Comunicação Fim-a-Fim a Alta Velocidade em Redes Gigabit

**Danilo M. Taveira, Igor M. Moraes, Daniel de O. Cunha,
Rafael P. Laufer, Marco D. D. Bicudo, Miguel E. M. Campista,
Aurelio Amodei Junior e Otto Carlos M. B. Duarte***

[1]Grupo de Teleinformática e Automação
PEE/COPPE-DEL/POLI
Universidade Federal do Rio de Janeiro

Resumo. *Com o aumento na largura de banda disponível e a possibilidade de transmissão de dados a Gigabits por segundo, os discos rígidos passam ser o principal gargalo em comunicações a altas taxas. Como solução, é mostrado que a utilização de diversos discos em paralelo possibilita taxas compatíveis com as exigências atuais das redes de comunicação.*

1. Introdução

Os enormes avanços tecnológicos obtidos na área de integração de circuitos permitiram ganhos extraordinários na capacidade de processamento dos computadores e na velocidade das redes de computadores. Embora uma rede com capacidade para transmitir a Gigabits por segundo permita um caminho de alta velocidade entre dois pontos, não é trivial o uso eficaz de toda esta taxa de transferência por uma aplicação. Alguns cuidados na escolha dos equipamentos de extremidade e na configuração dos protocolos a serem usados são essenciais para atingir as altas taxas de transmissão oferecidas.

Atualmente, um computador pessoal possui um alto poder de processamento com relógios da ordem de gigahertz. O barramento de memória, responsável pela comunicação direta entre o processador e a memória principal já possui relógios com centenas de megahertz. Desta forma, este barramento não representa mais um gargalo significativo no serviço a altas taxas. Os discos, entretanto, continuam sendo dispositivos com partes mecânicas, incapazes de acompanhar a velocidade dos componentes puramente eletrônicos.

A tecnologia RAID (*Redundant Array of Inexpensive Disks*) foi proposta inicialmente por Patterson *et al.* em 1988 [Patterson et al., 1988].No RAID nível 0, os dados são distribuídos uniformemente entre dois ou mais discos. Esse nível de RAID aumenta a taxa agregada obtida ao permitir que requisições de leitura e escrita aos discos sejam atendidas quase que simultaneamente.Entretanto, no RAID nível 0 nenhuma redundância é empregada.

O artigo está organizado da seguinte forma. A Seção 2 descreve os procedimentos adotados para a realização dos testes e os equipamentos envolvidos. Detalhes referentes à análise dos resultados são discutidos na Seção 3. Por fim, na Seção 4 são apresentadas as conclusões.

2. O Ambiente de Testes

Para avaliar o desempenho dos usuários finais em uma comunicação a altas taxas, desenvolveu-se um ambiente de testes. O ambiente é formado por dois computadores pessoais providos de interfaces de rede Gigabit Ethernet e interconectados através de um

*Este trabalho foi realizado com recursos do CNPq, CAPES, FAPERJ, FINEP, RNP e FUNTTEL.

comutador. Mais detalhes sobre a especificação de cada um dos equipamentos usados são apresentados na Tabela 1. A configuração do ambiente e dos equipamentos foi concebida para permitir a realização de testes com componentes da arquitetura de um computador pessoal, como as interfaces de rede e os discos rígidos, e também com os protocolos usados na comunicação.

Tabela 1: Especificação dos equipamentos usados no ambiente de testes.

Equipamento	**Especificação**
Computadores A e B	Processador Intel Pentium IV com relógio de 3,2 MHz
	Placa mãe Intel SE7210TP1-E com interface de rede Gigabit Ethernet integrada
	Memória RAM Kingston DDR 400 MHz Dual-Channel, 1 GB
	Disco rígido Seagate SCSI-320 10000 rpm (36GB)
	Sistema operacional Debian Linux 3.1 com núcleo 2.4.20
Comutador	Comutador D-Link DGS-3308TG com 6 portas Gigabit Ethernet

Os testes realizados envolvem a verificação experimental da vazão máxima obtida em uma rede Gigabit e a determinação do gargalo da comunicação a altas taxas. Em um primeiro cenário, busca-se determinar a taxa máxima de transmissão de dados entre os dois computadores. Os testes usam o protocolo de transporte UDP (*User Datagram Protocol*). O tráfego é gerado com a ferramenta `Iperf` [Iperf, 2005].

Além disso, busca-se analisar situações onde o objetivo é servir conteúdo armazenado em disco a taxa de um Gigabit por segundo. Nestes experimentos, um computador é usado como servidor de conteúdo a altas taxas. A especificação deste computador é a mesma apresentada na Tabela 1. O objetivo dos testes é comprovar que um computador provido de apenas um disco rígido não consegue transferir dados a um Gigabit por segundo. A partir daí, busca-se mostrar que o RAID nível 0 é uma alternativa para garantir a transferência de dados a altas taxas.

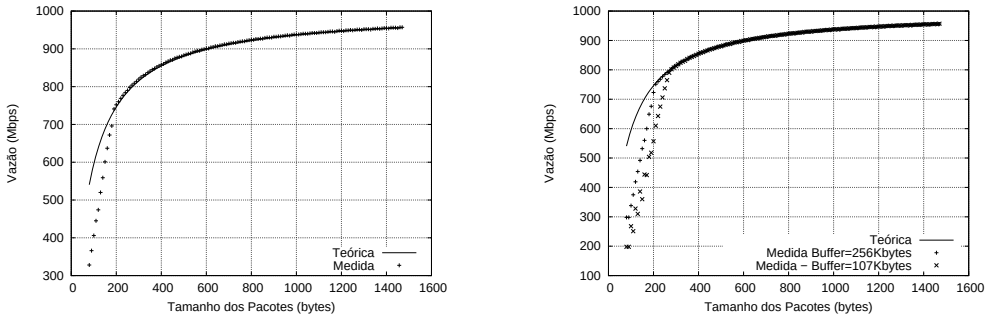
Para determinar a capacidade de transferência dos discos rígidos são consideradas três situações. Na primeira, o computador possui apenas um disco. Na segunda situação, considera-se que o computador está equipado com dois discos operando em RAID, nível 0. Por fim, na terceira o computador trabalha com três discos em RAID0. A ferramenta `IOMeter` [IOMeter, 2005] é usada nos testes com os discos rígidos

Nos testes também são analisados dois modos de implementação do RAID nível 0. O RAID implementado por software e o RAID híbrido. A implementação por software é em geral mais flexível do que a implementação que utiliza placas controladoras dedicadas. No entanto, o preço pago pela flexibilidade é o maior consumo de processamento, já que mais operações precisam ser feitas a cada leitura ou gravação. O suporte ao RAID é nativo no núcleo 2.4.20 do Linux. Entretanto, é necessário usar um conjunto de ferramentas, chamado de `Raidtools`, para criar a matriz de discos. A implementação híbrida está disponível na placa controladora RAID integrada à placa mãe do computador usado nos testes. No RAID híbrido, a controladora só implementa por hardware as funções básicas de acesso ao disco que são usadas pelas interrupções da BIOS. Estas interrupções permitem que todos os sistemas operacionais acessem o disco de uma forma padronizada. Só após acessarem os discos, os sistemas operacionais acessam diretamente a controladora RAID. Além disso, após o sistema operacional carregar o *driver* da controladora, todas as funções de RAID são implementadas por software. Dessa forma, a implementação híbrida pode ter um desempenho similar à implementação por software, ou até mesmo pior, caso o *driver* não seja otimizado.

3. Resultados

Inicialmente mediu-se a taxa máxima de transferência de dados obtida com o uso da rede Gigabit. O protocolo de transporte UDP é usado neste teste e os dados a serem transmitidos são lidos diretamente da memória RAM.

A Figura 1(a) mostra a taxa de dados obtida em função do tamanho do pacote enviado. Comparando a curva experimental com a curva teórica, nota-se que o computador consegue transmitir os dados em uma taxa próxima da taxa máxima teórica, a partir de um dado tamanho de pacote. Para pacotes abaixo de 200 bytes, é possível notar que a diferença é bastante expressiva. Esta queda ocorre por que é necessário um tempo mínimo para que cada pacote possa ser gerado.



(a) Taxa máxima de transmissão. (b) Taxa máxima de recepção.

Figura 1: Taxa de transferência da rede em função do tamanho do pacote.

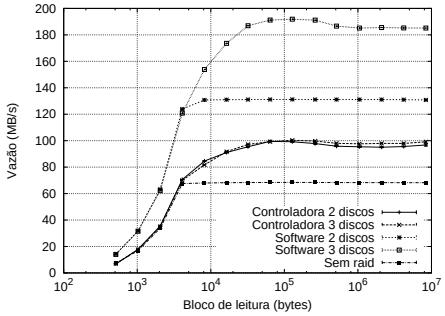
Pode-se observar na figura 1(b) que a taxa de dados recebida é menor que a taxa de dados transmitida. Esse fato indica que o tempo necessário na recepção de um pacote é superior ao tempo necessário para a sua geração. Isto se comprova com uma análise mais atenta da Figura 1(b), onde se pode perceber que o aumento do *buffer* do receptor diminui este efeito. Os resultados demonstram, assim, que o processamento e o barramento não são gargalos significativos na comunicação em alta velocidade, desde que os dados não sejam enviados em pacotes muito pequenos.

Após a analise da transmissão a altas taxas de dados lidos diretamente da memória, são realizados testes com o intuito de verificar o impacto do acesso ao disco no serviço de dados. O número de discos e o tipo de implementação do RAID nível 0 são variados. Também são avaliadas as implementações de RAID por software e híbrida, implementada pela controladora.

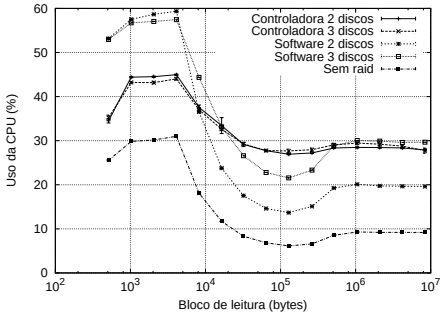
A Figura 2 mostra os resultados experimentais para a taxa de leitura fornecida pelos discos rígidos ao lerem diferentes tamanhos de blocos de dados seqüenciais. Dados seqüenciais são aqueles que estão armazenados de forma contígua no disco rígido.

Pode-se ver na Figura 2(a) que o RAID0 usando a controladora existente na placa mãe é menos eficiente para a leitura do que, o implementado por software. O melhor desempenho do RAID0 por software indica que os seus *drivers* são mais otimizados. De acordo com resultados obtidos, a vazão máxima dos discos só é alcançada quando são lidos blocos grandes de dados. Isso ocorre, pois para blocos pequenos, a probabilidade dos dados estarem em apenas um disco é grande, impossibilitando acessos simultâneos. Na Figura 2(b) é mostrado o uso da CPU durante o teste. Para blocos de leitura grandes, a diferença entre os resultados obtidos pelas duas implementações é significativamente menor.

Para analisar a velocidade de escrita em disco, são realizados testes semelhantes aos descritos anteriormente para a leitura. A Figura 3 mostra os resultados destes testes.

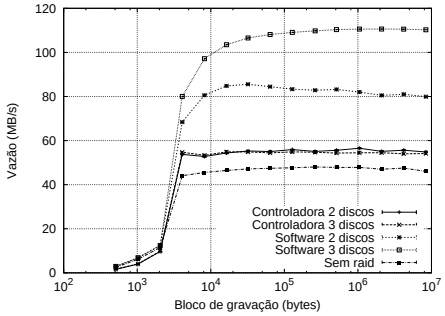


(a) Taxa de leitura.

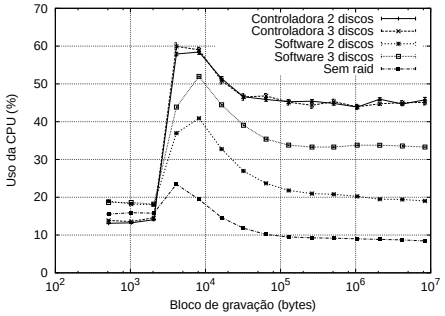


(b) Uso da CPU durante os testes.

Figura 2: Leitura de blocos seqüenciais.



(a) Taxa de escrita.



(b) Uso da CPU durante os testes.

Figura 3: Escrita de blocos seqüenciais.

No teste de escrita, é possível observar novamente que o desempenho do RAID0 implementado por software foi superior ao híbrido. A Figura 3(b) mostra o uso da CPU durante o teste de escrita seqüencial. Nessa situação, o uso da CPU pela implementação por software é menor se comparado ao uso da CPU usando-se o RAID0 da controladora. Os resultados mostram também que o custo da operação de escrita dos dados é maior que o de leitura, já que a capacidade máxima de escrita é menor do que a de leitura. Também é possível observar que mesmo com três discos em RAID0, não é possível obter uma taxa de 1 Gbps.

4. Conclusão

Este artigo analisou as principais limitações às comunicações em altas taxas impostas pela atual arquitetura dos computadores pessoais.

O resultados demonstram que o nó emissor é capaz de ocupar todo o meio de acordo com o tamanho dos pacotes utilizados. Foi também demonstrado que os discos rígidos constituem o principal gargalo nessas comunicações. Os testes demonstraram que o uso do RAID nível 0 consegue atender às exigências de comunicações a altas taxas e que a forma com que ele é implementado impacta significativamente os resultados.

Referências

IOMeter (2005). IOMeter. `http://www.iometer.org/`.

Iperf (2005). Iperf - The TCP/UDP Bandwidth Measurement Tool. `http://dast.nlanr.net/Projects/Iperf/`.

Patterson, D. A., Gibson, G. e Katz, R. H. (1988). A case for redundant arrays of inexpensive disks. Em *ACM SIGMOD'88 - International Conference on Management of Data*, páginas 109–116.