

AVALIAÇÃO DO DESEMPENHO DAS TRANSFERÊNCIAS DE ARQUIVO
E PROCESSAMENTO DE TRANSAÇÕES NAS REDES LOCAIS

Claudio Cesar de Sá (1)
DMAE/CCT/UECe

Marcos Antonio Gonçalves Brasileiro (2)
DEE/GRC/UFPb

José Antão Beltrão Moura (3)
DSC/GRC/UFPb

Resumo:

Este trabalho investiga a integração de serviços em Redes Locais de Computadores, de duas importantes aplicações. Resultados finais de desempenho para as aplicações são apresentados quando submetidas à variação de duas velocidades do meio físico de transmissão, analisando assim o comportamento desta mistura de tráfegos distintos na sub-rede.

1 - Introdução

A tecnologia das Redes Locais (RLs) tem despertado interesse em diversos segmentos da informática atual. O espectro do uso das RLs é dos mais amplos, seguindo da automação de escritórios, ambientes de desenvolvimento de software à automação industrial. As RLs se prezam pelo suporte na agilização do sistema de informação das empresas em seu processo de tomada de decisão, utilizando-se de serviços computacionais distribuídos a nível de software e hardware. São alguns exemplos destas aplicações disponíveis aos usuários: planilhas e correio eletrônico, transferências de documentos ou créditos/débitos em contas correntes, sistemas de reservas em geral (alocação de recursos), interações com um sistema operacional remoto, ... etc. Há aplicações de uso mais específico, como é o caso do compartilhamento de uma impressora "laser" remota para o novo logotipo das cartas expedidas pela empresa, ou um servidor de arquivos que disponha de áreas reservadas em discos "winchesters" ou óticos, para trabalhos de paginação na memória principal para uma estação remota de trabalho.

Diante desta gama de aplicações estruturadas sobre RLs, tráfegos característicos são produzidos na sub-rede de comunicação. Dois tipos importantes de tráfegos são observados nas RLs, estes são gerados a partir das seguintes aplicações: Transferências de Arquivos (TAs) e Processamento de Transações (PTs). Estas aplicações serão postas à prova para duas velocidades de sub-rede, 100 Kbps e 1 Mbps.

Até então, estudos de avaliação e desempenho de aplicações em RLs, se restringiam às aplicações homogêneas [1,2]. Ao investigarmos estações de trabalho de usuários em TAs e PTs interagindo com um Servidor de Ar

quivos (SA) remoto através da RL, lançamos mão de uma técnica para analisar este sistema denominado de Modelagem Hierárquica (MH) [3].

Na seção 2 é descrito o ambiente de estudo, a operação destes dois serviços baseados em um SA remoto, e o modelo hierárquico empregado na análise. Em razão da quantidade de detalhes envolvidos nestes modelos, estes serão descritos resumidamente na seção 3. Na seção 4, os resultados finais (tempos de resposta a nível de usuário final) para TAs e PTs são obtidos para suas velocidades de sub-rede.

2 - Ambiente, Aplicações e Modelo Hierárquico

2.1 - Ambiente de Estudo

O sistema a ser estudado é mostrado na figura 1. Este ambiente constitui-se de duas classes de usuários (uma classe em TAs, e a outra em PTs), estes interagem com um Servidor de Arquivos (SAs) remoto através da sub-rede de comunicação. O SA é um sistema que gerencia o uso de discos rígidos, bem como os acessos de pedidos de serviços dos usuários às aplicações residentes neste servidor. Tratando-se de projetos em RLs, os SAs apresentam algumas vantagens no tocante a: compartilhamento equânime das informações, manutenção e gerenciamento de dados, crescimento modular, custo reduzido por usuários ligados ao SA, maiores detalhes ver referência [4].

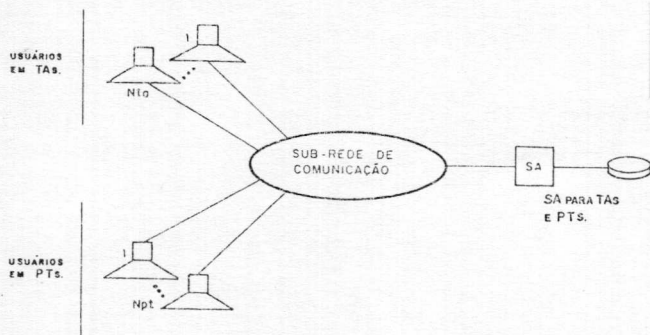


Fig. 1 - Ambiente de Estudo
(Usuários de TAs e PTs, sub-rede e SA)

2.2 - As Aplicações

As duas aplicações (TAs e PTs) a serem apreciadas, são eleitas pela razão de apresentarem um fluxo típico de pacotes trafegando na sub-rede, que representam muitas e significativas aplicações existentes para RLs. Estas aplicações podem se apresentar como um serviço de suporte às tarefas específicas do usuário, ou como um serviço de uso final por parte da estação de trabalho. Por exemplo, a recuperação de um texto "longo" perdida pelo usuário junto ao SA, o processo ocorre como se uma transferência de arquivo típica estivesse sendo realizada, um transporte unidirecional e contínuo de "grandes" pacotes de dados (o tamanho dos pacotes de dados a serem trafegados na sub-rede é determinado pelo protocolo de rede). Observa-se que as TAs estão presentes no encaminhamento deste texto, como um serviço de auxílio direto a esta aplicação. Com isto, caracteriza-se uma TA como sendo uma sequência limitada de pacotes encaminhando-se unidirecionalmente entre uma fonte e um sorvedouro, que são as Estações de Trabalho do Usuário (ETs) e o SA.

Por outro lado, PTs relacionam-se com transações interativas simples entre as ETs e o SA remoto. A característica de uma transação, é um simples pacote de pergunta ao SA, um processamento neste recurso, e um outro pacote de retorno como uma resposta. O tamanho destes dois tipos de pacotes é da ordem de algumas centenas de bits, que correspondem entre 5 e 20% do tamanho dos pacotes de dados das TAs. Exemplos deste serviço: consultas bancárias em geral, interações com o sistema operacional do SA remoto, consultas à itens de arquivos, início e término de sessões de sistemas, "passwords", ... etc, são aplicações que se acomodam na classe de PTs. Vale a observação feita para TAs, o PT pode se apresentar como uso final ao usuário, ou estar encapsulado a serviços mais sofisticados ou específicos do usuário.

2.3 - O Modelo Hierárquico

Após a sucinta exposição do ambiente e aplicações de estudo, temos de encontrar uma técnica apropriada para analisar este sistema. Tendo o ambiente uma população máxima de N_{ta} usuários em TAs e N_{pt} usuários em PTs, considera-se este sistema como sendo fechado para efeito de análise, uma consideração inicial importante para elaboração de um modelo tratável.

A metodologia aqui empregada é a Modelagem Hierárquica (MH), também chamada de Decomposição [3]. A idéia da MH consiste em dividir o sistema completo em subsistemas (níveis, submodelos) mais simples e de menor complexidade. Feita esta divisão, aplica-se um método de solução mais ajustado e eficiente a cada um dos níveis do modelo, encontrando-se resultados de cada nível que representam as partes internas do sistema, e seguindo uma composição adequada das relações entre os níveis, obtém-se medidas finais de interesses do modelo completo.

Por se tratar de um sistema fechado, as variáveis internas ou de controle estão relacionadas (ligadas, coesas) entre si. Sendo assim, o processo da divisão do modelo em partes menos complicadas, é de grande relevância no emprego da MH. A separação das partes funcionais mais proeminentes do modelo, é um indicativo a ser seguido, apesar da questão "arte de modelar sistemas". Portanto, na abstração do sistema segundo sua funcionalidade interna, é a idéia sugerida na divisão do modelo em submodelos.

No tocante ao processo de resolução do modelo completo, o caminho seguido é o "bottom-up". Resolve-se o nível mais inferior (nível 1, ou inicializador), que não depende de nenhum resultado anterior; seus resultados são os parâmetros de relação entre os submodelos) são passados como entradas ao nível seguinte. Resolvendo o nível 2 com uma técnica apropriada às características deste nível, os resultados obtidos são passados ao nível seguinte, e assim sucessivamente até o nível mais superior. Resultados deste nível são as medidas (valores, resultados) procuradas sobre o modelo completo do sistema. Neste trabalho será feita uma comparação entre medidas finais a nível de usuário, isto é: $t_3^1(j, N_{ta})$ --> tempo médio de resposta para uma transferência de arquivo completa, $t_2^2(i, j, W)$ --> tempo médio de resposta para uma transação completa.

Para o sistema em análise, é construído um Modelo Hierárquico em três níveis, capaz de captar as principais funções internas mais relevantes da operacionalidade do ambiente em questão. Este modelo é representado esquematicamente na figura 2. Os trabalhos das referências [1,2], a técnica da MH é também utilizada.

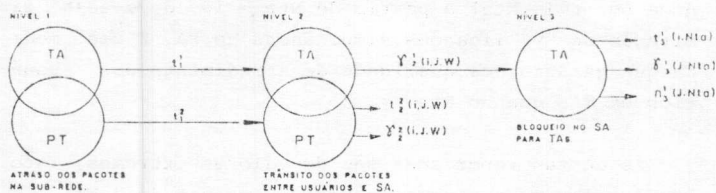


Fig. 2 - Representação do Modelo Hierárquico

Na figura 2, os círculos figuram simbolicamente as aplicações TAs e PTs. Observa-se que apenas 2 níveis são suficientes para analisar PTs, enquanto que às TAs precisa-se de 3 níveis. Destemodo, as medidas de interesse final para PTs são obtidas no nível 2, e para TAs no nível 3. As setas entre os círculos (níveis), mostram as interfaces e parâmetros passados e recebidos entre dois níveis adjacentes. A interação entre dois círculos num mesmo nível, destaca uma solução integrada para as duas aplicações. Quanto à notação usada temos que os índices dos parâmetros representam o seu nível de origem, e o expoente define a aplicação. Por convenção: TAs é a aplicação um (1), e PTs é a aplicação dois (2).

3 - Descrição dos Modelos

Na seção anterior argumentou-se sobre o modelo hierárquico construído em três (3) níveis para resolução do sistema em estudo. Em virtude da quantidade de detalhes e considerações matemáticas a serem feitas a cada um dos níveis do modelo, decidiu-se em omitir esta descrição extensiva neste trabalho, indicando a referência [5] ao leitor interessado. Deste modo, optou-se por uma abordagem resumida a respeito das funções de cada nível do modelo.

O primeiro nível retrata o atraso que os pacotes sofrem na sub-rede de comunicação. O atraso considerado é o tempo de transferência dos pacotes no meio físico, e o tempo de espera para o acesso a este recurso compartilhado. Há pacotes de PTs e TAs presentes na sub-rede, portanto, tempos de transmissões distintos para os pacotes destas duas aplicações heterogêneas. Por se tratar de um sistema fechado e conseqüentemente o número de pacotes é invariável (pois cada usuário ativo só pode realizar uma tarefa por vez, ou está em uma TA, ou em PT), ocasionando uma dependência probabilística do número de pacotes presentes nas ETs, ou sub-rede, ou Sa. Desta forma, a solução deste nível ficou na estimativa dos tempos de transmissões dos pacotes de TAs e PTs na sub-rede, sem poder calcular os tempos de filas ou acessos desta entidade neste nível.

No segundo nível, há um tratamento do trânsito de pacotes entre as ETs e o SA via sub-rede. O encaminhamento

destes pacotes entre estas duas entidades (ETs e SA), é modelado através de um sistema de múltiplas cadeias de filas fechadas, com uma população fixa dos pacotes de TAs e PTs em circulação. Neste nível será possível encontrar o atraso fim-a-fim, isto é, o tempo de travessia para os dois tipos de pacotes entre o SA e as ETs. A técnica empregada neste nível é a simulação discreta, para isto foi construído um simulador na linguagem GPSS [5].

Calculando-se o tempo dispendido por um pacote de PT realizar uma viagem de ida e volta ao SA a partir da ET, temos o tempo médio de resposta de uma transação $t_2^2(i,j,w)$, que por sua vez é uma medida final a nível de usuário. O nível 2 faz considerações a respeito dos atrasos que os pacotes sofrem nas ETs, e quanto à disputa do acesso ao SA. Quanto às TAs, é calculado o tempo de transferência entre SA - ETs de um certo número médio de pacotes por arquivo (np). Por razões típicas das TAs, é imposto um limite máximo de conexões simultâneas (y) entre as ETs e o SA, este problema é abordado no nível 3.

Caso o número de usuários ativos em TAs (Nta) seja maior que "y", um bloqueio ocorrerá no SA no momento que este limite for atingido. Esta "perda" de novos pedidos para TAs por parte das ETs, independe do número de usuários ativos em PTs (Npt). Esta rejeição no SA a partir de "y" transferências presentes imposta sobre os Nta usuários, é analisada através de um modelo analítico de filas fechadas com população finita igual a Nta, e "y" servidores em paralelo. Utilizando as Cadeias de Markov encontra-se uma expressão para o tempo médio de resposta de uma TA $t_3^1(j,Nta)$. Outras medidas de interesse podem ser obtidas para TAs e PTs na análise dos níveis, mas para o estudo proposto o "tempo médio final das aplicações" é suficiente. A exposição dos resultados feita na seção 4, é segundo um conjunto de valores atribuídos aos parâmetros do modelo hierárquico.

4 - Resultados

Alguns problemas são pertinentes na escolha de valores aos parâmetros do modelo, maiores detalhes desta discussão [5]. O critério de atribuição destes valores teve um caráter prático, ou seja, um ambiente operacional segundo o conhecimento da tecnologia presente. As velocidades do meio físico (C) que serão comparadas são iguais a 100 Kbps e 1 Mbps. Os demais valores são os seguintes:

- Tamanho médio de um pacote de TAs: 8000 bits.
- Tamanho médio de um pacote de PTs: 500 bits.
- Atraso médio por pacote no servidor da ET (TA ou PT): 200 ms.
- Atraso médio por pacote no servidor do SA (TAs): 50ms.
- Atraso médio por pacote no servidor do SA (PTs): 30ms.
- Número médio de pacotes por arquivo (np): 40.
- Número máximo de conexões de TAs no SA (y): 10.

- Tamanho da janela para controle de fluxo em TAs (W): 1 e 4.
- Tempo médio de pensamento para usuários em TAs: 5 segundos.
- Tempo médio de pensamento para usuários em PTs: 0 - *
- Intervalo de confiança para simulação discreta: 10%.

(*) - Este valor foi adotado por três razões:

- 1 - Sobrecarregar o sistema com este tipo de pacote.
- 2 - Considerando um número fixo de viagens deste pacote, podemos caracterizar tipos de transações em particulares.
- 3 - Podemos considerar novos usuários presentes, mas com "think time" diferente de zero.

As figuras 4 e 5 quantificam os valores de $t_2^2(i, j, W)$ para C igual a 100 Kbps e 1 Mbps respectivamente. Apesar das figuras não serem muito claras na escala em que estão reportadas, o crescimento das curvas é exponencial. Melhores valores para $t_2^2(i, j, W)$ foram obtidos com C igual a 1 Mbps, é claro que ao aumentarmos a disponibilidade de um recurso físico, um melhor desempenho é registrado.

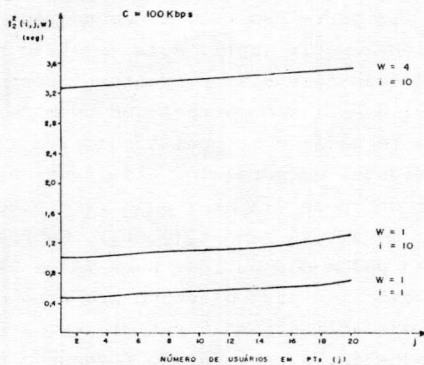


Figura 3

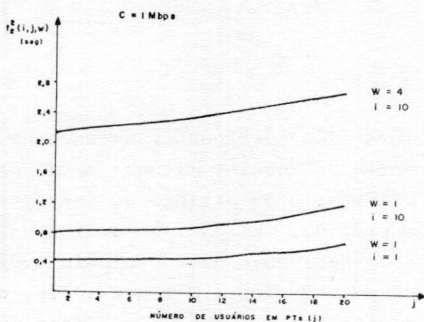


Figura 4

Observa-se também que o número de usuários presentes em TAs (i) e o tamanho da janela (W), ocasionam um sensível aumento de pacotes presentes no sistema, implicando na queda dos níveis de $t_2^2(i, j, W)$. Esta concorrência em igualdade aos recursos compartilhados do sistema (SA e sub-rede), resulta numa degradação generalizada para TAs e Pts.

Passando ao nível 3 do modelo, obtemos $t_3^1(j, Nta)$

nas figuras 5 e 6 para C igual a 100 Kbps e 1 Mbps respectivamente. Nestas duas figuras a estabilidade atinvida por $t_3^1(j, Nta)$ a partir de $Nta \approx 10$, deve-se a limitação de "y" ligações simultâneas no SA. É uma medida que garante uma qualidade de atendimento aos usuários em TAs quando $Nta > y$.

As curvas reportadas são de valores extremos, isto é, os melhores e piores resultados para $t_2^2(i, j, W)$ e $t_3^1(j, Nta)$, curvas intermediárias se encontram dentro dos limites dos resultados mostrados, sem qualquer anormalidade ou distorção.

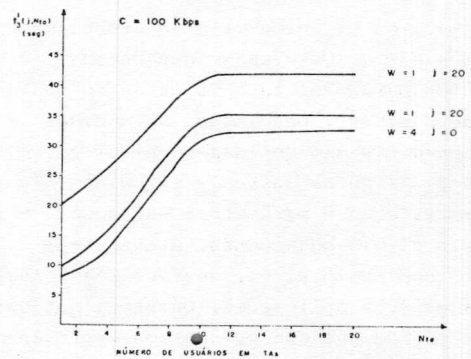


Figura 5

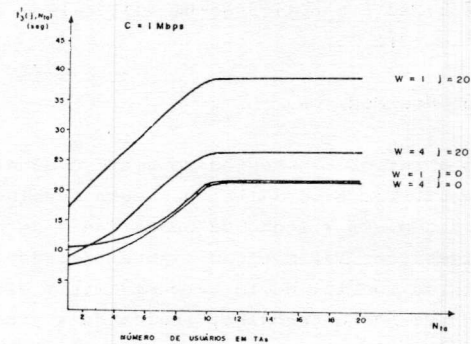


Figura 6

Das figuras 5 e 6, fica evidente os melhores valores para $t_3^1(j, Nta)$ quando C = 1 Mbps, pois são grandes quantidades de pacotes a serem transportados unidirecionalmente sem interrupções no fluxo do arquivo. A perturbação dos usuários em PTs (j), é registrada nestas duas figuras. São pequenos pacotes que trafegam bidirecionalmente entre as ETs e SA via sub-rede, que degradam os resultados de desempenho para TAs. São pacotes "pequenos" e "grandes" que disputam em igualdade os recursos da RL interferindo-se mutuamente. Atrasos devido a recursos não-físicos, como por exemplo, o servidor de protocolos para comunicação fim-afim, estão considerados nos parâmetros do modelo.

5 - Conclusões

O sistema computacional analisado se utiliza de um Servidor de Arquivos [4] para hospedar as aplicações de Transferência de Arquivos (TAs) e Processamento de Transações (PTs) tornando-as disponíveis aos usuários das Redes Locais (RLs). Estas duas aplicações represen

tam um considerável número de serviços existentes nas RLs, razão de sua importância de estudo.

O ambiente em questão, teve a velocidade do meio de transmissão (C) como parâmetro variável, de valor igual a 100 Kbps e 1 Mbps. São valores suficientes para investigarmos a viabilidade da integração de serviços nas RLs.

Como técnicas de análise foi empregada a Decomposição [3], e um modelo hierárquico estruturado em três níveis foi construído para encontrar medidas representativas ao usuário final. As medidas de desempenho abordadas neste trabalho foram: $t_2^2(i, j, W)$ --> tempo médio de resposta para uma transação completa, e $t_3^1(j, Nta)$ --> tempo de resposta para transferência de um arquivo completo.

Devido ao tipo de tráfego gerado por TAs, foi constatado que o valor de C é determinado no desempenho desta aplicação, e um valor maior para C implicou em melhores tempos de respostas para TAs. Esta sensibilidade não se manifestou de maneira mais acentuada para PTs.

O tamanho da janela (W) empregado em TAs, melhorou as medidas desta aplicação para $W > 1$, uma conclusão também encontrada em [1]. A transferência em lote de um arquivo, sem muitas interrupções para controle dos pacotes entre fonte e destino, implica em melhores resultados para TAs. Mas uma interferência deste parâmetro sobre os tempos de respostas de PTs ficou registrada pelo aumento da degradação quando $W = 1$.

A perturbação do tráfego entre os dois tipos de pacotes (TAs e PTs) de características operacionais distintas, foi observada de maneira mútua. Na variação do número de pacotes presentes na sub-rede, verifica-se a oscilação ocorrida nos níveis de desempenho destas duas aplicações. A ausência de uma disciplina discriminada para os dois tipos de pacotes nos recursos compartilhados (SA, sub-rede, protocolos), agravou ainda mais o desempenho destas aplicações.

Enfim, para o conjunto de valores propostos aos parâmetros do modelo, as medidas de desempenho destas aplicações se mostraram aceitáveis e bem comportadas ao usuário final, tornando viável e atrativo um projeto de RLs com integração de tráfegos heterogêneos, com um SA remoto centralizado aos usuários.

6 - Referências Bibliográficas

[1] - Moura, J. A. B., Hierarchical Modelling of Local Area Networks under File Transfer, Tese de Ph.D., Depart. de Eng. Elétrica, Univ. of Waterloo, Ontario, Canadá, 1982.

[2] - Mussi, J. M. F., A Model for a Network-Based File Server in Interactive Operation, tese de M.A.Sc., Depart. de Eng. Elétrica, Univ. of Waterloo, Ontario, Canadá, 1985.

[3] - Courtois, P. J., Decomposability: Queueing and Computer Systems Applications, Academic Press, New York, 1977.

[4] - Svobodova, L., "File Servers for Network - Based Distributed Systems", Computing Surveys, n.4, v. 16, dec. 1984.

[5] - Sá, C. C., Modelagem de Aplicações em Redes Locais, dissertação de mestrado do Departamento de Engenharia Elétrica da Universidade Federal da Paraíba. Apresentada em agosto de 1987 em Campina Grande - Paraíba.

Endereços dos autores:

(1) UECe/CCT/DMAE
Av: Paranjana, 1700
Bairro: Itaperi
60.175 - Fortaleza - Ce
Fone: (085) - 245.1133 - Ramal 2

(2) e (3) - GRC/UFPb
Av: Aprígio Veloso, 882
PRAI - CAMPUS II - CP 10.032
58.100 - Campina Grande - Pb
Fone: (083) - 322.1948 (direto)/321.722 ramal 549.