Anais do I WORCAP, INPE, São José dos Campos, 25 de Outubro de 2001, p. 82-84

SIMULAÇÃO DE FILAS *GI/G/m*

José Olimpio Ferreira[*, 1,2], Solon Venâncio de Carvalho [**, 1]

(1) Área de Pesquisa Operacional
Laboratório Associado de Computação e Matemática Aplicada
Instituto Nacional de Pesquisas Espaciais (INPE)

(2) Departamento de Ciência da Computação
Universidade Católica de Goiás (UCG)

(*)Doutorado, e-mail: olimpio@nucleo.inpe.br; (**) Orientador, e-mail: solon@lac.inpe.br

Resumo

A avaliação de desempenho de um sistema de fila *GI/G/m* é um clássico problema difícil. Estes sistemas de filas com múltiplos servidores são notoriamente difíceis de serem analisados analiticamente, sendo que uma solução analítica exata somente é possível para casos especiais, tais como as *filas M/M/m, M/D/m, M/K$_2$/m, GI/H$_2$/m, GI/E$_K$/m, GI/Ph/m, Ph/D/m e Ph/Ph/m*. Os modelos de simulação por computadores, apesar de exigirem muitos recursos de hardware e de consumirem muito tempo na execução das rodadas da simulação, permitem considerações mais próximas das situações reais. Este trabalho trata da estimação de medidas de desempenho, em equilíbrio, de filas *GI/G/m* através de simulação. Também é tratada a verificação de aproximações de filas *GI/G/m* por filas *Ph/Ph/m*.

Palavras-Chave: Simulação,Processos Estocásticos, Teoria de Filas, Sistemas de Filas

1 - Introdução

Os problemas de fila tiveram basicamente sua origem no trabalho de Erlang em telefonia no início do século. E, atualmente, a aplicação de teoria de fila na análise de performance de sistemas de computação, comunicações, produção industrial, transporte, manutenção, entre outros, tem estimulado muitas pesquisas aplicadas em aspectos computacionais de modelos de filas [7].

Este trabalho trata da obtenção de medidas de desempenho, em equilíbrio, de filas *GI/G/m* através de simulação. Estes sistemas de fila *GI/G/m* são considerados em isolamento, onde os clientes requerem um único serviço e abandonam o sistema após ter recebido o serviço [7].

Um modelo de simulação é um tipo particular de modelo matemático de um sistema. Na simulação, um computador é usado para avaliar um modelo numericamente por um período de tempo de interesse, e dados são coletados para estimar as verdadeiras características de interesse do modelo.

Neste estudo foi adotado o modelo de simulação a eventos discretos, pois o sistema de fila *GI/G/m* evolui através do tempo mas o estado de suas variáveis muda somente nos instantes em que ocorre a chegada de um cliente ou o término do serviço a um cliente (eventos).

A realização de um estudo de simulação completo e seguro envolve a observância de alguns passos, que são consensuais na literatura sobre simulação: formulação do problema; estabelecimento de objetivos; codificação do modelo; verificação; validação; projeto experimental; realização e análise das execuções; execuções adicionais; documentação do programa e relatório dos resultados [2,3].

Existem dois tipos de estimadores a considerar com relação as medidas de performance de uma fila *GI/G/m*. O primeiro tipo se refere a estimativas que são obtidas a partir dos dados de saída da simulação através da média aritmética ordinária e o segundo tipo se refere a estimativas obtidas através da média aritmética ponderada pelo tempo (freqüentemente, referida como média integrada no tempo) [2].

A precisão e exatidão do modelo programado necessárias à verificação foi feita através da comparação entre as estimativas das medidas de performance oriundas do modelo simulado e os valores destas medidas obtidas por modelos analíticos correspondentes, para configurações de filas que possuem soluções analíticas exatas [1], tais como: *M/M/m, D/D/m, M/Ph/m, Ph/M/m, Er/Ek/m* e *Ph/Ph/m*. E em última hipótese sempre é possível recorrer à verificação da taxa de utilização efetiva do sistema ou à verificação do número de servidores efetivamente ocupados, estimados dos dados de saída da simulação, para verificar a precisão e exatidão do modelo programado.

Foi adotado neste experimento de simulação a estratégia de replicações independentes com descarte de uma porção inicial de cada replicação, e com tamanhos de execuções das replicações suficientemente longos para atenuar o vício [6].

2 – Descrição dos Experimentos

Foram consideradas 8 configurações - cenários de filas *Gm/Gm/2*, através da especificação de $\rho = 0{,}75$ e $c_x^2 \in \{0{,}5;\ 1;\ 2\}$, onde, $\rho = \lambda E(S)/m$ é a taxa de utilização do servidor, e c_x^2 é o coeficiente de variação ao quadrado dos processo de chegada ou de serviço.

O tempo médio de serviço foi considerado sempre igual a 1, ou seja, $E(S) = 1$, o que implica em $\rho = \lambda/m$. A unidade de medida do tempo de simulação foi o minuto, exceto quando explicitada outra unidade.

O experimento de simulação para cada configuração dos sistemas de filas previstos consistiu de 10 replicações com fluxos de números aleatórios (sementes) diferentes para cada replicação.

Cada execução da simulação foi dividida em duas fases, a fase de inicialização e a fase da coleta de dados. Desta forma, o tempo T_I de inicialização da simulação, durante o qual a simulação executará sem que sejam coletados dados de saída da simulação, será determinado por (1) e multiplicado por um fator igual 10 para garantir uma adequada remoção do vício. Portanto tem-se que [4,5]:

$$T_I = 40b/a^2(1-\rho)^2 \text{ , onde } \quad a = -m \text{ e } b = m\left(c_A^2 + c_S^2\right) \qquad \textbf{(1)}$$

Desta forma o período (ou tempo) total de execução *TE*, considerando uma largura absoluta ε especificada (precisão estatística especificada), foi determinado por [4,5]:

$$T_E = \frac{8\left(c_A^2 + c_S^2\right)z_{\beta/2}^2}{m\rho^2\varepsilon^2(1-\rho)^2} \qquad \textbf{(2)}$$

O tamanho total da simulação foi dividido em 10 macro-replicações de tamanhos $T_E/10$ e cada uma destas macro-replicações foi dividida em 10 micro-replicações independentes de tamanhos $\frac{T_E/10}{10}$, onde cada uma destas micro-replicações independentes tem semente geradora, própria, de números aleatórios e período de inicialização T_I. O tamanho total da simulação foi acrescido de 25% e de 50% e colhem-se dados para o tempo original T_E, para o tempo $1{,}25T_E$ e para o tempo $1{,}5T_E$.

Como o modelo programado de simulação transforma $I \geq 0$ entradas em $J = 5$ saídas é importante salientar que o nível de significância β = 0,05 especificado para a análise estatística foi repassado por $\beta_J = 0{,}05/5 = 0{,}01$, para cada saída $J \in \{1,2,3,4,5\}$, com base na desigualdade de *Bonferroni*. Isto implica que maiores discrepâncias entre os valores esperados das medidas de performance e as estimativas oriundas da simulação são aceitáveis e, consequentemente, mantém a probabilidade total de erro do tipo 1 no experimento abaixo do valor β = 0,05.

A análise dos resultados da experimentação consistiu de: (1) verificação da precisão relativa especificada de 2% ($\varepsilon \leq 0{,}02$) para um nível de confiança, com grau de confiança da ordem de 95% ou superior, das medidas de desempenho, em equilíbrio, das diferentes configurações – cenários - replicações do sistema de filas *GI/G/m* previstos no projeto de experimento; (2) análise dos resultados da aplicação de testes estatísticos padrões, como o teste *t* de Student, quando os valores verdadeiros das medidas em observação são conhecidos; (3) averiguação da probabilidade de cobertura verificada nos dados de saída da simulação, quando os valores verdadeiros das medidas em observação são conhecidos.

3 – Resultados e Conclusões

Com relação aos experimentos efetuados visando a verificação e validação do modelo de simulação foram obtidos os seguintes resultados com relação ao tamanho total da simulação, T_E. Para 4 destes experimentos foi obtida, com os valores originalmente previstos, uma precisão relativa menor ou igual a 2% ($\varepsilon \leq 0{,}02$). Mas para os experimentos referentes às filas $E_2/H_2^b/5$ e $H_2^b/E_2/5$ a precisão relativa hl_r para o processo L_q foi de 0,021372 e 0,020570, respectivamente e, para o processo W_q foi de 0,021261 e 0,020337, respectivamente. Para estas filas foram necessários tamanhos totais de execução da simulação 25% maiores que os originalmente previstos, para a obtenção da precisão relativa de 2%. Os outros dois experimentos se referem às filas *D/D/1* e *D/D/2* os resultados obtidos foram exatamente iguais aos valores verdadeiros das características destas filas de interesse para este estudo. A probabilidade de cobertura observada (P_{co}) foi da ordem de 100% para 5 destes experimentos, com exceção daquele relativo à fila *M/M/2* com taxa de utilização de 20%, para o qual a probabilidade de cobertura para a estimativa do número de servidores ocupados ($\hat{NSO}$) foi de 90%. O tamanho da execução deste experimento foi revisto, tomando-se um período total de simulação 25% maior que o originalmente previsto. E com este novo período total de execução da simulação obteve-se 100% de cobertura. A probabilidade de cobertura observada (P_{co}) não se aplica para aos experimentos relativos às filas *D/D/1* e *D/D/2*.

Os resultados dos 8 experimentos de simulação previstos para a verificação e validação do modelo de simulação da fila *GI/G/m* em equilíbrio pode ser verificado que os experimentos relativos às filas *D/D/1* e *D/D/2* com ρ = 80% produziram estimativas, das medidas de performance destas filas, exatamente iguais aos valores verdadeiros destas características, obtidas de um modelo analítico exato. Com relação às demais filas

o modelo programado da simulação produziu estimativas suficientemente precisas (a meia largura, hl_r, do intervalo de confiança é menor ou igual a 2%) e exatas (a $Bias_r$ é praticamente desprezível). A probabilidade de cobertura observada, P_{co} , é de 100% contra o valor projetado de 95%. Estas estimativas passaram pelo teste de hipóteses (não há evidências de que elas difiram do valor verdadeiro destas características, em equilíbrio).

Desta forma pode-se concluir que o modelo programado (código) para a simulação em equilíbrio de filas *GI/G/m* é uma implementação suficientemente precisa e exata do modelo conceitual do sistema de filas *GI/G/m*. Em outras palavras, para os valores de projeto da simulação em equilíbrio de filas *GI/G/m*, o modelo programado (código) de simulação funciona como o pretendido.

Foram efetuadas as execuções dos 8 experimentos de simulação fila *Gm/Gm/2* e analizados os resultados. Para 4 destes experimentos foi obtida uma precisão relativa menor ou igual a 2% ($\varepsilon \leq 0{,}02$) para um nível de confiança da ordem de 95%, com os valores originalmente previstos. Mas para os outros 4 experimentos foram necessários tamanhos totais de execução da simulação maiores que os originalmente previstos, para a obtenção da precisão relativa de 2%. Para a fila *Gm/Gm/2* com $c_T^2 = 0{,}5$ e $c_S^2 = 0{,}5$ e a fila *Gm/Gm/2* com $c_T^2 = 0{,}5$ e $c_S^2 = 1$ foram necessários tempos totais de execução da simulação 50% maiores que os valores inicialmente previstos. E para a fila *Gm/Gm/2* com $c_T^2 = 0{,}5$ e $c_S^2 = 2$ e a fila *Gm/Gm/2* com $c_T^2 = 2$ e $c_S^2 = 2$ foram necessários tempos totais de execução da simulação 25% maiores que os valores inicialmente previstos.

Para todas as estimativas das medidas de performance previstas no experimento, pode-se notar, que as meias larguras relativas do intervalo de confiança foram inferiores a 0,02, conforme a especificação de projeto para uma precisão relativa de 2% ($\varepsilon \leq 0{,}02$) e um nível de confiança da ordem de 95%.

Desta forma, a Equação (1) para a determinação do período de inicialização se mostrou adequada para a remoção do vício. A Equação (2) quando multiplicada por um fator de no máximo 1,5 se mostrou suficiente para a obtenção de estimativas dentro de uma precisão estatística relativa especificada das configurações de filas *GI/G/2* previstas nos experimentos. Mas parece razoável que fossem realizados experimentos com outras configurações de filas *GI/G/m* para a obtenção de uma conclusão mais segura, com processos de chegadas e ou de serviços distribuídos segundo a distribuição de Weibull, por exemplo: *W/W/m*, *Gm/W/m, W/Gm/m.*

Parece importante observar da equação (2), que o período de execução T_E é crescente para $0{,}5 \leq \rho < 1$ e decrescente para $0 \leq \rho \leq 0{,}5$. Consequentemente o período de execução T_E é o mesmo para ρ = 0,1 e ρ = 0,9. Questiona-se, então, se a equação em (2) é adequada para a determinação do período de execução T_E para $0 \leq \rho \leq 0{,}5$, pois estes períodos de execução T_E parecem longos demais para esta faixa de variação de ρ. Não seria suficiente um período de execução T_E da ordem daquele para ρ = 0,5 ou menor do que este período para a faixa de taxas de utilização do sistema $0 \leq \rho \leq 0{,}5$.

Referências

[1] Kleijnen, J. P. C. European Journal of Operational Research, 1:145 (1995)

[2] Law, A. M.; Kelton, W. D. Simulation modeling and analysis, McGraw-Hill, New York, 1982

[3] Poole, T.G.; Szymankiewicz, J. Using simulation to solve problems, MacGraw Hill, London, 1977.

[4] Whitt, W. Management Science, 11:1341 (1989a)

[5] Whitt, W. Simulation run length planning. **In**: Proceedings of Winter Simulation Conference. Washington, 1989. Heidelberg, P.; MacNair, E. A.; Musselman, K. J. ed., New York, 1989b, 106-112**.**

[6] Whitt, W. Management Science, 6:645 (1991)

[7] Wolff, R. W. Stochastic modeling and the theory of queues, Prentice Hall, Englewood Cliffs, 1989.