WSクラスタを用いたCORErouterプロトタイプの構成

2Bb－6

三栄 武　小倉 毅　丸山 充　高橋 直久
NTTソフトウェア研究所

1 はじめに

我々は、CORErouterと呼ぶ、応答性・信頼性・サービス性・可用性・保守性・安全性の高いネットワーク制御機能およびネットワークサービス機能を備えた高並列IPルータの研究を進めている[2]。現在は、その第一段階として、CORErouterの開発に必要となる基礎的データの収集とルータの論理シーケンスの検証を目的とした、プロトタイプCORErouter-Iを作成中である。本プロトタイプは、複数台のワークステーションを高速な相互結合網で接続したWSクラスタを構成する。本稿では、プロトタイプの構成および、プロトタイプが使用する相互結合網の基礎的評価について述べる。

2 並列ルータの構成

2.1 物理構成

文献[2]で述べたように、柔軟でスケーラブルなIPルータを実現するため、プロトタイプCORErouter-Iは、以下の4種類の機能を持つ汎用ワークステーション(WS)を汎用相互結合網で接続した、機能分散型WSクラスタの構成をとる。

- **ルーティング プロセッサ (RP):** BGP (Border Gateway Protocol) などのルーティングプロトコルの処理機能を持ち、ルータ間で経路情報の交換を行なう。
- **ルートサーチプロセッサ (RSP):** RPから経路情報の更新通知を受信し、ルーティングテーブルの作成・更新を行なうルーティングテーブルの管理機能、および、IFPから宛先IPアドレスを受け付け、フォワードすべきIFPをルーティングテーブルから検索して回答するルーティングテーブル検索機能を持つ。
- **回線インタフェースプロセッサ (IFP):** 外部ネットワークと接続するインタフェースを持ち、IPパケットのフォワード機能およびフィルタリング機能を持つ。IPパケットのフォワード機能は、外部ネットワークから受信したIPパケットを、RSPの検索結果に従ってIFP間で転送し、他ネットワークに送信する。この時、RSPから得た検索結果のキャッシングし、次回以降のフォワーディングの高速化を図る。フィルタリング機能は、ルータ管理者の設定と一致するIPパケットに対して、フォワーディングの許可 / 不許可の処理を行なう。
- **サービスプロセッサ (SP):** ファイアウォール、DNS、proxyなど高機能なネットワークサービスを処理する。

CORErouter-Iが使用する汎用相互結合網は、8個の入出力インタフェースを持つクロスバ型スイッチであり、全二重640Mbpsのデータ転送能力を持つ。また、スイッチの制御ソフトウェアには、スイッチ標準APIとFM[1](Fast Messages)、およびネットワークドライバがある。スイッチ標準APIとFMは、ユーザがスイッチを直接操作してデータ転送を行なうためのインタフェースを提供し、ネットワークドライバは、イーサネット等と同様なIP層のデバイスに、スイッチを仮想化する。ここで、FMは、スイッチ標準APIよりレイテンシの小さいデータ転送を目的としてIllinois大学で開発されたソフトウェアである。

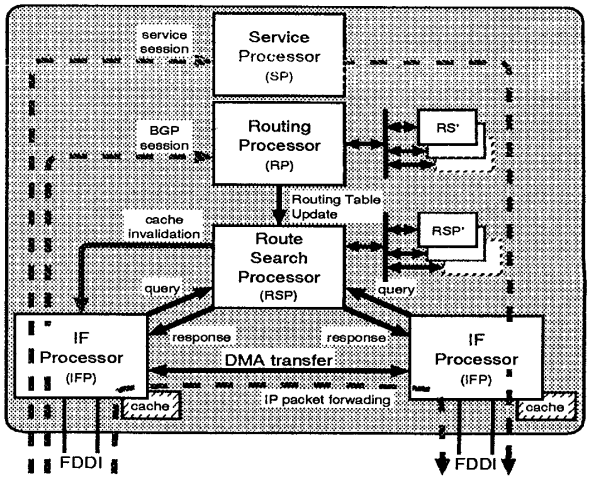

図1: CORErouter-Iの論理構成

2.2 論理構成

CORErouter-I の論理構成を図1に示す。CORErouter-Iは、各IFPを介して、ネットワークとIPパケットの送受信を行なう。IFPは、受信したIPパケットのヘッダ部から宛先IPアドレスを読み出し、その値に基づき、それぞれ以下の手順でIPパケットを処理する。

宛先IPアドレスが本ルータ以外の場合: IFPは、宛先IPアドレスが自キャッシュにあるか判定する。キャッシュになければ、RSPに問い合わせを行い、検索結果をキャッシュする。IFPは、フィルタリングを行なった

Design of a Parallel IP Router Prototype *CORErouter-I*.
Takeshi MIEI (take@slab.ntt.jp), Tsuyoshi OGURA (ogura@slab.ntt.jp), Mitsuru MARUYAMA (mitsuru@ntt-20.ntt.jp), Naohisa TAKAHASHI (naohisa@slab.ntt.jp)
NTT Software Laboratories
9-11, Midori-Cho 3-Chome Musashino-Shi, Tokyo 180 Japan.

後、フォワード先IFPにIPパケットを転送し、外部ネットワークに送出する。

宛先IPアドレスが本ルータの場合: IFPは、ルーティングプロトコルの場合にはRPに、それ以外の場合にはSPにIPパケットをフォワードする。RPはIPパケットを処理し、経路情報の更新をRSPに通知する。RSPは、更新通知に従ってルーティングテーブルを更新するとともに、IFPに更新したキャッシュエントリの無効化を通知する。また、SPは必要な処理を行ない、結果をIFPに送る。

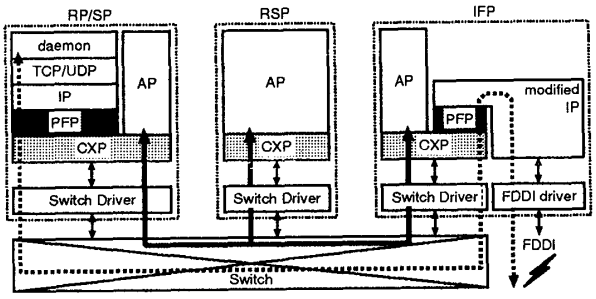

図2: CORErouter-Iの内部通信プロトコル

この時、CORErouter-Iの内部通信に使用するプロトコルを図2に示す。CXP (Communication eXchange Protocol)は、スイッチのデバイスドライバを直接操作し、高速なメッセージパッシングのインタフェースを提供する。CORErouter-Iを構成する各WSはCXPを用いて情報交換を行なう。PFP (Packet Forwarding Protocol)は、IFP-IFP間およびIFP-RP間のIPパケットフォワーディングのインタフェースを提供する。図1, 2において、実線矢印はCXPによる通信を示し、破線矢印はPFPによる通信を示している。

3 相互結合網の評価

前節で述べたCXPにおいて、スイッチ制御を実現するために現在は、前述のスイッチ標準APIまたはFMを用いることができる。我々は、CORErouter-I作成における基礎的評価として、スイッチ標準APIとFMを用いた場合の、帯域幅とレイテンシを測定した。図3は、2台のWS間でデータ転送を行ない、パケットの大きさを変化させた場合の帯域幅の変化を、スイッチ標準APIとFM各々について測定した結果である。また、図4は、往復のレイテンシについての測定結果を示している。

これらの結果から、スイッチ標準APIではパケットサイズに比例して帯域幅が大きくなるのに対して、FMではパケットサイズが1000byte以上であれば、130Mbps程度の速度が得られることが分かる。また、FMでは、スイッチ標準APIの1/8～1/3程度のレイテンシに抑えられていることが分かる。この結果に従い、レイテンシが小さく、かつ、パケットサイズが小さい場合にも大きな帯域幅が得られるFMを用いて、CORErouter-IのCXPの実現を進めている。

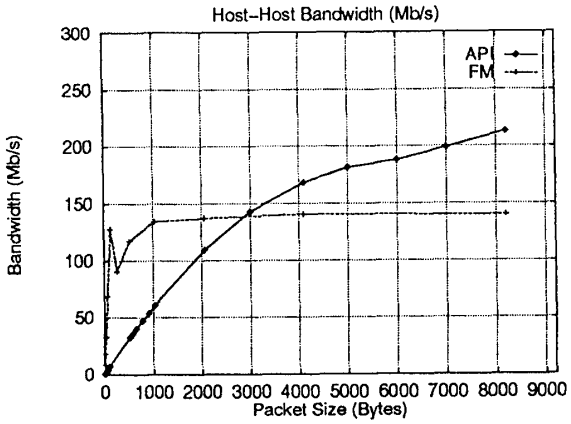

図3: スイッチ標準APIとFMを用いた場合の帯域幅

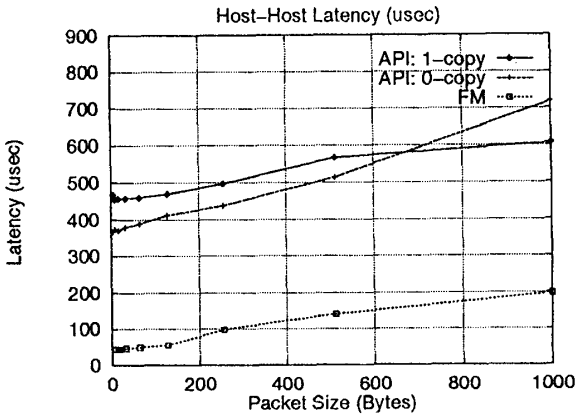

図4: スイッチ標準APIとFMを用いた場合のlatency

4 おわりに

現在、CORErouter-Iは、2台のIFP間で1対1のIPパケットフォワーディング機能を実装し、80Mbpsのフォワーディング性能を確認した段階である[3]。今後は、IFPの多重化および、ルーティングテーブルのキャッシュ機構などCORErouterで使用する各種アルゴリズムの検証を進めていく予定である。

謝辞 本研究を御支援下さる後藤滋樹広域コンピューティング研究部長、ならびに日頃御討論いただく超並列プログラミング研究グループの皆様に感謝いたします。

参考文献

1) "Illinois Fast Messages (FM)", http://www-csag.cs.uiuc.edu/projects/comm/fm.html, 1995.

2) 高橋, 丸山, 三栄, 小倉, "柔軟でスケーラブルな高性能ルータ CORErouter の基本構想", 第52回情処全国大会, 2B-05, 1996.

3) 丸山, 三栄, 小倉, 高橋, "WSクラスタを用いたCORErouterプロトタイプの評価", 第52回情処全国大会, 2B-07, 1996.