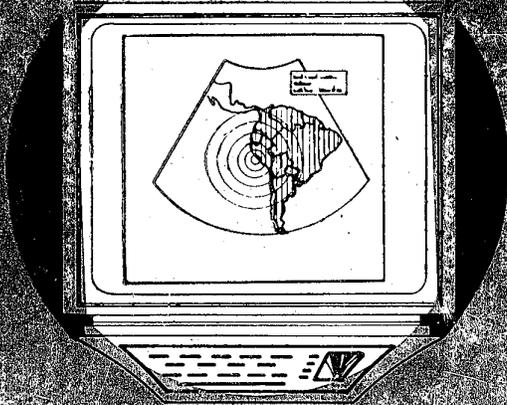


 **PANEL
INFO '82**



IX CONFERENCIA LATINOAMERICANA DE INFORMATICA

Centro Latinoamericano de Estudios de Informática (CLEI)
Asociación Peruana de Computación e Informática (APCI)
Universidad Nacional de Ingeniería (UNI)



**16 AL 20 DE AGOSTO DE 1982
LIMA - PERU**

**III
Convención Peruana
de
Computación e
Informática**

004.06
C748
V.1



**Anales de la
IX Conferencia Latinoamericana
de Informática**

Centro Latinoamericano de Estudios de Informática (CLEI)

y

III Convención Peruana de Informática

Organizado por:

**Asociación Peruana de Computación é Informática (APCI)
Universidad Nacional de Ingeniería (UNI)**

Patrocinado por:

Oficina Intergubernamental para la Informática (IBI)

Federación Internacional del Proceso de la Información (IFIP)

UNESCO

16 al 20 de Agosto de 1,982

LIMA - PERU

Daniel A. Menascé
Virgílio A.F. Almeida

Resumo

O processo de planejamento de capacidade dos recursos de computação requer uma ferramenta adequada para prever o desempenho das várias alternativas de configuração de uma instalação de processamento de dados. Este trabalho apresenta como ferramenta um modelo analítico, o qual é baseado na teoria das filas e na análise operacional, cujos fundamentos são também apresentados. A partir de parâmetros que caracterizam a carga de trabalho, o "Software" e o "Hardware" de um sistema de computação o modelo calcula as medidas representativas do desempenho deste sistema.



IX CONFERENCIA LATINOAMERICANA DE
INFORMATICA

16 al 20 de Agosto de 1982
Lima - Perú

1. Introdução

O processo de planejamento de capacidade dos recursos de computação requer uma ferramenta adequada para prever o desempenho das várias alternativas de configuração de uma instalação de processamento de dados.

Este trabalho apresenta como ferramenta um modelo analítico, o qual é baseado na teoria das redes de filas e na análise operacional. A partir de parâmetros que caracterizam a carga de trabalho, o "Software" e o "Hardware" de um sistema de computação, o modelo calcula medidas representativas do desempenho deste sistema.

2. O Planejamento de Capacidade

Ao se elaborar o plano de recursos para a área de informática, surgem, frequentemente, questões como:

- a atual configuração do sistema é suficiente para suportar o crescimento previsto?
- quanto tempo o atual sistema pode ser mantido sem alterações?
- o que acontecerá ao sistema "on-line" ao se adicionar mais 20 terminais?
- o que acontecerá ao se instalar a nova versão do "software" de gerência de banco de dados, que requer mais 200K de memória?

Enfim, estas questões podem ser resumidas na seguinte:

Qual a melhor estratégia de crescimento, em termos de custo/desempenho, para a configuração existente?

A principal finalidade do planejamento de capacidade é prover respostas às questões tais como as acima apresentadas.

2.1- Conceituação

O planejamento de capacidade é um processo de determinação do nível adequado dos recursos de "hardware", necessários ao atendimento dos requisitos de computação, e à manutenção dos níveis de serviços estabelecidos.

À luz desta definição, os elementos críticos que caracterizam o planejamento de capacidade são:

- . nível de serviço oferecido ao usuário
- . carga de trabalho corrente e prevista
- . utilização dos recursos do sistema

Ao nível de serviço estão associadas diretrizes como:

- . 90% dos tempos de resposta nos terminais de aplicações devem estar abaixo de 3 segundos.
- . "jobs" "batch" que utilizam menos de 1 minuto da UCP devem ter um "turnaround time" máximo de 1 hora, etc.

A carga de trabalho existente é o conjunto de todas aplicações processadas na instalação. Estas aplicações, em função da natureza, podem ser agrupadas em categorias, por exemplo:

- . transações de terminais
- . "jobs" "batch" de menos de 1 minuto de CPU
- . "jobs" de banco de dados, etc.

A utilização dos recursos representa o percentual de ocupação de cada dispositivo do sistema, para processar a carga da instalação. Por exemplo:

- . 80% de utilização de UCP
- . 35% de utilização do canal 1
- . 70% de utilização da unidade de disco, etc.

A capacidade de um sistema não pode ser caracterizada somente pela utilização dos recursos, pois esta medida nada indica sobre o nível de serviço oferecido.

Implícita no conceito de nível adequado de recursos de "hardware", está a idéia de um sistema balanceado, isto é, uma configuração onde a relação custo x benefício seja a mais favorável:

2.2- A Importância do Planejamento de Capacidade

As principais razões para a existência do planejamento de capacidade são:

- . Riscos para a empresa - o uso crescente de sistema de computação torna as empresas sensíveis ao funcionamento da área de informática. A impossibilidade de atender novos usuários, a degradação do nível de serviço, podem significar prejuízos para a empresa. O planejamento de capacidade é um meio de evitar tais riscos.

. Imagem externa - Uma característica importante de muitas aplicações "on-line" é que o desempenho do sistema, principalmente o tempo de resposta, causa impacto direto ao público. Quando o desempenho é fraco, os clientes manifestam insatisfação, e a imagem da empresa é prejudicada. A preservação de uma boa imagem é preocupação permanente da alta administração da empresa. Com o crescimento das aplicações "on-line" voltadas para o usuário final, fica evidenciada a constante necessidade de se manter o desempenho do sistema dentro de limites aceitáveis.

A elaboração de uma estratégia de alterações do "hardware", consistente com os padrões de desempenho estabelecidos, é uma das atividades do planejamento de capacidade.

Tempo para instalação de equipamentos - A experiência tem mostrado que o tempo de importação solicitado pelos fabricantes varia entre 6 e 14 meses. Portanto, a prática de esperar que o desempenho de um sistema atinja níveis insatisfatórios, para então iniciar um processo de aquisição, torna-se altamente perigosa. O planejamento de capacidade fornece instrumentos capazes de prover, em tempo hábil, alternativas de crescimento para uma instalação.

2.3- Uma Ferramenta para o Planejamento de Capacidade

O processo de planejamento de capacidade requer uma ferramenta para a previsão de desempenho.

A partir das estimativas de carga e dos níveis de serviço, deve-se obter com esta ferramenta as curvas de desempenho do sistema:

Esta previsão pode ser feita através de várias técnicas. A escolha da técnica é baseada em três parâmetros; complexidade, precisão e custo.

Os modelos analíticos são técnicas de solução que estabelecem relação entre parâmetros descritores do sistema e determinados critérios de desempenho. Estas relações são expressas através de equações, que são analiticamente resolvíveis.

Das disciplinas matemáticas utilizadas para construir modelos analíticos, as mais importantes são a teoria das filas e análise operacional [4]. Um sistema de computação pode ser visto como um sistema de múltiplos recursos, onde os recursos são a UCP, memória, canais e dispositivos de E/S. Os programas demandam serviços destes recursos. A maior parte dos problemas de desempenho está relacionada aos atrasos causados pela disputa de recursos. A teoria das filas e a análise operacional provem uma estrutura matemática para tratar esses problemas. A seguir, serão apresentados os fundamentos teóricos do modelo apresentado neste artigo.

3. Análise Operacional

Análise Operacional [1] é um ramo da matemática que trata das relações entre os atributos de sistemas reais ou hipotéticos. Estes atributos devem ser mensuráveis. O método operacional é baseado em um conjunto de conceitos, que corresponde direta e naturalmente às propriedades observadas nos sistemas de computação. As suposições e definições utilizadas por este método refletem o ponto de vista prático dos analistas engajados nos estudos empíricos de desempenho de sistemas de computação. Neste item, os termos "rede de filas" e "servidor" serão substituídos, respectivamente, por sistema e dispositivo.

3.1- Áreas de Aplicação

Dentro da análise do desempenho de sistemas de computação, as principais áreas de aplicação do método operacional são:

cálculo do desempenho - os métodos operacionais podem ser usados para calcular quantidades que não foram medidas, mas que poderiam ser.

verificação de consistência - a não verificação de leis ou teoremas operacionais indicará que os dados utilizados são inconsistentes.

previsão de desempenho - os métodos operacionais podem ser usados para estimar o desempenho futuro do sistema.

3.2 - Princípios

Todas as equações e resultados da análise operacional são derivados dos seguintes princípios operacionais:

- todas as quantidades devem ser definidas de modo que possam ser precisamente mensuráveis, e todas afirmações estabelecidas devem ser diretamente testáveis. Os resultados só serão válidos se forem obtidos de acordo com os princípios acima.
- o sistema deve ter um fluxo balanceado, isto é, o número de chegadas a um dado dispositivo, num período T de observação, deve ser igual ao número de saídas deste dispositivo no mesmo período T. Para ser modelado, o sistema deve ser conservativo, o que significa que, num dado período de observação, o número de entradas em um determinado estado do sistema deve ser igual ao número de saídas deste estado.
- os dispositivos devem ser homogêneos, isto é, as rotas dos "jobs" dentro do sistema devem depender do comprimento das filas nos dispositivos do sistema. Além disto, o tempo médio de serviço em um dado dispositivo não deve depender do comprimento das filas dos outros dispositivos.

3.3- Análise das Redes de Filas

Sistemas de computação podem ser modelados através de redes de filas. Um conceito básico na análise e solução de redes de filas é o de estado da rede, que é a descrição das localizações e status dos "jobs" da rede. Um estado da rede é formado pela especificação do estado de cada componente da rede. Portanto, a descrição do estado de uma rede com M servidores é dado por M-tuplas ordenadas:

$$\underline{S} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_M)$$

onde \underline{x}_i é uma descrição da fila i. A forma de \underline{x}_i depende da disciplina de atendimento e da distribuição do tempo de serviço na fila i. Se existem classes de "jobs" na rede, então \underline{x}_i conterá informações sobre as classes. Uma característica de \underline{x}_i é que ele sempre contém o número de "jobs" na fila i.

Na análise do desempenho, o objetivo não é encontrar a solução das redes, mas sim a determinação de medidas que caracterizam o comportamento da rede.

3.5- Descrição do Modelo

O modelo analítico utilizado neste trabalho baseia-se numa rede de filas, cujas características estão descritas abaixo.
Seja uma rede fechada de filas onde:

- M - é o número de dispositivos da rede
 r - é uma classe qualquer de "jobs", $r=1,2,\dots,R$.
 n_{ir} - é o número de "jobs" da classe r presentes no dispositivo i .
 N_r - é o número total de "jobs" da classe r presentes na rede,

$$N_r = \sum_{i=1}^M n_{ir}.$$
 N - é o número total de "jobs" presentes na rede,

$$N = \sum_{r=1}^R N_r.$$
 Y_{ir} - é o uso médio do dispositivo i por "job" da classe r , ou seja, o tempo médio total que um "job" da classe r usa o dispositivo i .
 X_{ir} - taxa de partida de pedidos da classe r no dispositivo i . A taxa de partida do sistema ("throughput"), referente à "jobs" da classe r , é designada por X_{or} .
 U_{ir} - é a utilização do dispositivo i por "jobs" da classe r .

Os dispositivos da rede considerada têm um comportamento independente da carga a que estão submetidas.

3.5.1- Solução da Rede

Um estado da rede é dado pelo vetor \underline{n} :

$$\underline{n} = (n_1, n_2, \dots, n_M)$$

onde \underline{n}_i é outro vetor dado por:

$$\underline{n}_i = (n_{i1}, n_{i2}, \dots, n_{iR})$$

tal que
$$\sum_{i=1}^M \sum_{r=1}^R n_{ir} = N.$$

Dentro do período T de observação, $T(\underline{n})$ é o tempo total que a rede permaneceu no estado \underline{n} . A probabilidade, ou fração de tempo, que o sistema se encontra no estado \underline{n} é dada por:

$$p(\underline{n}) = T(\underline{n})/T \quad (1)$$

tal que
$$\sum_{\underline{n}} p(\underline{n}) = 1,$$
 onde $\sum_{\underline{n}}$ é a somatória de todos os possíveis estados \underline{n} .

A solução da rede [2] é dada por:

$$p(\underline{n}) = \frac{1}{G} \prod_{i=1}^M Y_{i1}^{n_{i1}} Y_{i2}^{n_{i2}} \dots Y_{iR}^{n_{iR}} \quad (2)$$

Para se obter a expressão (2) não foi assumida nenhuma hipótese quanto à

distribuição do tempo de serviço dos servidores da rede [2].

Na expressão, G é uma constante de normalização definida como:

$$G = \sum_{n \in S(N, M, R)} \prod_{i=1}^M Y_{i1}^{n_{i1}} Y_{i2}^{n_{i2}} \dots Y_{iR}^{n_{iR}} \quad (3)$$

onde $S(N, M, R) = \{ \underline{n} = (n_{11}, n_{12}, \dots, n_{1R}, n_{21}, \dots, n_{2R}, \dots, n_{nR}) \}$

$$\sum_{i=1}^M n_{ir} = N_r \quad \& \quad \sum_{r=1}^R N_r = N \quad \& \quad n_{ir} \geq 0 \quad \forall i, r$$

O cálculo de G pode ser feito através do seguinte algoritmo:

$$G_i(K_1, \dots, K_R) = G_{i-1}(K_1, \dots, K_R) + Y_{i1} G_j(K_1^{-1}, K_2, \dots, K_R) + \dots + Y_{iR} G_i(K_1, \dots, K_{R-1}, K_R^{-1}) \quad (4)$$

onde K_1, K_2, \dots, K_R representam o número de "jobs" das classes $1, 2, \dots, R$ respectivamente.

A expressão (4) permite, iterativamente, o cálculo eficiente da constante de normalização G .

As condições iniciais para o uso da expressão (4) são:

$$G_i(K_1, K_2, \dots, K_R) = 1, \text{ se } K_1 = K_2 = \dots = K_R = 0 \quad (5)$$

$$G_0(K_1, K_2, \dots, K_R) = \begin{cases} 1, & \text{se } K_1 = K_2 = \dots = K_R = 0 \\ 0, & \text{caso contrário} \end{cases} \quad (6)$$

Finalmente, as expressões (4), (5), e (6) permitem o cálculo de:

$$G = G_M(N_1, N_2, \dots, N_R)$$

3.5.2 - Cálculo de Medidas de Desempenho

Algumas medidas de desempenho, que caracterizam o comportamento de um sistema de computação, podem ser calculadas em função da constante de normalização G .

Sejam:

$U_i(N_1, \dots, N_R)$ - utilização do dispositivo i quando no sistema existem N_1, N_2, \dots, N_R "jobs" das classes $1, \dots, R$ respectivamente.

$U_{ir}(N_1, \dots, N_R)$ - a utilização do dispositivo i pelos "jobs" da classe r , quando no sistema existem N_1, \dots, N_R "jobs" das classes $1, \dots, R$, respectivamente.

$X_0(N_1, \dots, N_R)$ - "Throughput" do sistema, quando no mesmo existem N_1, \dots, N_R "jobs" das classes $1, \dots, R$, respectivamente.

$X_{Or}(N_1, \dots, N_R)$ - "Throughput" da classe r , quando no sistema existem N_1, \dots, N_R "jobs" das classes $1, \dots, R$, respectivamente.

Como exemplo temos:

$$U_i(N_1, \dots, N_R) = \frac{Y_{i1} G(N_1-1, N_2, \dots, N_R) + \dots + Y_{iR} G(N_1, \dots, N_{R-1}, N_R-1)}{G(N_1, \dots, N_R)} \quad (7)$$

$$U_{ir}(N_1, \dots, N_r, \dots, N_R) = \frac{Y_{ir} G(N_1, \dots, N_r-1, N_{r+1}, \dots, N_R)}{G(N_1, \dots, N_R)} \quad (8)$$

$$X_{Or}(N_1, \dots, N_R) = \frac{G(N_1, \dots, N_r-1, N_{r+1}, \dots, N_R)}{G(N_1, \dots, N_R)} \quad (9)$$

A soma dos "throughputs" por classe fornece o "throughput" total do sistema :

$$X_0 = \sum_{r=1}^R X_{Or} \quad (10)$$

$$X_0(N_1, \dots, N_R) = \frac{G(N_1-1, N_2, \dots, N_R) + \dots + G(N_1, \dots, N_{R-1}, N_R-1)}{G(N_1, \dots, N_R)} \quad (11)$$

4. Especificação do Modelo PC/1

4.1- Finalidade

O modelo, também chamado de PC/1, é uma ferramenta auxiliar no processo de planejamento de capacidade [3]. Dentro deste enfoque, o seu uso é apropriado para:

- . prever o impacto de novas aplicações
 - . avaliar alternativas de "hardware"
 - . determinar a substituição de componentes (UCP, memória, E/S) do sistema.
 - . avaliar o impacto da instalação de novos terminais, etc.
- A figura 1 apresenta um diagrama da função deste modelo.

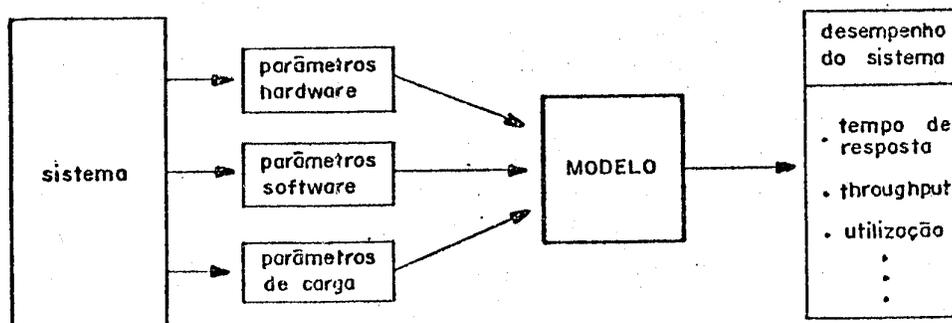


Figura 1 - Função do Modelo

4.2- Conceitos

O modelo PC/1 vê um sistema de computação como um conjunto composto de cargas de trabalho e de recursos. Os conceitos de desempenho surgem quando estes recursos processam estas cargas de trabalho.

4.3- Cargas de Trabalho

A carga de trabalho, composta de "jobs" e transações, é o conjunto das aplicações processadas pelo sistema. Os termos "job" e "transação" referem-se à unidade de trabalho que chega ao sistema provenientes de alguma fonte externa, recebem uma quantidade de processamento e abandonam o sistema. O termo "job" é mais usado no processamento "batch", enquanto o termo "transação" é mais comum no processamento interativo.

A importância do conceito de carga de trabalho reside no fato de que as medidas que caracterizam o desempenho de um sistema são calculadas por carga de trabalho.

As cargas de trabalho, de acordo com a modalidade do uso do sistema, são classificadas em três tipos; "batch", "time sharing" e transação. À cada tipo de carga está associado um conjunto de parâmetros, que caracteriza completamente esta carga.

4.3.1- Cargas "Time Sharing"

A natureza das cargas "time sharing", ou interativas, está mostrada na figura 2.

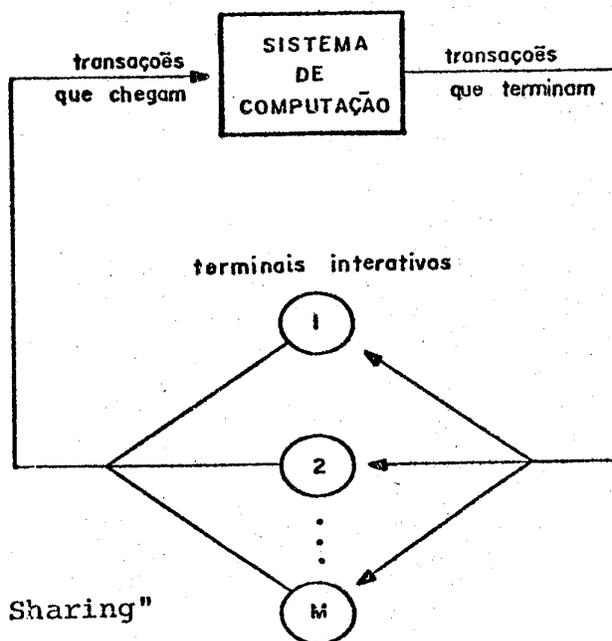


Figura 2 -
Carga "Time - Sharing"

Cada carga "time sharing" é associada a um conjunto de terminais interativos. Os usuários dos terminais geram transações, que chegam ao sistema

e são processadas. Quando a transação termina, o usuário recebe a resposta no terminal e gera nova transação, que repetirá o ciclo. Assim, a carga "time sharing" é caracterizada pelo número de terminais a ela associados, pelo tempo de pensar e pelo consumo dos recursos do sistema.

O tempo de pensar é definido como o intervalo médio de tempo entre a resposta a uma dada transação e o início, pelo usuário, da transação seguinte.

O consumo de recurso é definido pelo tempo total médio que um "job" típico desta carga gasta em cada dispositivo do sistema.

As principais medidas de desempenho, calculadas para esta classe, são o tempo de resposta por transação e a taxa de processamento ("throughput").

4.3.2- Carga Transação

A carga transação tem um processamento interativo similar ao descrito para a carga "time sharing". A diferença entre estes dois tipos de carga situa-se na maneira de especificá-las. Na carga "time sharing", a especificação da chegada de transações é feita em função do número de terminais e do tempo de pensar. A especificação da carga transação é feita, diretamente, através do número de transações que chegam ao sistema por unidade de tempo (figura 3).

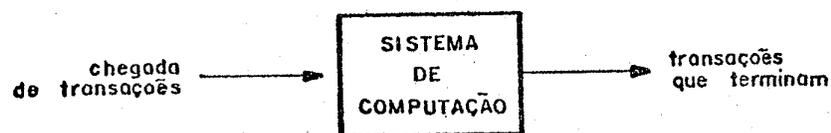


Figura 3 - Carga Transação

O consumo de recursos do sistema e a taxa de chegada caracterizam a carga transação.

Como a taxa de chegada é um parâmetro de entrada para o modelo, e como a taxa de saída ("throughput") deve ser igual à de chegada (o sistema é conservativo), a principal medida de desempenho calculada para a carga transação é o tempo de resposta.

4.3.3- Carga "Batch"

No caso da carga "batch", tem-se interesse de analisar o sistema em situações de carga intensa.

Nestas situações, existe sempre, pelo menos, um "job" na fila, aguardando a vez de ser carregado na memória. A figura 4 representa esta situação de carga intensa. O "loop" fechado, conforme mostra a figura, indica que

cada "job" que termina é, imediatamente, substituído por outro idêntico.

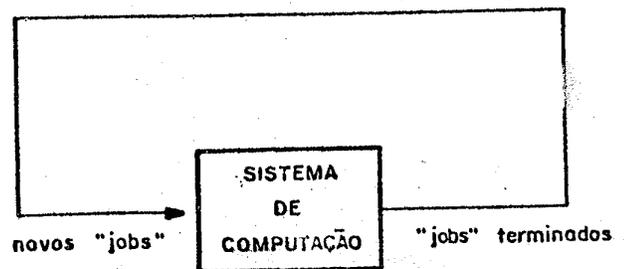


Figura 4 - Carga "Batch"

A carga "batch" é caracterizada pelo nível médio de multiprogramação e pelo consumo de recursos.

As medidas de desempenho, calculadas para esta carga, são o tempo de retorno dos "jobs" "batch" e a taxa de processamento.

4.3.4- Múltiplas Classes

Para se obter uma análise mais detalhada do sistema, o modelo permite que a carga de trabalho total do sistema seja subdividida em até seis cargas. Este limite é apenas uma restrição da atual implementação do programa, e não uma restrição do modelo analítico.

4.3.5- Dispositivos

Se for seguido o caminho que um "job" percorre ao ser processado por um sistema de computação, será observado que este "job" sempre estará em um dos seguintes estados:

- . recebendo serviço de algum recurso ativo do sistema, isto é, UCP, E/S, etc.
- . esperando na fila por algum recurso ativo.
- . esperando na fila por algum recurso passivo (memória, tabelas, etc.), o qual deve ser adquirido para que o "job" passe a receber serviço de algum recurso ativo.

No âmbito do modelo, os recursos ativos são chamados de dispositivos ou servidores.

4.4- Especificação das Entradas do Modelo PC/1

As entradas para o modelo descrevem, de uma maneira simples e precisa, o sistema de computação. Esta descrição é composta da especificação das cargas de trabalho e do consumo dos recursos do sistema, e baseia-se na descrição dos dados de entrada do BEST/1 [5].

4.4.1- Descrição das Cargas

Os descritores de carga (figura 6) são usados para especificar as seguin

tes informações:

- . tipo da carga de trabalho: "batch" (BP), transação (TP) ou "time sharing" (TS)
- . nome usado para identificar a carga nos relatórios de saída.
- . informações associadas ao nível de multiprogramação. No caso "batch", especifica-se o nível médio; nos outros casos (TP e TS), especifica-se o nível máximo de multiprogramação.
- . no caso de carga transação, especifica-se a taxa de chegada de transações ao sistema (trans./hora).
- . no caso de carga "time sharing", especifica-se o número de terminais associados à carga e o tempo médio de pensar.
- . no caso de existir prioridade, deve-se especificá-la. O modelo permite que seja associado uma prioridade de UCP à carga de trabalho. Esta prioridade é do tipo "preemptive resume", que é usualmente encontrada nos sistemas operacionais.

DESCRICAO DO SISTEMA (DS)

.....CARGA 1.....DESCRITORES.....
 BP CARGA-BATCH
 3.8 NIVEL MEDIO DE MP

.....CARGA 2.....DESCRITORES.....
 TP DATA-BASE-TRANS
 4000.0 TRANSACOES/HORA
 4.0 MAXIMO NIVEL DE MP

.....CARGA 3.....DESCRITORES.....
 TS CARGA-TIME-SHARING
 30.0 NO. DE TERMINAIS
 25.0 TEMPO DE PENSAR SEG
 6.0 MAXIMO NIVEL DE MP

DISPOSITIVO	CARGA 1	CARGA 2	CARGA 3
1 CPU 370/16	13000	205	320
2 DISCO PAG.	1099	83	92
3 DISCO PAG.	1940	147	175
4 FITA 3420/	395	0	0
5 FITA 3420/	496	0	0
6 3350 SYS.1	747	0	0
7 3350 SYS.2	204	0	0
8 3330 ARQ.1	4042	13	80
9 3330 ARQ.2	12461	0	0
10 3330 ARQ.3	12462	0	0
11 3330 ARQ.4	2474	64	42
12 3330 ARQ.5	2474	62	46
13 3330 ARQ.6	3034	150	100
14 3330 SPOOL	3035	150	100
15 3330 SPOOL	1073	27	200
16 3330 WORK	1075	29	0
17 3330 WORK	1727	49	0
18 3330 WORK	1727	52	0

Figura 6

No exemplo da figura 6, a carga "batch" tem um nível médio de multiprogramação igual a 3.8, sendo que este nível deve se referir a situações de carga intensa. A carga transação tem um nível de multiprogramação que varia de 0 a 4, e recebe transações à uma taxa de 4000 por hora. A carga "time sharing" possui 30 terminais, um tempo médio de pensar de 25 seg. e um nível de multiprogramação máximo igual a 6.

4.4.2- Descrição do Consumo dos Recursos

O segundo conjunto de parâmetros de entrada especifica a quantidade total de tempo que um "job"/transação gasta em cada dispositivo do sistema. Estes parâmetros são fornecidos por carga de trabalho. Deve ser ressaltado que esta quantidade, chamada de consumo de recurso, refere-se somente ao tempo que o "job"/transação gasta recebendo serviço do dispositivo. O tempo de permanência nas filas por dispositivo não deve ser incluído nesta quantidade.

Por exemplo, na figura 6 um "job" típico da carga "batch" consome, ao ser processado, 13000 mseg. de CPU, 1099 mseg. do dispositivo 2, etc. Uma transação típica da carga "time sharing" consome 320 mseg. de CPU, 92 mseg. do dispositivo 2, etc.

Estas quantidades, chamadas de consumo de recursos, são, em geral, obtidas através de monitores de "software" e do "software" de contabilização dos sistemas de operacionais.

4.5- Saídas do Modelo PC/1

A partir da descrição do sistema, o modelo calcula, por carga de trabalho, as principais medidas que caracterizam o desempenho de sistemas de computação.

Os principais relatórios fornecidos pelo modelo são:

- . resultados principais (figura 7)
- . utilização do sistema
- . utilização da memória (figura 8)

RESULTADOS PRINCIPAIS (RP)

CARGA	TEMPO DE RESPOSTA (SEG)	TAXA DE PRCC. (JOBS/HR)	(CPU
1 CARGA-BATCH	133.97	102.F/ HORA	36.8
2 DATA-BASE-TRANS	3.17	4000.F/ HORA	22.7
3 CARGA-TIME-SHARING	3.23	3825.F/ HORA	34.0
UTILIZACAO TOTAL DA CPU			92.6

Figura 7

RELATÓRIO DE UTILIZAÇÃO DA MEMÓRIA (UM)

	CARGA 1	CARGA 2	CARGA 3
NIVEL MAXIMO DE MULTIPROG.	4.0	4.0	6.0
NIVEL MEDIO DE MULTIPROG.	3.800	3.861	2.666
FILA MEDIA P/ MEMORIA	0.000	31.047	0.326
UTILIZACAO DA MEMORIA	0.950	0.965	0.611
P (0 NA MEMORIA)	0.000	0.005	0.035
P (1 NA MEMORIA)	0.000	0.015	0.100
P (2 NA MEMORIA)	0.000	0.022	0.156
P (3 NA MEMORIA)	0.200	0.024	0.178
P (4 NA MEMORIA)	0.800	0.931	0.164
P (5 NA MEMORIA)	0.000	0.000	0.129
P (6 NA MEMORIA)	0.000	0.000	0.236
P (0 NA FILA P/ MEMORIA)	1.000	0.099	0.857
P (1 NA FILA P/ MEMORIA)	0.000	0.026	0.058
P (2 NA FILA P/ MEMORIA)	0.000	0.025	0.036
P (3 NA FILA P/ MEMORIA)	0.000	0.024	0.021
P (4 NA FILA P/ MEMORIA)	0.000	0.023	0.012
P (5 NA FILA P/ MEMORIA)	0.000	0.023	0.006
P (MAIS QUE 5)	0.000	0.777	0.006

Figura 8

4.6- Exemplos

Para ilustrar a utilização do modelo PC/1, serão apresentados alguns exemplos que mostram como o PC/1 pode prover respostas às questões do tipo: "qual será ...?"

4.6.1- Instalação de Novos Terminais

No momento de se prever novos terminais para um sistema, surgem sempre questões sobre o futuro desempenho desse sistema. Por exemplo, qual será o novo tempo de resposta ao se instalar no sistema (o mesmo apresentado na figura 6) 15 novos terminais. A figura 9 apresenta as respostas a essa questão.

*** RESULTADOS PRINCIPAIS ***

CARGA	TEMPO DE RESPOSTA	TAXA DE PROC.	% CPU
1 CARGA BATCH	165.17	82. P/ HORA	29.9
2 CARGA TRANSACAO	5.55	4000. P/ HORA	22.7
3 CARGA TIME-SHARING	6.15	5199. P/ HORA	46.2
UTILIZACAO TCTAL DA CPU			98.9

Figura 9

5. Validação

Os modelos analíticos, baseados na teoria das redes das filas, tem sido usados com sucesso na previsão de desempenho de sistemas de computação. As principais medidas, tempo de resposta, taxa de processamento e utilização, têm sido obtidas com erros de no máximo 10%.

Referências

1. BUZEN, J.P. - "Operational Analysis: The Key to the new generation of performance prediction Tools", in Proc. IEEE COMPCON, IEEE, New York, 1976.
2. ROODE, J.D. - "Multiclass operational analysis of Queueing networks models", Performance of Computer Systems, North Holland Publishing Company, 1979.
3. ALMEIDA V., MENASCÉ A. - "Uma ferramenta para o Planejamento de Capacidade, Anais do XIV Congresso Nacional de Informática, São Paulo, Brasil, 1981.
4. MENASCÉ, D.A. e Virgílio A.F.Almeida - "Computing Performance Measure of Computer Systems with variable degree of Multiprogramming", Proceedings of the XII CMG Conference, New Orleans, USA, dezembro/1981.
5. BUZEN, J.P. - "BEST/1 - design of a Tool for computer System capacity planning", in Proc 1978 AFIPS National Computer Conf., Vol 47, AFIPS Press, Montvale, N.J., pp 447-455, 1978.