

Dimensionamento de Sistemas Multimicroprocessadores para
Redes de Computadores

Armando S. Barbosa Jr.

Lucas A. Moscato

Escola Politécnica da Universidade de São Paulo

Resumo:

O uso de máquinas construídas com múltiplos microprocessadores vem se intensificando cada vez mais nas mais diversas áreas de aplicação, inclusive na de equipamentos para uso em redes de computadores. As máquinas assim construídas podem constituir redes confinadas, com necessidades e soluções próprias, ou sistemas fortemente acoplados característicos.

Nas fases iniciais de desenvolvimento de um destes sistemas torna-se bastante desejável uma avaliação da capacidade de processamento que será requerida pelo equipamento para execução de suas funções. Falk e Mc Quillan desenvolveram um trabalho com este objetivo para aplicação em nós de comutação de pacotes (Falk 77). Este artigo baseia-se neste trabalho, discutindo-o e estendendo-o em alguns aspectos considerados relevantes pelos autores, pela experiência obtida no projeto e construção de dois sistemas multimicroprocessadores de grande porte distintos: uma central de comutação de pacotes (RUGG 80), (BARB 82) e um equipamento PAD - "Packet Assembler/Disassembler" (CAMP 83A) (CAMP 83B).

I - O PROBLEMA

O projeto de uma rede de computadores envolve uma série de questões de dimensionamento que têm que ser solucionadas nas várias etapas do desenvolvimento. Entre estas estão o dimensionamento das capacidades dos canais de comunicação, o hardware dos nós de comutação e número destes nós, juntamente com a topologia a eles associada.

Todas estas questões têm que ser solucionadas levando-se em conta uma série de parâmetros condicionados ao desempenho que se quer obter da rede (tempos de trânsito de mensagens, vazão), à confiabilidade, ao custo e às características de tráfego dos usuários.

Por outro lado, tem sido cada vez maior a utilização de sistemas construídos com múltiplos microprocessadores para utilização como elemento de comutação do sub-sistema de comunicação. A utilização de tais sistemas tem se tornado cada vez mais atraente em virtude do custo relativo cada vez mais baixo dos microprocessadores, das facilidades que estas famílias de componentes têm oferecido para a construção de sistemas de multiprocessamento e, talvez o mais importante, em virtude da aplicação envolver uma grande quantidade de tarefas repetitivas e de baixa conectividade, o que limita o projeto do software de tais sistemas dentro de um nível de complexidade encorajador. Além disso, essa filosofia de projeto possibilita o aumento da disponibilidade e a realização de sistemas com reais características de modularidade, que permitem fácil configuração conforme as necessidades de cada aplicação. Desta forma, muitas máquinas multiprocessadores têm sido construídas para implementar nós de comutação de pa -

cotes, "packet assembler/disassemblers" e redes locais
(BARB 82) (RUGG 80) (CAMP 83A) (CAMP 83B) (JENN 76)
(BUX 79).

Algumas das questões que surgem no início do desenvolvimento de um destes sistemas são: Qual será o número de processadores necessário para atender à demanda prevista, dentro dos parâmetros de desempenho especificados? Qual é o "overhead" de comunicação entre processadores? Até que ponto se poderá aumentar o número de processadores para se obter uma melhoria do desempenho? É desejável, nesta altura, um método que permita obter estimativas de pelo menos alguns destes parâmetros, de modo a dirigir decisões de projeto que venham a ser necessárias no decorrer do desenvolvimento.

O dimensionamento da capacidade de processamento é, contudo, apenas um dos itens de dimensionamento a serem resolvidos no transcurso do projeto. Os demais itens, que podem se igualar em importância, dependendo da aplicação, são o dimensionamento dos buffers de comunicação (SCHW 76), da capacidade de entrada e saída (FALK 77), da memória de programa e da memória de massa. Este artigo concentra-se na discussão de dimensionamento da capacidade de processamento, e toma como base um trabalho previamente elaborado por Falk e Mc Quillan (FALK 77).

Como o enfoque é na construção de máquinas com múltiplos microprocessadores, a discussão é feita com vistas a obter uma estimativa do número de processadores de uma determinada potência necessário para implementar um equipamento que possibilite alcançar os requisitos de desempenho especificados para a rede, independentemente da arquitetura física utilizada para a sua construção.

O processo de obtenção desta estimativa é iterativo. No início não são introduzidos parâmetros quantitativos referentes ao desempenho da máquina. Apenas um crité

rio qualitativo é utilizado para determinar a condição de ocupação dos processadores considerada adequada para a aplicação. O sistema, desta forma dimensionado, deve, então, ter o seu desempenho avaliado usando um método qualquer de análise, seja exato, aproximado ou simulação. A figura 1 esquematiza o processo iterativo para o dimensionamento da capacidade de processamento

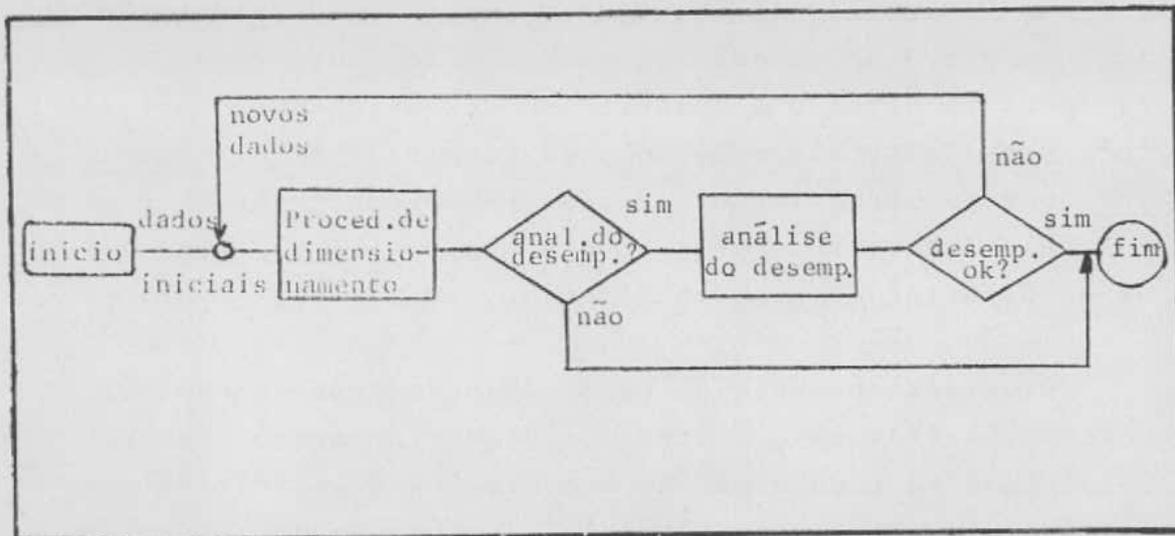


Figura 1 - O Processo de Dimensionamento

Nesta figura os dados iniciais representam os parâmetros necessários para alimentar o procedimento de dimensionamento. Em primeira instância estes dados se baseiam em experiências anteriores dos projetistas ou em valores obtidos de outros sistemas, e constituem, basicamente: o "overhead" de comunicação entre processadores, os coeficientes de utilização dos processadores e as demandas de processamento por unidade de dados.

A partir destes dados o procedimento de dimensionamento produz uma estimativa do número mínimo de processadores capaz de executar as tarefas previstas para o sistema.

Nesta altura, dependendo de quão definido já está o

projeto, pode ser desejável aos projetistas verificar mais precisamente se os parâmetros heurísticos introduzidos inicialmente realmente se traduzirão nos índices de desempenho objetivados para o sistema. Para tanto, contudo, é necessário que já tenha sido tomada uma série de decisões que dizem respeito não só à arquitetura física da máquina, mas também à arquitetura lógica, com o respectivo mapeamento das tarefas do sistema nos processadores. Estes dados, juntamente com as estimativas de demanda de processamento das tarefas, possibilitam levar a cabo um procedimento de análise do desempenho. Este procedimento, como já foi citado, pode constituir-se por uma metodologia analítica exata, pela utilização de métodos aproximados ou numéricos, ou por simulação. Convém, contudo, ressaltar-se que não faz sentido, nas fases primárias do desenvolvimento, a utilização de modelos muito elaborados para a análise, uma vez que as incertezas introduzidas pelas hipóteses e estimações dos dados para o dimensionamento de forma alguma justificariam os altos custos e tempos de resolução destes modelos. Estes devem representar o sistema real de forma tão fiel quanto possível, mas dentro de limites de complexidade coerentes com as simplificações do dimensionamento. Mais uma vez a própria experiência dos projetistas e o seu bom senso serão os fatores norteadores para estabelecer qual o modelo aplicável em cada caso.

Os dados obtidos da análise assim elaborada podem satisfazer ou não os índices de desempenho especificados para o projeto. No primeiro caso o processo de dimensionamento é encerrado e o projeto é continuado até que surja alguma alteração substancial que justifique o disparo de um novo processo de dimensionamento. No segundo caso é necessário reiniciar o processo de dimensionamento com novos dados, numa tentativa de alcançar os padrões de desempenho desejados. Estes novos dados podem ser conseguidos por des- de uma simples redução dos fatores de utilização dos processadores até por alterações substanciais de estrutura do software, ou até do hardware.

II - O Procedimento de Dimensionamento

Em uma rede de computadores o sub-sistema de comunicação é constituído por um conjunto de equipamentos de comunicação (nós de comutação de pacotes, PAD'S, interfaces de comunicação nas redes locais e centrais de comutação nas redes locais tipo PABX) e pelos enlaces de comunicação que interconectam estes elementos entre si e aos equipamentos usuários.

O equipamento de comunicação pode ser modelado, de forma geral, como um nó pelo qual passam tres tipos de fluxo, conforme apresenta a figura 2 (FALK 77). Dependendo da aplicação um ou dois dos fluxos podem não estar presentes.

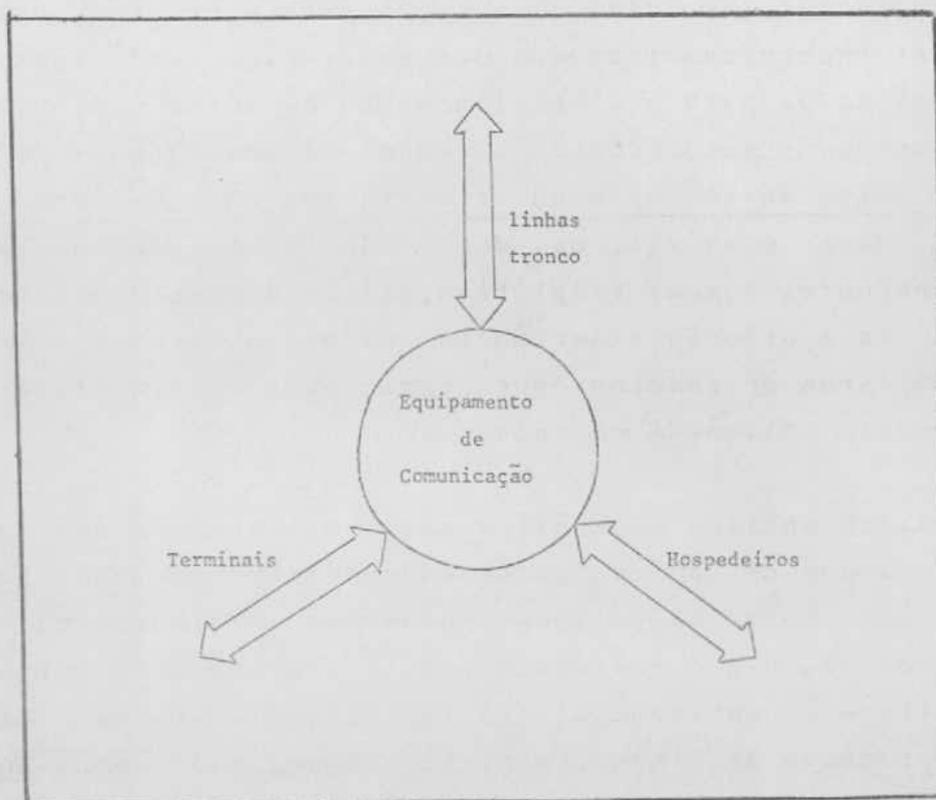


figura 2 - Tipos de Fluxo no Equipamento de Comunicação (FALK 77)

Os tres tipos de fluxos de dados que podem ser manipulados pelo equipamento de comunicação de sub-rede são, conforme mostra a figura 2:

- a) O tráfego gerado pela comunicação do equipamento com outros equipamentos da sub-rede. Chamamos a este fluxo tráfego de linhas tronco.
- b) O tráfego dos sistemas hospedeiros, computadores usuários da rede.
- c) O tráfego gerado por equipamentos terminais da rede, diretamente conectados ao equipamento de comunicação.

A tarefa primária do equipamento de comunicação é o manuseio destes tres tipos de tráfego e constitui, nos casos práticos, a parcela significativa da demanda de processamento. Desta forma, as tarefas secundárias do equipamento não serão consideradas para o dimensionamento.

Chamamos de λ_H a taxa em bits por segundo, em "full - duplex", devida aos equipamentos hospedeiros. Da mesma forma chamamos de λ_T a taxa devida aos terminais e λ_{TR} a taxa de vida às linhas tronco.

A largura de faixa P do equipamento de comunicação, em bits por segundo, é, para cada tipo de fluxo:

$$P = \frac{S}{T} \quad (\text{bits/s})$$

onde:

- S: tamanho da unidade de dados, em bits.
- T: tempo gasto no manuseio da unidade, em segundos.

O tempo de manuseio das unidades de dados, t , é normalmente constituído por duas parcelas resultantes dos tipos de processamento associados a elas: uma parcela independente do tamanho da unidade de dados e uma parcela que é função do tamanho da unidade de dados:

$$t = t_1 + t_2 \quad (s)$$

Desta forma, as máximas vazões que podem ser suportadas por uma capacidade de processamento P , para cada fluxo, é:

$$\lambda_{HMAX}(S_H) = \frac{S_H}{t_{1H} + t_{2H}(S_H)}, \quad \lambda_{TRMAX}(S_{TR}) = \frac{S_{TR}}{t_{1TR} + t_{2TR}(S_{TR})},$$

$$\lambda_{TMAX}(S_T) = \frac{S_T}{t_{1T} + t_{2T}(S_T)},$$

onde S_H , S_{TR} e S_T são os tamanhos das unidades de dados, em bits, para os tráfegos de hospedeiros, de troncos e terminais, respectivamente.

Agora, para calcularmos a capacidade de processamento P' , em múltiplos de P , para suportar λ_{TR} bps de tráfego de troncos, λ_H bps de tráfego de hospedeiros e λ_T bps de tráfego de terminais, devemos ter:

$$P' \geq \frac{\lambda_T}{\lambda_{TMAX}} + \frac{\lambda_H}{\lambda_{HMAX}} + \frac{\lambda_{TR}}{\lambda_{TRMAX}}$$

A unidade do valor resultante deste cálculo é a capacidade de processamento P , que foi tomada como base para a obtenção de λ_{TMAX} , λ_{HMAX} e λ_{TRMAX} . Desta forma, o valor P' pode ser encarado como o número de processadores de capacidade P necessários para suportar os tráfegos descritos.

Contudo, neste cálculo ainda não foram considerados tres fatores importantes:

- a) A capacidade de processamento utilizada por funções não diretamente associadas ao tratamento das unidades de dados. É o processamento consumido pela necessidade de comunicação entre os processadores, com todos os seus controles, sincronismos e protocolos, pelo roubo de ciclos das UCP'S pelos DMA'S e pelo uso de recursos compartilhados.
- b) Um sistema multiprocessadores é um sistema de filas em que o fator de utilização não deve ser alto, para que os tempos de espera estejam dentro de limites aceitáveis.
- c) O mapeamento das tarefas nos processadores nem sempre possibilita uma distribuição completamente uniforme da carga. Desta forma, teremos processadores que serão mais requisitados do que outros.

Para compensar a existência destes fatores definimos K_S , K_ρ e K_μ , de forma que:

$$P'' = K_S \cdot K_\rho \cdot K_\mu \cdot P'$$

onde

P'' : capacidade de processamento necessária

K_S : fator de compensação para as funções de comunicação entre processadores.

K_ρ : fator de redução da utilização dos processadores.

K_μ : fator de uniformidade

Desta forma,

$$P'' \geq K_S \cdot K_\rho \cdot K_\mu \cdot \left(\frac{\lambda_T}{\lambda_{TMAX}} + \frac{\lambda_H}{\lambda_{HMAX}} + \frac{\lambda_{TR}}{\lambda_{TRMAX}} \right)$$

III - A Análise do Desempenho

No procedimento de dimensionamento, o fator K_ρ foi introduzido de forma a reduzir o fator de utilização ρ de cada processador da máquina, de modo a se garantir tempos de espera em filas para processamento dentro dos limites desejáveis. Surge, então, o problema de que critério usar nesta redução. Definimos $K_\rho = 1/\rho$, onde ρ é o fator de utilização desejável nos processadores do sistema. A prática mostra que, em geral, este ρ deve situar-se entre 0,6 e 0,8, para as arquiteturas mais utilizadas. Para uma estimativa mais cuidadosa pode-se especificar o tempo médio desejável de espera nas filas utilizando como unidade o tempo de serviço. Existem algumas curvas que facilitam a obtenção de ρ a partir deste dado e do coeficiente de variação do tempo de serviço (MART 72).

A adoção destes critérios não garante, contudo, que o desempenho do sistema estará dentro dos índices desejados. Cumpre, desta forma, efetuar-se uma verificação através de um procedimento de análise do desempenho.

Como já foi abordado, a introdução de inúmeras simplificações e hipóteses no dimensionamento não justifica a utilização de modelos ou metodologias complexas para a análise do desempenho, nesta fase do projeto. Não se pode perder de vista, nesta altura, que o objetivo desta metodologia é o de obtenção de uma estimativa da capacidade de processamento de um sistema em desenvolvimento, embora um certo grau de detalhamento seja necessário para possibilitar a identificação de pontos críticos do sistema.

A experiência do projetista será, talvez, o fator determinante para a escolha do método a ser utilizado: exato, aproximado ou simulação, mas nos parece que para os objetivos visados os métodos aproximados são os que apresentam melhor compromisso entre a confiabilidade dos resultados e o tempo/custo de solução, sem descartar, contudo, a utilização de simulação com um modelo simplificado.

A utilização de métodos aproximados para a solução de sistemas de redes de filas permite, com facilidade, a utilização de algoritmos iterativos que possibilitam a análise da grande maioria dos sistemas práticos. A construção de programas com esta finalidade se apresenta como uma solução muito viável, cômoda e desejável, mas, de uma forma geral, qualquer método com que os projetistas tenham familiaridade pode ser utilizado para a validação dos dados obtidos no procedimento de dimensionamento (BARB 84) (KUEH 79) (CAMP 81).

Nas fases de definições de projeto nem sempre será possível levar a cabo esta validação, por falta de dados. Para a análise de desempenho é necessário que o sistema já esteja estruturalmente definido, tanto na parte física quanto lógica, para que o mapeamento das tarefas nos processadores possa ser estabelecido.

Quando os dados de desempenho obtidos da análise não estiverem dentro dos limites especificados para o sistema, todo o procedimento de dimensionamento e análise deve ser repetido, alimentado, desta vez, com dados revisados e com alterações estruturais, se necessário. Em alguns sistemas uma simples redução da utilização dos processadores, expresso em K_p , pode levar o sistema à faixa aceitável, mas em muitos outros casos alterações mais substanciais podem se mostrar necessárias.

IV - Aplicação a um Nô de Comutação de Pacotes

Este ítem apresenta o dimensionamento da capacidade de processamento de uma central X.25 de comutação de pacotes desenvolvida no laboratório de Sistemas Digitais da Escola Politécnica da USP. Esta central de comutação foi construída utilizando-se uma máquina de arquitetura distribuída tendo por rede de interconexão um anel do tipo DLCN (BARB 82) (RUGG 80). A figura 3 apresenta a arquitetura deste sistema

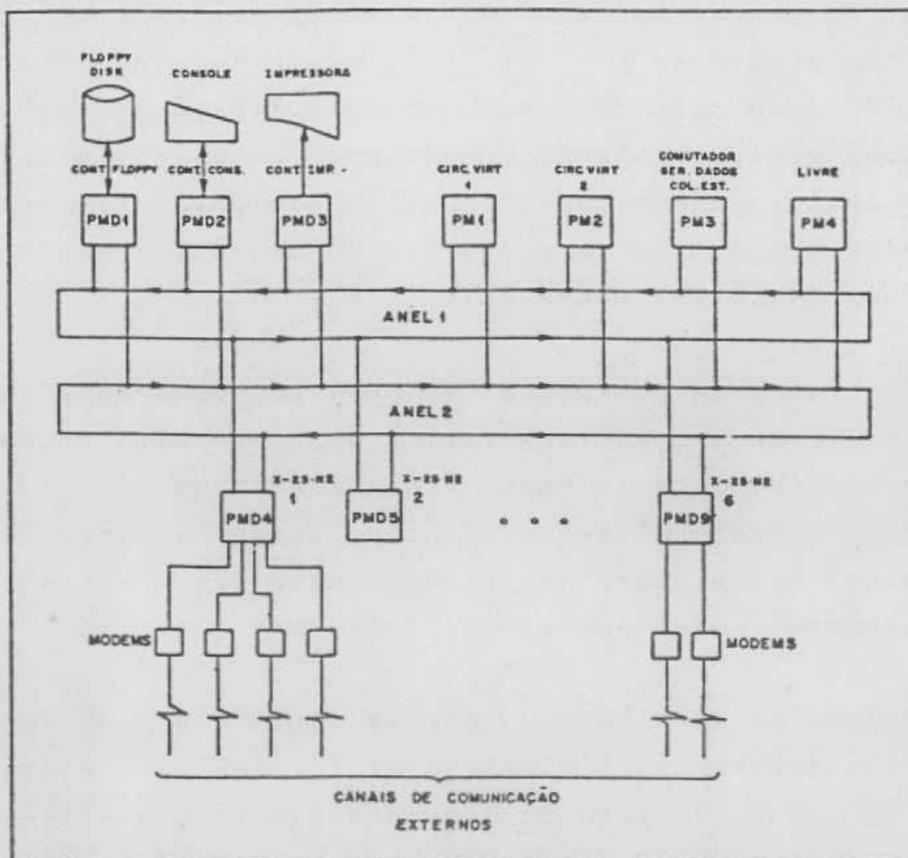


figura 3 - Arquitetura física do nó de comutação de pacotes

O processador utilizado em cada módulo do nó de comutação é o 8085, da Intel, operando com clock de 3MHz. Os tempos calculados a seguir são baseados nas características destes microprocessador.

Os primeiros valores a serem calculados são λ_{HMAX} e λ_{TRMAX} . A taxa λ_{TMAX} não é necessária, neste caso, porque o nó de comutação não aceita conexão direta de terminais.

Para o cálculo de λ_{TRMAX} a função $t_{TR} = t_{1TR} + t_{2TR}(S_{TR})$ pode ser escrita como:

$$t_{TR} = t_{1TR} + K_1 \cdot S_{TR},$$

onde K_1 representa o tempo de processamento por bit. Mas como este valor $K_1 \cdot S_{TR}$, é muito pequeno em relação à parte independente do tamanho do pacote, t_{TR} pode ser aproximado por t_{1TR} apenas:

$$t_{TR} \approx t_{1TR}$$

Então,

$$\lambda_{TRMAX} = \frac{S_{TR}}{t_{1TR}}$$

No caso de λ_{HMAX} t pode ser explicitado como:

$$t_H = t_{1H} + K_2 \cdot \frac{S_H}{S_{TR}} + K_3 \cdot S_H,$$

onde K_2 é o tempo de processamento na origem e destino associado a cada unidade de dados componente de uma mensagem e K_3 é o tempo de processamento por bit. Como no caso anterior ($K_3 \cdot S_H$) pode ser desprezado por ser muito pequeno em comparação aos demais fatores.

Desta forma:

$$\lambda_{HMAX} = \frac{S_H}{t_{1H} + K_2 \frac{S_H}{S_{TR}}}$$

Considerando-se uma média de 4 pacotes por ligação estabelecida, $S_H = S_{TR}$ (a rede não oferece serviços de montagem/desmontagem de mensagem), e $S_{TR} = 1000$ bits, obtem-se, pela análise da arquitetura lógica do sistema e pela média de instruções executadas em cada caso:

$$t_{1TR} = 19,7 \text{ ms}, \quad t_{1H} = 16,2 \text{ ms}, \quad K_2 = 14,3 \text{ ms}$$

Destes resultados obtemos: $\lambda_{HMAX} = 32$ Kbps e $\lambda_{TRMAX} = 50$ Kbps.

Finalmente é necessário definir os valores para K_s , K_ρ e K_μ .

Para definição de K_s vamos assumir que 10% da capacidade dos processadores é consumida pelo sistema de comunicação, portanto $K_s = 1,1$. Para o coeficiente de uniformidade K_μ vamos atribuir o valor 1,05 o que significa reservar uma capacidade adicional de 5% para compensar os desbalanceamentos de carga. Para K_ρ vamos estabelecer que em média qualquer transação espere pelo processamento tres vezes o tempo de serviço. Isto fornece um $\rho = 0,7$ (assumindo coeficiente de variação 0,64) e $K_\rho = 1,43$

Então,

$$P'' \geq 1.65 \left(\frac{\lambda_{TR}}{50K} + \frac{\lambda_H}{32K} \right)$$

Atribuindo, agora, valores a λ_T e λ_H e considerando P'' como o menor inteiro que satisfaz a inequação, obtemos os valores apresentados na tabela 1. Nesta tabela temos representados o número mínimo de processadores 8085 necessários para atender diversos valores de demanda de λ_{TR} e λ_H . Por exemplo, para atendimento de

um tráfego de 1 Mbps nas linhas troncos e 200 Kbps de tráfego de hospedeiros o número de processadores necessário é de 44. Como dado comparativo, em (JENN 76) foi apresentado um projeto de um nó de chaveamento de pacotes e circuitos implementado com múltiplos processadores. A vazão máxima especificada foi de 1,2 Mbps e seus autores previam a utilização de 36 processadores com tempo médio de execução de uma instrução de $2\mu s$ (cerca de 35% mais poderoso que o 8085), o que se aproxima muito do resultado obtido no nosso exemplo.

		$\lambda_H \rightarrow$						
		1	5	10	50	100	200	500 (Kbps)
λ_{TR} ↓	5	1	1	1	3	6	11	26
	10	1	1	1	3	6	11	27
	20	1	1	2	4	6	11	27
	50	2	2	3	5	7	12	28
	100	4	4	4	6	9	14	30
	150	5	6	6	8	11	16	31
	200	7	7	8	10	12	17	33
	300	10	11	11	13	16	21	36
	400	14	14	14	16	19	24	39
	500	17	17	18	20	22	27	43
1000 (Kbps)	34	34	34	36	39	44	59	

Tabela 1 - Número de Processadores 8085 necessários para suportar o tráfego ($\lambda_H + \lambda_{TR}$)

V - Conclusões

A utilização da metodologia de dimensionamento a apresentada é bastante ampla nas diversas fases do desenvolvimento de um sistema multimicroprocessadores. Ela funciona como um elemento auxiliar nas tomadas de decisões necessárias no decorrer do projeto e nas definições estruturais do hardware e, principalmente, do software. Ela se mostra especialmente útil nos sistemas de médio e grande porte e, embora o trabalho inicialmente elaborado por Falk e Mc Quillam visasse somente nós de comutação de pacotes, a sua aplicação se estende a uma vasta gama de equipamentos de comunicação de dados, embora a filosofia de análise seja aplicável a virtualmente qualquer equipamento multimicroprocessadores.

O presente artigo concentrou-se no procedimento de dimensionamento propriamente dito, apenas comentando em linhas gerais diretrizes a serem seguidas na fase de análise de desempenho. Neste aspecto, o objetivo foi estender o desenvolvimento apresentado em (Falk 77) para aplicação geral, explicitando ainda as dependências dos "overheads" destes sistemas e do desempenho desejado. Um sistema que automatize os procedimentos de análise de desempenho mostra-se bastante desejável, uma vez que uma maior velocidade na análise dos modelos se traduziria em maior flexibilidade na aplicação do método como um todo. Na época de redação deste trabalho uma primeira versão de tal sistema estava sendo desenvolvida, utilizando métodos aproximados para a solução.

Bibliografia

- (BARB 82) - BARBOSA, A.S., Ruggiero, W.V., Moscato, L.A., Campos, E.G.L., Stiubiener, S., "Rede Local no Laboratório de Sistemas Digitais da Escola Politécnica da Universidade de São Paulo", anais do 2º Simpósio Latino-Americano sobre Redes de Computadores, São Paulo, 1982.
- (BARB 84) - BARBOSA, A.S., "Dimensionamento e Análise de Desempenho de Sistemas Multimicroprocessadores para Redes de Computadores", Dissertação de Mestrado, Escola Politécnica da USP, em preparação.
- (BUX 79) - BUX, W, Kühn, P., Kümmerle, K., "Throughput Considerations in a Multi-Processor Packer Switching Node", IEEE Transactions on Communications, vol. com-27, NO. 4, april 1979.
- (CAMP 81) - CAMPBELL, C., Curso de Análise de Desempenho de Sistemas de Computação - notas de aula, Escola Politécnica da USP, 1981.
- (CAMP 83A)- CAMPOS, E.G.L., Barbosa, A.S., "Um PAD - Packet Assembler/Disassembler para Redes de Comutação de Pacotes", XVI Congresso Nacional de Informática-Anais, São Paulo, outubro 1983.
- (CAMP 83B)- CAMPOS, E.G.L., Barbosa, A.S., "Projeto de Implementação de um Equipamento PAD", 1º Simpósio sobre Redes de Computadores - Anais, Porto Alegre, 1983.

- (FALK 77) - FALK, G. Mc Quillan, J., "Issues in sizing store and forward communication switches", Computer Networking Symposium Proceedings, december 1977.
- (JENN 76) - JENNY, C.J., Kümmerle, K., "Distributed Processing Within an Integrated Circuit/ Packet-Switching Node", IEEE Transactions on Communications, vol. com.-24, NO.10, october 1976.
- (KUEH 79) - KUEHN, P.J., "Approximate Analysis of General Queuing Networks by Decomposition", IEEE Transactions on Communications, vol. com.-27, NO. 1, january 1979.
- (MART 72) - MARTIN. J., "System Analysis for Data Transmission", Prentice Hall, 1972.
- (RUGG 80) - RUGGIERO, W.V., Barbosa, A.S., "Implementação de uma central X.25 de comutação de pacotes através da aplicação de uma arquitetura distribuída multimicroprocessadores", VII Seminário Integrado de Software e Hardware, 1980.
- (SCHW 76) - SCHWEITZER, P.J., Lam, S.S., "Buffer Overflow in a Store-And-Forward Network Node", IBM Journal of Research and Development, november 1976.