A-9　複数NICを用いた通信負荷分散によるバンド幅性能向上と並列処理への効果

福永隆文*　上瀧剛**　村上悟**
菊池泰紀**　芦原評**　梅野英典**

1．はじめに

近年，PC/AT 互換機等（以後 PC）の高性能化及び Fast Ethernet や Gigabit Ethernet 等による通信性能の向上にともない PC をクラスタ結合した並列システムが広まりつつある．しかしながら，通信の多いアプリケーションでは通信オーバヘッドが大きくなり性能が向上しないケースも見受けられる．

我々は TCP/IP アーキテクチャを用いた Fast Ethernet 上に複数 NIC を用いて通信バンド幅を向上させる仕組み（以後 NIC 分散）を組み込み，バンド幅の向上が並列処理に与える効果を評価する．当通信負荷分散の仕組みは，Linux の Channel Bonding ドライバ（以後 Bond ドライバ）へ変更を加えることにより実装した．

複数 NIC を用いて通信バンド幅を向上させる仕組みとしては PM/Ethernet[1] の Network Trunking があるが，我々が提案する方式は TCP 上で実現している点が異なる．

2．NIC 分散処理

2．1　NIC 分散処理の概要

複数 NIC を用いた NIC 分散方式のアーキテクチャを図 1 に示す．eth0～eth3 は NIC を示す．送信処理では，今回修正を加えた Bond ドライバが各 NIC ドライバをラウンドロビン方式で呼びだすことにより，送信データを複数の NIC へ交互に割り当てて送信する．1 枚の NIC 使用時には，NIC 送信中は次の送信メッセージが存在しても送信できないため待ちが発生するが，複数 NIC を使うことにより送信中以外の NIC を利用して送信できる．各 NIC から送信される Ethernet フレームの宛先アドレスは，送信側 NIC 毎に対応する受信側 NIC のアドレスに書き換えるので，Switch により対応する NIC へ送信される．

2．2　NIC 分散処理の実装

我々は Linux2.4.2 上に NIC 分散を実装した．カーネル及び NIC ドライバは変更せず，Bond ドライバの修正を行った．このドライバは Linux に実装されているダイナミックロードが可能なドライバであり，送信側の負荷分散のみを行う．宛先 Ethernet アドレスは同一であり，受信側では特定の NIC が受信し，負荷分散は行われない．この Bond ドライバの送信関数内に宛先 Ethernet アドレスをラウンドロビン方式で交互に変更する仕組みを組み込んだ．送信元の NIC と送信先の NIC は送信先アドレステーブルで対応づけられている．送信先の NIC

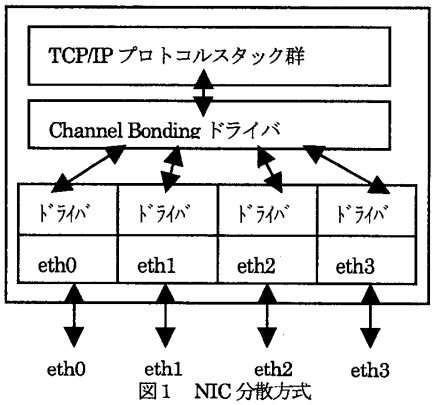

図 1　NIC 分散方式

表 1　送信先アドレステーブルの項目

項目	説明
送信先識別子	送信先識別用
NIC 番号	送信元の NIC に付けた番号で，eth0 が 1，eth1 が 2，以下連番となる
変更アドレス	送信元 NIC と通信を行う送信先 NIC の Ethernet アドレス

アドレスは事前に送信先アドレステーブルに登録する必要がある．テーブルの項目を表 1 に示す．Bond ドライバの送信関数内で各 NIC のドライバをラウンドロビンで呼びだしているが，その呼び出しの前に当テーブルを検索し，宛先アドレスを送信元 NIC に対応する送信先 NIC のアドレスに書き換える．

TCP/IP 処理は受信側が特にオーバヘッドが高いが，今回のコード追加は送信側にのみコードを追加した．これにより，オーバヘッドの増加による通信への影響はある程度押さえられる．

3．評価

バンド幅の測定結果，Channel Bonding 方式との比較結果及び NAS 並列ベンチマークの性能評価結果を示す．測定環境は Celeron 533MHz,,256MB SDRAM ﾒﾓﾘ（環境 1）と Pentium IV 1.6GHz,,256MB SDRAM ﾒﾓﾘ（環境 2）を用いた．1MB は 1,024×1,024 バイトで計算した．

3．1　TCP/IP バンド幅性能

TCP/IP 処理はコネクション型でありオーバヘッドが高い．今回 NIC 分散のためコードを追加するのでオーバヘッドはさらに高くなる．そこで，NIC 分散によるバンド幅性能向上と TCP/IP 処理オーバヘッドが NIC 分散に与える影響を示す．

図 2 は環境 1 を用いた場合の NIC 分散とバンド幅の関係であり，図 3 は環境 2 を用いた場合の結果である．図 2 より CPU の処理能力が低い場合，NIC 2 枚までは分散による効果が見られるが，それ以上 NIC 枚数を増やしても性能向上が見られな

*　熊本県立技術短期大学校
**　熊本大学　工学部

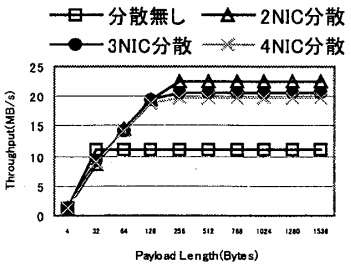

図2 バンド幅性能（環境1）

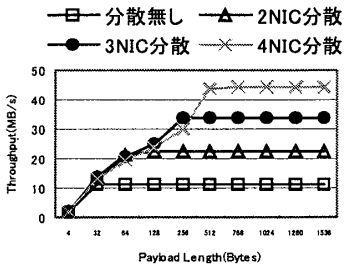

図3 バンド幅性能（環境2）

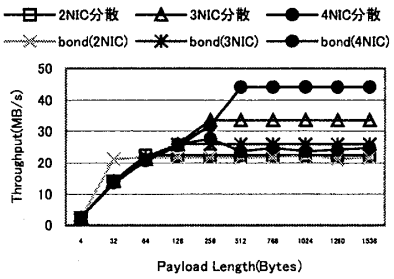

図4 バンド幅性能比較

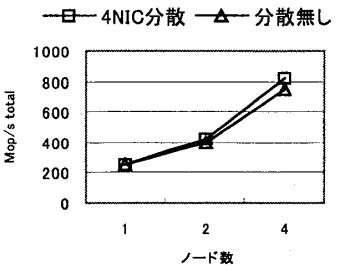

図5 LUクラスW

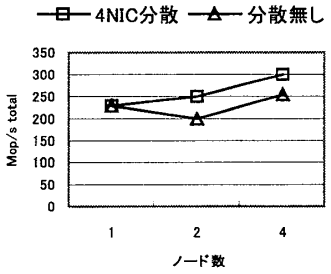

図6 MGクラスW

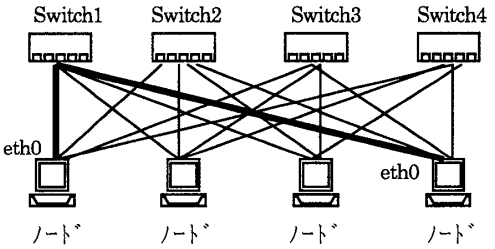

図7 Switchとの接続形態

いことが分かる．原因はCPUアイドル率が2枚に分散した時点でほぼ0%に達しているためである．これに対し図3から分かるようにCPUの処理能力が高い場合，枚数にほぼ比例して性能が向上している．CPUの処理能力向上に伴いTCP/IP処理のオーバヘッドは相対的に低くなっていくと考えられる．

3．2 Channel Bondingとの比較

図4にNIC分散方式とChannel Bonding方式のバンド幅の比較結果を示す．測定に用いた環境は前述の環境2に加えSwitchを用いた．

Channel Bonding方式で2枚のNICを使った場合は1,024バイトメッセージで22.1MB/sの性能を示した．NIC分散方式と同程度である．Fast Ethernetの物理的性能である11.9MB/sの1.8倍以上の性能を実現した．しかし，3枚以上のNICを使った場合の性能向上は見られない．受信側CPUの負荷率が60%程度であることからCPUの受信処理がボトルネックになっているわけではなく，受信側NICによる処理がボトルネックになっていると考えられる．

図4の結果よりNIC分散方式は，3枚NICの場合1,024バイトメッセージでChannel Bondingの1.29倍の性能を示し，4枚NICの場合1.88倍の性能を示している．

3．3 NAS並列ベンチマーク

NIC分散方式を組み込んだ場合と組み込まない場合についてNAS並列ベンチマーク2.3の性能比較を行った．今回は通信が比較的多いLUとMGについて測定した．クラスはWを使用した．NIC分散方式は4枚のNICを用いた．結果を図5，図6に示す．

結果よりLU,MGともに台数増加による速度向上が見られる．また，NIC分散方式を利用した場合と利用しなかった場合を比較すると，NIC分散方式が良い性能を示すことが分かる．MGクラスWでは4ノードの時，分散しなかったときの1.16倍の性能を実現している．LUクラスWでは4ノードの時1.09倍の性能を実現している．

4． Switch接続方式

当NIC分散方式は送信側の各NICに受信側の各NICを対応させ，その間で並列に送受信を行うため，図7の接続形態が可能である．この接続形態を用いれば，複数Switchを必要とする場合もSwitch間を接続する必要がなく，Switch間リンクのバンド幅による制限を受けず，さらに複数のSwitchに通信負荷を分散するため，各Switch内のバンド幅の制限も受けにくい．

5．まとめ

本稿ではNIC分散による通信バンド幅の性能向上とそれが並列処理に与える効果について述べた．コードの変更はChannel Bonnding方式で用いられているBondドライバに対してのみ行った．カーネルや各NICドライバに対しては変更を行っておらず，既存のTCP/IP等の標準のプロトコルもサポートし，また，多くのNICで利用可能である．

通信バンド幅では，4枚のNICを使って分散した場合は分散しない場合の4倍近い性能を実現できることが分かった．NAS並列ベンチマークにおいては，わずかではあるが，NIC分散方式を用いた場合が分散しなかった場合より良い性能を示した．

参考文献

1) 住元真司，堀 敦史，手塚宏史，原田 浩，高橋俊行，石川 裕：既存OSの枠組みを用いたクラスタシステム向け高速通信機構の提案，情報処理学会論文誌，Vol.41,No.6,pp.1688-1696 (2000).