 2005, pp.488-495.

ングによる分散計算環境を動的に構築できることが分かる。

5 むすび

本稿では、WDM ネットワークにおいて光パススイッチングを利用して分散計算環境を構築するための二つの高機能光パス設定技術について紹介した。ランク計算型光パス設定方式では、ランク計算を利用することで複数ドメインを経由する光パスを効率よく構築できることを示した。光パスを用いた光リング構築技術では、RSVP-TE を拡張した制御シグナリングによって、多地点間通信のための光リングを動的に構築できることを示した。これらの方式を利用することで、光パススイッチングによる大規模な分散計算環境を動的に構築することが期待できる。

謝辞

本稿を執筆するに当たり、議論した情報通信研究機構超高速フォトニックネットワークグループ原井洋明主任研究員に深謝する。

- 9 S.Xu and H.Harai, "Optical Ring Services in GMPLS Based Mesh Networks: An Implementation of Optical GRID", in Proc., Optical Fiber Communication Conference & Exposition and the National Fiber Optical Engineers Conference (OFC/NFOEC 2006), Mar. 2006, No.JThB74.
- 10 L.Berger, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", IETF RFC 3471, Jan. 2003.
- 11 L.Berger, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", IETF RFC 3473, Jan. 2003.



橘 拓至

奈良先端科学技術大学院大学(元情報通信部門超高速フォトニックネットワークグループ専攻研究員) 博士(工学)
フォトニックネットワークアーキテクチャ、光パスネットワーク



徐 蘇鋼 (Xu Sugang)

新世代ネットワーク研究センターネットワークアーキテクチャグループ有期研究員(旧情報通信部門超高速フォトニックネットワークグループ専攻研究員) 博士(工学)
超高速ネットワーク制御及び設計