AVALIAÇÃO DO U-NET EM CLUSTERS COM REDE MYRINET

Paulo A. Geromel e Sergio T. Kofuji
[geromel, kofuji]@lsi.usp.br

Laboratório de Sistemas Integráveis - Departamento de Engenharia Eletrônica
Escola Politécnica da Universidade de São Paulo
Av. Prof. Luciano Gualberto, trav. 3, 158. CEP. 05508-900 – São Paulo – SP – Brasil
Tel. (011) 818-5667 Fax. (011) 211-4574

Abstract: Clusters of computers and workstations interconnected by high speed networks are widely used in high performance computing. Although high speed networks have the best latency and throughput, these features are not properly used by conventional communication protocol implementations. A possible solution for such a problem is found in Light-Weight Protocols. These protocols are stated in a User-Level aproach, where the communication tasks are directly processed in user space, instead of being processed inside operating system kernel, using a communication library. This paper evaluates the performance of the U-Net protocol. Using computer platforms with Intel Pentium processors interconnected by a Myrinet Network, features like latency, throughput and concurrency support are evaluated. The results from these evaluations indicate that applying light-weight protocols is a proper way to use features available in new network technologies.

Keywords: Light-weight Protocols, U-Net Protocol, Myrinet Network.

1. Introdução

Um *cluster* de computadores [1] consiste de um conjunto de computadores completos e independentes interligados através de rede e utilizado como uma plataforma de computação unificada. Em outras palavras, o usuário desse tipo de sistema distribuído não tem noção que existem múltiplos processadores; tudo se passa como se fosse um único processador.

Devido ao fato dos computadores utilizados estarem interligados por rede local, a comunicação em um *cluster* é muito mais lenta que em sistemas de processamento maciçamente paralelos (MPP - *Massively Parallel Processing*). Essa característica dificulta a utilização dos *clusters* em aplicações que necessitam de uma rápida e confiável troca de informações para sincronizar operações ou compartilhar dados.

Este problema tem sido enfrentado com a utilização de redes locais de alta velocidade. Os desenvolvimentos recentes de redes como as redes padrões ATM (*Asynchronous Transfer Mode*) [2], *Fast* Ethernet [3] (IEEE 802.3u), Gigabit Ethernet [4] (IEEE 802.3z) e SCI [5] (IEEE 1596), e redes rápidas de solução proprietária, como a Myrinet (Myricom, Inc) [6], ServerNet [7] (Tandem Computers) e *Memory-Channel* [8] (Digital Equipment Corp.), contribuíram de forma significativa para isso.

No entanto, o incremento de largura de banda e a redução nos tempos de latência proporcionados pelas redes locais de alta velocidade não têm sido adequadamente aproveitados pelas aplicações que são executadas nos *clusters* de computadores.

O principal problema em relação à dificuldade de aproveitamento das características disponibilizadas pelas redes de alta velocidade reside no fato que todas as funções de comunicação com as redes locais encontram-se hoje concentradas no *kernel* do sistema operacional. O caminho percorrido por uma mensagem entre a aplicação e a interface de rede, dentro do *kernel*, envolve diversas cópias dessa mensagem nos múltiplos níveis de abstração

existentes. O *overhead* resultante desse processamento limita o máximo aproveitamento da largura de banda de comunicação e causa um aumento nos tempos de latência.

Diversos estudos têm sido realizados com a finalidade de reduzir os *overheads* de processamento e adaptar os protocolos de rede para aproveitarem melhor a maior largura de banda disponibilizada pelas redes de alta velocidade. Nessa linha, podemos destacar os trabalhos em Mensagens Ativas [9], Mensagens Rápidas [10], U-Net [11,12], *Sockets* Rápidos [13] e Interface BIP [14].

A maior parte desses trabalhos é baseada no desenvolvimento de protocolos leves de comunicação, implementados diretamente no espaço de usuário através de bibliotecas de comunicação. Com esse tipo de abordagem, conhecida como **biblioteca a nível de usuário** (*user-level interface library*), evita-se o tratamento convencional realizado pelos diversos níveis de *kernel* e *software* de protocolo de rede, obtendo-se uma melhoria considerável no desempenho das redes de alta velocidade.

2. Rede Myrinet e Arquitetura U-Net

2.1 Rede Myrinet

A rede Myrinet [15] é uma rede local comercial de alto desempenho, desenvolvida pela empresa Myricom, Inc., com características que a tornam bastante atraente para utilização em sistemas do tipo *clusters* de computadores e que necessitam de uma rede de alta velocidade para troca de mensagens. Entre suas principais características técnicas, podemos destacar: (1) canais de comunicação com controle de fluxo e controle de erro; (2) comutadores *cut-through* de baixa latência, (3) interface de rede que pode mapear a rede, selecionar rotas e também tratar o tráfego de pacotes e (4) *software* flexível que permite comunicação direta entre os processos de usuário e a interface de rede. Uma rede local Myrinet é composta de conexões ponto a ponto, *full-duplex* que interligam os computadores e comutadores. Os comutadores com múltiplas portas podem ser interligados diretamente tanto a outros comutadores como a interfaces de rede, em qualquer topologia.

2.1.1 Interface de Rede Myrinet

A Figura 2-1 apresenta o diagrama de blocos da interface de rede Myrinet e os detalhes internos da pastilha VLSI proprietária LanAI na qual a interface é baseada.

Essa micro-arquitetura provê uma interface flexível e de alto desempenho entre um barramento genérico chamado E-Bus e a conexão Myrinet. A memória SRAM é utilizada para armazenar o programa de controle Myrinet (MCP - Myrinet *Control Program*) e para armazenamento de pacotes. Encontra-se disponível uma área de memória de 32 *bit*, acessada duas vezes em cada período de relógio, uma vez pelo barramento E-bus e uma vez pelo processador ou pela interface de pacotes. Como os acessos originados pelo barramento E-Bus não são arbitrados, a pastilha LanAI se parece com uma memória síncrona e pode ser posicionada como um barramento escravo em qualquer barramento de 32-*bit*, tanto de memória como de E/S. Adicionalmente, quando a máquina DMA é utilizada para prover endereços, o LanAI pode atuar como um barramento mestre para transferir blocos de dados entre o barramento E-Bus e a memória SRAM. A máquina DMA também calcula o *checksum* Internet dos dados transferidos.

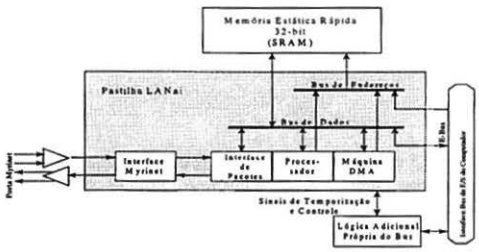

Figura 2-1 - Diagrama de Blocos da Interface de Rede Myrinet.

2.2 Arquitetura U-Net

A Arquitetura U-Net (*User-Level Network Interface*) [11], tem como idéia central retirar em parte ou completamente o processamento de protocolos do *kernel* e transferi-lo para a aplicação do usuário. A principal vantagem desse tipo de abordagem é permitir que cada aplicação possa controlar diretamente e de forma mais eficiente a utilização da rede. A Figura 2-2 ilustra a arquitetura básica dessa implementação.

A implementação desse tipo de arquitetura apresenta quatro aspectos fundamentais:

- multiplexação da rede entre diversos processos
- garantia de que diversos processos que utilizam a rede não interfiram entre si
- gerenciamento limitado de recursos de comunicação sem utilizar o *kernel*
- projeto de uma interface de programação eficiente e versátil com a rede

Outro ponto que merece destaque é o fato que a U-Net é uma arquitetura de comunicação que passa para o domínio da aplicação do usuário a ilusão que ela é a proprietária exclusiva da interface de rede, podendo usufruir o máximo desta interface, independente do tipo de interface utilizada.

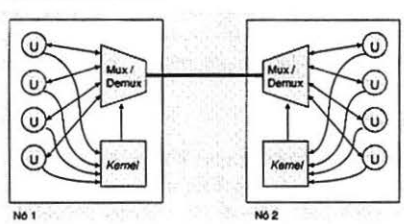

Figura 2-2 - Arquit. da Implementação U-Net

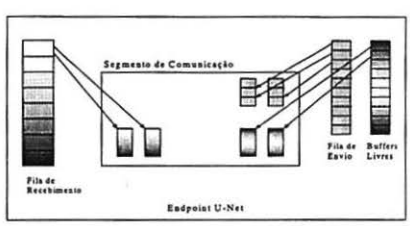

Figura 2-3 - Esquema de um *Endpoint* U-Net

2.2.1 Características do Projeto da Arquitetura U-Net

A arquitetura U-Net é composta de três blocos principais: (i) *Endpoints* que funcionam como um controle da aplicação para a rede, (ii) *Segmentos de comunicação* que são as regiões de memória que mantém o conteúdo das mensagens e (iii) as *filas de mensagens* que armazenam os descritores para as mensagens que são enviadas e recebidas.

Como se pode observar na Figura 2-3, os blocos (ii) e (iii) encontram-se contidos no bloco (i). Dessa forma, cada processo que deseja acessar a rede deve inicialmente criar um ou mais *endpoints*, e associar um segmento de comunicação e um conjunto de filas de mensagens enviadas, recebidas e livres.

Dependendo da profundidade da implementação e do hardware utilizado, os componentes U-Net manipulados pelos processos podem ser: (i) o próprio hardware utilizado na interface de rede, (ii) posições de memória que são interpretadas pelo sistema operacional ou (iii) uma combinação das duas características anteriores.

3. Objetivo

O objetivo deste artigo é apresentar os resultados da avaliação de desempenho da interface U-Net trocando informações entre dois computadores através de uma rede Myrinet. A análise é feita com a comparação dos resultados obtidos em três diferentes plataformas. Cada plataforma é formada por dois computadores iguais, interligados através de um comutador Myrinet. O que diferencia cada uma das plataformas é o fato dos processadores utilizados serem de diferentes tecnologias. Pretende-se com isso, identificar o impacto que a mudança de plataforma provoca sobre o desempenho do protocolo de comunicação. Em seguida, para uma das plataformas é avaliado o desempenho do protocolo U-Net no aspecto relativo ao suporte à concorrência de processos.

4. Análise de Desempenho

4.1 Ambiente de Teste

O ambiente de teste é formado por computadores Intel Dual Pentium II 333 MHz, Intel *Single* Pentium II 300 MHz e Intel Pentium Pro 200 MHz executando o sistema operacional Red Hat 5.0 e Linux *Kernel* 2.0.32. Os computadores estão equipados com placa de interface de rede Myrinet-SAN/PCI *Interface* modelo M2M-PCI32B com processador LanAI4.1 e 512 KB de memória, e interligados através de um comutador Myrinet de oito portas, modelo M2M-Dual-SW8 (*Dual 8 Port Myrinet SAN Switch*). Define-se assim, três plataformas:

a) Plataforma 1 - Formada pelos computadores que utilizam microprocessador Intel Dual Pentium II 333 MHz. Esses computadores receberam os nomes delta.spa e aurigae.spa.
b) Plataforma 2 - Formada pelos computadores que utilizam microprocessador Intel *Single* Pentium II 300 MHz. Esses computadores receberam os nomes sirius.spa e betelgeuse.spa.
c) Plataforma 3 - Formada pelos computadores que utilizam microprocessador Intel Dual Pentium Pro 200 MHz. Esses computadores receberam os nomes aquila19.spa e aquila20.spa.

4.2 Caracterização dos Computadores Utilizados

A caracterização dos computadores utilizados em cada plataforma foi feita com auxílio do *software lmbench* [16] e encontram-se apresentadas na Figura 4-1. Este *software* disponibiliza um conjunto de ferramentas para *micro-benchmarking*, permitindo realizar medidas de desempenho em uma determinada máquina e identificar os pontos críticos dos blocos componentes de um sistema. Os *benchmarks* enquadram-se em duas classes: largura de banda e latência. Os pontos críticos são reproduzidos nos *benchmarks* que medem a latência do sistema, a largura de banda de transferência de dados entre processador e cada um dos seguintes blocos: memória, rede e sistema de arquivo.

As medidas de largura de banda procuram mostrar o desempenho do sistema em relação a transferência de dados dentro da memória interna do computador. As medidas são feitas em relação aos seguintes aspectos: função *read()*, biblioteca *bcopy()*, função *bcopy()* manual, escrita e leitura direta na memória (sem cópia) através da interface *mmap()*, através de *pipes* e *sockets* TCP. Os *benchmarks* de largura de banda podem ser divididos em três tipos:

velocidade de memória (largura de banda de memória, *overhead* do sistema operacional (largura de banda IPC) e reutilização de *cache* (largura de banda de E/S com *cache*).

As medidas de latência procuram mostrar com que rapidez o sistema pode fazer uma determinada operação. Os tempos de latência avaliados são relativos aos seguintes itens: criação de processos, custo de acesso ao sistema operacional básico, chaveamento de contexto, comunicação entre processos, latências de sistema de arquivos e latências de mapeamento de memória.

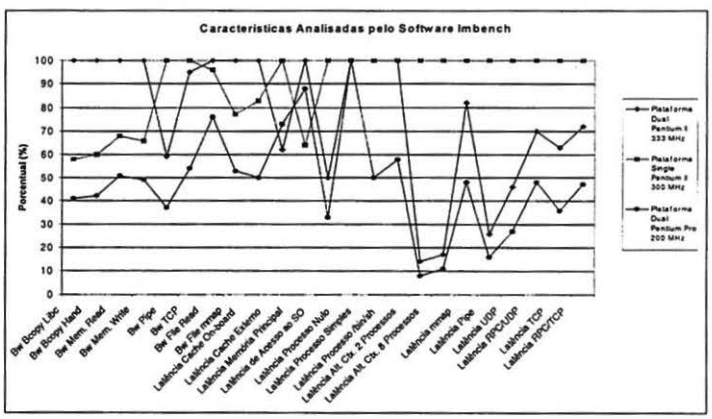

Figura 4-1 - Comparativo das Características Analisadas pelo Software lmbench em Relação à Plataforma de Melhor Desempenho.

4.3 Caracterização da Interação entre Computador e Interface de Rede

A avaliação específica do desempenho entre o computador e a interface de rede Myrinet foi feita com a utilização do programa de teste *hswap*, escrito pelo fabricante da interface de rede. Este programa disponibiliza valores de largura de banda para a transferência de mensagens entre a memória do computador e a memória da interface de rede, e da memória da interface de rede para a memória do computador, de dois modos diferentes: *Programmed I/O* (PIO) e *Direct Memory Access* (DMA).

No caso do modo PIO, o computador realiza a transferência de mensagens entre as regiões de memória através de instruções de seu próprio microprocessador. No modo DMA, o computador utiliza a máquina DMA existente na interface de rede, preocupando-se apenas em determinar os endereços de origem e destino da transferência e disparar seu início. Na seqüência são apresentados os resultados de largura de banda obtidos para cada uma das plataformas nos dois modos mencionados.

4.3.1 Avaliação dos Resultados

Os dados coletados para essa caracterização [17] mostram que para transferências de *bytes* com tamanho inferior a 128 *bytes*, o modo de transferência PIO é superior ao modo de transferência DMA, independente do sentido do deslocamento dos dados.

Nas transferências com tamanho maior que 128 *bytes*, o modo de transferência DMA leva vantagem em relação ao modo de transferência PIO, quando as transferências ocorrem da placa de rede para o computador. Quando o sentido de deslocamento dos dados é do computador para a placa de rede, o modo PIO é vantajoso para quantidade de *bytes* de até

410~800, dependendo da plataforma utilizada. Nas transferências de tamanho superior a 410~800 *bytes*, o desempenho do modo DMA é expressivo, alcançando valores de largura de banda maiores que 100 MBps para grandes quantidades de *bytes*, independente da plataforma utilizada.

4.4 Caracterização do Desempenho do Protocolo U-Net

O protocolo U-Net é utilizado através das primitivas de comunicação disponibilizadas por uma biblioteca, que deve ser montada juntamente com o programa de aplicação. Dessa forma o tráfego de comunicação da aplicação é feito pelos *endpoints* U-Net e não mais pelos caminhos convencionais do sistema operacional

O programa de teste utilizado para a avaliação de desempenho consiste em um módulo escrito em Linguagem C, capaz de transmitir e receber mensagens de tamanhos variados. Esse programa permite avaliar de forma bastante precisa os tempos envolvidos na transmissão das mensagens e, a partir do tempo obtido e dos tamanhos de mensagens enviados, apresentar valores de latência e largura de banda. Para cada uma das plataformas, o programa foi executado variando-se o tamanho das mensagens exponencialmente entre 2 e 4096. Os testes foram repetidos para cada uma das diferentes plataformas e no momento de suas realizações não existiam outros tipos de tráfego sendo desenvolvidos nas interfaces de rede e nem outras aplicações sendo executadas pelos computadores. Adicionalmente, para efeito de comparação, foi também realizada uma avaliação de desempenho dos computadores de uma das plataformas quando em comunicação através do protocolo TCP/IP, utilizando as interfaces Myrinet. Os dados para essa avaliação foram coletados com o *software* Netperf [18].

4.4.1 Avaliação dos Resultados

O primeiro aspecto a ser destacado é a confirmação que o desempenho do protocolo U-Net comparado com o protocolo TCP/IP através da rede Myrinet é bastante superior em termos de latência de mensagens, conforme observa-se na Figura 4-2. O protocolo U-Net é cerca de 20 vezes mais rápido para transmitir mensagens pequenas (32 *bytes*) e cerca de 5 vezes mais rápido para transmitir mensagens maiores (4096 *bytes*).

Em relação ao desempenho comparativo entre as diferentes plataformas, observa-se a partir dos dados coletados que, para tamanhos de mensagem variando entre 4 e 32 *bytes*, a diferença de latências entre a plataforma Dual Pentium Pro 200 MHz e a plataforma Dual Pentium II 333 MHz é da ordem de 1% e para mensagens maiores (4096 *bytes*) essa diferença não chega a 5%. A análise comparativa da plataforma com processador *Single* Pentium II 300 MHz deve ser vista de forma diferenciada, porque para pequenos pacotes de mensagens os tempos de latência de comunicação obtidos foram superiores àqueles das outras duas plataformas. A explicação para esse fato é creditada principalmente à latência de sua memória principal, que é quase 50% maior que as obtidas para as outras duas plataformas, conforme resultados do *benchmark* apresentado na Figura 4-1.

Embora a latência seja a principal preocupação em qualquer desenvolvimento de protocolos leves, a manutenção da largura de banda não pode ser esquecida. Em relação a esse aspecto, o protocolo U-Net consegue obter um desempenho de 458 Mbps para transferência de mensagens de 4096 *bytes* na plataforma Dual Pentium II 333 MHz. Esse valor corresponde a 43% do valor teórico da largura de banda disponibilizada pela rede Myrinet.

O desempenho do protocolo U-Net, comparado com o protocolo TCP/IP foi bastante superior, conforme pode-se observar na Figura 4-3. No caso de transferência de mensagens de pequenas, 100 *bytes*, o ganho de desempenho foi de cerca de 280 vezes e para mensagens maiores (4096 *bytes*), cerca de 2,5 vezes.

Como já ocorrido no caso da avaliação dos valores de latência, a diferença de desempenho entre as plataformas não foi muito significativa. Entre a plataforma Dual Pentium Pro 200 MHz e a plataforma Dual Pentium II 333 MHz, o ganho médio de desempenho foi da ordem de 3%, alcançando 5% para pacotes de 4096 *bytes*.

Observa-se também na Figura 4-3 que à medida que o tamanho das mensagens aumenta, os mecanismos de transferência de dados implementados pelo protocolo tornam-se mais eficazes. Isso ocorre porque o tempo efetivo utilizado na transferência das mensagem ultrapassa o tempo fixo de configuração dos mecanismos de transferência. Superado esse custo fixo observa-se um crescimento escalar bastante rápido.

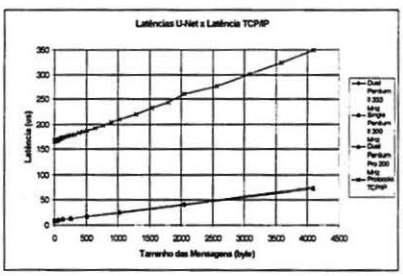

Figura 4-2 - Latências U-Net comparadas com a Latência TCP/IP.

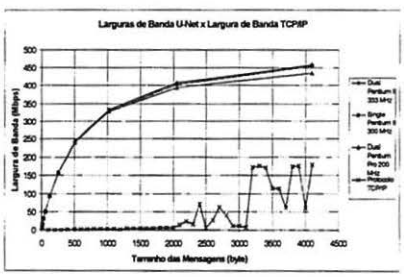

Figura 4-3 - Larg. de Banda U-Net comparada com a Larg. de Banda TCP/IP.

Considerando que os resultados alcançados nas diferentes plataformas foram muito semelhantes, reforça-se a idéia que o caminho crítico para a melhora de desempenho do protocolo U-Net em redes Myrinet encontra-se na interação entre o computador e a interface de rede e nos algoritmos de tratamento implementados no código dos protocolos leves.

4.5 Avaliação do Suporte à Concorrência de Processos

O suporte à concorrência de processos ou aplicativos é uma característica bastante desejada em qualquer sistema que envolva comunicação entre computadores. Esse tipo de característica permite que diversos processos sendo executados simultaneamente em uma mesma máquina possam transferir mensagens de forma simultânea para diferentes processos em outra máquina. Essa característica bastante comum nos protocolos de comunicação convencionais, ainda não encontra-se amplamente implementada na maioria dos protocolos leves. No caso do protocolo U-Net, um de seus grandes méritos é justamente a possibilidade de compartilhar a interface de rede entre diversos processos de forma simultânea, garantindo que um processo não interfira no outro. O protocolo U-Net utiliza diferentes *endpoints* para cada processo iniciado e na transmissão da mensagem associa um rótulo ao cabeçalho da mesma. Esse rótulo permite ao computador que recebe a mensagem identificar o *endpoint* de destino.

Com o intuito de analisar o comportamento do protocolo U-Net no aspecto de suporte à concorrência de processos foram realizados diversos testes para se determinar os valores de latência e largura de banda na plataforma Dual Pentium II 333 MHz. Esses testes foram realizados com os mesmos programas utilizados na avaliação do protocolo U-Net executado em processo único. A diferença nesse caso, é que cada um dos processos simultâneos transmite ou recebe mensagens através de uma porta de serviço diferente.

4.5.1 Análise dos Resultados

Em sistemas de alto desempenho, as aplicações típicas de concorrência procuram aproveitar as características SMP dos computadores utilizados, submetendo em cada computador um processo para cada processador. O estudo realizado utilizou computadores Dual Pentium II 333 MHz que dispunham de dois processadores cada um. Os gráficos apresentados da Figura 4-4 até a Figura 4-9 mostram o desempenho do protocolo U-Net para algumas quantidades de processos concorrentes, comparados com o valor de referência obtido para um único processo.

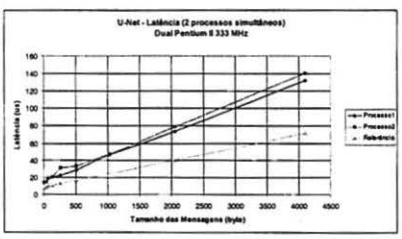

Figura 4-4 - Latências U-Net com 2 Processos Simultâneos.

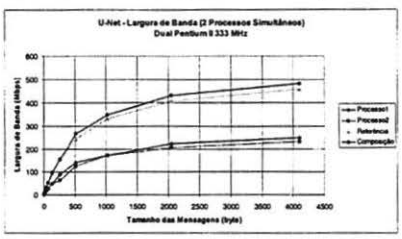

Figura 4-7 - Largura de Banda U-Net com 2 Processos Simultâneos.

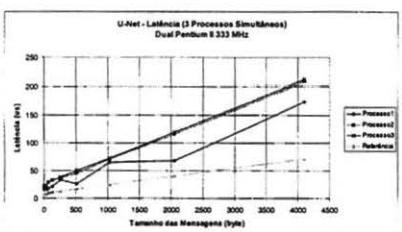

Figura 4-5 - Latências U-Net com 3 Processos Simultâneos.

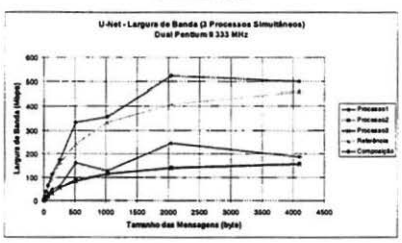

Figura 4-8 - Largura de Banda U-Net com 3 Processos Simultâneos.

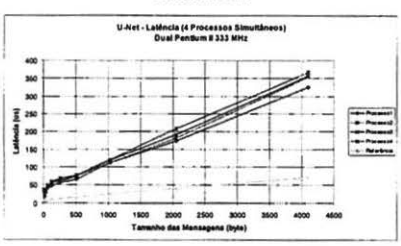

Figura 4-6 - Latências U-Net com 4 Processos Simultâneos.

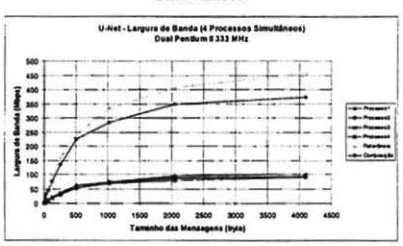

Figura 4-9 - Largura de Banda U-Net com 4 Processos Simultâneos.

No caso de dois processos simultâneos, as latência individuais de cada processo foram praticamente multiplicadas por dois em relação ao desempenho de um único processo e a largura de banda individual foi ligeiramente superior a metade da largura de banda de um único processo. O ganho total na largura de banda para dois processos simultâneos foi de 5%, alcançando uma largura de banda total de 481 Mbps para mensagens de 4096 *bytes*.

A avaliação de três processos simultâneos apresentou resultados semelhantes ao caso de dois processos, com as latências individuais multiplicadas praticamente por três e as larguras de

banda individuais de cada processo em torno de um terço do valor da largura de banda para um único processo. Novamente, a composição das três larguras de banda resultou em um desempenho da largura de banda total 9% superior a resultado de um único processo, alcançando praticamente 500 Mbps para mensagens de 4096 *bytes*.

A surpresa principal aconteceu em relação ao desempenho para quatro processos simultâneos, onde embora as latências tenham sido cerca de quatro vezes maior que a latência de um único processo, o desempenho da largura de banda individual de cada processo foi inferior a um quarto do valor da largura de banda de um único processo. Com isso a largura de banda resultante foi cerca de 18% menor que aquela obtida para um único processo, totalizando 374 Mbps para mensagens de 4096 *bytes*.

O resultado obtido não significa que o protocolo não suporte concorrência maior que três processos simultâneos. Os valores obtidos indicam que conforme as características de comportamento desejadas é necessário fazer um ajuste fino em relação à quantidade de total de descritores de comunicação previstos, bem como a quantidade total de memória alocada para a área de *buffer*. Esse ajuste deverá ser uma solução de compromisso entre a quantidade de memória alocada para a interface de comunicação e quantidade de processos simultâneos prevista para a plataforma.

4.6 Algumas Propostas para Melhoria de Desempenho

Um aspecto observado na implementação atual da biblioteca U-Net é que a mesma não faz uso do modo PIO na transferência de mensagens entre o computador e o LanAI e vice-versa, utilizando sempre o modo de transferência DMA. Conforme os resultados obtidos em nossa caracterização, para mensagens de tamanho menor que 128 *bytes*, o modo de transferência PIO é sempre mais vantajoso, permitindo um maior aproveitamento da largura de banda. Uma alteração bastante interessante seria a utilização dos dois modos de transferência de uma forma controlada, baseada no tamanho das mensagens. Uma idéia em estudo é incorporação de um pequeno *buffer* no descritor de transmissão para permitir a transferência de mensagens pequenas. Outro ponto que merece atenção é relativo à recepção de mensagens. A versão de implementação analisada não faz uso de mecanismos de interrupção do lado do computador. Em outras palavras, após a transmissão de uma mensagem, todo o processamento fica bloqueado até a recepção da resposta. Esse ponto deve ser necessariamente melhorado para se obter um melhor desempenho.

5. Conclusões

Esse estudo mostrou que a essência do problema de comunicação em sistemas de alto desempenho encontra-se em permitir que os processos e aplicações, executados pelos computadores, aproveitem de uma forma plena as características que são disponibilizadas pelas redes de alta velocidade. Em particular, deseja-se conseguir uma baixa latência de comunicação quando a rede transporta mensagens de pequeno tamanho, e uma largura de banda muito próxima dos limites teóricos da rede utilizada quando do transporte de mensagens grandes.

O protocolo U-Net foi implementado sobre três plataformas de computadores com processadores Intel Pentium de diferentes tecnologias e velocidades, interligados através de rede Myrinet, e avaliado em relação à características de latência e largura de banda para situações de tráfego entre dois computadores iguais. Desta avaliação, concluiu-se que no caso da combinação U-Net / Myrinet, pouco ganho de desempenho é alcançado com a mudança de plataforma e que a atenção para se conseguir qualquer melhoria, seja de latência ou largura de

banda, deve ser concentrada na melhoria dos mecanismos de tratamento da transmissão e recepção de mensagens, existentes entre a interface de rede e o computador onde a interface está instalada. No melhor caso estudado foi conseguido um resultado de latência da ordem de 7,6 μs para mensagens de 4 *bytes* e uma largura de banda de 458 Mbps para mensagens de 4096 *bytes*.

Outro aspecto avaliado foi o suporte a concorrência de processos oferecido pela interface U-Net. Verificou-se que em relação aos tempos de latência, os valores obtidos para cada processo foram praticamente multiplicados pela quantidade de processos existentes, enquanto que a largura de banda resultante da soma individual dos processos aumentou 5% e 9% para os casos de 2 e 3 processos, respectivamente. No caso onde foram avaliados quatro processos simultâneos, a largura de banda resultante diminui 18%. Os estudos realizados no código da implementação mostram que o mesmo não foi suficientemente ajustado para mais que dois processos simultâneos, necessitando de uma reavaliação em relação a tamanho de *buffers* e áreas de memória alocadas.

6. Referências

[1] PFISTER, G.F. **In Search of Clusters.** 2 ed., Upper Saddle River, Prentice Hall PTR, 1998.

[2] PRYCKER, M. **Asynchronous Transfer Mode:** Solution for Broadband ISDN. 3. Ed., London, Prentice Hall, 1995.

[3] STALLINGS, W. **High-Speed Networks:** TCP/IP and ATM Design Principles. Upper Saddle River, Prentice Hall, 1998.

[4] GIGABIT **Gigabit Ethernet** Gigabit Ethernet Alliance, August, 1996. (White Paper).

[5] GUSTAVSON, D.B. The Scalable Coherent Interface and Related Standard Projects. **IEEE Micro**, v.12, n.1, p.10-22, February, 1992.

[6] SEITZ, C. Myrinet: A Gigabit-per-second Local Area Network. In: HOT INTERCONNECTS II, 1994, Stanford University, Stanford, CA, August, **Proceedings**.

[7] HORST, R.W. Tnet: A Realiable System Area Network. **IEEE-Micro**,v.15, n.1, p.37-45, February 1995.

[8] GILLETT, R.B. Memory Channel network for PCI. **IEEE Micro**, v.16, p.12-18, February, 1996.

[9] von EICKEN, T; CULLER, D.E.; GOLDSTEIN, S.C.; SCHAUSER, K.E. Active Messages: a Mechanism for Integrated Communication and Computation. In: ANNUAL INTERNATIONAL SYMPOSIUM ON COMPUTER ARCHITECTURE, 19., Gold Coast, Australia, May, 1992. **Proceedings**. 1992. p.256-266.

[10] PAKIN, S.; LAURIA, M., CHIEN, A. High Performance Messaging on Workstations: Illinois Fast Message (FM) for Myrinet. In: SUPERCOMPUTING '95, San Diego, California, 1995. **Proceedings.**

[11] von EICKEN, T.; BASU, A.; BUCH, V.; VOGELS, W. U-Net: A User-Level Network Interface for Parallel and Distributed Computing. In: ACM SYMPOSIUM ON OPERATING SYSTEMS PRINCIPLES, 15., Copper Mountain, Colorado, December, 1995. **Proceedings**. 1995. p.40-53.

[12] BASU, A.; WELSH, M.; von EICKEN, T. Incorporating Memory Management into User-Level Network Interfaces. In: ANNUAL INTERNATIONAL SYMPOSIUM ON COMPUTER ARCHITECTURE, 24., 1997. **Proceedings**.

[13] RODRIGUES, S.H.; ANDERSON, T.E.; CULLER, D.E. High-Performance Local Area Communication With Fast Sockets. In: USENIX '97, 1997. **Proceedings**.

[14] PRYLLI, L.; TOURANCHEAU, B. **BIP:** a new protocol designed for high performance networking on Myrinet. Lyon, Ecole Normale Supérieure de Lyon, Laboratoire de l'Informatique du Parallélisme, September, 1997.

[15] BODEN, N.J; COHEN, D.; FELDERMAN, R.E.; KULAWIK, A.E.; SEITZ, C.L.; SEIZOVIC, J.N.; SU, W.K. Myrinet – A Gigabit-per-Second Local-Area Network. **IEEE-Micro**, v.15, n.1, p.29-36, February, 1995.

[16] McVOY, L.; STAELIN, C. lmbench: Portable Tools for Performance Analysis. In: USENIX ANNUAL TECHINICAL CONFERENCE, San Diego, California, 1996. **Proceedings**.

[17] GEROMEL, P.A. **Protocolos Leves de Comunicação Para Sistemas de Alto Desempenho.** São Paulo, 1998. 113p. Dissertação (Mestrado) - Escola Politécnica, Universidade de São Paulo.

[18] HEWLETT-PACKARD **Netperf:** A Network Performance Benchmark. Revision 2.1, Hewlett-Packard Company, Information Networks Division, February, 1996.