

UNIVERSIDADE FEDERAL DA PARAIBA
CENTRO DE CIENCIAS E TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELETRICA
MESTRADO EM ENGENHARIA ELETRICA

MODELAGEM DE APLICACOES
EM REDES LOCAIS

016
871.3.11.75(043)
522

CLAUDIO CESAR DE SA

CAMPINA GRANDE - PB
AGOSTO - 1987

CLAUDIO CESAR DE SA

MODELAGEM DE APLICACOES

EM REDES LOCAIS

Dissertação apresentada no Curso de
MESTRADO EM ENGENHARIA ELETRICA da
Universidade Federal da Paraíba, em
cumprimento às exigências para obtenção
do Grau de Mestre.

Area de Concentração: PROCESSAMENTO DA INFORMAÇÃO

MARCOS ANTONIO GONÇALVES BRASILEIRO

Orientador

JOSE ANTONIO BELTRAO MOURA

Co-orientador

CAMPINA GRANDE - Pb

AGOSTO - 1987

MODELAGEM DE APLICAÇÕES
EM REDES LOCAIS





S111m Sa, Claudio Cesar de
Modelagem de aplicacoes em redes locais / Claudio Cesar
de Sa. - Campina Grande, 1987.
81 f.

Dissertacao (Mestrado em Engenharia Eletrica) -
Universidade Federal da Paraiba, Centro de Ciencias e
Tecnologia.


1. Sistema Distribuido em Tempo Real 2. Processamento da
Informacao 3. Dissertacao I. Brasileiro, Marcos Antonio
Goncalves, Dr. II. Moura, Jose Antao Beltrao, Dr. III.
Universidade Federal da Paraiba - Campina Grande (PB)

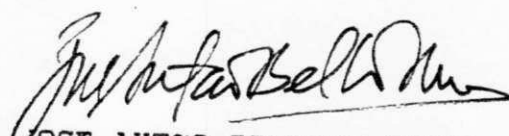
CDU 621.3.011.75(043)

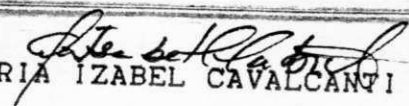
MODELAGEM DE APLICAÇÕES
EM REDES LOCAIS

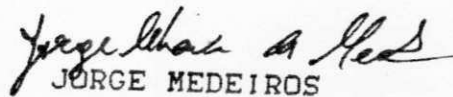
CLAUDIO CESAR DE SA

DISSERTAÇÃO APROVADA EM 14/08/1987


MARCOS ANTONIO GONÇALVES BRASILEIRO
Orientador


JOSE ANTAO BELTRAO MOURA
Co-orientador


MARIA IZABEL CAVALCANTI CABRAL
Componente da Banca


JORGE MEDEIROS
Componente da Banca

AGRADECIMENTOS

Agradeço aos Drs. J. A. B. Moura e M. A. G. Brasileiro pelo estímulo e dedicação em conduzirem este trabalho. Pelas idéias, sugestões e significativos comentários no texto.

A CAPES, CNPq e ITEEL/GRC pelo apoio financeiro.

A Universidade Estadual do Ceará.

Ao Departamento de Sistemas e Computação da UFPb por ter-me propiciado um clima de pesquisa.

A MSc. M.I.C. Cabral por ter-me auxiliado no andamento deste trabalho.

A minha esposa Ana Maria que com carinho soube acompanhar-me nos momentos difíceis desta jornada.

Aos meus pais Cesar e Sige que sempre acreditaram em mim.

A Lucia Helena que idealizou e incentivou o meu retorno ao mestrado.

Aos colegas do mestrado do DSC/UFPb.

Aos amigos e parentes.

A Deus.

RESUMO

Uma Rede Local é utilizada por seus usuários para se comunicarem entre si, ou com um dispositivo de uso compartilhado. Um destes dispositivos é o Servidor de Arquivos. Baseado neste servidor, podemos oferecer várias aplicações em diversos níveis de sofisticação ao usuário.

Aplicações como transferência de arquivos e processamento de transações, têm sido amplamente utilizadas como serviço de suporte a outras aplicações mais particulares como: submissão remota de tarefas, processamento de textos, consulta a banco de dados, interação com um sistema operacional remoto, etc.

Este trabalho propõe avaliar os níveis de desempenho das aplicações de transferência de arquivos e processamento de transações, quando disponíveis em Servidores de Arquivos remotos aos usuários. O interesse principal está em estudar a degradação do desempenho provocada pelas interferências entre estas duas aplicações de naturezas distintas, quando num mesmo ambiente de Rede Local.

SUMARIO

	Página
1 - INTRODUÇÃO	
1.1 - Redes Locais	1
1.2 - Avaliação e Desempenho de Redes Locais	3
1.3 - Modelagem Hierárquica	5
1.4 - Contribuição do Trabalho	6
1.5 - Organização do Trabalho	7
2 - AMBIENTES DE INTERESSE	
2.1 - Apresentação dos Ambientes de Interesse	8
2.2 - Servidores de Arquivos em Redes Locais	11
2.3 - As Aplicações	13
2.3.1 - Transferência de Arquivos	14
2.3.2 - Processamento de Transações	16
2.4 - Atendimento no Servidor de Arquivos	17
3 - UM MODELO HIERARQUICO PARA APLICAÇÕES EM REDES LOCAIS	
<hr/>	
3.1 - O Modelo Hierárquico	20
3.1.1 - Modelo do Nível 1	23
3.1.2 - Modelo do Nível 2	24
3.1.3 - Modelo do Nível 3	31
3.2 - Solução do Modelo Hierárquico	33
3.2.1 - Solução do Nível 1	34
3.2.2 - Solução do Nível 2	36
3.2.3 - Solução do Nível 3	41

4 - RESULTADOS E ANALISES

4.1 - Escolha dos Parâmetros	46
4.2 - Resultados do Nível 2	49
4.3 - Resultados do Nível 3	57

5 - EXTENSÕES DOS MODELOS E CONCLUSÕES

5.1 - Caso de População Única	63
5.2 - Outros Casos	67
5.3 - Conclusões	67
5.3.1 - Sugestões para Futuras Pesquisas	70

6 - ANEXOS

72

7 - REFERENCIAS BIBLIOGRAFICAS

76

LISTAS DAS FIGURAS

	Página
Fig. 1.1 - Componentes básicos de uma RL	1
Fig. 2.1 - Dois SAs correspondentes às duas aplicações	9
Fig. 2.2 - Um SA para as duas aplicações	10
Fig. 2.3 - Operação básica para TAs	15
Fig. 2.4 - Operação básica para uma transação	16
Fig. 3.1 - Representação conceitual do Modelo Hierárquico	22
Fig. 3.2 - Representação das três entidades de interesse por onde os pacotes transitam	24
Fig. 3.3 - Modelo de uma cadeia de filas simples	25
Fig. 3.4 - Modelo de múltiplas cadeias fechadas para o ambiente 1	27
Fig. 3.5 - Modelo de múltiplas cadeias fechadas para o ambiente 2	28
Fig. 3.6 - Modelo de filas com bloqueio para TA no SA	32
Fig. 3.7 - Modelo de um servidor único representando o meio físico	35
Fig. 3.8 - Diagrama de transição de estados do sistema de bloqueio no SA para TAs	43
Fig. 4.1 - Tempo médio de resposta para PTs, $t_2^2(i, j, W)$	50
Fig. 4.2 - Tempo médio de resposta para PTs, $t_2^2(i, j, W)$	51

Fig. 4.3 - Tempo médio de resposta para PTs, $t_2^2(i, j, W)$	52
Fig. 4.4 - Tempo médio de resposta para PTs, $t_2^2(i, j, W)$	53
Fig. 4.5 - Vazão média de um CV para TAs, $\gamma_2^1(i, j, W)$	54
Fig. 4.6 - Vazão média de um CV para TAs, $\gamma_2^1(i, j, W)$	55
Fig. 4.7 - Tempo médio de resposta para TAs, $t_3^1(j, Nta)$	58
Fig. 4.8 - Tempo médio de resposta para TAs, $t_3^1(j, Nta)$	59
Fig. 4.9 - Tempo médio de resposta para TAs, $t_3^1(j, Nta)$	60
Fig. 4.10 - Tempo médio de resposta para TAs, $t_3^1(j, Nta)$	61
Fig. 5.1 - População única para as duas aplicações	64
Fig. 5.2 - Diagrama de transições de estados dos usuários	65

ABREVIATURAS

- RLs : Redes Locais
TAs : Transferências de Arquivos
PTs : Processamentos de Transações
PAM : Protocolo de Acesso ao Meio
MH : Modelagem Hierárquica
SAs : Servidores de Arquivos
PI : Processamento independente/autônomo
FCFS : "First - Come - First - Served"
RR : Pacotes de reconhecimento ("acknowledge" ou "receive - ready")
CVs : Circuitos Virtuais
SA1 : Servidor de Arquivo #1
SA2 : Servidor de Arquivo #2
W : Tamanho da janela para TAs

1 - INTRODUÇÃO

1.1 - Redes Locais

De acordo com [STAL 84a], uma Rede Local (RL) é uma rede de comunicação que propicia a interconexão de uma variedade de equipamentos (microcomputadores, impressoras, servidores de arquivo, etc) em uma área geográfica limitada. Uma RL é um sistema distribuído que possui algumas características típicas como:

- a) Altas taxas de transferência de dados (50 Kbps a 1 Gbps)
- b) Baixas taxas de erro (1 bit errado a cada 10^{-8} a 10^{-12} bits transmitidos)
- c) Limitação geográfica (0,1 a 10 Km)
- d) Compartilhamento de recursos (hardware e software) não disponíveis localmente.

Uma RL é formada por usuários (assinantes, clientes), e uma sub-rede de comunicação, ver figura 1.1.

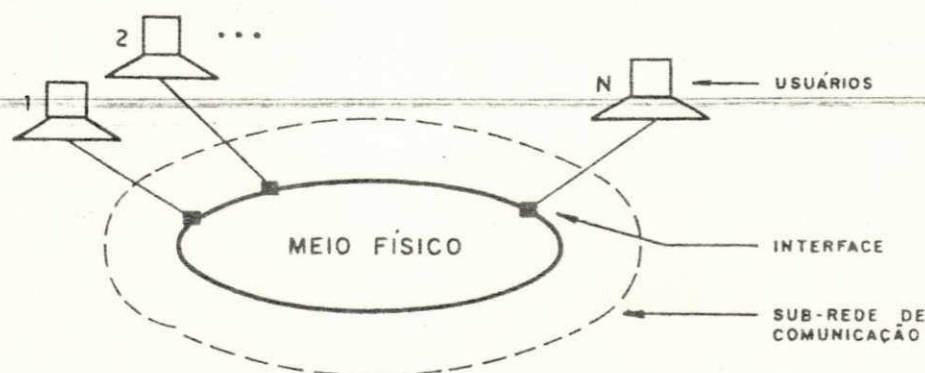


Fig. 1.1 - Componentes básicos de uma RL

Uma sub-rede de comunicação é formada por um meio físico (ou de transmissão) e interfaces de rede, também chamadas de nós de comunicação. Estes são responsáveis pelo envio e recebimento confiável de mensagens em geral, entre dois usuários através da sub-rede. O meio de transmissão é o suporte físico da comunicação (este pode ser um par trançado, cabo coaxial, e fibra ótica, etc); e está disposto segundo uma topologia de rede, geralmente em anel ou barramento [GAMS 86a, SOAR 86, CLAR 78, IEEE 82].

As interfaces permitem a conexão das estações de trabalho do usuário ao meio físico. Elas têm como função o controle, gerenciamento e monitorização do acesso a este meio compartilhado. O acesso e transmissão da mensagem na sub-rede são disciplinados por um protocolo de acesso ao meio físico. Este protocolo permite uma confiabilidade no tráfego de mensagens entre as interfaces fonte-destino. Fazendo uma extensão desta idéia a níveis mais elevados da comunicação, surge a necessidade de gerenciar o diálogo dos níveis físicos até um nível de aplicação do usuário. O nível de aplicação, é o nível mais elevado de uma comunicação fim-a-fim entre dois usuários. É neste nível que os serviços são oferecidos aos usuários.

O número de aplicações em RLs já é considerável e continua a crescer. Áreas como automação de escritórios, controle de processos, computação distribuída, processamento de voz e imagem, são alguns exemplos para o uso de RLs. Algumas aplicações como submissão remota de tarefas, processamento de textos, usam como serviço de suporte a Transferência de Arquivos (TAs). Aplicações como verificação de saldos bancários, interações com algum sistema operacional, têm como serviço de suporte o Processamento de Transações (PTs). Em razão da importância destas duas aplicações, um estudo na área de avaliação e desempenho destes dois serviços é a proposta deste trabalho. Na seção a seguir, é apresentado um resumo dos estudos realizados nesta área em RLs.

1.2 - Avaliação e Desempenho em Redes Locais

A avaliação e desempenho em RLa, têm como objetivo uma previsão de resultados no funcionamento da rede, segundo alguns parâmetros de projeto. Até então, muitas pesquisas estavam voltadas à modelagem e análise de protocolos de acesso ao meio de transmissão [MOUR 83, STAL 84b, BUX 81a, LAM 80]. Ou seja, as medidas de desempenho obtidas são referentes à sub-rede de comunicação. As medidas importantes neste nível físico são: atraso médio de pacotes entre duas interfaces e vazão média dos pacotes na sub-rede.

Em [BUX 81b], é realizada uma análise comparativa entre quatro tipos de protocolos de acesso ao meio (PAM) em RLa. Um modelo de filas com um servidor simples é usado para mostrar as características do PAM e topologia da rede. O meio físico foi modelado por um servidor de filas. As disciplinas de enfileiramento neste servidor, caracterizam o PAM e a topologia da rede. Os fregueses deste sistema de filas são representados pelos pacotes que trafegam na sub-rede [KLEI 75, SAUE 81]. Os parâmetros levados em consideração foram: comprimento do pacote, atraso de propagação no meio, velocidade do meio, taxa média de geração de pacotes nas interfaces, número de estações e comprimento do meio físico. Analisando os resultados em [BUX 81b], registra-se a importância destes parâmetros no desempenho de PAM e topologias nas RLa.

Há muitos trabalhos nesta área que utilizam modelos de filas para estudo de performance em RLa [LAM 80, TOBA 80]. Mas estes esforços têm-se concentrado a nível de sub-rede, ou seja, os níveis mais baixos de uma comunicação em RLa. Considerando níveis mais elevados, isto é, diálogo entre assinantes na rede, poucos trabalhos têm sido feitos. Alguns estudos na área de avaliação e desempenho de aplicações em RLa, estão em [MOUR 82, WONG 83, MUSS 85, LANT 85, BUX 84, WONG 80]. Em [MUSS 85], um modelo hierárquico [WONG 80] é usado para descrever um ambiente de

usuários realizando transações interativas com um servidor remoto via RL. Esta aplicação é uma sequência de interações entre um assinante e um servidor remoto, tendo-se um determinado número de pacotes entre estas duas entidades. Há um fluxo bidirecional de pacotes através da sub-rede.

Neste trabalho serão estudadas duas importantes aplicações (TA e PT), quando disponíveis aos usuários em um mesmo ambiente. Estas duas aplicações geram dois tipos de tráfegos na sub-rede. Portanto, é de interesse se conhecer a perturbação causada entre estas duas aplicações quando residentes na mesma RL.

Os procedimentos para avaliação e desempenho de aplicações em RLs sugerem soluções por modelos analíticos ou simulação discreta [MOUR 82]. Realizando um tratamento analítico, um modelo muito complexo pode surgir em razão dos detalhes a serem modelados, tornando esta solução matemática inviável. Quanto a simulação discreta, pode tornar-se proibitiva devido a extensão dos programas a serem desenvolvidos, e o aspecto "custo de máquina" (CPU) que é elevado. Adiantando, neste trabalho será empregada uma técnica híbrida (mista), isto é, uma parte do modelo será resolvida por simulação discreta, e na outra parte será empregado um tratamento analítico.

O estudo proposto abrange todos os níveis de comunicação fim-a-fim entre dois usuários. Este complexo processo de diálogo, que envolve várias camadas (fases) na comunicação, o que nos sugere a dividir este processo de diálogo em partes. Esta tarefa consiste em fazer uma divisão funcional das partes mais relevantes deste processo. Com isto, a idéia é de construir um modelo em módulos estruturais relacionados entre si, abstraindo algumas entidades funcionais importantes do sistema. Como resultado, obter-se-á uma melhor compreensão deste sistema de diálogo fim-a-fim. Esta metodologia é também conhecida como "modelagem hierárquica" ou "decomposição". O emprego da modelagem hierárquica é atrativa pelo fato de dividir um sistema em partes

menores, permitindo um melhor entendimento funcional do problema, e soluções mais simples.

Por estas razões, à modelagem hierárquica será empregada neste trabalho. Maiores detalhes sobre esta técnica, será o assunto da seção seguinte.

1.3 - Modelagem Hierárquica

A concepção de modelagem hierárquica (MH) ou decomposição, é representar um modelo completo através de submodelos (subsistemas, camadas, níveis) mais simples [COUR 75, COUR 77, COUR 85].

O objetivo desta decomposição, é a obtenção de uma solução mais eficiente aos submodelos que compõem o modelo geral. Estas soluções podem ter características típicas do nível em análise. Estes níveis tendem a ser independentes entre si, e com particularidades internas de funcionamento bem próprias. Por conseguinte, estes níveis possuem uma interface de ligação muito fraca com seus níveis adjacentes. Suas interfaces servem para passar e/ou receber valores de seu nível superior e inferior, respectivamente.

Esta hierarquização do modelo completo, vai de um nível mais baixo, a um nível mais elevado, sendo uma construção do tipo "bottom-up". O processo de resolução se inicia do nível inferior até o superior. Medidas de interesse a respeito da funcionalidade do modelo (sistema), são obtidas na solução de cada nível. Os resultados provenientes do último nível, são as medidas de interesse final a respeito do modelo. Concluindo, alguns aspectos importantes a ponderar quanto ao uso desta técnica:

- a) Fornece uma melhor visão estrutural do sistema, bem como uma melhor compreensão e clareza de suas partes em funcionamento.

- b) Ocorre uma combinação de métodos na solução de seus níveis, ou seja, um tratamento discriminado no estudo de cada nível.
- c) Propicia uma construção modular que permite acomodar futuras alterações para o modelo.
- d) No presente momento, tem-se algumas experiências bem sucedidas no emprego desta técnica [MOUR 82, MUSS 85, SCHW 78].
- e) As soluções dos submodelos, se apresentam de forma mais eficientes e simples.

Como ficará evidenciado no capítulo 2, o grau de complexidade da solução do modelo em estudo, é grande. Conseqüentemente, devemos optar por métodos de soluções que sejam precisos, eficientes e rápidos. Considerando as características da modelagem hierárquica reportadas acima, temos que esta se adapta como uma possível metodologia a ser usada na solução do modelo em consideração (integração do tráfego de PTs e TAs, em RLs).

1.4 - Contribuição do Trabalho

Com o avanço tecnológico das Redes Locais (RLs), e a diversidade de suas aplicações, um estudo sobre avaliação e desempenho das aplicações voltadas ao usuário final se faz necessário. Em estudos realizados na área [MOUR 82, MUSS 85], uma só aplicação por vez foi analisada.

Neste trabalho, será feita uma análise de duas importantes aplicações, a Transferência de Arquivos (TAs) e Processamento de Transações (PTs), disponíveis simultaneamente numa mesma RL. Esta integração de dois tipos de tráfegos em RLs, é uma situação mais realística que os dois estudos supracitados. Esta perturbação de um tipo de tráfego sobre outro, origina a motivação deste trabalho.

Por se tratar de uma área pouco explorada, estudos ainda são necessários para que projetistas em geral (especialmente de aplicações e RLs), tenham informações adicionais quando na elaboração de projetos.

1.5 - Organização do Trabalho

O capítulo 2 apresenta as entidades envolvidas neste estudo. É feita uma abordagem dos ambientes que serão explorados, das duas aplicações e sobre servidores de arquivos.

No capítulo 3 se apresenta um modelo hierárquico em 3 níveis. Este modelo, é construído com base nas partes mais relevantes do processo de comunicação entre um grupo de usuários e um servidor remoto, conectados através de uma RL. Uma descrição detalhada de cada nível do modelo, bem como suas respectivas soluções são apresentadas neste capítulo.

No capítulo 4 se apresenta os resultados dos modelos propostos no capítulo 3. Os resultados obtidos são reportados em gráficos, sendo que uma análise é realizada sobre a apresentação destas figuras.

No capítulo 5 é apresentado um ambiente mais realístico. É a situação de uma única população de usuários, realizando probabilisticamente uma das duas aplicações. Neste capítulo é feita apenas uma análise, pois resultados não foram obtidos; em seguida apresenta-se conclusões finais, e comentários deste trabalho.

2 - AMBIENTES DE INTERESSE

No capítulo 1, foi situado o contexto em que este trabalho se encontra. Neste capítulo, serão descritos os ambientes de interesse para estudo, bem como as aplicações já citadas. Estas aplicações normalmente residem em algum tipo de servidor compartilhado, e neste trabalho, o dispositivo considerado é o Servidor de Arquivos (SA). As configurações propostas para estudo são aqui apresentadas com o envolvimento de usuários, SA e aplicações. Após a descrição destes ambientes, será feita uma abordagem sobre SAs, e sobre alguns aspectos operacionais e de projeto. Finalizando, as aplicações de interesse (TA e PT), são definidas e especificadas. Com isto, formaliza-se um relato a respeito das aplicações, usuários, SAs e ambientes a serem investigados.

2.1 - Apresentação dos Ambientes de Interesse

Segundo as características das aplicações, TA e PT, geralmente elas podem estar disponíveis em servidores remotos. Para estas aplicações em RLs, os servidores usuais apropriados são os Servidores de Arquivos (SAs) [SVOB 84], isto devido a sua flexibilidade operacional em poder existir muitos sistemas (serviços) neles instalados.

Quanto aos usuários (clientes, assinantes), estes se utilizam de estações de trabalho ("workstation"), sistemas dotados de capacidade interna de processamento (por exemplo: microcomputadores pessoais, terminais específicos, etc), podendo estar associado a algum tipo de armazenamento secundário (por exemplo: unidades de discos flexíveis). São considerados usuários ativos aqueles que se utilizam da sub-rede para interagirem com o SA, operando em uma das duas aplicações.

Estes usuários, por simplificação, se encontram engajados em uma só aplicação por vez. Portanto, um usuário está em uma das seguintes atividades: TA, PT, ou em processamento independente ou autônomo (PI). Neste último caso, eles se encontram num estado inativo ou desconectado do SA, logo, o usuário não se utiliza da sub-rede.

Na figura 2.1 é mostrado um dos ambientes de interesse [BROW 84, SWIN 79]. É o caso onde há um SA dedicado para cada tipo de aplicação, isto é, uma separação física de onde residem os serviços. Há uma população de N_{ta} usuários em TAs, e N_{pt} usuários em PTs. Sendo assim, o número total de usuários ativos se utilizando da sub-rede é: $N_{ta} + N_{pt}$.

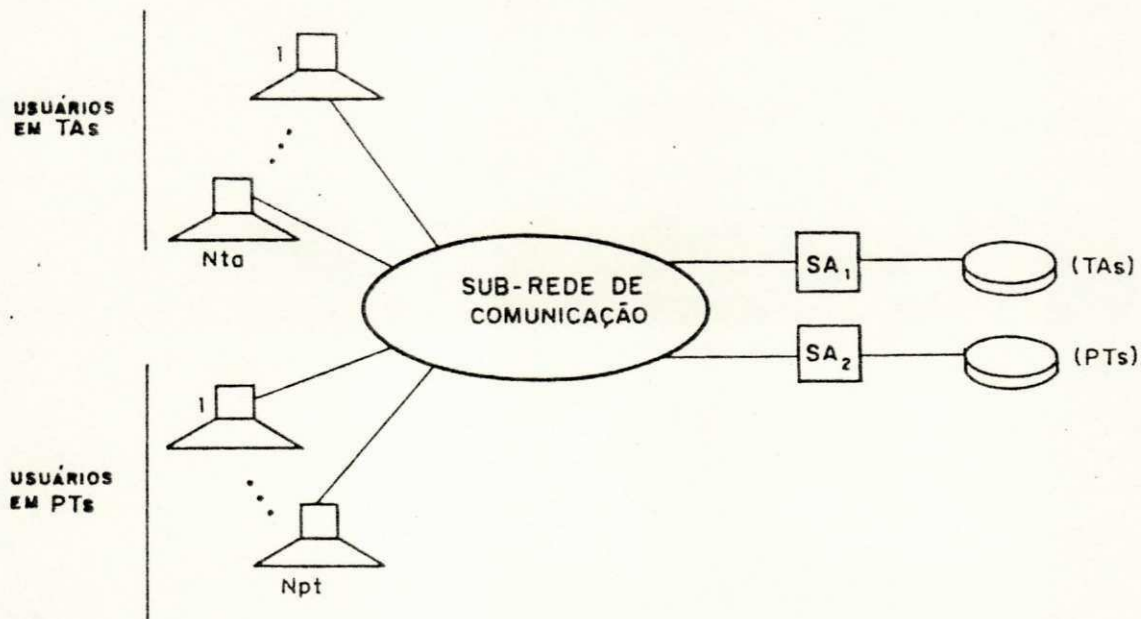


Fig. 2.1 - Dois SAs correspondentes às duas aplicações

O outro ambiente é o da figura 2.2 [MITC 82, DELL 82]. É o caso onde há um SA híbrido, que suporta as duas aplicações simultaneamente. Neste caso, o SA oferece os dois serviços em um mesmo sistema computacional.

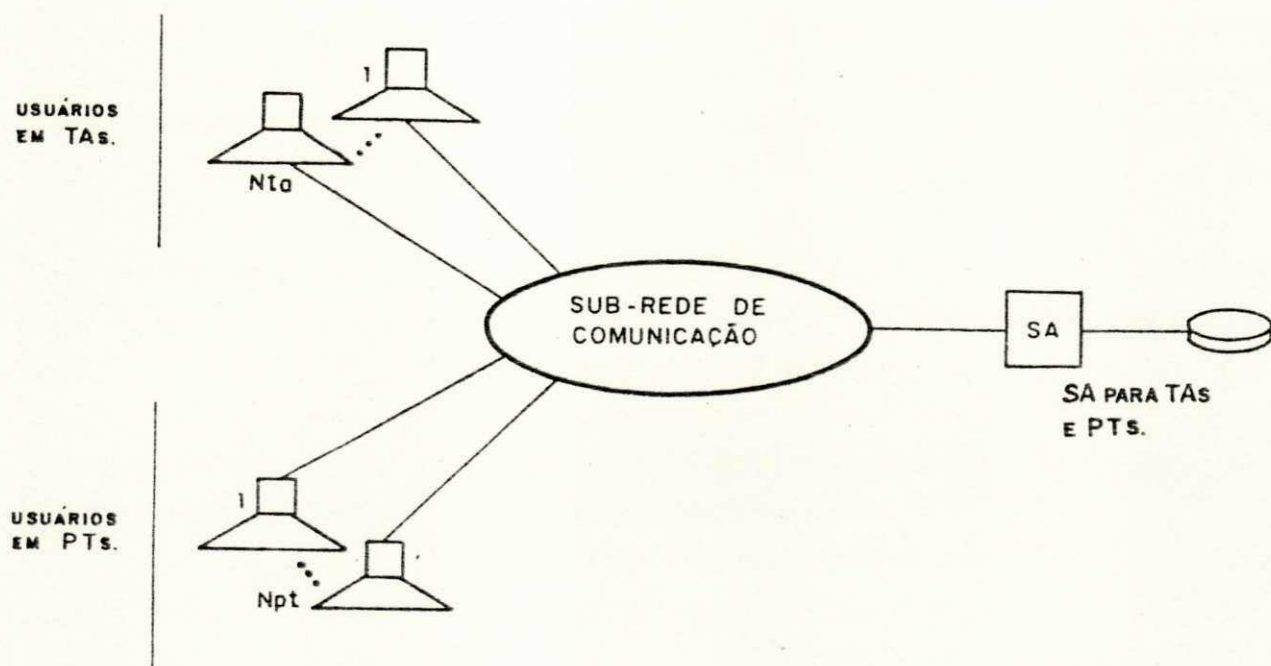


Fig. 2.2 - Um SA para as duas aplicações

A análise das perturbações causadas entre as aplicações de TAs e PTs, será desenvolvida sobre estes dois ambientes supracitados. Ao longo deste trabalho, serão feitas referências às figuras 2.1 e 2.2 como configurações ou ambientes 1 e 2, respectivamente. Um outro ambiente relevante, e uma possível solução, são descritos no capítulo 5. Na seção a seguir, é mostrada uma abordagem sobre SAs, justificando seu uso como um sistema computacional que suporta aplicações em RLs.

2.2 - Servidor de Arquivos em Redes Locais

Muitos sistemas computacionais distribuídos têm sido alvo de projetos e pesquisas nos últimos anos; alguns já disponíveis comercialmente aos usuários de RLs. Este é o caso de SAs [STAN 84], ver figuras 2.1 e 2.2.

O SA é um sistema para serviços de arquivos em geral, capaz de atender, de modo eficiente e compartilhado, informações comuns em discos rígidos (tipo "winchester"), a usuários a ele conectados. Como visto na seção 2.1, a ligação entre SA e usuários normalmente é realizada através de uma RL.

O SA é constituído por discos de alta capacidade de armazenamento (na ordem de centenas de megabytes) e um administrador, que são programas responsáveis pelos controles de acessos [MITC 82]. Portanto, um serviço típico de um SA é o armazenamento de grandes quantidades de dados, os quais são compartilhados entre os usuários autorizados.

A complexidade dos SAs variam com seu uso, desde simples armazenamentos secundários até um nível mais alto, em que residem alguns tipos de Sistemas Gerenciadores de Banco de Dados. E entre estes dois extremos de implementação que os SAs se encontram [BIRR 80].

Alguns pontos indicam o SA como uma boa alternativa de projeto em sistemas distribuídos [SVOB 84]:

- 1 - Num ambiente de processamento distribuído, existe o problema de se manter uma versão consistente e atualizada das informações compartilhadas por muitos usuários. Este problema é contornado com o uso de um SA centralizado.
- 2 - Um grande disco no SA tem um custo de administração, manutenção e implementação, menor do que vários pequenos discos nas estações do usuário.
- 3 - Compartilhamento de algumas aplicações em RLs de modo

equânime pelos usuários.

As características de um projeto sobre SAs têm como variáveis: confiabilidade, desempenho, compartilhamento de recursos, extensibilidade e funcionalidade.

Quanto a sua confiabilidade, é desejável que o SA tenha serviços como: controle de redundâncias, recuperações automáticas de arquivos, sobrevivência da informação em caso de falhas, proteção contra usos indevidos (acessos não autorizados ou usuários maliciosos), reinicializações automáticas, etc. Em resumo, baixas tolerâncias às falhas em geral.

No tocante ao desempenho, espera-se que devido a sua centralização de acessos concorrentes e simultâneos, não dê origem a um ponto de gargalo. Fazendo com que o sistema se torne lento, elevando assim os tempos de resposta.

No que tange ao compartilhamento de recursos, deseja-se que com o SA ocorra um aumento de produtividade para os usuários a um menor custo. Entende-se custo por informação armazenada e trafegada na sub-rede.

A extensibilidade são as alterações possíveis de hardware e software em geral, sem detrimento dos serviços já existentes no sistema.

Quanto à funcionalidade dos SAs aos usuários, está confinada a complexidade dos serviços oferecidos. Em níveis mais baixos, somente operações primitivas podem interagir com o SA. As invocações destas primitivas no manuseio de arquivos são do tipo: criar, abrir, incluir, alterar, ler, etc. Aqui o servidor funciona como um armazenador secundário. Neste caso, a aplicação é desenvolvida na estação de trabalho, pois, apenas estas primitivas padrões são reconhecidas pelo SA.

Num grau de sofisticação mais elevado, o SA pode causar a sensação ao usuário de um sistema dedicado. Este nível de abstração (transparência ao usuário), faz com que o SA possa realizar gerenciamentos de memória virtual, oferecendo serviços de paginação aos usuários. Nesta situação, as implementações são mais simples nas estações do usuário.

Nas referências [SVOB 84, MACE 86] há uma abordagem sobre alguns SAs existentes. São discutidas questões de projeto, como: controle e concorrência de acessos, granularidades dos acessos, atomicidade das operações, chamadas remotas, interfaces com o usuário, SAs distribuídos, congestionamentos e resolução de impasses, diversos sistemas operacionais interagindo com o servidor, etc. Com base em alguns pontos citados acima, conclui-se a grande complexidade do projeto de um SA [CHRI 84, DELL 82].

Neste trabalho, os SAs oferecem os serviços de TAs e PTs, embora os detalhes de seu funcionamento interno não sejam aqui pertinentes, podem ser encontrados em [CHRI 84]. Mas, a maneira de como será feito este atendimento dos pedidos para TAs e PTs é tocante a este trabalho, e este o será assunto da seção 2.4. Na seção a seguir, serão descritas as aplicações de interesse.

2.3 - As Aplicações

Juntamente ao crescente uso das RLs, está o desenvolvimento de aplicações a este sistema distribuído. As aplicações se apresentam numa versão final de um serviço desejado pelo usuário. É uma apresentação terminal (externa) de uma tarefa requisitada pelo usuário. Como exemplo, cita-se o correio eletrônico, a submissão remota de tarefa, sistema telecompras, reservas de passagens aéreas (estas duas últimas provavelmente se utilizam dos serviços de redes à longa distância), interações com o sistema operacional de uma máquina remota, processamentos de voz e imagem, etc.

Numa abrangência conceitual destas duas aplicações supracitadas, o suporte básico em alguns casos é a transferência de arquivos e o processamento de transações. Devido a frequente presença destes dois serviços embutidos em outras importantes aplicações de RLS mais sofisticadas, explica-se a razão de suas escolhas para estudo.

A TA é uma sequência unidirecional (origem/destino) de pacotes "longos", enquanto PT são pacotes "curtos" (algumas centenas de bits, de 5% a 20% do tamanho dos pacotes de TAs) trafegando de modo bidirecional [SOAR 86].

São dois tipos de pacotes típicos que trafegam entre usuários e SA, via sub-rede. Concluindo, são tráfegos distintos de muita importância nas aplicações de RLS, portanto, se faz necessário um estudo mais aprofundado sobre suas interferências (um tipo de tráfego sobre o outro, e vice-versa) quando num mesmo ambiente. A seguir, uma descrição funcional destas duas importantes aplicações.

2.3.1 - Transferência de Arquivos

Aplicações como controle de estoque, submissão remota de tarefas, atualizações bancárias (matriz/filial, à noite), arquivos com correspondências expedidas/recebidas nos últimos 12 meses (automação de escritório em geral), etc, normalmente são serviços que envolvem uma transferência volumosa de dados entre dois usuários remotos.

Para os exemplos acima citados, a TA encontra-se embutida num nível mais baixo (menos sofisticado), como um serviço de suporte básico a estas aplicações; em outros casos, pode-se apresentar em um nível final de aplicação, por exemplo: uma típica/pura transferência de arquivo. Esta prestação de serviços de base a outras aplicações mais específicas, faz com que estas se apresentem como uma "casca externa" às TAs. Um

diagrama típico de seu funcionamento entre usuário e SA é mostrado na figura 2.3.



Fig. 2.3 - Operação básica para TAs

Sua inicialização se processa com um pedido de conexão do usuário ao SA. Uma vez habilitado em atender esta requisição, se estabelecerá um Circuito Virtual (CV) através da sub-rede entre o usuário e o SA. Este caminho virtual permanecerá ativo durante toda a transferência do arquivo, sendo liberado no final da transmissão do mesmo.

Quanto ao envio do arquivo, ele é fragmentado (segmentado) em pacotes de tamanho característico ao estipulado pelo protocolo de rede [GAMS 86a]. Na medida que os pacotes de dados vão chegando na estação de trabalho, esta vai gerando pacotes de reconhecimento em sentido contrário, segundo o mecanismo de janela deslizante do X-25 [GAMS 86a]. Tem-se neste caso, um controle de fluxo fim-a-fim. Quando forem elaborados os modelos, este assunto voltará a ser discutido, e outros detalhes no decorrer do trabalho serão reportados.

2.3.2 - Processamento de Transações

As aplicações bancárias típicas como: verificação de saldos, saques, depósitos, cartões de crédito, abre/fecha sessão remoto, etc, caracterizam-se num grupo de transações simples do tipo pergunta-resposta ("inquiry-response"). São transações de mensagens pequenas (na ordem de centenas de bits) tanto para o pedido (consulta) como para resposta. As aplicações que possuem as características de um simples pedido ao servidor remoto, um processamento, e uma simples resposta, se enquadram em PTs.

Um típico exemplo para PTs, são as interações entre a estação de trabalho e o sistema operacional do SA. São requisições do tipo: "login", "logout", "time" do sistema, verificar disponibilidade de área em disco, alguma ajuda a comandos, etc.

Os pacotes de "inquiry" e "response", são pequenos em relação ao tamanho dos pacotes de TAs. Normalmente, não estão sujeitos a um controle de fluxo rigoroso. Portanto, um serviço de Datagrama oferecido pelo protocolo de rede atenderá as exigências desta aplicação. Um diagrama de funcionamento para PTs é mostrado na figura 2.4.

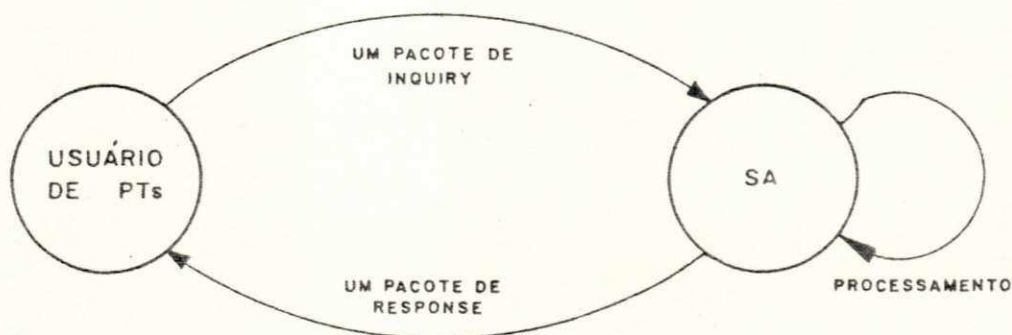


Fig. 2.4 - Operação básica para uma transação

Com base na figura 2.4, o ciclo de uma transação é definido como um pacote de "inquiry", um processamento remoto, e um pacote de "response" como retorno. Devido à simplicidade no funcionamento desta aplicação, ela se preza por um tempo de resposta aceitável ao usuário, e com uma boa confiabilidade.

No serviço Datagrama não há qualquer ligação lógica entre o usuário e SA, mas como será visto mais adiante, uma conexão lógica vai existir a nível de modelo. Este número de ligações entre o servidor e muitos usuários, é um aspecto funcional da capacidade de atendimentos concorrentes ao SA, que é assunto da próxima seção para as duas aplicações.

2.4 - Atendimento no Servidor de Arquivos

Na seção anterior foram especificadas as duas aplicações de interesse. Nesta seção, será descrito o comportamento do SA diante das requisições de serviços destas aplicações. Um aspecto interessante de qualquer máquina de uso compartilhado, é quanto a sua capacidade de atendimentos simultâneos/concorrentes, com um tempo de resposta satisfatório. O atendimento a uma tarefa a ser realizada, é como o SA gerenciará esta requisição. Em geral, os SAs possuem uma limitação em sua capacidade de atendimentos simultâneos. Supondo que este limite seja de aceitar até "y" pedidos de conexões ao SA, as seguintes situações podem ocorrer:

- 1 - Quando o valor de "y" não foi atingido, isto implica que um novo pedido de conexão de mais um usuário será atendido.
- 2 - Se o número de conexões no momento já é "y", um novo pedido poderá ser enfileirado no "buffer" de entrada (área de "spool"), ou ser rejeitado. No caso de rejeição, há uma perda do pedido do usuário, e este fará uma nova tentativa no futuro.

A explicação que se segue, é sobre o funcionamento do SA quando requisitado a atender os pedidos de TAs e PTs, nos ambientes 1 ou 2, respectivamente, figuras 2.1 e 2.2. Todos atendimentos realizados pelo SA, ocorrem na ordem histórica de suas chegadas, isto é, os pedidos mais antigos serão atendidos primeiro. Um sistema de fila do tipo FCFS ("First - Come - First - Served"), uma suposição válida para PTs e TAs. No ambiente 2 onde o SA é híbrido, não há critérios de prioridades quanto à ordem de atendimento para pedidos de TAs ou PTs. O procedimento do FCFS é mantido, apesar dos tempos de serviços dos pacotes de TAs e PTs serem diferentes. Este assunto, será discutido em detalhes mais adiante.

Devido às naturezas das aplicações, diferentes procedimentos serão tomados pelo SA. Para o caso de TAs, um novo pedido será rejeitado quando o servidor já estiver operando com "y" conexões em TAs. Neste caso ocorre uma perda do pedido do usuário. Enquanto para PTs, não há rejeição (perdas) de pedidos, isto é, não há limites de requisições na fila de entrada do SA aos usuários em PTs.

As requisições para TAs, serão aceitas até um limite máximo de "y" conexões lógicas para TAs. Independentemente das conexões já existentes ou que possam vir a ocorrer para PTs. A partir de "y" conexões existentes em TAs, os novos pedidos para TAs serão rejeitados, exceto para PTs onde não há limite de conexões para seus usuários. Seus pedidos entram numa fila, e são processados assim que o SA estiver livre, respeitando o seu lugar histórico (FCFS) na fila, em igualdade com os pedidos de TAs. Em outras palavras, para PTs não ocorrem perdas de pedidos, e são independentes das "y" conexões em TAs que possam existir. É um tipo de pedido "persistente", isto é, que não desiste do serviço no SA.

Mediante todas estas características descritas (ambientes, aplicações, e SA), tem-se um trabalho bastante árduo na obtenção

de um modelo monolítico para solução do sistema. Portanto, uma divisão por partes do problema (MH) tem como objetivo suavizar a tarefa de investigação deste sistema. No capítulo 3, será descrito em detalhes um modelo hierárquico, e este modelo é proposto para estudar os ambientes 1 e 2, considerando as características aqui mencionadas.

3 - UM MODELO HIERARQUICO PARA APLICACOES EM REDES LOCAIS

No capítulo 2, foi feita uma apreciação sobre SAs, ambientes e aplicações de interesse. Sobre estas entidades descritas é que se vai construir um modelo para estudo. Este modelo é estruturado em três níveis, onde cada nível modela uma parte funcional relevante aos ambientes em consideração. Esta estruturação em níveis, ou MH, oferece algumas vantagens neste tipo de investigação, como já descrito na seção 1.3, esta será a técnica aqui usada.

O capítulo se inicia com algumas considerações gerais para o modelo. Em seguida, os três níveis propostos para estudo são apresentados, e cada nível do modelo será descrito em detalhes para os ambientes 1 e 2, onde uma solução adequada a cada nível é sugerida, levando-se em consideração as naturezas das aplicações 1 e 2, e suas medidas relevantes de desempenho.

3.1 - O Modelo Hierárquico

Na construção de um modelo hierárquico, algumas aproximações são necessárias. Aqui reportaremos aproximações próprias deste trabalho, como também outras de carácter mais geral sugeridas em [REIS 82]:

- 1 - Cada usuário está engajado em um só tipo de aplicação. Visto nas figuras 2.1 e 2.2, há uma população de usuários que realizam TAs, e outra em PTs.
- 2 - Os usuários não podem ter atividades pendentes, isto é, enquanto um usuário estiver realizando uma transação ou uma transferência de arquivo, não podem existir novas submissões de pedidos ao servidor.
- 3 - Os pacotes transmitidos na rede, são livres de erros. Como já escrito, as taxas de erros em RLs são muito baixas [SHOC 80].

- 4 - Os atrasos das fases de conexão e desconexão do CV são desprezados. Para o caso das TAs, são atrasos muito pequenos comparados à fase de transferência dos pacotes de dados. Em PTs, o serviço escolhido é o Datagrama, portanto, estes atrasos não existem.
- 5 - Não há problemas de "estouro" de buffers nas interfaces das estações de trabalho. O tamanho mínimo de um buffer, é igual ao tamanho da janela utilizada na interação entre usuário e SA. Esta suposição é válida para TAs e PTs.
- 6 - Todos os pacotes têm prioridades iguais quando trafegam através da sub-rede. No ambiente 2, onde o SA é compartilhado pelas duas aplicações, também não há critérios quanto às prioridades de atendimentos, isto é, os pacotes das duas aplicações, são atendidos indistintamente na ordem de suas chegadas. Resumindo, é um sistema de filas do tipo FCFS ("First-Come-First-Served") com um servidor único, porém, com duas classes de pacotes sem prioridades.
- 7 - Erros do tipo: endereço de destino, parâmetros inválidos, acessos não autorizados, etc, não serão considerados.

Como já mencionado, o método da decomposição consiste em abstrair as partes funcionais mais relevantes do sistema, e construir submodelos mais elementares. Para construir o modelo, é preciso reproduzir as partes mais importantes do processo de comunicação entre o usuário e SA. Para o sistema em estudo, um modelo conceitual é sugerido e representado na figura 3.1.

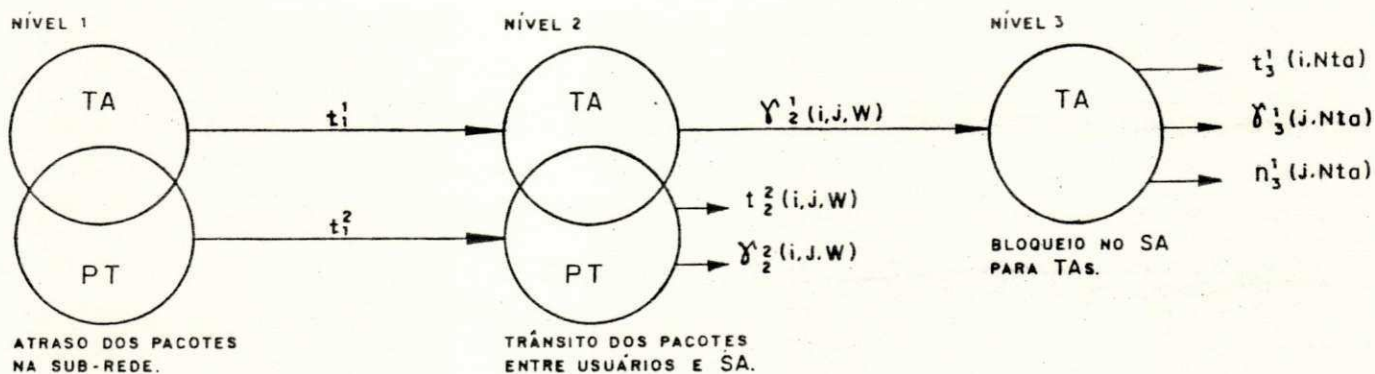


Fig. 3.1 - Representação conceitual do Modelo Hierárquico

Os círculos simbolicamente representam as aplicações no modelo. Como se observa, o modelo consiste em três níveis para análise de TAs, e dois níveis para PTs. É importante deixar claro que estes níveis não se relacionam com os sete níveis funcionais do protocolo de comunicação do Modelo de Referência da ISO [TANE 81a]. Tratando-se de modelagem, procura-se retratar os aspectos mais importantes do sistema. Neste modelo, medidas finais de interesse para TA são obtidas no nível 3, e para PT no nível 2. Nos níveis 1 e 2 da figura 3.1, tem-se uma intersecção dos dois círculos, isto quer dizer que, a solução do nível será tratada de modo integrado para as duas aplicações.

As setas entre os círculos (níveis), indicam as interfaces e os parâmetros passados de um nível para o outro adjacente. Os índices dos parâmetros (as letras), representam o seu nível de origem, e o expoente define a aplicação. Por comodidade e convenção, TA será chamada de aplicação 1 (um), e PT será a aplicação 2 (dois). A seguir, cada nível da figura 3.1 será detalhado, bem como suas soluções.

3.1.1 - Nível 1 do Modelo

Este nível se refere aos atrasos dos pacotes das aplicações 1 e 2 (TAs e PTs respectivamente), na sub-rede de comunicação. O atraso aqui considerado, é apenas o tempo de transmissão do pacote na sub-rede. Os atrasos devido às colisões, erros, propagação dos sinais, etc, serão desprezados, uma vez que estes atrasos são pequenos em relação ao tempo final de uma aplicação. Esta comunicação fim-a-fim, envolve outros tipos de atrasos mais consideráveis no tempo de aplicações do usuário final. Por exemplo, os atrasos dos protocolos de alto-nível na sub-rede, que são mais relevantes nos tempos finais das aplicações [GAMS 86b], estes atrasos serão considerados nos parâmetros dos modelos aqui desenvolvidos.

Há quatro tipos de pacotes trafegando na sub-rede. Para TAs, há pacotes de dados e reconhecimento ("acknowledge" ou "receive-ready" - RR). Em PTs, há pacotes de "inquiry" e "response". No caso de TAs, os pacotes de reconhecimento (RR) são gerados segundo às chegadas dos pacotes de dados. A cada W pacotes de dados recebidos, um pacote de RR é enviado como retorno à fonte. Este número W , é também chamado de "tamanho da janela". Os pacotes de RR, normalmente são muito menores que os pacotes de dados. Uma aproximação é feita no sentido de aumentar proporcionalmente o tamanho dos pacotes de dados, para que se possa desconsiderar o efeito do tráfego em sentido contrário dos pacotes de RR. Esta contabilização do atraso de RR nos pacotes de dados, foi utilizada em [REIS 79].

Uma aproximação semelhante é feita para o caso de PTs, pois os pacotes de "inquiry" e "response" são pequenos (curtos). Neste caso, o atraso será contabilizado de modo único. Será um atraso igual ao dobro de um pacote de pergunta e resposta. Por conseguinte, as medidas de interesse fornecidas pelo nível 1 são:

- t_1^1 : Atraso médio de um pacote de dados com um acréscimo proporcional do pacote de reconhecimento, para TAs na sub-rede.
- t_1^2 : Atraso médio de um pacote de pergunta e resposta para PTs na sub-rede.

Como usar estes valores e sua interface com o nível 2, será descrito na subsecção 3.2.1. A seguir, será descrito o nível 2 do modelo.

3.1.2 - Modelo do Nível 2

O nível 2 analisa o encaminhamento dos pacotes entre as estações de trabalho do usuário e SA. Das figuras 2.1 e 2.2, vemos que os pacotes devem trafegar através de três entidades: na estação de trabalho, na sub-rede e SA. Sendo estas duas últimas, entidades compartilhadas. Neste nível, deseja-se analisar o comportamento (trânsito) dos pacotes quando se movimentam entre a estação do usuário e o SA, como mostrado na figura 3.2.



Fig. 3.2 - Representação das três entidades de interesse por onde os pacotes transitam

Os pacotes trafegam tanto do usuário ao SA, como em sentido contrário. Por exemplo, em uma TA, os pacotes são enviados de modo unidirecional, mas o SA pode ser tanto o receptor como o

emissor deste arquivo em relação ao usuário. Por convenção, o SA será o receptor (sorvedouro, destino) de pacotes. Assim sendo, a estação de trabalho será o transmissor (fonte, origem).

No trânsito dos pacotes entre usuário e SA, via sub-rede, estes sofrem atrasos nestas três entidades. Considerando estes atrasos na estação do usuário, na sub-rede e SA, apresenta-se como modelo para figura 3.2, um sistema de filas fechadas como mostra a figura 3.3.

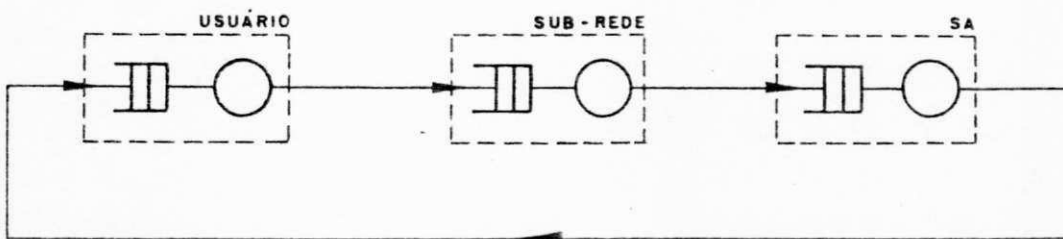


Fig. 3.3 - Modelo de uma cadeia de filas simples

Esta idéia de representar uma conexão lógica entre um usuário e SA, por uma cadeia de filas, é sugestivo para os dois grupos (classes) de usuários. No caso de TAs, os pacotes são unidirecionais pela sua natureza. O serviço oferecido pela camada de rede [TANE 81a] é o Circuito Virtual (CV). É adequado utilizar esta conexão lógica no encaminhamento dos pacotes do arquivo, entre o usuário e SA. O controle de fluxo é realizado através do mecanismo de janela deslizante, e este procedimento exerce uma limitação no número de pacotes presentes entre estas duas entidades. Tratando-se de modelagem, o que se faz é limitar uma população máxima de pacotes contidos nesta cadeia fechada, igualando-se ao tamanho da janela, W . A idéia desta abordagem, tem como origem [REIS 79].

No caso de PTs, os pacotes são bidirecionais. Há um envio de um pacote de pergunta ao SA, e um de resposta de retorno ao usuário. Este procedimento caracteriza uma transação interativa. Lembrando da aproximação feita na subseção anterior (onde dois pacotes pequenos são transformados em um único pacote), a representação da viagem de ida e volta de uma transação é feita por apenas um pacote entre o usuário e SA. Neste caso, a camada de rede oferece o serviço Datagrama. Mas como há um pacote apenas, o modelo da figura 3.3 também é usado para ilustrar o caminho do pacote de PTs.

Resumindo, o trânsito dos pacotes da aplicação 1 será modelado como uma cadeia de filas fechadas, com uma população (pacotes) finita igual a W . Idem para aplicação 2, exceto quanto à população da cadeia, que é igual a 1 (um). Relembrando que não há admissão de novos pedidos para TAs ou PTs enquanto existir uma tarefa pendente (em andamento, execução), portanto, a população das cadeias (conexões lógicas) de cada usuário permanecem fixas. São cadeias fechadas de população finita e invariável.

Fazendo uma extensão das idéias expostas acima aos ambientes 1 e 2, resulta nos modelos das figuras 3.4 e 3.5 respectivamente. São modelos com múltiplas cadeias fechadas para os ambientes 1 e 2 (figuras 2.1 e 2.2 respectivamente), válidos para os dois grupos de usuários.

Na figura 3.4, o SA1 é utilizado pelos "i" usuários que realizam TAs. Idem para o SA2, que interage com os "j" usuários em PTs. Não se deve esquecer que no caso de TAs, o SA pode manipular até "y" conexões simultâneas. Portanto, "i" é menor ou igual a "y". Aqui só são representadas as possíveis conexões lógicas simultâneas ($N_{ta} \leq y$) de TAs com o SA. Para o caso de $N_{ta} > y$, será um assunto a ser tratado no nível 3. No caso de PTs, não há limitação de usuários interagindo com o SA, por conseguinte, $j = 0, 1, 2, 3 \dots N_{pt}$.

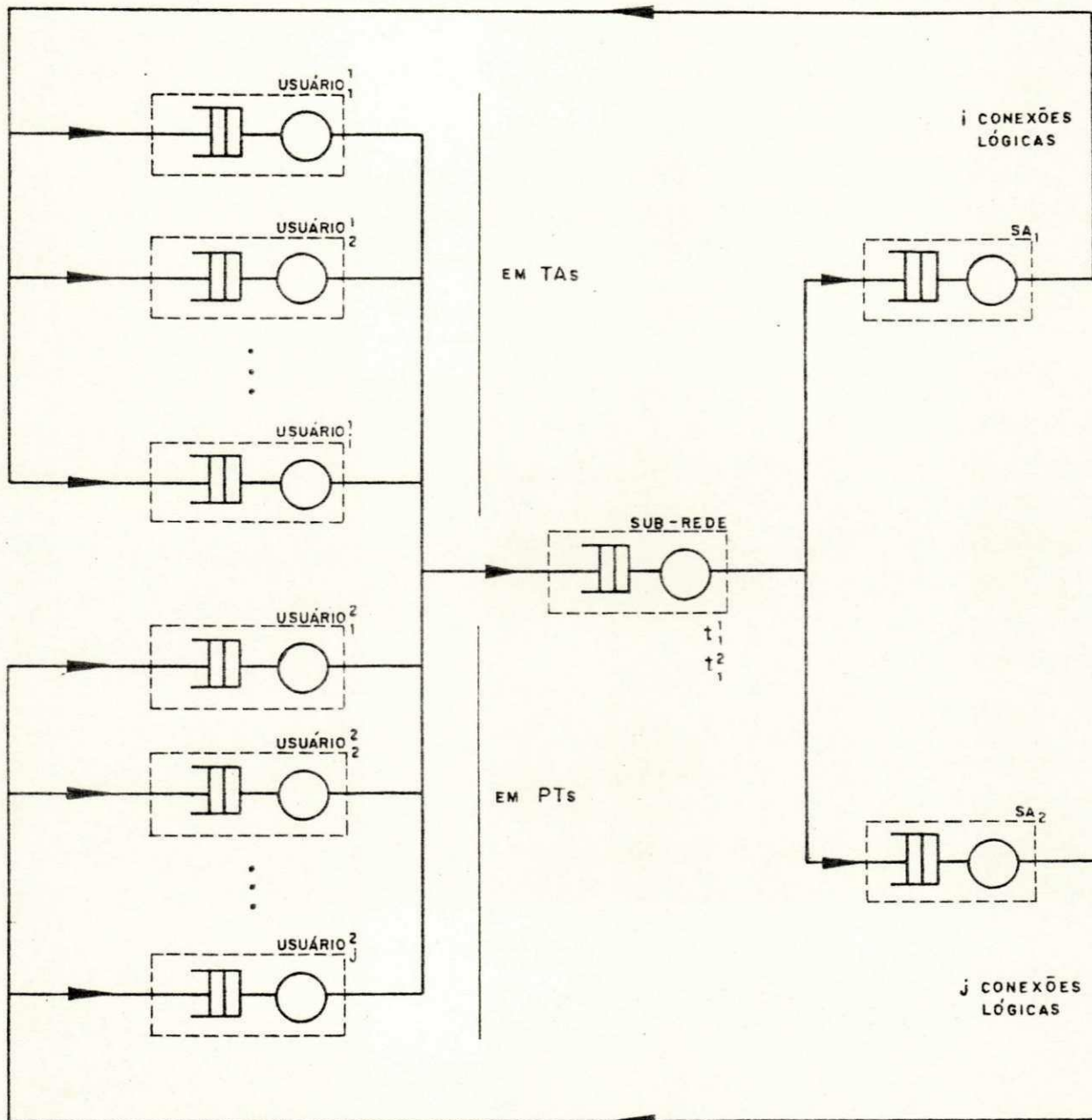


Fig. 3.4 - Modelo de múltiplas cadeias fechadas para o ambiente 1

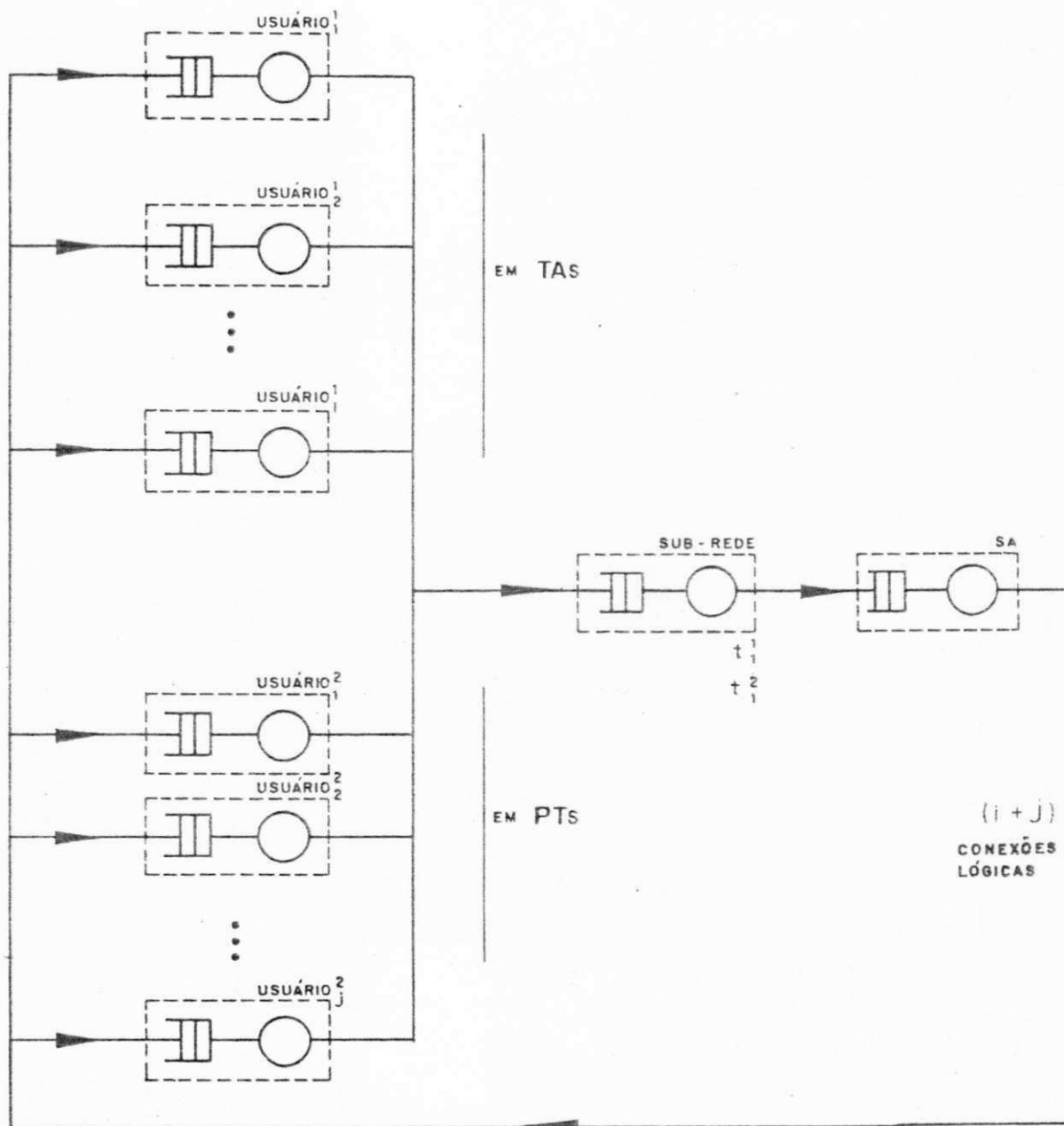


Fig. 3.5 - Modelo de múltiplas cadeias fechadas para o ambiente 2

Outro detalhe das figuras 3.4 e 3.5, é que todos "i" usuários, direcionam o fluxo de seus pacotes ao SA. Ou seja, todas as TAs estão ocorrendo no sentido do usuário ao SA. Em experimentos feitos em [MOUR 82] por simulação, mostrou-se que a interferência no sentido dos pacotes entre usuário \leftrightarrow SA e sentido contrário, é um efeito praticamente nulo. Sendo assim, um sentido único da estação do usuário ao SA é adotado por clareza, convenção e comodidade.

O tempo de serviço no servidor usuário dos modelos (figuras 3.4 e 3.5), representa os atrasos dos pacotes devido a protocolos de alto-nível da comunicação fim-a-fim, da interface, e da própria estação de trabalho. É prematuro em dizer que este atraso possui uma distribuição exponencial, mas, é uma suposição que será necessária na resolução deste sistema.

O atraso dos pacotes no servidor SA, será caracterizado pela aplicação, pois, os pacotes têm um tratamento distinto ao chegar no SA. Sendo assim, o atraso será típico para cada pacote da aplicação 1 e 2. Esta observação é também válida para o servidor da sub-rede.

A suposição que estes tempos de serviço nos servidores usuários, sub-rede e SAs, possuem uma distribuição exponencial, é uma hipótese razoável e necessária. A respeito deste assunto, será enfatizado na subseção 3.2.2.

Da seção anterior, as medidas de performance fornecidas ao nível 2 são t_1^1 e t_1^2 . Os valores de t_1^1 e t_1^2 correspondem ao tempo médio de serviço no servidor da sub-rede, para um pacote de TA e PT respectivamente, nas figuras 3.4 e 3.5. Neste nível, algumas medidas de desempenho podem ser obtidas, como por exemplo: tempo e comprimento médio de pacotes nas filas dos servidores deste sistema. Outras medidas serão passadas ao nível 3 (ver figura 3.1).

No caso de TAs, a medida de interesse é $\gamma_2^1(i, j, W)$, que é a vazão média de um CV (uma cadeia simples) dado "i" conexões simultâneas, com janela de tamanho igual a W, e "j" usuários PTs. A unidade de $\gamma_2^1(i, j, W)$, é em arquivos por unidade de tempo. Por se tratar de CVs idênticos para os "i" usuários em TAs, a hipótese é que todos trabalhem com o mesmo valor para W.

Considerando o início de uma TA, ocorre o envio de um primeiro pacote do usuário ao SA, através de um CV. Este primeiro pacote não enfrenta nenhuma fila causada por pacotes de sua própria conexão lógica. Assim que este pacote chega ao seu destino, um reconhecimento é enviado em seguida, e os demais pacotes do arquivo começam a serem transmitidos. O tempo médio de circulação de um pacote quando ele não precisa entrar em nenhuma fila, causada por pacotes de sua própria cadeia, é dado por $f_2(i, j, W)$.

Considerando que o tamanho médio de um arquivo qualquer, seja igual a "np" pacotes, falta contabilizar o tempo gasto no envio dos "np-1" pacotes restantes. Definindo $g_2(i, j, W)$, como sendo o tempo médio entre-chegadas de pacotes do arquivo no servidor SA em um dado CV, pode-se obter o tempo médio gasto na transferência de um arquivo completo. Utilizando-se a Lei de Little, a vazão é definida como sendo o inverso do tempo de circulação de um arquivo completo, portanto, $\gamma_2^1(i, j, W)$ é dado por:

$$\gamma_2^1(i, j, W) = \frac{1}{[f_2(i, j, W) + (np-1)g_2(i, j, W)]} \quad (3.1)$$

Os índices i, j, e W, indicam o estado de operação da rede, mostrando que num dado instante, há "j" usuários em PTs, e "i" usuários em TAs com CVs de características idênticas, por conseguinte todos com o mesmo W.

Quanto a PTs, as medidas de interesse final para esta aplicação são obtidas neste nível, ver figura 3.1. Como indicado na referida figura, pode-se obter $t_2^2(i, j, W)$, que é o tempo médio de resposta de uma transação (PT), dado i , j , e W . Outra interessante medida, é a vazão média de uma estação de usuário em PTs, $\gamma_2^2(i, j, W)$, dado em transações por unidade de tempo.

Como se obter todas estas medidas, será assunto da seção 3.2.2. Estas medidas são as mais importantes em PTs, outras medidas de desempenho podem ser facilmente obtidas, como por exemplo: o número médio ou tempo médio de pacotes de PTs, ou de TAs, presentes nas filas do SA e sub-rede. Aqui a aplicação 2 é finalizada, restando um tratamento para TAs no nível 3. Na subseção a seguir, será descrito o modelo do nível 3.

3.1.3 - Modelo do Nível 3

Este nível refere-se à particularidades da aplicação 1 e SA. Um usuário em TAs pode se encontrar em dois estados alternantes: um dos estados seria o de conexão com o SA (um estado ativo), realizando uma transferência de arquivo, o outro, seria um estado de desconexão (um estado inativo), ou em processamento autônomo/independente (PI). Entre uma transferência e outra, o usuário passa e permanece no estado PI, por um tempo médio de $1/\lambda$ segundos. Este tempo é definido como tempo de pensamento ("think time").

Como já mencionado, no caso de TAs, o SA vai impor um limite máximo de "y" conexões simultâneas com seus usuários. Esta limitação no SA, é com o objetivo de não sobrecarregá-lo em suas tarefas, pois uma TA envolve pacotes "longos", e em grande número. Esta é uma aplicação crítica no sentido de monopolizar o uso do SA, e este limite "y", é uma tentativa de impedir uma degradação do sistema nos atendimentos destas duas aplicações (TA e PT). Sendo assim, aos usuários de TAs, há um bloqueio (perdas) de novos pedidos quando $i = y$. Com esta idéia de limite no SA, e

alternâncias dos estados de um usuário de TAs, constroi-se um modelo de filas fechadas com bloqueio para representar estas características. Esta interação entre os N_{ta} usuários de TAs, e o SA, é ilustrada na figura 3.6.

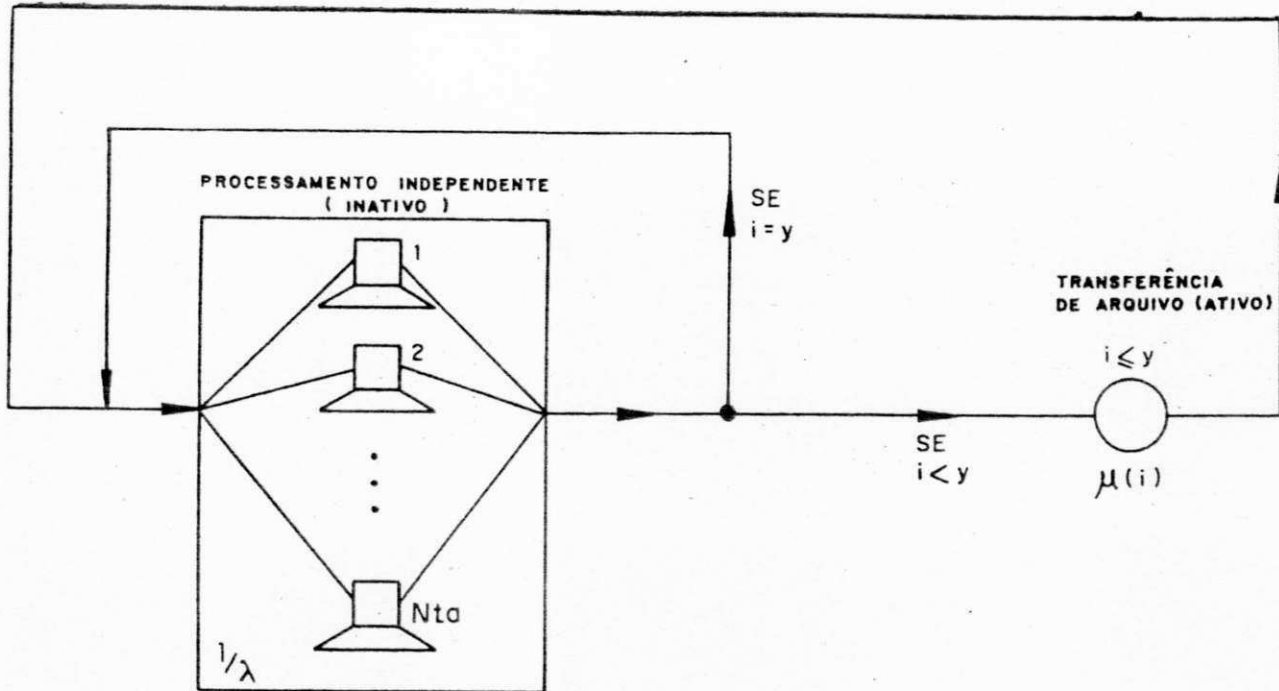


Fig. 3.6 - Modelo de filas com bloqueio para TA no SA

Um novo pedido de conexão do usuário de TAs, é rejeitado quando o SA já estiver manipulando as "y" conexões, então, ocorre um retorno ao estado de PI, ver figura 3.6. Este procedimento é válido aos ambientes 1 e 2, e independe do número de usuários presentes em PTs interagindo com o SA. Esta operacionalidade do SA, foi descrito na seção 2.4.

Da figura 3.1, tem-se que as medidas de interesse final são: $t_3^1(j, Nta)$, é o tempo médio de resposta de uma transferência completa de um arquivo, $\chi_3^1(j, Nta)$, é a taxa média de liberação de conexões no SA, dado em CVs por unidade de tempo, e $n_3^1(j, Nta)$, o número médio de transferências em andamento no SA, tudo isto para um dado j , W e Nta . A taxa média de serviço no SA é $\mu(i)$, que por sua vez é dependente do número "i" de transferências presentes no SA. Esta taxa é dada por:

$$\mu(i) = i \cdot \chi_2^1(i, j, W) \quad (3.2)$$

O valor de $\chi_2^1(i, j, W)$ aqui utilizado, é fornecido pelo nível 2 do MH, ver figura 3.1. Maiores detalhes de como se obter as outras medidas supracitadas, será assunto da subseção 3.2.3. Com a descrição deste nível para TA, finaliza-se a apresentação dos três níveis do modelo hierárquico. Na seção a seguir, serão apresentadas as soluções dos níveis aqui descritos.

3.2 - Solução do Modelo Hierárquico

Nestes três níveis propostos, um método de soluções exatas ou aproximadas será utilizado. Esta é uma das idéias da MH, uma solução apropriada a cada tipo de problema. Para o modelo conceitual em questão (ilustrado na figura 3.1), uma solução híbrida (mista) [SCHW 78] será aplicada ao problema, como uma conjunção de dois ou mais métodos, de naturezas diferentes, em busca de uma solução eficiente/adequada ao modelo. Os métodos aqui empregados são: simulação discreta e modelagem analítica. Nesta seção, estes métodos são propostos para resolver cada um dos níveis em estudo.

3.2.1 - Solução do Nível 1

Como já citado, neste nível serão considerados apenas os tempos de transmissões dos pacotes no meio físico, t_1^1 para TAs, e t_1^2 para PTs. Este atraso "t", é dado pela razão entre o comprimento do pacote "L", e a velocidade de transmissão da sub-rede, "C".

Em TAs, os pacotes de reconhecimento estão contabilizados juntos aos pacotes de dados, de tal forma, que o tempo médio efetivo de um pacote de TAs no servidor da sub-rede, t_1^1 , é dado por:

$$t_1^1 = X_d + X_a \quad (3.3)$$

Onde X_d e X_a , são os tempos médio de transmissão de um pacote de dados e reconhecimento ("acknowledge"), respectivamente. Caso $\bar{W} > 1$, implica que X_a será uma parte proporcional de um pacote de reconhecimento distribuído entre os \bar{W} pacotes de dados.

Em PTs, os pacotes de "inquiry" e "response", são somados a fim de fornecerem o tempo médio de serviço efetivo no servidor da sub-rede para uma transação completa, este tempo t_1^2 é dado por:

$$t_1^2 = X_i + X_r \quad (3.4)$$

Onde X_i e X_r são os tempos médios de transmissão dos pacotes de pergunta e resposta de uma transação, respectivamente.

Algumas considerações são necessárias para que as expressões (3.3) e (3.4) sejam válidas, são elas:

- a) Inicialmente se observa a independência do número de interfaces desejando o acesso ao meio de transmissão. Em [MOUR 82], é mostrado que este atraso devido à concorrência das interfaces desejando transmitir seus pacotes (tempo de acesso ao meio físico), é pequeno em relação ao atraso de

transmissão dos mesmos.

- b) Atrasos devidos à propagação de sinais, retransmissões, erros, etc, são pequenos em relação aos valores finais da aplicação. Concluindo, há alguns atrasos que não são computados.

Outros aspectos a respeito do atraso na sub-rede poderiam ser levantados a fim de melhorarem a precisão do modelo neste nível. Em [BUX 81b, SAST 85], são abordados alguns itens que influenciam no desempenho da sub-rede. Mas, o atraso mais significativo na transmissão dos pacotes no meio físico, é aqui considerado.

Existem " $i+j$ " estações de trabalho em TAs e PTs, desejando enviar seus pacotes através do meio físico, ou seja, há uma concorrência destas " $i+j$ " interfaces no acesso/uso da sub-rede. Sendo assim, um modelo com um servidor único se viabiliza na abstração das características do uso compartilhado da sub-rede. Este sistema de filas foi utilizado em [BUX 81b], e é representado na figura 3.7.

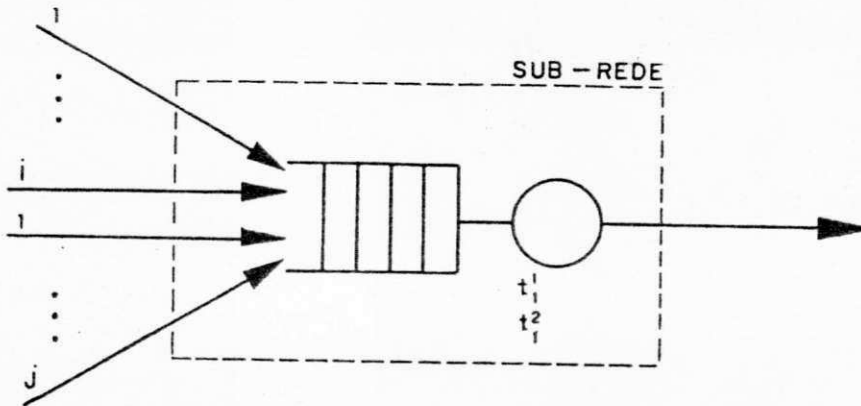


Fig. 3.7 - Modelo de um servidor único representando o meio físico

Como se observa nesta figura, existe um servidor único com dois tempos de serviços distintos, um para cada tipo de pacote. A disciplina de serviço adotada no servidor vai caracterizar o protocolo de acesso ao meio e a topologia da RL em consideração [BUX 81b]. A disciplina de atendimento neste servidor (e também nos demais servidores dos modelos das figuras 3.4 e 3.5) é o FCFS sem prioridades para as duas classes de pacotes. Independente do tipo de pacote que se apresentar na interface para transmissão, será atendido na sua ordem cronológica de chegada. A disciplina no servidor aqui adotada, é adequada à RLs, onde os acontecimentos dos eventos ocorrem sequencialmente no tempo [MUSS 85]. Quanto à solução deste sistema de filas com duas classes de fregueses (pacotes), é um tanto complexa, pois envolve uma dependência de estados das "i+j" estações de trabalho e seus pacotes. Considerando as figuras 3.4 e 3.5, o servidor sub-rede aqui discutido faz parte de um grande sistema de filas fechadas, de difícil solução. Há ainda o envolvimento dos servidores das estações dos usuários, e SAs. Nestas circunstâncias, esta será uma questão a ser resolvida no nível 2 com uma técnica adequada.

Finalizando esta subseção, resta dizer que os pacotes de TAs têm um atraso médio de transmissão no canal de comunicação igual a t_1 . No caso de PTs, este atraso é t_2 . Na subseção a seguir, é apresentada a solução do nível 2, que resolve a questão deste atraso total dos pacotes na sub-rede.

3.2.2 - Solução do Nível 2

O sistema a ser resolvido neste nível é mostrado nas figuras 3.4 e 3.5, e refere-se aos ambientes 1 e 2, respectivamente. Da subseção anterior, o servidor da sub-rede tem os tempos médio de serviços fornecidos pelo nível 1, t_1 e t_2 . Os tempos de serviço nos servidores da estação do usuário e SA, serão típicos para cada aplicação. Nestes servidores, a disciplina de serviço também é o FCFS. Uma suposição bastante razoável, realística, e não comprometedora ao mérito desta modelagem.

Para que este sistema de filas com múltiplas cadeias fechadas admita uma solução sob a forma de produto [BASK 75], algumas suposições se fazem necessárias, tais como:

- a) Tempo de atendimento nos servidores possuem uma distribuição exponencial.
- b) A atividade dos servidores devem ser independentes entre si.
- c) Os servidores possuem atendimento do tipo FCFS.

Estas são as principais premissas para que exista uma solução exata ou aproximada. Na referência [KELL 75], há um tratamento analítico sobre o assunto, fornecendo uma idéia da complexidade destes tipos de sistemas. Em [SAUE 81], é feita uma análise sobre os principais métodos analíticos para resolução de sistemas de filas fechadas.

Nas referências [MOUR 82, MUSS 85], um sistema de filas semelhante ao aqui descrito, foi resolvido através da Análise do Valor Médio (AVM) aproximado [REIS 79, REIS 80]. Devido à extensibilidade deste algoritmo heurístico a duas classes de fregueses (pacotes) em sistemas de múltiplas cadeias fechadas, o AVM foi inicialmente empregado em [SA 87a]. Mas, validando os resultados obtidos com a simulação discreta, constatou-se sensíveis erros quando a população das cadeias eram maiores que um ($W > 1$). Devido a esta imprecisão do AVM, abandonou-se este algoritmo em função do grande número de cadeias e populações, com duas classes de fregueses, a serem analisadas. Vale ressaltar que o AVM aproximado é uma poderosa ferramenta na análise de sistemas de múltiplas filas fechadas. Diante desta ineficiência do AVM na análise do trânsito dos pacotes entre as estações dos usuários e SA, decidiu-se pelo emprego da simulação discreta, para isto foi construído um programa simulador na linguagem GPSS [STRA 84]. O código do programa fonte para o ambiente 1, encontra-se nos anexos deste trabalho. Para o ambiente 2, o simulador é adaptado com pequenas simplificações, resultando em um programa mais

simples. A respeito das medidas de desempenho obtidas neste nível, tem-se as seguintes para cada uma das aplicações:

1 - Transferências de Arquivos

A medida de interesse neste nível é $\chi_2^1(i, j, W)$. Para isto, precisa-se encontrar $f_2(i, j, W)$ e $g_2(i, j, W)$ que está relacionado com $\chi_2^1(i, j, W)$ através da equação 3.1.

Uma cadeia é definida pelos servidores da estação do usuário (fonte geradora de pacotes), sub-rede, e SA (receptor dos pacotes). Um caminho lógico através destes três servidores constitui-se numa cadeia simples, abstraindo um CV.

O conjunto de servidores da cadeia "c", é definido por $SRV(c) = \{ \text{usuário}_1, \text{sub-rede}, \text{SA} \}$, onde $c = 1, 2, 3 \dots i$, e $i = 1, 2, 3 \dots y$. Definindo $ts(c)(i, j, W)$ como sendo o atraso médio de um freguês (pacote) da cadeia "c", no servidor $s(c) \in SRV(c)$ dado i, j , e W , para qualquer $c = 1, 2, 3 \dots i$. Sendo assim o valor de $f_2(i, j, W)$ é dado por:

$$f_2(i, j, W) = \sum_{s(c) \in SRV(c)} ts(c)(i, j, \vec{W}_c) \quad (3.5)$$

Onde $\vec{W}_c = (W_1, W_2, W_3, \dots, W_c = 1, \dots, W_i)$, é o vetor população das "i" conexões em TAs, com $W_c = 1$ na cadeia em questão. Lembrando que se está analisando o atraso do primeiro pacote na transferência do arquivo, numa cadeia em particular. As demais cadeias, permanecem com o mesmo W . Esta simplificação é devido às características de que todos CVs são idênticos. Esta questão já foi citada anteriormente. Quanto ao valor de $g_2(i, j, W)$, é dado por:

$$g_2(i, j, W) = \frac{\sum_{s(c) \in SRV(c)} t_{s(c)}(i, j, \vec{W})}{W} \quad (3.6)$$

Similar a $f_2(i, j, W)$, com $c = 1, 2, 3 \dots i$, e $i = 1, 2, 3 \dots y$, para um dado j e W . O tempo médio de circulação de qualquer pacote na cadeia "c" é igual a $\sum_{s(c) \in SRV(c)} t_{s(c)}(i, j, \vec{W})$, aqui o vetor população W , é dado por: $\vec{W} = (W_1, W_2, W_3, \dots, W_i)$. Cada índice de W , corresponde a população de cada cadeia de TAs, como se observa, estas "i" cadeias de TAs possuem o mesmo W . Sendo assim, uma outra definição para $f_2(i, j, W)$, seria o tempo médio do primeiro pacote de dados a chegar a seu destino, e um reconhecimento de retorno. Enquanto $g_2(i, j, W)$, é o intervalo médio entre-chegadas de pacotes num dado servidor, numa cadeia "c", dado i, j , e W .

2 - Processamento de Transações

Nesta aplicação, há apenas um pacote (frequês) circulando na cadeia "c", isto é, entre a estação do usuário e SA, via sub-rede. Existem "j" cadeias representando os caminhos lógicos entre os usuários e SA, onde $j = 1, 2, 3 \dots Npt$.

Uma viagem de ida e volta do pacote ao SA, define o tempo de resposta de uma transação completa. Pode existir o caso onde um certo número destas viagens, define uma transação, esta situação é estudada em [MUSS 85]. Neste trabalho, será excluído (não considerado) o tempo de pensamento ("think time") para os usuários de PTs. Esta consideração do intervalo de tempo entre uma transação e outra ser igual a zero, tem dois objetivos: um deles é sobrecarregar o sistema com PTs (análise de um pior caso), e o outro, é tornar esta abordagem extensível ao modelo utilizado em [MUSS 85].

Devido à simplicidade desta aplicação, a existência de apenas um pacote circulando na cadeia, e o SA com número ilimitado para este tipo de acesso, medidas finais de interesse já podem ser obtidas. O tempo médio de resposta para esta aplicação, $t_2^{(i,j,W)}$, é igual ao tempo médio de circulação do pacote na cadeia "c", $t_{s(c)}^{(i,j,W)}$, e é dado por:

$$t_2^{(i,j,W)} = \sum_{s(c) \in SRV(c)} t_{s(c)}^{(i,j,W)} \quad (3.7)$$

onde $c = 1, 2, 3 \dots j$, e $j = 1, 2, 3 \dots Npt$.

O índice W refere-se ao vetor população das "i" conexões em TAs que estão presentes. O valor de $\gamma_2^{(i,j,W)}$ é facilmente obtido a partir da Lei de Little:

$$\gamma_2^{(i,j,W)} = \frac{W}{t_2^{(i,j,W)}} \quad (3.8)$$

como $W = 1$

$$\gamma_2^{(i,j,W)} = \frac{1}{t_2^{(i,j,W)}} \quad (3.9)$$

Neste nível, novas medidas interessantes podem ser obtidas, como por exemplo, a taxa de utilização do servidor da sub-rede, ρ . Há uma contribuição do tráfego das duas aplicações, portanto, ρ é definido como:

$$P = P_1 + P_2 \quad (3.10)$$

onde P_1 e P_2 são contribuições dos fluxos de pacotes das aplicações 1 e 2 respectivamente, empregando diretamente a Lei de Little, suas expressões são:

$$P_1 = i \cdot np \cdot \chi_2^1(i, j, W) \cdot t_1^1 \quad (3.11a)$$

e

$$P_2 = j \cdot \chi_2^2(i, j, W) \cdot t_1^2 \quad (3.11b)$$

Observar que todas as "i" cadeias em TAs presentes, junto as "j" cadeias em PTs contribuem diretamente aos valores de P_1 e P_2 , respectivamente. Desta maneira, investiga-se o percentual de uso no meio físico devido as aplicações 1 e 2. Esta análise é extensível e válida para o SA. Na subseção a seguir, serão mostradas as soluções no nível 3 para TAs.

3.2.3 - Solução do Nível 3

Nesta subseção, se apresenta uma abordagem quanto à questão do SA limitado ao número de conexões lógicas para TAs. O estudo que se segue, é baseado na figura 3.6 da subseção 3.1.3. Nesta figura, se ilustra os dois estados possíveis que um usuário em TA pode se encontrar.

Estamos supondo que o tempo de permanência no estado P1, é um tempo com uma distribuição exponencial com média igual a $1/\lambda$. Recordando a aproximação do nível 2, que todos os servidores possuem um tempo médio de serviço com uma distribuição

exponencial. Conclui-se que, o tempo médio de duração de uma conexão (uma transferência completa) com o SA, também possui esta distribuição; pois ocorre uma superposição de funções de distribuição de probabilidades.

Estes são os dois estados que um usuário pode se encontrar, em PI ou realizando uma transferência. Ambos possuem um tempo médio de permanência em seus estados, com uma distribuição exponencial e independentes entre si. São suposições que nos levam a uma tratabilidade matemática.

A taxa de partida do estado PI, para um pedido de conexão ao SA é λ , e ocorre se $i < y$. Neste caso, o pedido será aceito e uma transferência será processada. Quanto à taxa de retorno no sentido SA ao estado PI, ou seja, a taxa de liberação de CVs, tem uma frequência igual a $\mu(i)$. O valor de $\mu(i)$ é dado pela equação (3.2).

A fim de se encontrar $t_3^1(j, Nta)$, $\gamma_3^1(j, Nta)$ e $n_3^1(j, Nta)$, precisa-se encontrar $p_{(i)}$, que é a probabilidade de equilíbrio de "i" transferências presentes no SA. Pelo fato de se ter um sistema de dois estados, cujos tempos médios de permanências têm distribuição exponencial, e com um número finito de estados, uma solução através das cadeias de Markov [KLEI 75, SAUE 81] é adequada na busca de uma expressão para $p_{(i)}$.

Com base na figura 3.6, há duas situações a se analisar:

- 1 - Quando $Nta < y$, isto é, não há bloqueio de novos pedidos.
- 2 - Quando $Nta > y$, ocorrerá bloqueio de um novo pedido a partir de $i = y$.

Construindo diagramas de transição de estados para estes dois casos, e utilizando os conceitos de cadeias de Markov, nos leva a figura 3.8, que ilustra os possíveis estados deste sistema.

Montando-se as equações de balanço global, e igualando-se os fluxos das entradas e saídas de cada estado, encontra-se $p_{(i)}$, obtido através da seguinte expressão:

$$P_{(i)} = P_{(0)} \binom{N_{to}}{i} \cdot \lambda^i \cdot \prod_{l=1}^i \frac{1}{\gamma_2^l(l, j, W)} \quad (3.12)$$

para $i = 0, 1, 2, 3 \dots \min(y, N_{to})$.

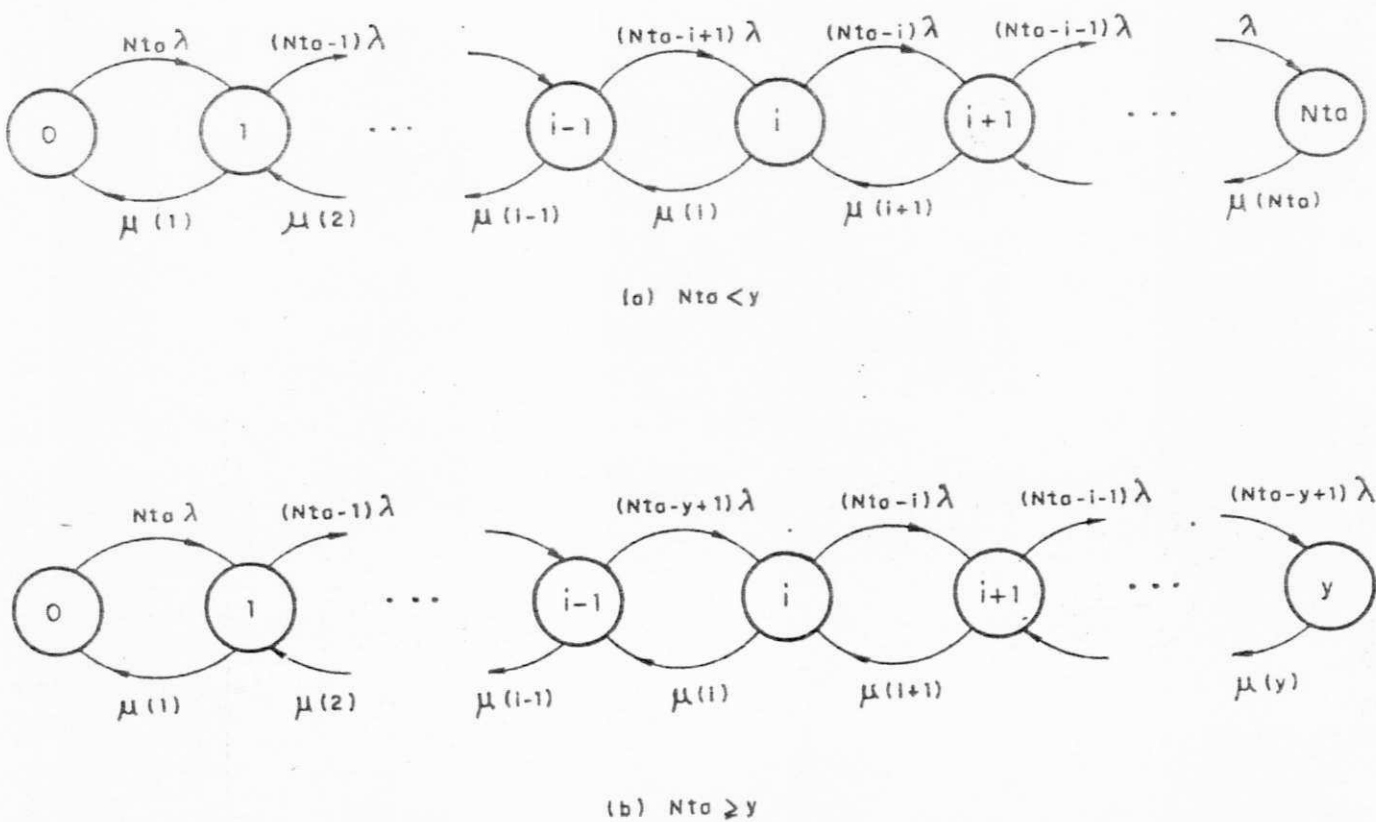


Fig. 3.8 - Diagrama de transição de estados do sistema de bloqueio no SA para TAs

Calculando-se a constante de normalização $p_{(0)}$, temos:

$$p_{(0)} = \left[\sum_{i=0}^{\min(y, Nta)} \binom{Nta}{i} \cdot \lambda^i \cdot \prod_{l=1}^i \frac{1}{\delta_2^1(l, j, W)} \right]^{-1} \quad (3.13)$$

O número médio de transferências de arquivo em progresso (andamento) para um dado j e Nta , é calculado como se segue:

$$n_3^1(j, Nta) = \sum_{i=1}^{\min(y, Nta)} i \cdot p(i) \quad (3.14)$$

Similarmente, a vazão média no SA para TAs é encontrada por:

$$\gamma_3^1(j, Nta) = \sum_{i=1}^{\min(y, Nta)} \mu(i) \cdot p(i) \quad (3.15)$$

Da Lei de Little, o tempo médio de resposta para uma transferência completa, é alcançado por:

$$t_3^1(j, Nta) = \frac{n_3^1(j, Nta)}{\gamma_3^1(j, Nta)} \quad (3.16)$$

Onde $n_3^1(j, Nta)$ e $\gamma_3^1(j, Nta)$, são dados pelas equações (3.14) e (3.15) respectivamente. Vale lembrar que, toda esta análise é válida para os dois ambientes em questão. Aqui finaliza-se o estudo de bloqueio no SA para a aplicação 1.

Os modelos aqui apresentados, junto as técnicas sugeridas para suas soluções, serão submetidos a apreciação no capítulo seguinte, encontrando resultados de desempenho das aplicações em estudo.

4 - RESULTADOS E ANALISES

Os modelos propostos para as duas aplicações, e seus ambientes, expostos no capítulo anterior, serão aqui utilizados. Como visto, dos três níveis apresentados para o modelo, são os níveis 2 e 3 que fornecem resultados de interesse final. O capítulo se inicia com uma abordagem a respeito dos valores escolhidos aos parâmetros dos modelos. Em seguida, serão reportados em gráficos os resultados do nível 2, dando ênfase a PTs, e por último, são apresentados os resultados para TAs no nível 3. Uma análise sobre os gráficos reportados acompanha esta exposição.

4.1 - Escolha dos Parâmetros

Algumas premissas devem ser levadas em conta quando se deseja escolher valores dos parâmetros de um modelo. Fazendo uma escolha arbitrária, resultados estranhos e imprevisíveis podem ser encontrados inesperadamente. Deve existir uma verdadeira relação entre o sistema real e o modelo, isto é, atribuir valores aos parâmetros de forma adequada. A idéia é que exista coerência entre a parte funcional do sistema real, e o modelo em estudo.

Há algumas dificuldades inerentes à atribuição de valores às entidades do modelo. O avanço tecnológico de software e hardware em RLs, faz com que certos valores se tornem obsoletos com o passar do tempo. Uma desatualização que se está sujeito diariamente. Uma tentativa é feita no sentido que esta análise seja abrangente e realista. A idéia é que se tenha valores de comum acordo com as muitas propostas existentes [SPEC 82, MOUR 82, MUSS 85, MITC 82, EISE 85, LAZO 86, BUX 84, CHRI 84], e que respeite-se as considerações feitas aos modelos dos sistemas estudados. Estas referências supracitadas, fornecem uma base sobre valores dos atrasos por pacotes nas estações de trabalho,

na sub-rede, e SAs.

Verifica-se uma dificuldade em se encontrar um valor de consenso as diversas propostas. Esta divergência muitas vezes é gerada por diferentes considerações na medida de um valor prático. Neste trabalho, utiliza-se valores de aprovação com as referências citadas, e respeitando as considerações do modelo.

De acordo com as figuras 3.4 e 3.5, há três tipos de servidores: as estações de trabalho do usuário, a sub-rede e SA. Relembrando que há duas classes de pacotes circulando através deste sistema de filas. Portanto, em alguns servidores os atendimentos são distintos, e isto ocorre devido às naturezas das aplicações. No servidor do SA, um pacote para TA sofre um atraso médio igual a 50 ms, e para PT, o atraso médio por pacote é igual a 30 ms. Esta escolha tem por base [MOUR 82], e estes atrasos estão em 50% abaixo do sugerido em [TOKU 84], mas dentro de limites propostos em [EISE 85, TANE 85].

Quanto ao servidor da sub-rede, existem alguns dados mais concretos a respeito da velocidade de transmissão no meio físico, "C". Para um meio físico do tipo par trançado, sua velocidade típica varia de 50 a 250 Kbps. No caso de cabo coaxial, pode variar de 100 Kbps a 10 Mbps. Neste trabalho, o valor utilizado para "C" foi igual a 100 Kbps e 1 Mbps. Em [SA 87b], foi investigado o caso onde "C" é igual a 10 Mbps, e foram encontrados níveis de desempenho praticamente idênticos a "C" igual a 1 Mbps.

No tocante ao atraso médio por pacote nos servidores das estações de trabalho dos usuários em TAs, este atraso é igual a 200 ms. Nas estações dos usuários de PTs, este atraso é igual a 400 ms, pois considera-se as duas passagens dos pacotes para esta aplicação (um pacote de ida ao SA, e outro de resposta). Este valor é utilizado em [MOUR 82] para TAs, que por sua vez se encontra próximo ao usado em [MUSS 85] para operações

interativas. Relembrando que este atraso nas estações de trabalho, contabiliza os atrasos devido a: protocolos de alto-nível, interfaces, sincronismos, conversores, etc.

Referente ao comprimento dos pacotes que trafegam na rede, há alguns fatores como: a tecnologia em uso pela RL (seus protocolos, sua topologia, buffers de interface, etc), as estações de trabalho, e SAs. Algo operacionalmente viável é trafegar pacotes de tamanho igual a unidade de manipulação de dados do SA. Estas unidades são chamadas em alguns sistemas de "páginas", que possuem valores típicos de 128 bits até 2 Kbytes [MITC 82]. Para TAs, o comprimento médio de um pacote (L1) é igual a 1000 bytes (8000 bits). O tamanho médio de um pacote de PT (L2), é igual a 125 bytes (1000 bits). Recordando que este valor equivale ao pacote de pergunta ao SA, e outro de retorno como resposta. Sendo assim, um pacote de PT para uma transação completa, é igual a 12,5% do tamanho de um pacote de TA.

Para TAs, temos ainda: tamanho de janela W, igual a 1 e 4; o tamanho médio por arquivo é igual a 40.000 bytes, portanto, "np" é igual a 40, uma vez que um pacote para TA é igual a 1000 bytes [MOUR 82]. O tempo de pensamento ($1/\lambda$) utilizado no nível 3, será igual a cinco (5) segundos, e número máximo de conexões simultâneas (γ), é igual a dez (10) [MOUR 82].

Quanto a simulação realizada no nível 2, os valores médios apresentados têm um nível de confiança igual a 90%. Há uma oscilação média de $\pm 5\%$ sobre os valores reportados. Em outras palavras, 90% das amostras obtidas, se encontram dentro do limite de tolerância de $\pm 5\%$ do valor médio, tornando o intervalo de confiança dos resultados igual a 10%. A escolha deste grau de precisão, tem por base [SAUE 81].

Da exposição acima, tem-se uma idéia das grandezas de valores dos parâmetros para os ambientes 1 e 2. Vale ressaltar que novos valores dos parâmetros podem tender e/ou

particularizar os resultados a um dado sistema. A seguir, apresentaremos as medidas de performance obtidas do nível 2 para os dois ambientes em questão.

4.2 - Resultados do Nível 2

Neste nível, a aplicação 2 é concluída com medidas de interesse final. É neste nível que com a simulação discreta se capta o atraso total do pacote na sub-rede. No nível 1 só foi considerado o atraso de transmissão por pacote, ou seja, o tempo de serviço no servidor sub-rede das figuras 3.4 e 3.5. Mas não é este atraso do pacote na sub-rede (devido a fila e serviço no servidor da sub-rede) que nos interessa, e sim resultados finais das aplicações.

Para PTs, a figura de performance apresentada é $t_2^2(i, j, W)$, que é o tempo médio de resposta de uma transação. Esta medida foi escolhida pela sua representatividade de uma boa medida de desempenho para avaliar aplicações em RLs. Outras medidas poderiam ser obtidas diretamente através do simulador, como por exemplo: o fator de utilização dos servidores, a fim de detectar algum ponto de gargalo no sistema, para estes cálculos, usa-se as expressões 3.10 e 3.11. Uma outra medida interessante, é o número médio de pacotes nas filas dos servidores, com o objetivo de realizar dimensionamentos de "buffers" nas interfaces dos dispositivos.

Em razão da grande quantidade de dados obtidos, somente os resultados mais relevantes serão apreciados. Uma curiosidade especial está em torno de valores máximos e mínimos, caracterizando assim, as melhores e piores medidas do sistema. Devido à relação direta entre $t_2^2(i, j, W)$, e $\gamma_2^2(i, j, W)$ para PTs, dado pela equação 3.9, $\gamma_2^2(i, j, W)$ terá sua apresentação dispensada.

Na figura 4.1, mostra-se $t_2^2(i,j,W)$ versus "j" para o ambiente 1, com $C = 1$ Mbps. Neste caso, todas as curvas são coincidentes, ou seja, não há interferência dos usuários em TAs (i), sobre PTs; os valores de "i" e W, não acarretam no aumento de $t_2^2(i,j,W)$. Isto deve-se ao fato dos SAs serem distintos (separados) para as duas aplicações, e a velocidade da sub-rede ser relativamente alta. Deste modo, não se verifica qualquer tipo de interferência da aplicação 1 sobre 2.

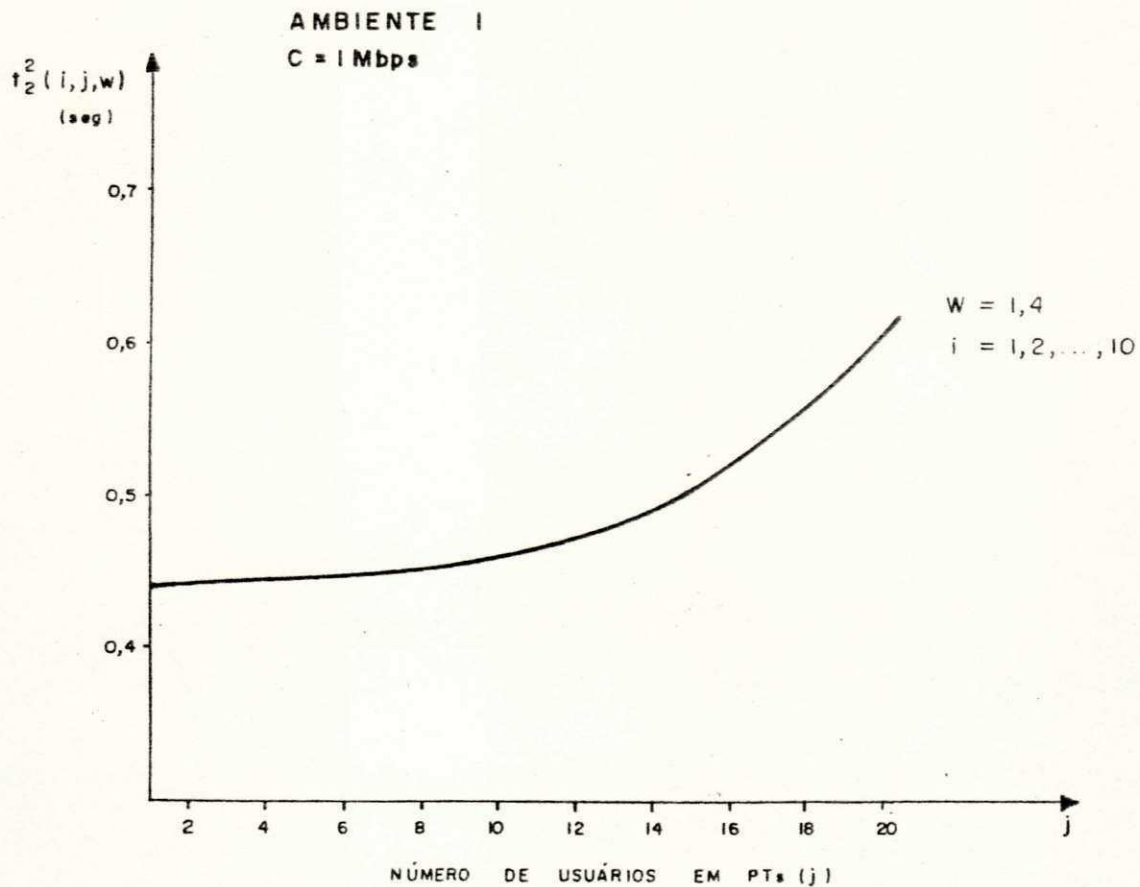


Fig. 4.1 - Tempo médio de resposta para PTs, $t_2^2(i,j,W)$

Na figura 4.2, o valor de "C" é dez vezes menor. Aqui, observa-se um aumento de $t_2^2(i,j,W)$ quando "i" e W crescem. E

redundante dizer que a variável "j", contribui para este crescimento. Nesta figura, percebe-se a interferência entre estas duas aplicações. Eis o primeiro registro do distúrbio causado pela aplicação 1, sobre 2. Aqui a sub-rede é menos veloz, e com isto os pacotes de TAs e PTs sofrerão um aumento no tempo de fila devido a se ter o servidor da sub-rede mais lento. Vale salientar que este desenho e os próximos, possuem uma escala oito vezes maior que a figura 4.1. Sendo assim, não fica muito visível o tipo do crescimento da curva para $t_2^2(i,j,W)$. Da figura 4.1, observa-se a tendência de um suave crescimento exponencial em função de i , j e W . Esta é uma constatação válida para as demais figuras de $t_2^2(i,j,W)$.

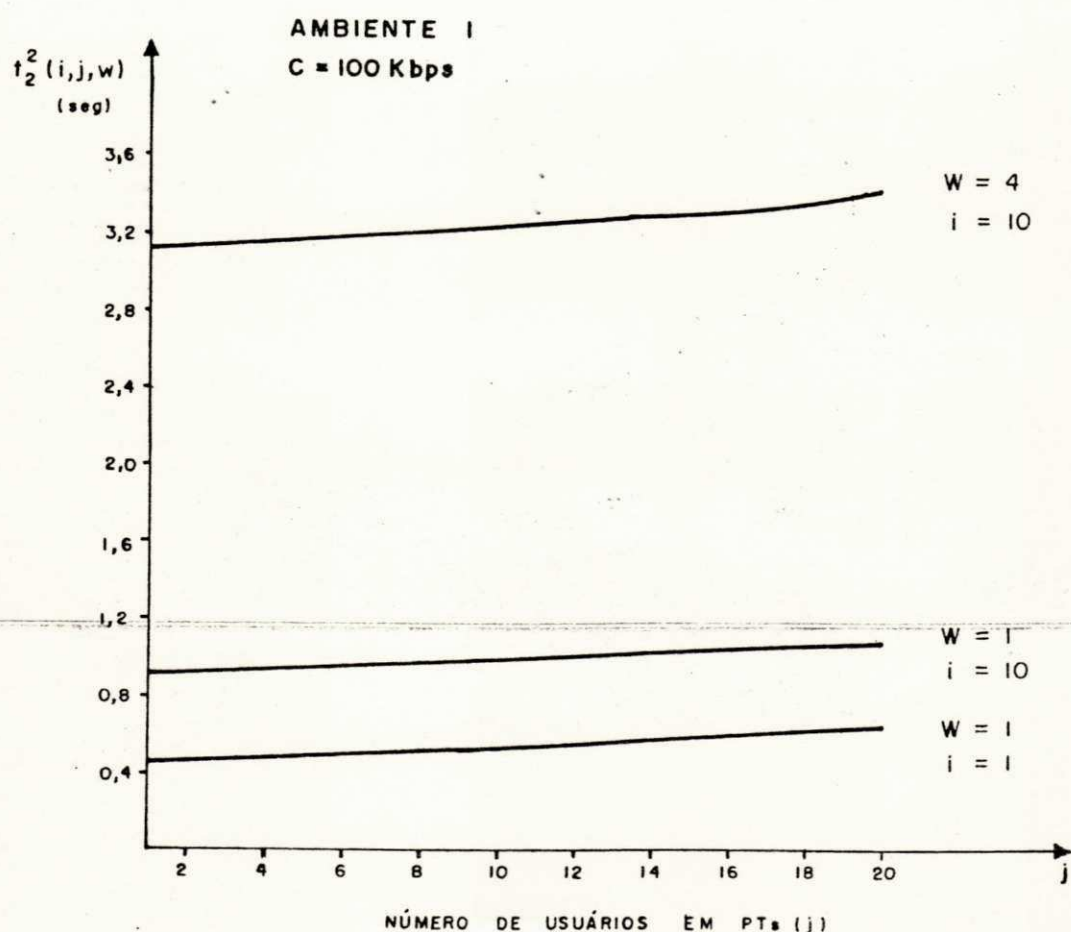


Fig. 4.2 - Tempo médio de resposta para PTs, $t_2^2(i,j,W)$

Passando para o ambiente 2, a figura 4.3 mostra $t_2^2(i,j,W)$ para $C = 1$ Mbps. Neste ambiente tem-se um SA único, concorrido por usuários em TAs e PTs. Aqui, duas facilidades (sub-rede e SA) são compartilhadas igualmente pelos dois tipos de pacotes, e garantidamente tem-se um sistema mais lento que o anterior. Com os valores de W iguais a 1 e 4, consegue-se ter uma idéia de como seriam os resultados para outros valores de W . Para curva $i = 1$ e $W = 4$, há uma coincidência com $i = 1$ e $W = 1$, por esta razão esta curva é omitida.

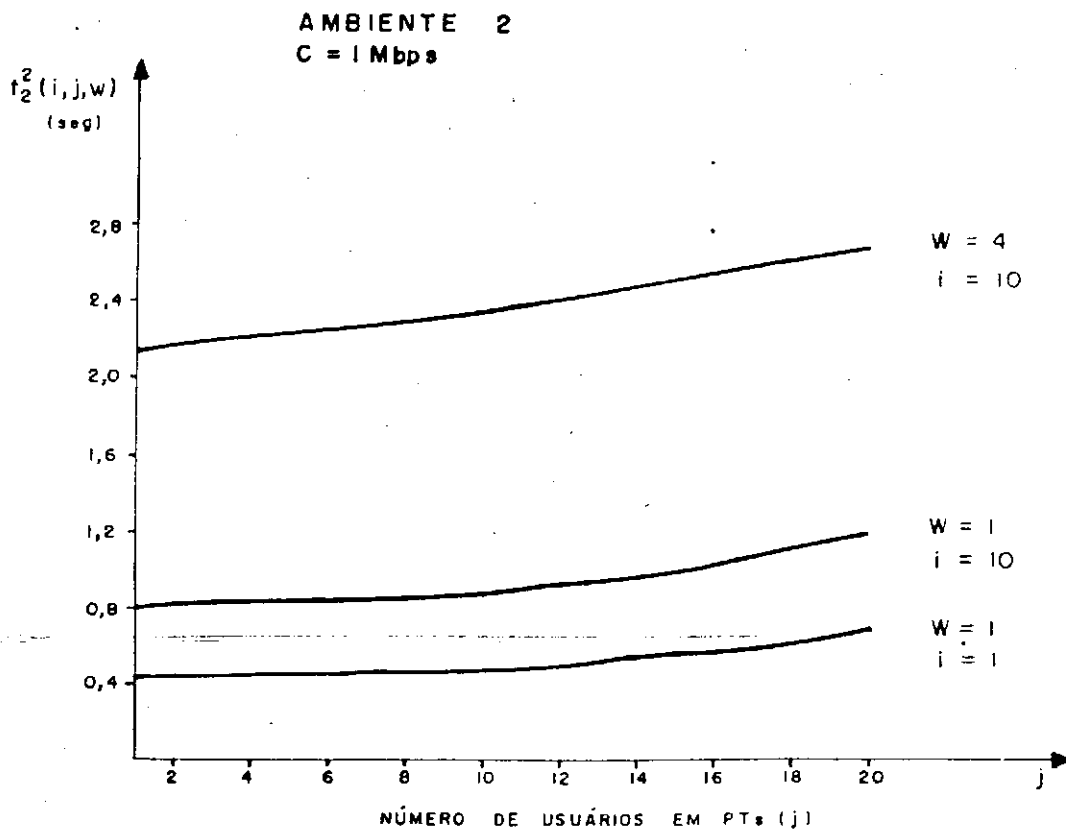


Fig. 4.3 - Tempo médio de resposta para PTs, $t_2^2(i,j,W)$

Comparando-se as figuras 4.1 e 4.3, é visível o problema de um SA centralizado para duas aplicações de naturezas distintas. Mais uma vez, vale salientar a perturbação sofrida pelos usuários em PTs (j), causada pelos usuários em TAs (i) com um dado W. O nível de crescimento destas curvas é mais acentuado que nos casos do ambiente 1. Lembrando dos tratamentos distintos que os pacotes sofrem nos servidores compartilhados, os pacotes de TAs dispendem mais tempo nos servidores, e com um W maior, implica numa queda de performance ainda maior para PTs nesta segunda configuração.

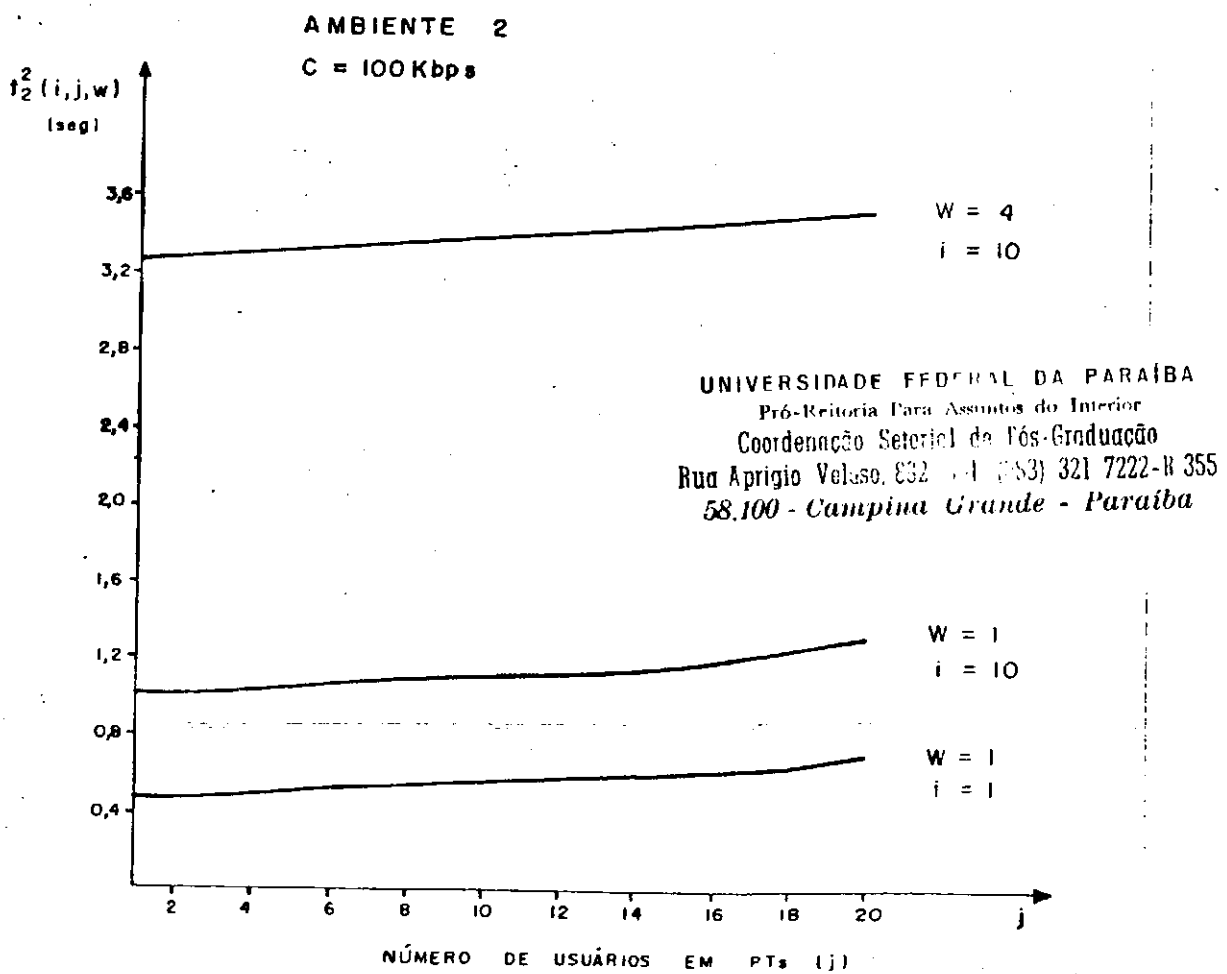


Fig. 4.4 - Tempo médio de resposta para PTs, $t_2^2(i,j,W)$

A figura 4.4 é relativa ao ambiente 2 com $C = 100$ Kbps. A velocidade da sub-rede é dez vezes mais lenta que o caso anterior. Aqui os níveis de $t_2^1(i, j, W)$ são piores que os da figura 4.3, bastante evidenciado para curva $i = 10$ e $W = 4$. É o caso onde o SA está atendendo simultaneamente a 10 (dez) TAs, com $W = 4$, e "j" usuários em PTs.

Para TAs, a medida obtida neste nível é $\gamma_2^1(i, j, W)$, número médio de TAs por segundo, para uma dada conexão lógica ou circuito virtual (CV). As figuras 4.5 e 4.6, apresentam $\gamma_2^1(i, j, W)$ para o ambiente 2, com "C" igual a 1 Mbps e 100 Kbps respectivamente.

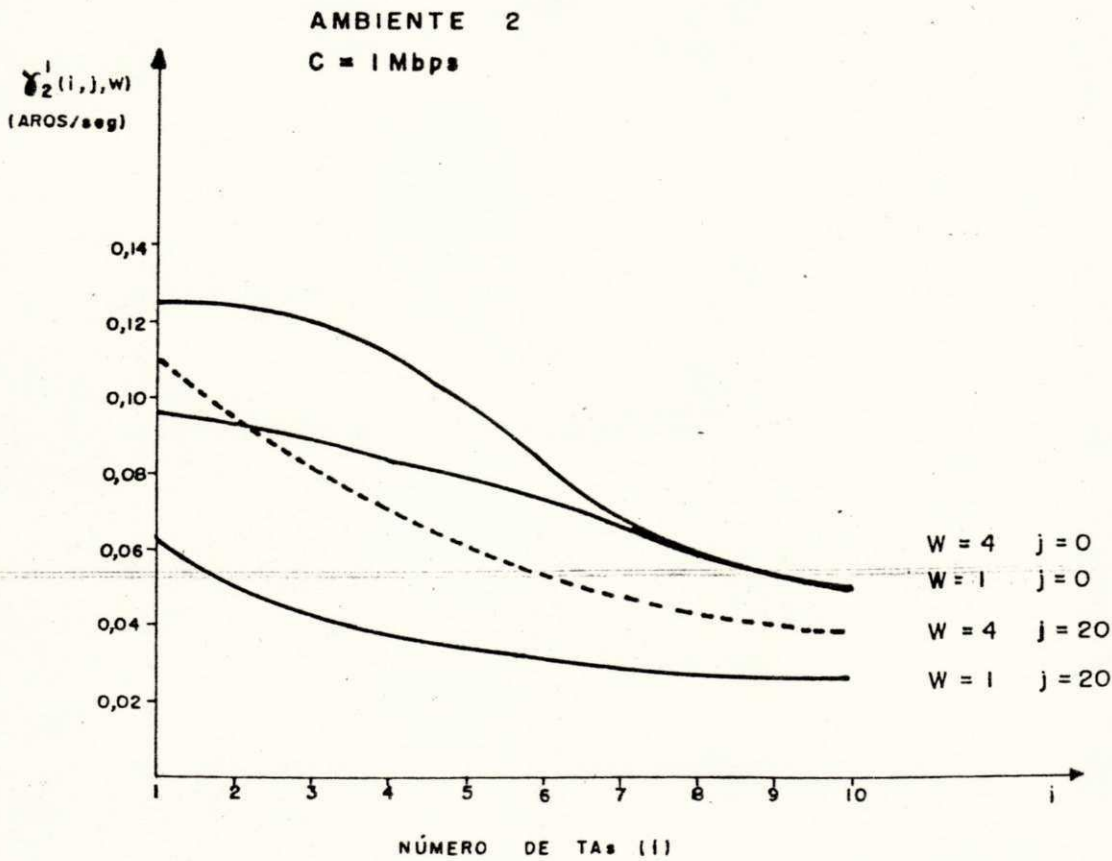


Fig. 4.5 - Vazão média de um CV para TAs, $\gamma_2^1(i, j, W)$

A curva $W = 1 : j = 20$, ilustra os piores resultados para $\gamma_2^1(i, j, W)$. Está claro que para $W = 4$, a transferência de um arquivo se processa de modo mais eficiente que $W = 1$, ou seja, para $W = 4$ são enviados 4 pacotes de dados para um pacote de reconhecimento como retorno. Há um uso mais intensificado do CV pelos pacotes de dados do arquivo. Nas curvas com $j = 0$, há melhores resultados para $\gamma_2^1(i, j, W)$ que para $j = 20$. Esta degradação de $\gamma_2^1(i, j, W)$, é devido ao aumento do número de usuários em PTs (j). Aqui, se destaca a perturbação da aplicação 2 sobre 1. A curva tracejada nas duas figuras, tem por objetivo ilustrar melhor esta degradação causada por "j".

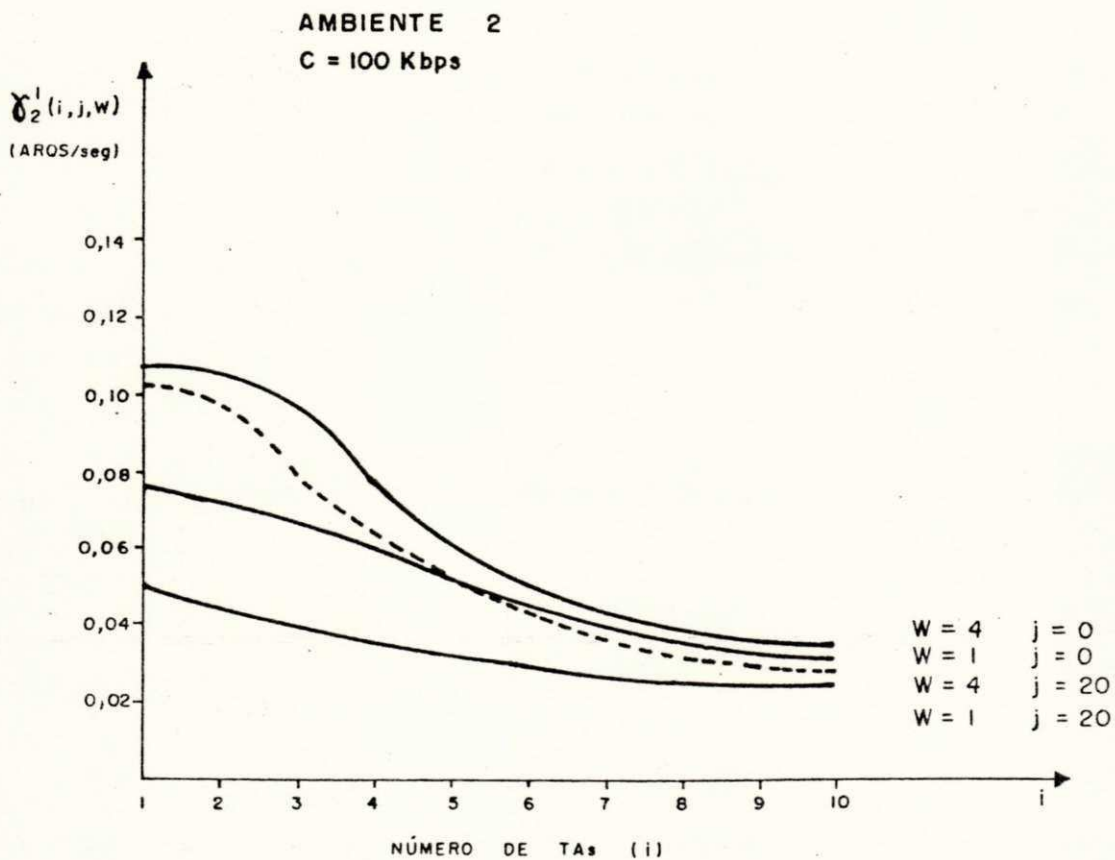


Fig. 4.6 - Vazão média de um CV para TAs, $\gamma_2^1(i, j, W)$

O cruzamento entre as curvas $W=4:j=20$ com $W=1:j=0$ nestas duas figuras, retratam inicialmente a degradação de "j" sobre as TAs. Mas observando o traçado (perfil) da curva $W=4:j=20$ nestas duas figuras, constata-se que são diferentes. O decaimento desta curva é mais acentuado ("rápido") para um "C" maior, cruzando $W=1:j=0$ na figura 4.5 num valor para "i" ($i \cong 2.3$) menor que o da figura 4.6 ($i \cong 5.1$). Isto significa que para o conjunto de valores nos parâmetros aqui utilizados, a curva $W=4:j=20$ se mostrou menos sensível no desempenho das TAs quando a velocidade da sub-rede for menor, ou seja, o seu comportamento tende acompanhar o da curva $W=4:j=0$ na figura 4.6. Algo que não ocorre na figura 4.5. A razão deste fato pode não ser tão evidente inicialmente. Sendo a sub-rede um recurso compartilhado mais veloz que o SA, esta não chega a acumular (conter, reter) os pacotes por muito tempo, a congestão de pacotes é menor que no SA. Desta forma, um "C" maior implica numa vazão maior, e conseqüentemente um acúmulo mais rápido de pacotes é registrado no SA (este servidor, provavelmente vai operar num estado de saturação), que é um servidor lento comparado a sub-rede. Portanto, este decaimento é mais acentuado (brusco) para $W=4:j=20$ na figura 4.5 devido a concorrência dos pacotes de TAs junto aos de PTs no SA, que ocorre numa taxa de chegada de pacotes maior neste servidor que o da figura 4.6. Finalizando, é bom deixar claro que é apenas o delineado da curva $W=4:j=20$ que é desigual nas figuras 4.5 e 4.6, mas os níveis de $\gamma_2^1(i,j,W)$ são melhores para um "C" maior, no caso 1 Mbps.

Como uma última observação a respeito das figuras 4.5 e 4.6, é quanto ao comportamento destas curvas para um valor de "i" muito maior que 10 (que é o valor estipulado para "y"). O que ocorre para um "i" $\gg 10$? Este aumento de "i", faz com que cresça o número de pacotes de TAs presentes no sistema (usuário, sub-rede, SA). Desta forma, todos os servidores podem atingir seus limites de saturação (problemas principalmente de filas na sub-rede e SA) com facilidade, tornando os valores finais de $\gamma_2^1(i,j,W)$ para $i \gg 10$ idênticos, isto é, independentes dos

valores de W (tamanho da janela para TAs) e " j ". Um valor muito grande de " i ", faz com que W e " j " não tenham mais influência sobre $\gamma_2^1(i, j, W)$, uma vez que há uma predominância de pacotes de TAs presentes no sistema. Vale lembrar, que esta constatação é verificável para o conjunto de valores aqui conferidos. Na seção seguinte, os resultados do nível 3 serão reportados.

4.3 - Resultados do Nível 3

Os valores de $\gamma_2^1(i, j, W)$ fornecidos a partir do nível 2, serão utilizados para encontrar medidas finais de TAs. As medidas que serão obtidas para os ambientes 1 e 2, variam em função de " j ", W e Nta . Esta afirmação pode ser verificada pelas equações da subseção 3.2.3. A medida aqui apresentada é o tempo médio de resposta de uma transferência de arquivo completa, $t_3^1(j, Nta)$. Os valores de $\gamma_3^1(j, Nta)$, e $n_3^1(j, Nta)$, são menos relevantes neste trabalho, porém obtidos com facilidade, e relacionam-se com $t_3^1(j, Nta)$, através da equação 3.16.

As figuras 4.7 a 4.10, referem-se aos ambientes 1 e 2 para " C " igual a 1 Mbps e 100 Kbps. Novamente, só foram retratadas as curvas mais representativas, isto é, com os melhores e piores resultados para $t_3^1(j, Nta)$. É interessante observar que as curvas com $j = 0$, se tornam válidas para os dois ambientes. Neste caso, onde não há usuários em PTs ($j = 0$), e apenas uma única população para TAs, faz com que os ambientes 1 e 2 fiquem idênticos.

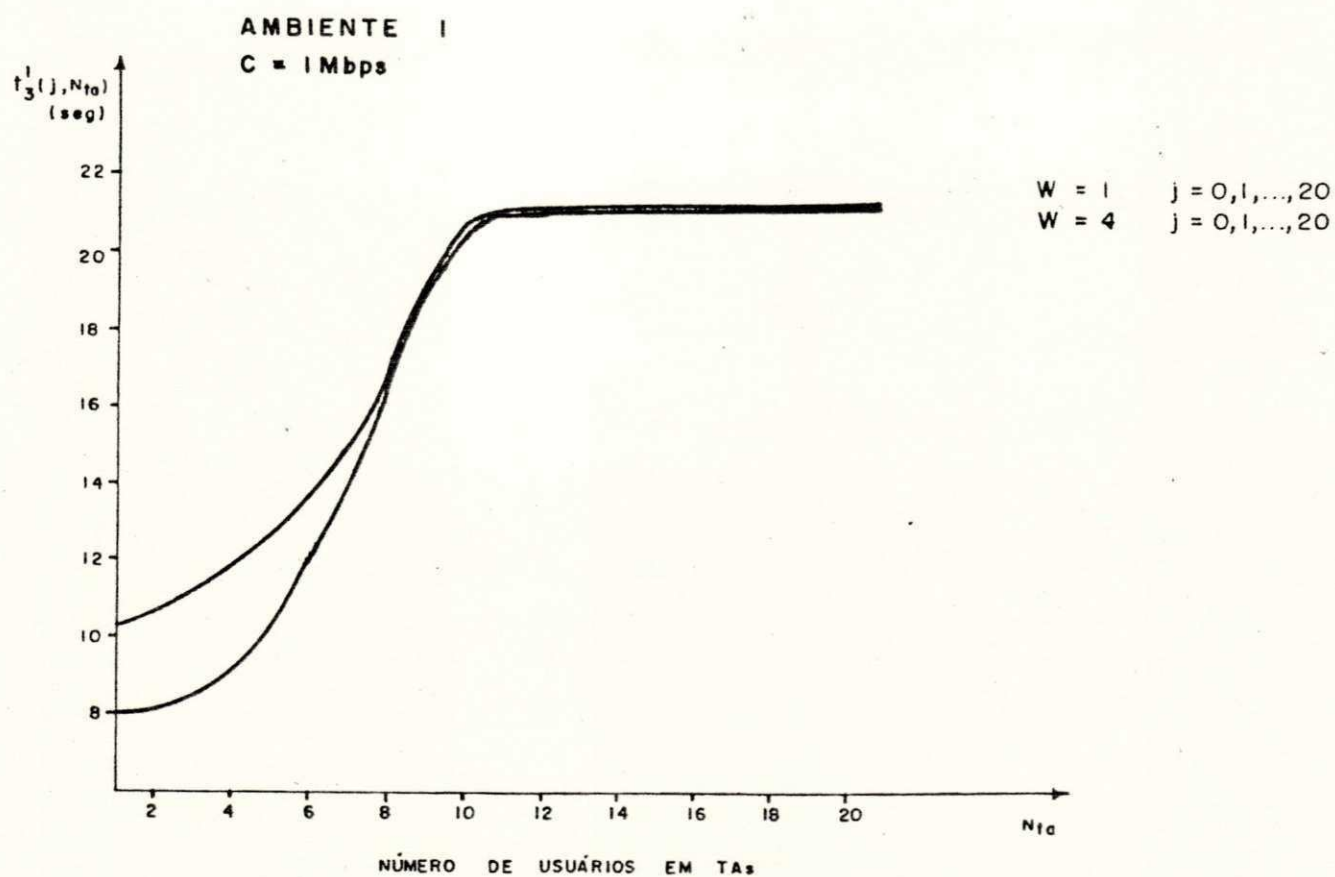


Fig. 4.7 - Tempo médio de resposta para TAs, $t_3^1(j, N_{ta})$

Em TAs, onde há um grande volume de dados a serem transportados, a velocidade da sub-rede é um parâmetro determinante no desempenho desta aplicação. Isto é comprovado para os dois ambientes, onde os melhores resultados foram verificados para $C = 1$ Mbps.

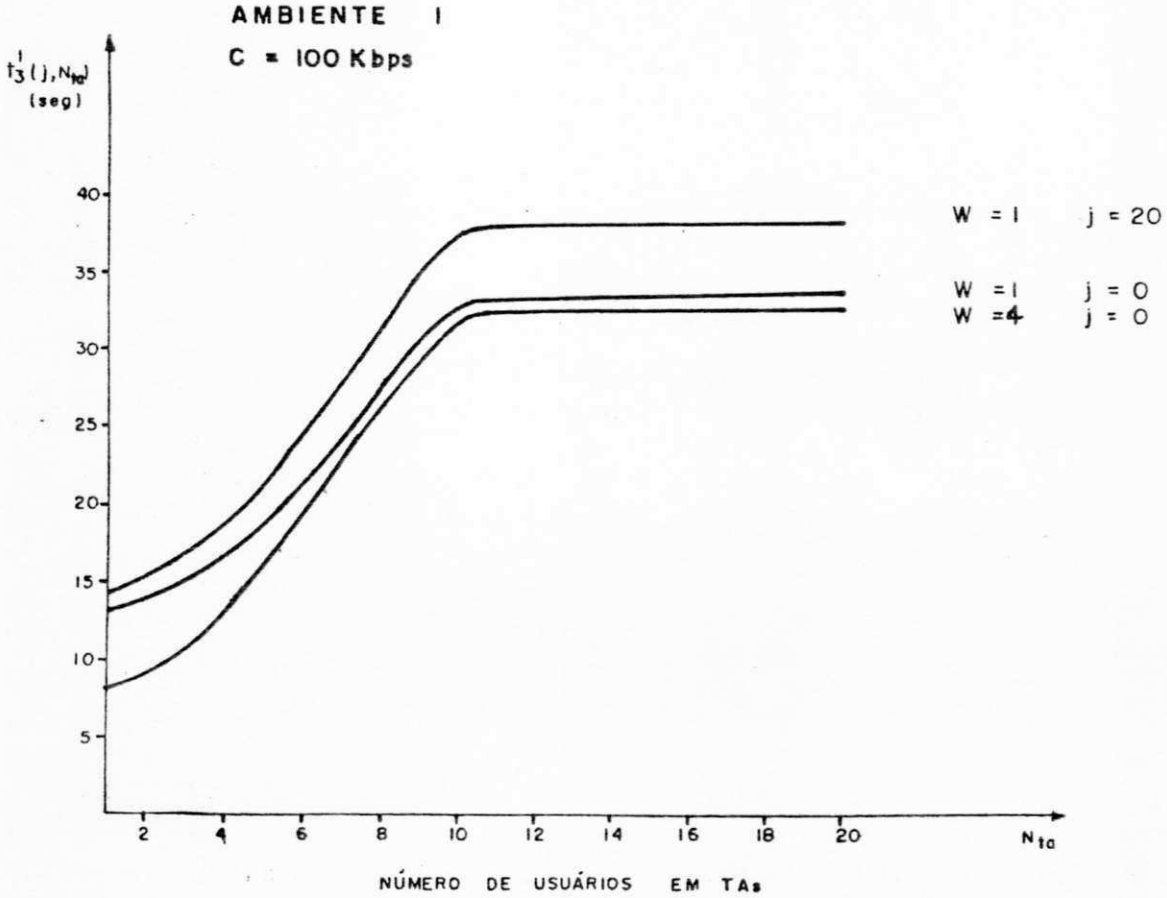


Fig. 4.8 - Tempo médio de resposta para TAs, $t_3^1(j, N_{ta})$

Na figura 4.9 ilustra a inteseção entre as curvas $W=4:j=20$ com $W=1:j=0$. Para este caso, o ponto de cruzamento entre estas duas curvas, tem como consequência os resultados do nível 2 da figura 4.5. Os valores de $\gamma_2^1(i, j, W)$ estão relacionados com $t_3^1(j, N_{ta})$, através das equações 3.12 a 3.16.

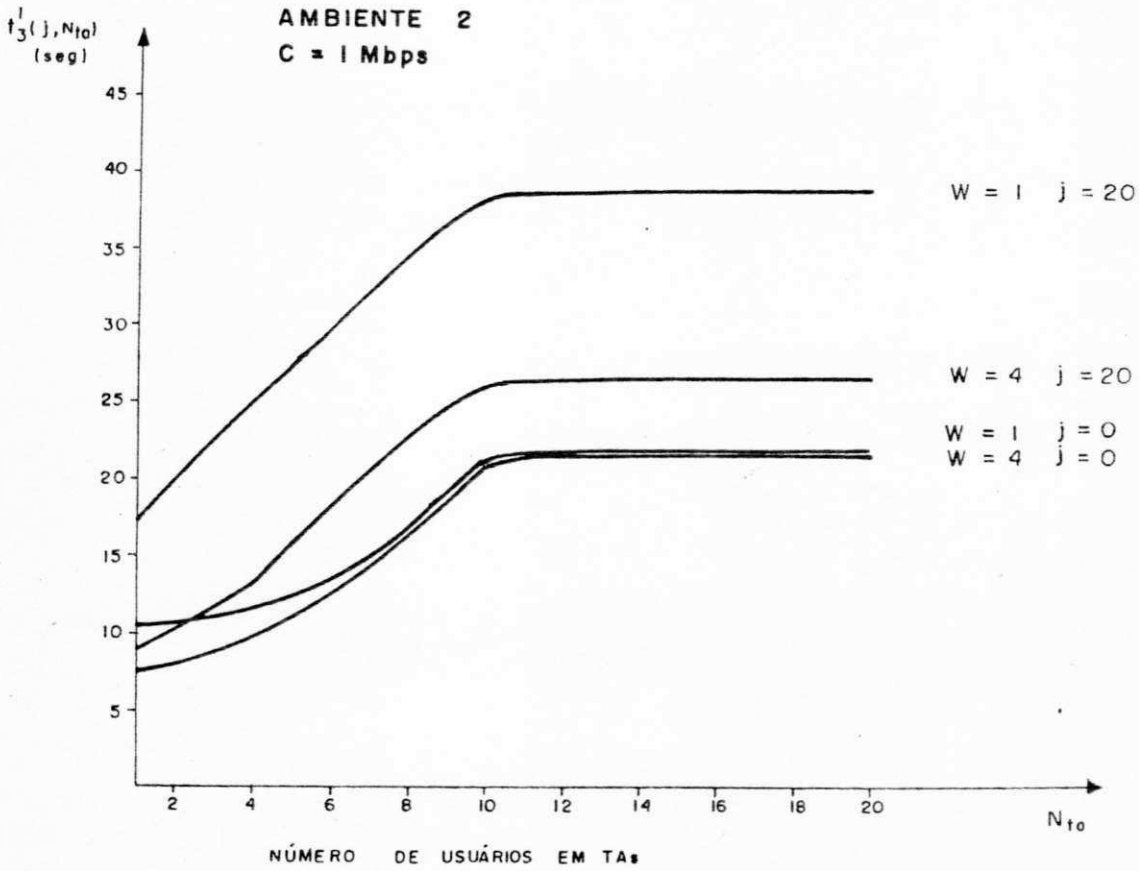


Fig. 4.9 - Tempo médio de resposta para TAs, $t_3^1(j, N_{ta})$

Outro importante parâmetro no desempenho é o valor de W . Com base nas figuras reportadas, melhores resultados foram obtidos para um W maior, ou seja, quanto maior o W , melhores são os níveis de desempenho para TAs. É o caso onde há um maior número de pacotes entre o SA e usuário, sem um reconhecimento em sentido contrário para os pacotes de dados. Evitando assim, muitas

interrupções na inversão de fluxo para o pacote de reconhecimento.

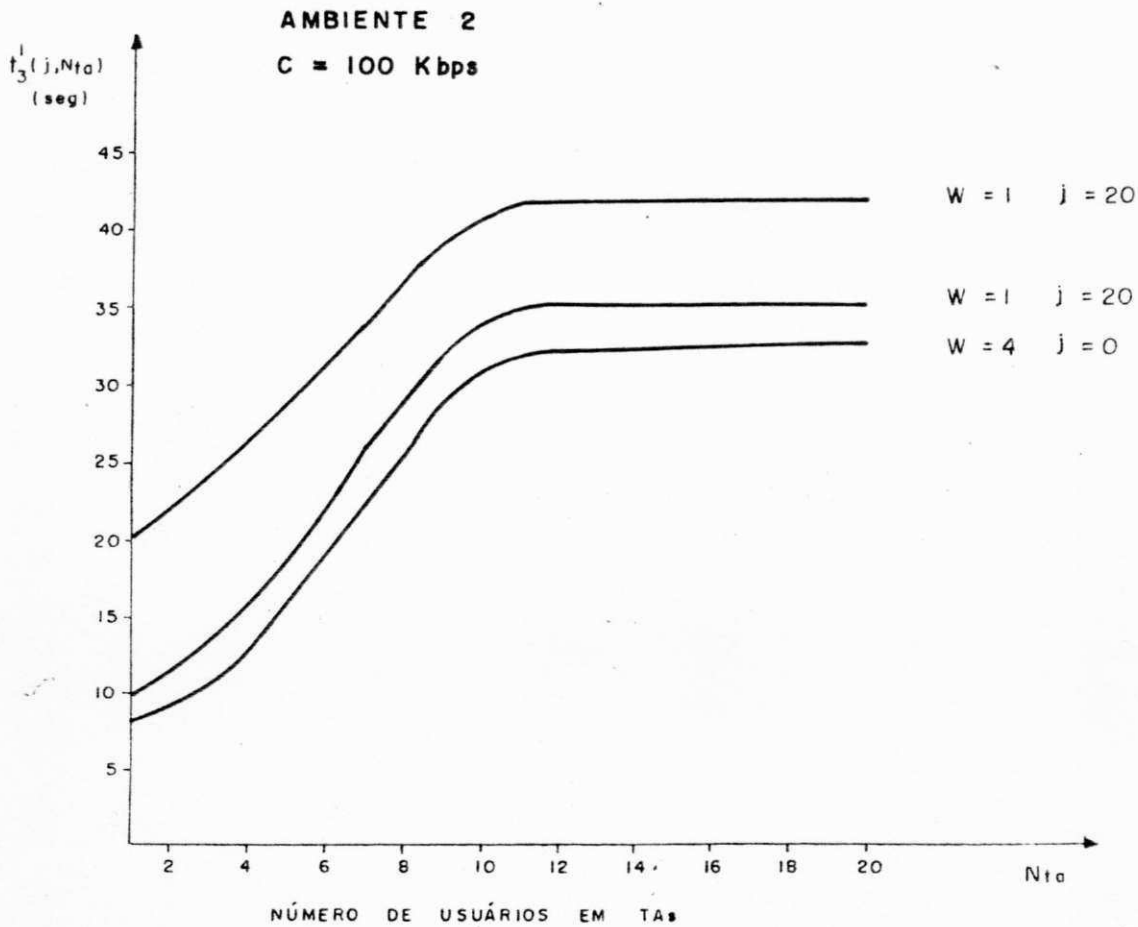


Fig. 4.10 - Tempo médio de resposta para TAs, $t_3^1(j, N_{ta})$

Um terceiro aspecto no desempenho, seria o número de usuários em PTs (j). Aumentando o "j", tem-se uma interferência direta em $t_3^1(j, N_{ta})$, exceto na figura 4.7 onde os SAs são separados, e a velocidade da sub-rede é suficientemente elevada, a fim de evitar a perturbação dos dois tipos de pacotes neste servidor. Esta é uma constatação feita na seção anterior.

Ficando aqui registrado o desequilíbrio (perturbação) causado por uma aplicação de pacotes pequenos e intermitentes, sobre uma outra de pacotes grandes, e de fluxo contínuo. Vale ressaltar que algumas curvas se mostraram muito próximas. Para se manter a clareza das figuras, algumas curvas foram suprimidas, porém, o comportamento destas, situam-se entre as curvas limites apresentadas, sem qualquer anormalidade (distorção).

O cruzamento existente na curva $W=4:j=20$, com a curva $W=1:j=0$, da figura 4.6, é também reflexivo neste nível. Porém, nada foi citado a respeito deste cruzamento, mas se sobrepusermos a figura 4.8 sobre 4.10, e observar as curvas supracitadas, verifica-se a interseção desta curva sobre o mesmo ponto encontrado na figura 4.6. Esta interseção não foi aqui plotada na mesma figura por questões de clareza e legibilidade.

Mais algumas observações importantes podem ser feitas:

- a) As curvas com $W = 1$ no ambiente 2, são bastantes sensíveis às variações de "j". Isto se deve a um problema de sobrecarga no SA. Servidores lentos em sistemas onde não há uma disciplina especial (adequada) no atendimento dos diversos tipos de requisições, uma queda na performance do sistema, é de se esperar.
- b) O valor de $t_3^1(j, Nta)$ tende a se estabilizar para $Nta > y$. Pois "y" é o número máximo de transferências simultâneas que o SA pode atender. Este limite do número de usuários em TAs no SA, mantém um padrão de atendimento para esta aplicação, bastante satisfatório. Neste ponto temos: $n_3^1(j, Nta) \rightarrow y$, e $\gamma_3^1(j, Nta) \rightarrow \mu(i)$, com $p(i) \rightarrow 1$.

Com estas conclusões, finaliza-se o estudo a respeito das interferências da aplicação 2, sobre a 1. No capítulo seguinte, inicialmente é feito um estudo para ambientes de população única, onde os usuários podem estar realizando TAs ou PTs de modo probabilístico. E por fim, é apresentada uma conclusão dos principais resultados deste trabalho.

5 - EXTENSOES DOS MODELOS E CONCLUSOES

Nos capítulos anteriores foram estudados os ambientes 1 e 2, que são configurações onde existe uma população definida para aplicação 1, e outra para aplicação 2. Neste trabalho, foi realizado uma interseção (cruzamento) de situações (estados) que estas duas populações podem assumir. Estas situações varrem quase todos os estados que se deseja analisar em uma situação operacional real. O número de usuários em TAs (i), e PTs (j), define os estados analisados neste trabalho. Uma combinação entre " i " e " j ", produz os possíveis estados para o sistema. As variações dos valores de " i " e " j ", foram as seguintes:

- a) Para aplicação 1, " i " variou de 1 até N_{ta} .
- b) Para aplicação 2, " j " variou de 0 até N_{pt} .

Portanto, o número de estados investigados foi da ordem de $N_{ta}(N_{pt} + 1)$. Como já definido, a hipótese é que a RL possua $N_{ta} + N_{pt}$ usuários nas aplicações 1 e 2 respectivamente. Neste capítulo, será mostrado o caso de população única para as duas aplicações é o caso onde um usuário pode realizar uma das duas aplicações, com uma probabilidade " p ". Finalizando este capítulo, é apresentada uma conclusão dos principais resultados obtidos neste trabalho.

5.1 - Caso de População Única

O caso de uma única população e dois SAs, é mostrado na figura 5.1. Esta análise é também válida para o caso de um servidor único. Observar que as duas aplicações estão disponíveis nos dois SAs.

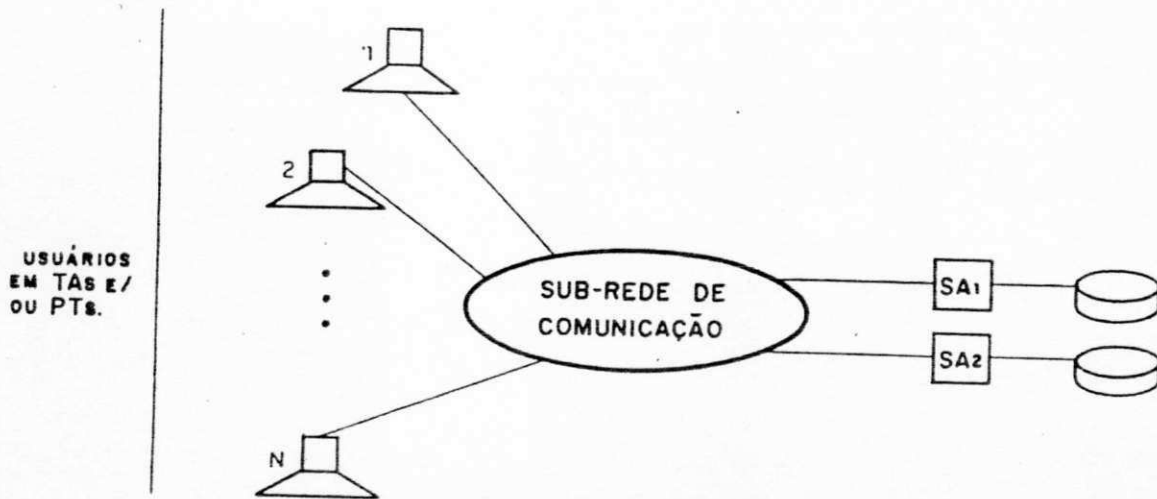


Fig.5.1 - População única para as duas aplicações

É uma situação mais real que os casos anteriores, pois um usuário não permanecerá indefinidamente realizando serviços na aplicação 1 ou 2. Destes "N" usuários, podem existir probabilisticamente "i" e "j" usuários, realizando TAs e PTs respectivamente. São muitos os estados que estes "i + j" usuários assumem com suas aplicações.

O tempo de um usuário num dado estado, é igual ao tempo de permanência (dispendido) em uma das duas aplicações. Ou seja, o tempo que um usuário permanece em um estado, garantidamente será o tempo necessário para ele completar uma transação, ou uma transferência de arquivo.

Com base nestas informações, pode-se construir uma cadeia de Markov [KLE1 75, SAUE 81] de duas variáveis para representar todos possíveis estados destes dois tipos de usuários. As taxas

de transições (mudança, frequência) entre os estados, são obtidas pelo inverso do tempo de permanência em uma das duas atividades. Isto é, as taxas de transições por usuário, são $\gamma_2^1(i,j,w)$ e $\gamma_2^2(i,j,w)$, para TAs e PTs respectivamente. Na figura 5.2, mostra-se um estado genérico (i,j) , onde existem "i" usuários em TAs, e "j" usuários em PTs. Ocorrendo um evento a partir do estado (i,j) , têm-se uma mudança (ocorre uma transição) para um dos estados vizinhos ilustrados na figura. Este evento, pode ser uma TA, ou um PT, que se inicializa ou se finaliza. Lembrar que os processos Markovianos assumem estados discretos. Por questões de legibilidade, vai se suprir o índice "W". Fica subentendido que ele é constante em um estudo.

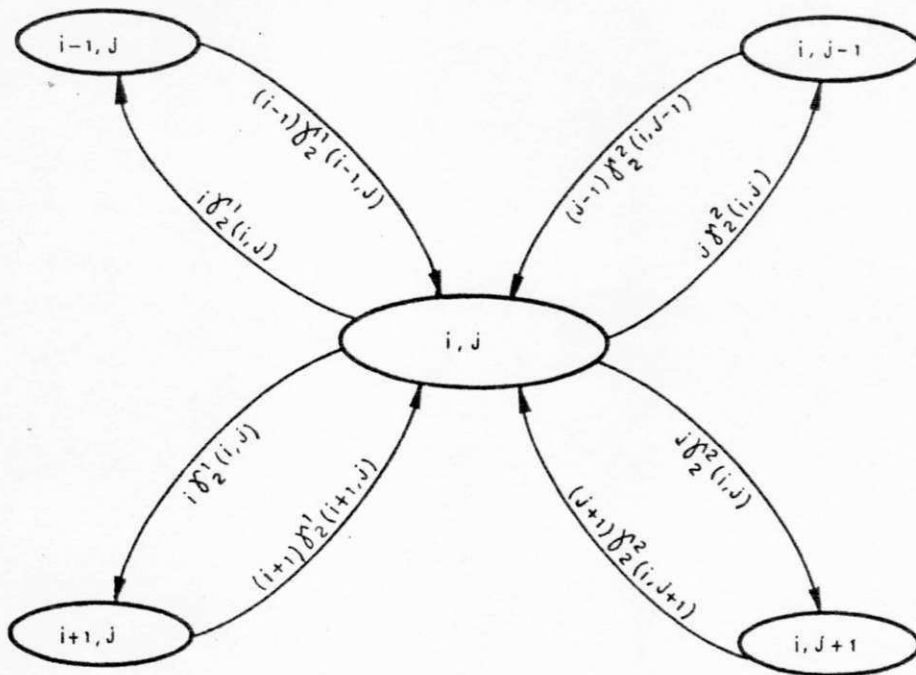


Fig. 5.2 - Diagrama de Transições de Estados dos Usuários

Uma equação genérica de balanço global para o estado (i, j) , é obtida igualando-se os fluxos das entradas aos das saídas, e é dado por:

$$2i\delta_2^1(i, j) + 2j\delta_2^2(i, j) = (i+1)\delta_2^1(i+1, j) + (i-1)\delta_2^1(i-1, j) + (j+1)\delta_2^2(i, j+1) + (j-1)\delta_2^2(i, j-1) \quad (5.1)$$

Observa-se que é uma expressão com 6 variáveis. Os valores destas variáveis são variáveis obtidas a partir do nível 2. Portanto, um conjunto de equações é resultante a partir de $i = 0, 1, 2, 3, \dots, y$ e $j = 0, 1, 2, \dots, Npt$.

A intenção desta análise é encontrar uma solução na forma de produto [BASK 75, SAUE 81] para qualquer estado. Com isto, obtém-se uma expressão para $p(i, j)$, que é a probabilidade de se ter "i" e "j" usuários realizando as aplicações 1 e 2, num dado estado.

Na tentativa de obter uma expressão analítica genérica, defronta-se com o número excessivo de variáveis. Pois, as taxas de transições são incógnitas variáveis que devem ser calculadas para todos os possíveis estados do sistema. Devido a esta complexidade, uma solução matemática é bastante difícil, podendo assim, partir para soluções através de sistemas de equações lineares, onde a recursividade computacional é da ordem de $(y+1)(Npt+1)6$, e que após todo este trabalho, chegar-se-ia a resultados de interesse idênticos aos que foram obtidos. Mas, vale ressaltar que seria um caso mais real e elegante de se estender o modelo.

Nos estudos anteriores, nem todos os estados foram investigados, faltaram os estados $(0, j)$. Mas, como o interesse maior seria estudar a interferência de pacotes curtos sobre pacotes longos, deixou-se estes estados sem análises, pois um estudo apenas sobre os usuários de PTs não teria sentido.

5.2 - Outros Casos

Situações onde possam existir grupos de usuários e SAs, são ambientes que podem ser propostos, que quando combinados com as duas aplicações em estudo, muitas variações de casos particulares poderão ser obtidos. Mas, não é objetivo deste capítulo investigar cada um destes casos, embora salienta-se que existem muitas outras situações que poderiam ser alvo para outras pesquisas.

Em [MOUR 82], são ilustrados alguns destes ambientes, bem como as diretrizes para suas soluções. Com certas modificações, é possível estender estas proposições para o caso de duas aplicações.

5.3 - Conclusões

O estudo da integração de aplicações em Redes Locais (RLs), tem sido objetivo de poucas pesquisas na área de avaliação de desempenho. Neste trabalho, considera-se duas importantes aplicações que podem estar simultaneamente presentes num mesmo ambiente de RL. Estas aplicações, Transferência de Arquivos (TAs) e Processamento de Transações (PTs), se encontram disponíveis aos usuários em algum tipo de servidor remoto de uso compartilhado. O servidor aqui considerado para estas aplicações, é o Servidor de Arquivos (SA).

As duas aplicações aqui estudadas, TAs e PTs, têm seu uso consagrado aos usuários de RLs. Elas são diferenciadas no tocante à operacionalidade e às características dos tráfegos por elas gerados. As TAs, se caracterizam por um fluxo unidirecional e contínuo de pacotes "longos" entre o SA e a estação de trabalho do usuário. Quanto a PTs, ocorre um fluxo bidirecional e intermitente de pacotes "curtos" entre o SA e a estação do usuário.

Estas duas aplicações foram apreciadas em dois ambientes: no primeiro há dois grupos de usuários (um grupo em TAs, e o outro em PTs), a sub-rede, e dois SAs, sendo que um dos SAs é utilizado pelos usuários em TAs, e o outro é reservado aos usuários em PTs; no segundo ambiente, há apenas um SA para as duas aplicações, atendendo simultaneamente às duas categorias de usuários.

Para a análise da integração destas duas aplicações nos ambientes supracitados, construiu-se um Modelo Hierárquico (MH) em três níveis. A idéia da solução para MHs, é utilizar-se de métodos distintos para cada nível do modelo. Nesta análise, foi utilizada uma solução híbrida (mista), onde os dois primeiros níveis do modelo foram resolvidos através da simulação discreta, enquanto que no terceiro nível, lançou-se mão de um método analítico. O primeiro nível deste modelo, refletiu o atraso de transmissão dos pacotes no meio físico da sub-rede. Para o segundo nível, foi analisado o atraso de circulação (trânsito) dos pacotes entre a estação do usuário e o SA. Já no terceiro nível, foi pertinente a uma particularidade do SA para os usuários em TAs. Aqui, o SA impôs um limite máximo de conexões simultâneas em TAs.

A partir do MH desenvolvido, foram obtidas algumas medidas de desempenho para TAs e PTs, estas medidas são: o tempo médio de resposta da aplicação, e sua vazão. Estas duas figuras de mérito são suficientes para demonstrar a perturbação existente entre estas duas aplicações, quando presentes numa mesma RL. Verificando a escassez e a necessidade de trabalhos nesta área, nos estimulou a uma pesquisa sobre a interferência entre estes dois tipos de tráfegos num mesmo ambiente, sendo esta, a contribuição deste trabalho.

De acordo com as figuras de desempenho apresentadas, algumas conclusões importantes podem ser extraídas, segundo um conjunto (universo) de valores atribuídos aos parâmetros dos modelos em estudo. Verificou-se que o tamanho da janela (W) usada nos

Circuitos Virtuais (CVs) para TAs, é responsável pelos níveis de desempenho das duas aplicações. Aumentando-se o valor de W , obtem-se um tempo de resposta melhor para os usuários em TAs. Isto deve-se ao fato que com um W maior, observa-se um uso mais intenso do CV no transporte dos pacotes de dados do arquivo, uma vez que os pacotes de reconhecimento em sentido contrário, ocorrem com menor frequência. Em contrapartida, pioram os tempos de resposta para PTs. Este acontecimento, tem como razão a disputa (concorrência) em igualdade nos recursos compartilhados na RL (sub-rede e SA), pelos dois tipos de pacotes. Devido às características das aplicações, os pacotes de TAs sofrem um atraso maior (são mais "lentos") nestes recursos, do que os pacotes de PTs. Portanto, não existindo nenhuma disciplina diferenciada nos atendimentos dos pacotes, na sub-rede e SA, os pacotes de PTs têm seu tempo degradado em razão de outros pacotes mais "lentos" no sistema. Fica aqui assinalada, a importância do parâmetro W das TAs, e sua interferência causada nas medidas de desempenho de PTs.

Uma outra influência nos tempos de respostas aos usuários em TAs, é causada pelos usuários presentes em PTs. Aumentando-se o número de usuários em PTs, conseqüentemente há um crescimento do número total de pacotes no sistema, gerando assim, uma queda nos níveis de desempenho das TAs. Esta ocorrência é bastante evidenciada no ambiente com um único SA, onde há tarefas das duas aplicações simultaneamente. Este fato, é a reação em sentido contrário da explicação do parágrafo anterior, isto é, são pacotes "pequenos" (rápidos em relação aos pacotes de TAs), que ficam congestionando (interferindo, perturbando) o tráfego dos pacotes "longos" de TAs, nos recursos comuns destas duas aplicações.

Um terceiro ponto a ser considerado, é a influência da velocidade da sub-rede (" C ") para TAs, nos dois ambientes avaliados. Com um valor de " C " maior, menores tempos de respostas para TAs foram obtidos. Este era um resultado já esperado, pois

devido à natureza desta aplicação, com muitos pacotes "longos" e de fluxo contínuo, a rapidez com que a sub-rede consegue fluir os pacotes, é um parâmetro importante para o desempenho de TAs. Quanto a PTs, o valor de "C" só foi crítico quando a sub-rede se encontrava num estado de sobrecarga, isto é, simultaneamente muitos usuários em TAs operando com W maior que um.

Observou-se que a sub-rede não é um recurso crítico em sistemas com RLs. Esta constatação, foi resultante das figuras de desempenho encontradas para estas duas aplicações, verificando que a sub-rede não congestionava o fluxo dos pacotes entre o SA e as estações do usuário. As sub-redes de comunicação nas RLs, são dispositivos (recursos) velozes, comparados às estações de trabalho, e Servidores de Arquivos (SA). Em função da tecnologia (hardware e software) atual, as velocidades destes dispositivos (Serv. de Arqs. e estações de trabalho), são em geral, os que mais comprometem o desempenho das aplicações nas RLs.

Neste trabalho, investigou-se a interferência entre os tráfegos de TAs e PTs nas RLs, ocasionando deste modo uma degradação no desempenho das duas aplicações avaliadas. Apesar deste fato, foi constatado que os níveis de desempenho destas aplicações, são aceitáveis ao usuário final, para o conjunto de valores propostos aos parâmetros utilizados nos modelos. Com estes resultados, se verifica a viabilidade da integração de aplicações em RLs.

5.3.1 - Sugestões para Futuras Pesquisas

- a) Elaborar um modelo mais detalhado das estações de trabalho e SAs. Este modelo deve captar em grande parte o comportamento real destes dispositivos. Por exemplo, poder-se-ia incluir no modelo, prioridades no atendimento das requisições ao SA, sendo este um aspecto de bastante

relevância na melhoria das figuras de desempenho, quando se tem diversos tipos (natureza distintas) de pedidos a serem atendidos num servidor.

- b) Realizar análises de um maior número de aplicações no mesmo ambiente. A heterogeneidade de tráfego gerada, seria uma situação mais próxima da realidade operacional de RLs.
- c) Na solução híbrida aplicada ao MH, é possível a ocorrência de erros na transferência dos parâmetros entre os níveis, uma vez que diferentes métodos foram aplicados na solução dos níveis do modelo. Cada nível tem suas imprecisões intrínsecas. Uma idéia seria investigar esta propagação de erros entre os níveis, comparando os resultados com outros tipos de soluções. Por exemplo, uma comparação da solução do modelo, unicamente obtida através de métodos analíticos, heurísticos, e simulação.

6 - ANEXOS

* Simulador em GPSS/360 para o ambiente 1, extensível ao
* ambiente 2, com pequenas alterações.

SIMULATE
RMULT 737 * Nova semente

* Função com Dist. Exponencial (FN1)

1 FUNCTION RN1,C24
0,0/.1,.104/.2,.222/.3,.355/.4,.509/.5,.69/.6,.915/.7,1.2/.75,1.3
/8.8,1.6/.84,1.83/.88,2.12/.9,2.3/.92,2.52/.94,2.81/.95,2.99/.96,
3.2/.97,3.50/.98,3.9/.99,4.6/.995,5.3/.998,6.2/.999,7/.9997,8

* Inicialização de algumas variáveis

INITIAL X1,1/X2,20/X3,0

* X1 : numeração inicial das cadeias de TAs

* X2 : numeração inicial das cadeias de PTs

* X3 : controle da população de cadeias de TAs

INITIAL X4,4 * Pop. máx. da cadeia de TAs
INITIAL X5,20 * Pop. das 20 cadeias de PTs
INITIAL X10,4 * Pop. final as cadeias de TAs

SUBRD EQU 77,F,Q * Sub-rede de comunicação
SVRQ1 EQU 88,F,Q * Serv. de Arq. #1
SVRQ2 EQU 99,F,Q * Serv. de Arq. #2

1 VARIABLE (X4/4)+1

* *
* Modelo para Transf. de Arquivos *
* *

GENERATE 1,,,X4,,6,F

* Atribui os parâmetros as transações de TAs

ASSIGN 1,80 * Atraso na sub-rede
ASSIGN 2,50 * Atraso no Serv. de Arq.
ASSIGN 3,200 * Atraso na Est. de Trabalho

* Numerar as cadeias de TAs, e suas populações
*

SAVEVALUE 3+,1
 TEST G X3,X10,SEGUE
 SAVEVALUE 3,1
 SAVEVALUE 1+,1

SEGUE TEST E X10,K4,SAL_1
 TEST E X1,V1,SAL_2

SAL_1 ASSIGN 4,X1
 ASSIGN 5,1
 TRANSFER ,INIC

SAL_2 ASSIGN 4,X1
 ASSIGN 5,4
 TRANSFER ,INIC

 *
 * Modelo para Proct. de Transações *
 *

GENERATE 1,,,X5,,6,F

* O valor de X5, vai variar de 2, 4, 6, ... 20
 * Atribui os parâmetros as transações de PTs

ASSIGN 1,10 * Atraso na sub-rede
 ASSIGN 2,30 * Atraso no Serv. de Arq.
 ASSIGN 3,400 * Atraso na Est. de Trabalho

* Numerar as cadeias de PTs, com população = 1 (um)

SAVEVALUE 2+,1
 ASSIGN 4,X2
 TRANSFER ,INIC

 * Início do Pgm. Principal *

INIC MARK 6

QUEUE *4
 SEIZE *4
 DEPART *4
 ADVANCE *3, FN1
 RELEASE *4

QUEUE SUBRD
 SEIZE SUBRD
 DEPART SUBRD
 ADVANCE *1, FN1
 RELEASE SUBRD

UNIVERSIDADE FEDERAL DA PARAÍBA
 Pró-Reitoria Para Assuntos do Interior
 Coordenação Setorial de Pós-Graduação
 Rua Aprigio Veloso, 882 - Tel. (083) 321-7222-R 355
 58.100 - Campina Grande - Paraíba

TEST L P4,K11,PUL_1 * Testar o pacote (TA ou PT).

* No SVRQ1 (Serv. de Arqvs. #1), para TAs

QUEUE SVRQ1
SEIZE SVRQ1
DEPART SVRQ1
ADVANCE *2, FN1
RELEASE SVRQ1

* Verifica a pop. da cadeia de TAs, 1 ou 4

TEST E P5,X10,PUL_2
TABULATE 11 * Contabiliza g2(i,j,W)
ASSIGN 6,KO
TRANSFER ,INIC

PUL_2 TABULATE 14 * Contabiliza f2(i,j,W)
ASSIGN 6,KO
TRANSFER ,INIC

* No SVRQ2 (Serv. de Arqvs. #2), para PTs.

PUL_1 QUEUE SVRQ2
SEIZE SVRQ2
DEPART SVRQ2
ADVANCE *2, FN1
RELEASE SVRQ2

TABULATE 22 * Contabiliza PTs.
ASSIGN 6,KO
TRANSFER ,INIC

* *

* Fim do Pgm. Principal *

* *

* *

* Saídas Desejadas *

* *

* Saída para g2(i,j,W), de TAs

11 TABLE MP6,0,1000,2

* Saída para f2(i,j,W), de TAs

14 TABLE MP6,0,1000,2

* Saída para t (i,j,W), de PTs

22 TABLE MP6,0,1000,2

* Controle das unidades de tempo

GENERATE 500000
SAVEVALUE 4+,W
TERMINATE 1
START 1
END

* W será 1 ou 4; pop. de TAs.

UNIVERSIDADE FEDERAL DA PARAIBA
Pró-Reitoria Para Assuntos do Interior
Coordenação Setorial de Pós-Graduação
Rua Aprigio Veloso, 882 Tel (083) 321-7222-R 355
58.100 - Campina Grande - Paraíba

7 - REFERENCIAS BIBLIOGRAFICAS

- BASK 75 - Baskett, F., Chandy, K.M., Muntz, R.R., e Palacios, F.G., "Open, Closed and Mixed Networks of Queues with Different Classes of Customer", JACM 22, n.2, apr. 1975.
- BIRR 80 - Birrel, A. D., e Needham, R. M., "A Universal File Server", IEEE Trans. Software on Engineering, v.6, n.5, sep. 1980.
- BROW 84 - Brown, M. R., Kolling, K., e Taft, E. A., "The Alpine File System", Tech. Rep. CSL 84-4, Xerox Palo Alto Research Center - Palo Alto, California, oct. 1984.
- BUX 81a - Bux, W., Closs, F., Janson, P., Kummerle, K., e Muller, H. R., "A Reliable Token-Ring for Local Area Communication", Proc. of NTC, 1981.
- BUX 81b - Bux, W., "Local-area subnetworks: A performance comparision", IEEE COM-29, n.10, v.29, oct. 1981.
- BUX 84 - Bux, W., "Performance Issues in Local-Area Networks", IBM Systems Journal, n.4, v.23, 1984.
- CLAR 78 - Clark, D. D., Porgram, K. T., e Reed, D.P., "A Introduction to Local Area Networks", Proc. of IEEE, v.66, n.11, nov. 1978.
- CHRI 84 - Christodoulakis, S., e Faloutsos, C., "Design Consideratisons for a Message File Server", IEEE Trans. on Software Engineering, v. SE-10, n.2, mar. 1984.
- COUR 75 - Courtois, P.J., "Decomposability, Instabilities, and Saturation im Multiprogramming Systems", CACM, n.9,

v.18, jul. 1977.

- COUR 77 - Courtois, P. J., Decomposability: Queueing and Computer Systems Applications, Academic Press, New York, 1977.
- COUR 85 - Courtois, P.J., "On time and spaces decomposition of computer structures", CACM, n.6, v.28, jun. 1985.
- DELL 82 - Dellar, Carl N. R., "A File Server for a Network of Low Cost Personal Microcomputers", Software - Practice and Experience, v. 12, dec. 1982.
- EISE 85 - Eisenhard, B., "Networks Disk an File Servers", Corvus Systems Inc, Proceedings of COMPCOM, 1985.
- GAMS 86a - Giozza, W.F., de Araújo, J.F.M., Moura, J.A.B. e Sauv , J. P., Redes Locais de Computadores: Aplica es e Tecnologias, McGraw-Hill, 1986.
- GAMS 86b - Giozza, W.F., de Ara jo, J.F.M., Moura, J.A.B. e Sauv , J. P., Redes Locais de Computadores: Protocolos de Alto N vel e Avalia o de Desempenho, McGraw-Hill, 1986.
- GELE 80 - Gelenbe, E. e Mitrani, I., Analysis and Synthesis of Computer Systems, Academic Press, 1980.
- IEEE 82 - IEEE Project 802 Local Network Standard, Draft D., 1982.
- KELL 75 - Kelly, F. P., "Networks Queues with Customer of Different Types", J. Appl. Prob., v.12, 1975.
- KLEI 75 - Kleinrock, L., Queueing Systems, vol 1: Theory, John Wiley & Sons Inc., New York, 1975.

UNIVERSIDADE FEDERAL DA PARA BA
 Pr -Reitoria Para Assuntos do Interior
 Coordena o Superior de P s-Gradua o
 Rua Apr gio Veloso, 862 - TEL (401) 321-7222-R 355
 58.100 - Campina Grande - Para ba

- LAM 80 - Lam, S.S., "A Carrier Sense Multiple Access for Local Networks", Computer Networks, v.4, feb. 1980.
- LANT 85 - Lantz, K. A., Nowicki, W. I. e Theimer, M. M., "An Empirical Study of Distributed Application Performance", IEEE Trans. Soft. on Eng., v.11, n.10, oct. 1985.
- LAVE 83 - Lavemberg, S.S., Ed., Computer Performance Modelling Handbook, Academic Press, New York, 1983.
- LAZO 86 - Lazowska, E.D., Zahorjan, J., Cheriton, D.R. e Zwaenepoel, W., "File Access Performance of Diskless Workstation", ACM Transactions on Computer Systems, n.3, v.4, aug. 1986.
- MACE 86 - Macedo, R.J.A., "Servidores de Arquivo: Uma Introdução", Anais da 1^a Semana de Informática da UFBA, março, 1986.
- LEIT 85 - Leitner, G., "Distributed Synchronization of Multiple Servers", Dep. of Computer Science, Columbia University, New York, Proceedings of COMPCOM, 1985.
- MITC 82 - Mitchell, J. G., e Dion, J., "A Comparison of Two Network-Based File Servers", CACM, v.25, n.4, apr. 1982.
- MOUR 82 - 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.
- MOUR 83 - Moura, J. A. B., e Sauvé, J. P., "Análise Assintótica do Desempenho de Redes de Dados Locais", Anais do Primeiro Simpósio Brasileiro de Redes de

Computadores, maio 1983.

- MUSS 85 - 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.
- REIS 79 - Reiser, M., "A Queueing Network Analysis of Computer Communication Networks with Window Flow Control", IEEE COM-27, n.8, v.27, aug. 1979.
- REIS 80 - Reiser, M., e Lavemberg, S., "Mean Value Analysis of Closed Multichain Queueing Networks", JACM 27, n.2, 1980.
- REIS 82 - Reiser, M., "Performance Evaluation of Data Communication Systems", Proceedings of the IEEE, v.70, n.2, feb. 1982.
- SA 87a - Sá, C. C., Brasileiro, M. A. G., Moura, J. A. B., "Modelagem de Aplicações em Redes Locais", Anais do V Simpósio Brasileiro de Redes de Computadores, São Paulo - SP, abril 1987.
- SA 87b - Sá, C. C., Pontes, R. C. M., "Redes Locais: Influência do Meio de Transmissão Sobre as Aplicações do Usuário", trabalho a ser submetido na 2^a Semana de Informática da UFBA, Salvador, Ba.
- SAST 85 - Sastry, A. R. K., "Maximum Mean Data in a Local Area Network with Specified Maximum Message Source Message Load", Proceedings of COMPCOM 1985.
- SAUE 81 - Sauer, C. H. e Chandy, K. M., Computer Systems Performance Modelling, Prentice-Hall, New Jersey, 1981.

- SCHW 78 - Schwertman, H. D., "Hybrid Simulation Models of Computer Systems", CACM, n.9, v.21, sep. 1978.
- SHOC 80 - Shoc, J.F. e Hupp, J.A., "Measured Performance of an Ethernet Local Network", CACM, n.2, v.23, dec. 1980.
- SOAR 86 - Soares, Luis Fernando G., Redes Locais, Rio de Janeiro - R.J., Editora Campus, 1986.
- SPEC 82 - Spector, A. Z., "Performing Remote Operations Efficiently on a Local Computer Network", CACM, n.4, v.25, apr. 1982.
- STAL 84a - Stallings, W., Local Networks: An Introduction, New York, Mac-Millan, 1984.
- STAL 84b - Stallings, W., "Local Network Performance", IEEE Communications Magazine, v.22, n.2, feb. 1984.
- STAN 84 - Stankovic, J. A., "A Perspective on Distributed Computer Systems", IEEE Transaction on Computer, n.12, v.33, dec. 1984.
- STRA 84 - Strack, J., GPSS: Modelagem e Simulação de Sistemas - LTC Editora S/A, Rio de Janeiro, 1984.
- STUR 80 - Sturgis, H. E., Mitchell, J.G., e Israel, J.E., "Issues in the design and use of a distributed file system", ACM SIGOPS Oper. Syst. Rev., v.14, n.3, jul. 1980.
- SVOB 84 - Svobodova, L., "File Servers for Network - Based Distributed Systems", Computing Surveys, n.4, v.16, dec. 1984.

- WIN 79 - Swinehart, D., McDaniel, G., e Boggs, D., "WFS: A Simple Share File System for a Distributed Environment", ACM SIGOPS Operating Systems Review, v.12, n.4, apr. 1979.
- TANE 81a - Tanenbaum, A. S., Computer Networks, Prentice Hall Inc., New Jersey - 1981.
- TANE 81b - Tanenbaum, A. S., "Network Protocols", Computing Surveys, n.4, v.13, dec. 1981.
- TANE 85 - Tanenbaum, A. S. e Renesse, R. V., "Distributed Operating Systems", Computing Surveys, n.4, v.17, dec. 1985.
- THOM 85 - Thomasian, A., e Bay, P. F., "Integrated Performance Models for Distributed Processing in Computer Communication Networks", IEEE Transaction on Software Engineering, v.11, n.10, oct. 1985.
- TOBA 80 - Tobagi, F. A., "Multiaccess protocols in packet communication systems", IEEE Trans. Comm. v.28 - 1980.
- TOKU 84 - Tokuda, H., Shosin: A Distributed Software Testbed, tese de Ph.D, University of Waterloo, Canadá, 1984.
- WONG 80 - Wong, J. W., Moura, J. A. B., e Field, J. A., "Hierarchical Modelling of Local Area Computer Networks", Proceedings of NTC, 1980.
- WONG 83 - Wong, J. W., Moura, J. A. B., e Field, J. A., "Hierarchical Modelling of File Transfers", Comput. & Elect. Eng., n.3, v. 10, 1983.

UNIVERSIDADE FEDERAL DA PARAIBA
 Pró-Reitoria para Assuntos do Interior
 Coordenação Setorial de Pós-Graduação
 Rua Aprígio Veloso, 832 - 71.707-321 - 7222-R 355
 58.400 - Campina Grande - Paraíba