

複数種 NIC による広帯域通信のための TCP 再送抑制手法*

東京電機大学 理工学部 情報システム工学科†

加藤 剛史 小野 真和 桧垣 博章‡ §

1 背景と目的

現在、LAN の構築にはイーサネットが広く利用されており、その帯域幅は 10Mbps から 1Gbps へと順次拡大されてきた。一方、ストリーミングのようにマルチメディアデータの配送に一定時間広帯域幅を必要とし続けるネットワークアプリケーションや、サイズの大きなファイルの FTP 転送のようにバースト的に広帯域幅を要するネットワークアプリケーションが存在する。これらのネットワークアプリケーションに対して十分な帯域幅を提供する手法のひとつとして、複数の通信路による配送を用いることによる広帯域 LAN の実現が考えられる。特に、IP 電話や IP テレビ会議といった P2P (Peer-to-Peer) 型のネットワークアプリケーションに広帯域通信を提供するためには、LAN に接続された各コンピュータに複数のイーサネット NIC を装着し、アプリケーションの通信要求に従って時々刻々変化する様々な通信相手との間に複数の通信路を用意することが求められる。我々は、ひとつの IP アドレスを複数の MAC アドレスに変換することが可能な拡張 ARP プロトコルを設計した。これによって、IP および UDP を用いた通信では広帯域な通信路を提供することが可能となったが、高信頼な通信路をアプリケーションに提供するために再送機構とフロー制御機構を導入した TCP においては、特に複数種 NIC が混在した場合に性能が劣化する。そこで本論文では、送信元コンピュータにおいて IP データグラム群の送信順序を制御することによって、TCP の再送機構が機能することを抑制し、TCP においても複数の通信路を用いてネットワークアプリケーションに広帯域幅を提供する手法を提案する。

2 関連研究

限られた帯域幅を持つ通信路を複数用いることによって、送信元コンピュータと送信先コンピュータとの間に広帯域通信を実現する方法は、これまでにも様々な研究開発がなされ、広く用いられてきた。しかし、これらの方法は、固定のコンピュータ対に対してのみ適用することが可能である。LAN に接続されたコンピュータが P2P 型のネットワークアプリケーションを実行する環境においては、各コンピュータに複数のイーサネット NIC を装着し、アプリケーションからの通信要求に従って時々刻々と変化する様々な通信相手との間に複数の通信路を用意しなければならない。

現在、イーサネット LAN では、リピータハブに変わり、スイッチングハブが広く利用されている。そこで、複数の NIC を装着したコンピュータを図 1 に示すようにスイッチングハブに接続し、複数の NIC から複数の NIC へとフレーム群を配送することによって、広帯域通信を実現することができる [1]。ここで、現在運

用されている LAN への導入を可能とするために、以下の条件を満たすことが求められる。

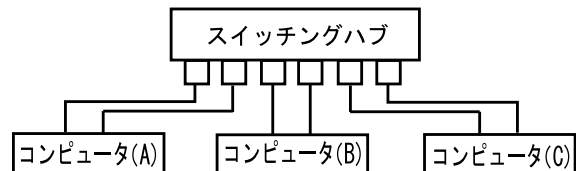


図 1: 複数 NIC を用いた広帯域通信

[要求条件]

1. 既存のネットワークアプリケーションへの変更が不要である。このためには、MAC アドレスの異なる複数の NIC に、同一の IP アドレスを与えることが求められる。
2. すべてのコンピュータに提案機構が導入されていることを前提としない。提案機構が導入されているコンピュータと導入されていないコンピュータとが混在していても、TCP/IP による通信が正しく行われるものとする。□

論文 [2] では、これらの条件を満足する拡張 ARP が提案されている。拡張 ARP を用いることによって、送信元コンピュータは、送信先コンピュータの IP アドレスに対応する複数の MAC アドレスを得ることができる。そこで、送信元コンピュータの各 NIC にいずれかの送信先 NIC を対応付け、送信元コンピュータの各 NIC から送信されるフレームの送信先を対応付けられた NIC とすることによって、複数の NIC を装着した送信先コンピュータが複数の NIC を用いてフレーム群を受信することが可能になる。また、帯域幅の異なる複数種の NIC が送信元コンピュータ、送信先コンピュータに装着されている場合には、可能な範囲で帯域幅の等しい NIC を対応付けることとし [3]、さらに各 NIC に対するフレームの配分率を動的に調節することによって、送信元コンピュータから送信先コンピュータへの帯域幅を拡大することを実現している [4]。ところが、これまでに提案された手法はすべて UDP による通信を前提としている。UDP と IP はいずれもパケット (ユーザデータグラムおよび IP データグラム) の配送順序保存を保証しないプロトコルである。送信元コンピュータと送信先コンピュータとの間に複数の通信路が存在する場合、伝達遅延や帯域幅、フレーム紛失率といった各通信路の特性を考慮せずにフレームを配分すると、たとえ各通信路が FIFO の特性を持っていて配送順序の保存を保証しているとしても、送信元コンピュータが送信した順に送信先コンピュータが受信するとは限らない。

3 提案手法

前章の手法を TCP に用いる場合には、以下の理由によって広帯域通信が提供されない。

* Avoidance of TCP Retransmission in Wider Bandwidth Communication with multiple NICs

† Tokyo Denki University

‡ Takeshi Katou, Masakazu Ono, and Hiroaki Higaki

§ {masa, kato, hig}@higlab.k.dendai.ac.jp

まず、下位層で配送順序が保証されないことにより、送信先コンピュータで受信された TCP セグメント群が長時間バッファリングされることにより、帯域幅が低下する。図 2 に示すように、狭帯域幅の通信路を用いて配送される TCP セグメント T_i を含むフレームが先に送信を開始されたにも関わらず、以降に送信を開始された TCP セグメント群 $T_{i+1}, \dots, T_j (j > i)$ を含むフレームの方が先に送信先コンピュータに受信されることが考えられる。このとき、TCP では $T_{i+1}, \dots, T_j$

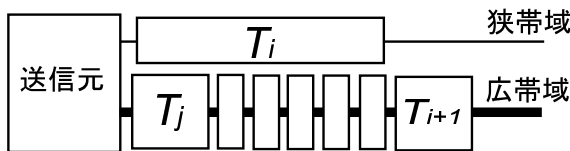


図 2: 配送順序の変動

を T_i が受信されるまでバッファに保存し、アプリケーションへの配送は遅延される。

つぎに、先着した TCP セグメント群に対する受信確認によって TCP の再送機構が機能して帯域幅が低下する。TCP では、累積的受信確認あるいは選択的受信確認を用いて、紛失したフレームに含まれる TCP セグメントを再送信している。ここでは、 $T_{i+1}, \dots, T_j$ の各 TCP セグメントの到着に対して、送信先コンピュータが受信確認の TCP セグメントを送信するが、このいずれにも T_i が未受信である情報が含まれている。これらを受信した送信元コンピュータは、閾値 k 回 (多くの実装が $k = 3$ としている) 以上の未受信情報を得ると、 T_i を再送信する。すなわち、 $j - i > k$ を満たすならば、 T_i が紛失していない場合でも再送信される。

さらに、TCP セグメントの再送信によって、TCP のフロー制御におけるウィンドウサイズが縮小され、配送中の状態にある TCP セグメント群に含まれるデータ量が低く制限され、帯域幅の拡大が実現されない。

TCP では配送効率を向上させるために、送信元コンピュータにおけるバッファリングが用いられている。そこで、アプリケーションから送信要求されたデータを含む TCP セグメント群の送信順序を、これらが送信される NIC の帯域幅に応じて変更することによって、送信先コンピュータにおける受信順序が送信元コンピュータのアプリケーションが配送要求した順序と等しい、あるいは大きく異ならないようにすることで、上記の問題を解決する。図 3 に示すように、送信元コンピュータのアプリケーションが配送要求した TCP セグメント群 $T_i, \dots, T_j (j > i)$ をこの順序で送信先コンピュータに受信させるために、TCP セグメント群 $T_{n_1}, \dots, T_{n_m}$ を狭帯域 NIC を用いて配送し、他の TCP セグメントを広帯域 NIC から送信するが、このとき、 $T_{n_i} (1 \leq i \leq m)$ を $T_{n_{i-s}}, \dots, T_{n_{i-1}}$ よりも先に送信する。これによって、先

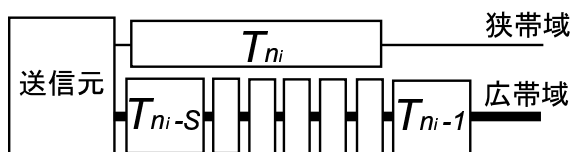


図 3: 送信順序の変更

に送信を開始された TCP セグメント T_{n_i} が受信される以前に、この TCP セグメントと同時あるいはこれ以降

に送信を開始された TCP セグメント群 $T_{n_{i-s}}, \dots, T_{n_{i-1}}$ の方が先に受信されることになり、広帯域通信が実現できる。なお、定数 s をスキップ数と呼ぶ。

4 評価

提案手法を Linux コンピュータに実装し、その有効性を実験によって確認した。実験には、PentiumIII 1GHzCPU と 192MB メモリを搭載した PC2 台と、スイッチングハブ LSW-GT-8W を用いた。各 PC には、1000Base-TNIC である PRO/1000-MT と 100Base-TXNIC である 3C905-TX を 1 枚ずつ装着した。オペレーティングシステムは、Linux(カーネルバージョン 2.2.17) である。図 4 にスキップ数の変化に対する帯域幅と再送信発生率を示す。スキップ数が 0 の場合

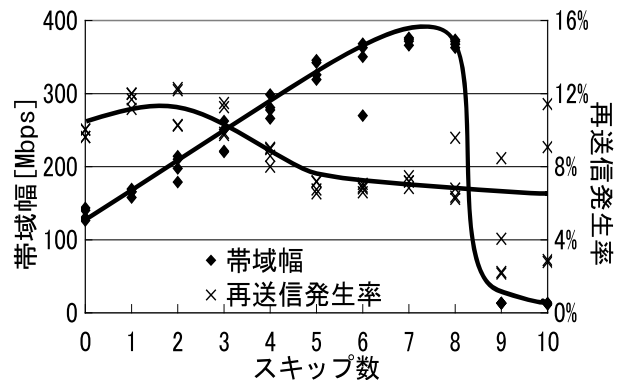


図 4: スキップ数に対する帯域幅と再送信発生率

が従来手法である。スキップ数が 8 のとき最大帯域幅 369Mbps が得られており、従来手法の 136Mbps に対して 271% の拡大が実現されている。また、再送信発生率は従来手法の 9.79% に対して提案手法では 7.17% となっており、26.8% の削減となっている。

5 まとめと今後の課題

本論文では、LAN に接続された複数の NIC を装着したコンピュータ間の通信を拡張 ARP に基づいて広帯域化する手法を TCP 通信に用いる場合に問題となる再送信要求の発生を抑制するために、狭帯域通信路で配送されるフレームの送信順序を早める手法を提案した。提案手法を Linux オペレーティングシステムに実装し、実験により、その有効性を示した。今後は、多様な NIC 数の組み合わせに対するスキップ数を定め、提案手法が一般に有効であることを示す。

参考文献

- [1] 出口, 桧垣, “複数 NIC とスイッチングハブを用いた広帯域通信機構の構築と評価,” 信学技報, Vol. 100, No. 670, pp. 129–134 (2001).
- [2] 林, 梅島, 桧垣, “複数 NIC とスイッチングハブを用いた広帯域通信機構の LINUX への実装,” 信学技報, Vol. 101, No. 639, pp. 33–38 (2002).
- [3] 加藤, 梅島, 森田, 桧垣, “複数種 NIC による広帯域 LAN のための ARP の拡張と実装,” 電子情報通信学会総合大会論文集, p. 351 (2003).
- [4] 中田, 杉木, 梅島, 桧垣, “複数種 NIC による広帯域通信のための配分率制御機構の実装,” 電子情報通信学会論文集, p. 327 (2004).