

SIMULAÇÃO DO SERVIÇO DE CORREIO ELETRÔNICO ATRAVÉS DE UM MODELO DE FILAS

Fábio E. Doy

Dep. de Estatística / IME / Univ. de São Paulo (USP)
São Paulo – SP
fdoy@ime.usp.br

Graça Bressan

Dep. de Eng. de Computação e Sistemas Digitais / EP
Universidade de São Paulo (USP)
São Paulo – SP
gbressan@larc.usp.br

Gustavo H. de A. Pereira

Dep. de Estatística / IME / Univ. de São Paulo (USP)
São Paulo – SP
ghapereira@yahoo.com

Marcos N. Magalhães *

Dep. de Estatística / IME / Univ. de São Paulo (USP)
São Paulo – SP
marcos@ime.usp.br

* *Corresponding author* / autor para quem as correspondências devem ser encaminhadas

Recebido em 11/2004; aceito em 07/2005 após 1 revisão
Received November 2004; accepted July 2005 after one revision

Resumo

Estudamos, através da simulação de um modelo de filas, o serviço de correio eletrônico (*e-mail*) realizado pela rede de computadores do Instituto de Matemática e Estatística da Universidade de São Paulo, com o objetivo de identificar características estatísticas e avaliar seu desempenho. Uma análise do fluxo de mensagens foi feita para obter informações sobre o processo de chegadas, o tempo de serviço e a proporção de sucessos na entrega, além de outras características. Com base nessas informações, um modelo de filas é proposto e simulado para avaliar os tempos de espera das mensagens nesse sistema. O efeito de alterações nos parâmetros do modelo é também estudado.

Palavras-chave: características estatísticas; modelo de filas; simulação; internet.

Abstract

We study, through simulation of a queuing model, the *e-mail* service provided by the Internet site of the Institute of Mathematics and Statistics of University of São Paulo, in order to identify statistics characteristics and to evaluate its performance. A data analysis of messages flow was done to obtain insight about the arrival process, the service time and the proportion of success, besides other characteristics. From this, a queuing model is proposed and simulated to evaluate waiting times of the messages in the system. The effect of changes in the parameters of the model is also studied.

Keywords: statistics characteristics; queuing model; simulation; internet.

1. Introdução

O serviço de correio eletrônico (*e-mail*) tem se tornado, nos últimos anos, uma forma de comunicação comum em várias camadas sociais. Não é raro solicitar a alguém que envie alguma notícia ou informação por *e-mail* o que, alguns anos atrás, seria feito por telefone. A tendência parece ser desse serviço se tornar cada vez mais acessível, tendo em vista sua rápida expansão em escolas do ensino médio e empresas. O presente trabalho continua a pesquisa iniciada em Doy *et al.* (1999) em que o serviço de *e-mail* do IME-USP foi avaliado. Naquele estudo, com dados referentes ao período compreendido entre 01/10/97 e 19/06/98, algumas características estatísticas foram observadas, um modelo de filas foi proposto e o serviço de *e-mail* avaliado através das médias de algumas medidas de desempenho. No trabalho atual, propomos um modelo de filas (ver Magalhães, 1996) simplificado e o simulamos para avaliar o desempenho do sistema. A alternativa de usar a simulação é motivada pela dificuldade técnica de um estudo analítico. Mais adiante voltaremos a essa questão.

Teoria das Filas tem sido utilizada para estudar o desempenho de sistemas de comunicação. Em Menascé & Almeida (2003) resultados dessa teoria são aplicados para responder diversas questões referentes ao planejamento e comportamento de redes. Com o avanço dos diversos tipos de serviço disponíveis pela Internet, estudos referentes ao tráfego na rede tornaram-se muito importantes para assegurar uma qualidade de serviço adequada. Algumas referências recentes são Garetto & Towsley (2003), Heyman & Lucantoni (2003) e Soule *et al.* (2004). No caso específico de serviço de *e-mail* não encontramos muitas referências. As mais próximas são os artigos de Bertolotti & Calzarossa (2001) e Charzinski (2003) que estudaram tráfegos de *e-mail*, via os protocolos SMTP e POP3. Apesar de alguns pontos comuns, esses estudos têm ênfase diferente daquela apresentada no nosso trabalho. Eles sugerem modelos probabilísticos para as chegadas e o tamanho das mensagens. Estudam também outras características buscando identificar as demandas de trabalho do servidor, originadas no serviço de mensagens.

A rede de computadores do IME-USP é mostrada na Figura 1. Ela se conecta à Internet através do *backbone* da rede USPnet (www.uspnet.usp.br) de 100 Mbps do Campus da Cidade Universitária da USP, com os usuários internos conectados em rede *Ethernet* de 10 Mbps. O serviço de *e-mail* é realizado pelo sistema *Qmail*, com configurações estabelecidas pelos administradores da rede.

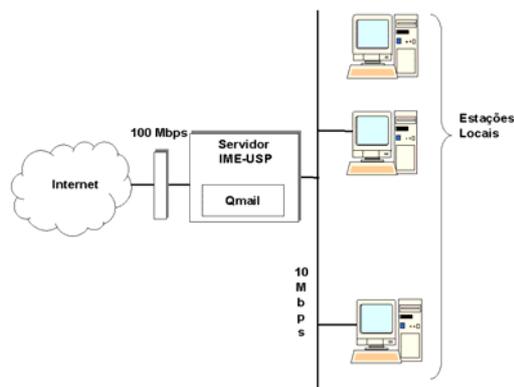


Figura 1 – Rede do IME-USP.

O restante deste artigo apresenta a seguinte organização: a próxima seção contém uma análise do conjunto de dados, buscando identificar características estatísticas relevantes. Na Seção 3, o modelo de filas é estabelecido e são apresentados os parâmetros e as distribuições de probabilidade necessárias para a simulação. Na Seção 4, avaliamos o efeito de alterações nos parâmetros do modelo. A conclusão é apresentada na Seção 5 e as referências vêm em seguida.

2. Análise Descritiva

Nesta seção, apresentamos a análise descritiva dos dados coletados em um dia de semana, considerado típico, de março de 2000. Dos dados de que dispúnhamos, descartamos os dias com situações especiais tais como ausência de aulas na pós-graduação, feriados e interrupções ocasionais da rede, entre outras. Dado o grande volume diário de mensagens, as informações desse dia típico são suficientes para as tarefas de estimação envolvidas na obtenção do modelo a ser simulado.

As informações sobre o fluxo das mensagens na rede foram obtidas a partir do *Log* diário do sistema *Qmail* e compreendem uma massa significativa de informações. Para o tratamento inicial das informações foram criados programas na linguagem *Awk* e a análise descritiva dos dados foi feita com o auxílio de planilhas e os pacotes estatísticos *S-plus* e *SPSS*.

Uma mensagem de *e-mail* tem sempre remetente e destinatário que podem ser de dois tipos. Um remetente (ou destinatário) é do tipo local se seu endereço eletrônico pertence à rede IME. Caso não pertença, o remetente (ou destinatário) será denominado remoto. Dessa forma, existem 4 possíveis direções de tráfego: LL, LR, RL e RR, com L e R indicando o usuário local e remoto, respectivamente.

Tendo em vista o interesse de estudar o sistema em momentos de maior frequência de mensagens, os dados analisados nesse trabalho se referem às chegadas de mensagens no período compreendido entre 9 e 19 horas, que foi identificado como o de maior incidência de mensagens. Isto ocasionou, possivelmente, um sub-dimensionamento no tráfego de origem remota, já que muitas dessas mensagens provêm do exterior e chegam ao IME fora do período acima mencionado, devido aos diferentes fusos horários existentes. De qualquer forma, essas chegadas encontram o sistema com pouca carga e, assim, do ponto de vista do estudo de desempenho do sistema é suficiente o período considerado.

Uma mensagem, ao chegar no sistema, pode estar endereçada a vários usuários simultaneamente. Diremos que houve um desdobramento de mensagem. Isso acontece nas listas dirigidas a um certo setor profissional ou interesse específico. Nesse caso, considera-se que ocorreram chegadas em grupo, todas com o mesmo instante inicial no sistema. O processamento, entretanto, seguirá individualizado pela rede. O volume de tráfego em cada direção é apresentado na Tabela 1. O tráfego de origem local representa mais de 75% do total de mensagens e num dia típico pode alcançar mais de 13 mil mensagens. O tráfego RR corresponde ao redirecionamento, por usuários locais, de suas mensagens para um servidor externo. Também pode ter ocorrido falha no sistema que não impediu o tráfego de passagem, que não é autorizado pelos administradores da rede.

Tabela 1 – Porcentagem de mensagens em cada direção.

| Direção | LL | LR | RL | RR |
|-------------|-------|-------|-------|------|
| Porcentagem | 56,2% | 21,8% | 18,4% | 3,6% |

As mensagens processadas podem resultar na ocorrência de um sucesso, caso que corresponde à entrega ao destinatário. Podem, também, resultar em fracasso, caso em que a mensagem é devolvida ao remetente. O fracasso na entrega da mensagem provém de uma condição que impossibilita a transmissão, tal como endereço inválido do destinatário. Existem situações em que uma mensagem adia sua saída do sistema, em geral, devido à uma impossibilidade temporária de acessar o destinatário. A Tabela 2 mostra, para cada direção, a proporção de adiamentos, falhas e sucessos na primeira passagem da mensagem pelo sistema.

As mensagens adiadas retornam para um novo processamento após um tempo de espera que depende de quanto tempo já transcorreu desde sua entrada no sistema. Antes do prazo máximo de permanência ser excedido, as mensagens podem fazer várias passagens até saírem do sistema. Os adiamentos ocorrem em número reduzido na maioria das direções, mas afetam as médias das medidas de desempenho.

Tabela 2 – Porcentagem de sucessos, falhas e adiamentos por direção na primeira passagem.

| Direção | Sucesso | Fracasso | Adiamento | Total |
|---------|---------|----------|-----------|-------|
| LL | 87,6% | 0,3% | 12,1% | 100% |
| LR | 95,8% | 1,5% | 2,7% | 100% |
| RL | 95,7% | 3,1% | 1,2% | 100% |
| RR | 95,1% | 1,0% | 3,9% | 100% |

Uma mensagem adiada sempre sairá do sistema em algum momento futuro e, então, haverá falha ou sucesso na sua entrega. O tempo de permanência de uma mensagem no sistema depende de sua origem, destino e também de ter sido entregue ou devolvida. As mensagens entregues, que resultam em sucesso, tendem a demorar mais no sistema, pois há necessidade de enviá-la ao destinatário. Por outro lado, mensagens que resultam em fracasso são, em geral, fruto de uma rápida verificação de algum equívoco de endereço e devolvidas ao remetente.

O tempo de permanência é contado a partir do momento em que o sistema registra a entrada da mensagem até sua saída da rede. A Tabela 3 apresenta o tempo médio total no sistema, em segundos, para mensagens nos diversos tipos de tráfego.

Tabela 3 – Tempo médio total no sistema para mensagens atendidas.

| Tipos de tráfego | Tempo médio total (segundos) |
|------------------|------------------------------|
| LLS | 79,5 |
| LLF | 7,0 |
| LRS | 137,4 |
| LRF | 43,6 |
| RLS | 11,7 |
| RLF | 2,1 |
| RRS | 134,0 |
| RRF | 44,8 |

Conforme constatamos de nossa análise descritiva, a origem, o destino e o resultado final (sucesso ou fracasso) interferem na passagem da mensagem pelo sistema. Na próxima seção, essas questões serão levadas em conta ao estabelecer a modelagem adequada.

3. Modelo de Simulação

O serviço de correio eletrônico apresenta dificuldades para um estudo analítico. Como indicado na análise descritiva, precisamos distinguir não só as diferentes direções das mensagens mas, também, os possíveis resultados. Também, ao chegar, uma mensagem pode ser desdobrada em várias outras devido às listas de interesse. Por sua vez, o adiamento no processamento de mensagens induz dependências no fluxo de usuários. Essa realimentação é, em geral, um problema difícil de ser estudado. Além disso, como veremos adiante, os tempos de serviço têm distribuição geral (não exponencial) para cada tipo de mensagem e resultado. Apesar de que técnicas analíticas ou aproximações possam ser tentadas, acreditamos que a complexidade envolvida no envio de mensagens justifica o emprego de simulação.

Para simular o serviço de correio eletrônico, estabelecemos um sistema de filas com um Centro de Recepção e dois Centros de Serviço (ver Figura 2). No Centro de Recepção a mensagem é classificada segundo a origem, o destino e o resultado de seu processamento. Nesse centro, são também desdobradas as mensagens originalmente enviadas a múltiplos usuários. Após passar pela recepção, a mensagem é enviada a um dos Centros de Serviço, de acordo com seu destino. Essa é uma característica física do sistema de processamento das mensagens e foi mantida na nossa modelagem. Assim, todas as mensagens com destino local são processadas no Centro de Serviço 1 e as com destino remoto vão ao Centro 2. Dessa forma, é possível, por exemplo, que um centro esteja com mensagens na fila de espera e o outro vazio. Os centros de serviço são considerados como tendo, em termos práticos, capacidade ilimitada de espera.

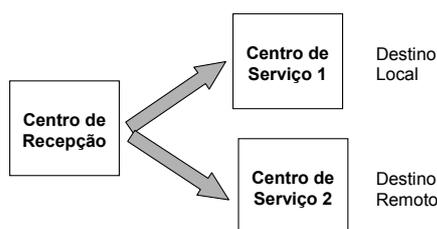


Figura 2 – Modelo para o sistema de *e-mail*.

Os Centros de Serviço 1 e 2 têm 10 e 20 servidores, respectivamente. O serviço em cada centro é feito em ordem de chegada a partir de uma fila única. O número de servidores em cada centro foi definido por um parâmetro de configuração do sistema *Qmail*. A Figura 3 apresenta um diagrama para o Centro de Serviço 1, que é análogo ao 2.

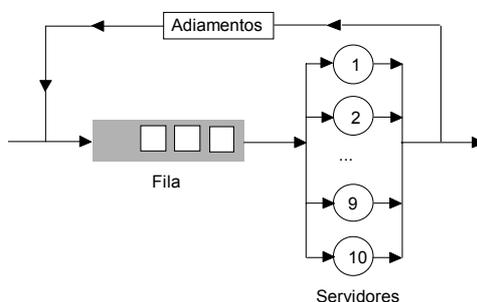


Figura 3 – Centro de Serviço 1 (destino local).

A classificação da mensagem é dada pela trinca ordenada (o,d,r) , representando, respectivamente, a origem, o destino e o resultado do processamento. O processo de chegadas das mensagens ao centro de recepção é assumido ser um processo de renovação com distribuição A_o , com o representando a origem da mensagem. No Centro de Recepção a mensagem é desdobrada de acordo com uma distribuição discreta de probabilidade D_o , também dependente da sua origem. A mensagem deixa o Centro de Recepção e é enviada ao Centro de Serviço 1 ou 2, de acordo com uma probabilidade condicionada à origem da mensagem. Essa probabilidade é representada por $p_{d|o}$. Com notação similar, definimos $p_{r|od}$ como a probabilidade condicional do resultado, dado que sabemos a origem e o destino da mensagem. Os tempos de atendimento nos centros de serviço dependem da classificação da mensagem. Exceto pelas dependências apontadas acima, os tempos entre sucessivas chegadas, entre sucessivos atendimentos e entre chegadas e atendimentos quaisquer são assumidos serem independentes.

Utilizando os dados disponíveis, buscamos estimar as probabilidades necessárias e ajustar funções de probabilidade para as diversas distribuições de interesse. Para as estimativas das probabilidades usamos as frequências relativas obtidas a partir da amostra. Para ajustar as distribuições, salientamos que os testes estatísticos de aderência, em amostras grandes, se tornam muito sensíveis às pequenas discrepâncias. Por essa razão é muito difícil que o teste aceite o ajuste a uma distribuição teórica de probabilidade. Como alternativa aos testes, utilizamos gráficos de probabilidade acumulada, conhecidos como gráficos $P-P$, disponíveis em vários pacotes estatísticos. Em seguida, usamos a opção *best fit* do *software* Arena, observando o comportamento do erro quadrático médio. Relatamos a seguir as escolhas feitas e os comentários pertinentes à cada caso.

Durante o período considerado, as chegadas locais e remotas tiveram número próximo de ocorrências. O intervalo médio entre chegadas foi igual a 11,4 e 11,5 segundos para as de origem local e remota, respectivamente. Existem diferenças de variabilidade, mas entre os vários ajustes testados, optou-se pelo modelo Exponencial pela simplicidade e por produzir um volume de mensagens bastante próximo do real. A Tabela 4 resume a estimação realizada.

Tabela 4 – Intervalo entre chegadas.

| | Distribuição entre chegadas sucessivas |
|---------------|--|
| Origem Local | $A_L \sim$ Exponencial ($\mu=1/11,4$); média = 11,4 seg. |
| Origem Remota | $A_R \sim$ Exponencial ($\mu=1/11,5$); média = 11,5 seg. |

O desdobramento das mensagens através das listas é um dos pontos mais delicados da estimação. Uma chegada desdobrada gera várias outras mensagens que serão processadas (de forma individual) pelo restante do sistema. Apesar de não ser típico, observamos em alguns dias desdobramentos de chegada local em até 1200 novas mensagens. Nos dados considerados, as listas têm número muito variado, com valores entre 1 e 21 para as de origem remota e de 1 a 235 para as de origem local. Existem ainda muitos valores intermediários com frequência muito baixa ou até mesmo inexistente. O conjunto de dados se comporta como se tivesse várias modas pontuais, com frequências relativamente altas ao redor desse valor. Isso poderia ser explicável, a partir de algumas listas que têm um certo número de componentes e que perdem, ou ganham adeptos eventuais, fruto de pedidos esporádicos de inclusão ou exclusão. Para enfrentar esses problemas, construímos uma distribuição empírica

de probabilidade nos inteiros de um certo intervalo real. Inicialmente, consideramos, separadamente, intervalos de valores observados. Para cada intervalo, usa-se a distribuição Uniforme Contínua ponderada pela frequência relativa do respectivo intervalo. Escolhido um número com essa estrutura probabilística, consideramos, então, o maior inteiro contido como o valor em que será desdobrada a mensagem. A escolha dos intervalos buscou incluir uma moda local, simétrica em relação aos limites do intervalo. Foram feitos ajustes nos intervalos de modo que a simulação pudesse ser validada. As Tabelas 5 e 6 apresentam as escolhas dos intervalos e as respectivas probabilidades, dependendo da origem da mensagem.

Tabela 5 – Desdobramento de chegada local.

| D_L | [1, 2) | [2, 3) | [3, 4) | [4, 8) | [8, 52) | [52, 150) | [150, 236) |
|-------|--------|--------|--------|--------|---------|-----------|------------|
| Prob. | 0,9159 | 0,0308 | 0,0079 | 0,0105 | 0,0165 | 0,0092 | 0,0092 |

Tabela 6 – Desdobramento de chegada remota.

| D_R | [1, 2) | [2, 3) | [3, 4) | [4, 12) | [12, 23) |
|-------|--------|--------|--------|---------|----------|
| Prob. | 0,9461 | 0,0138 | 0,0209 | 0,0176 | 0,0016 |

É sobre a mensagem desdobrada, cuja origem já é conhecida, que cabe completar sua classificação, indicando destino e resultado. Iniciamos, obtendo estimativas da probabilidade condicional do destino dado a origem. Fixado o destino, a mensagem é enviada ao correspondente centro de serviço. Por fim, usamos estimativas da probabilidade condicional do resultado dados à origem e o destino da mensagem. As estimativas estão nas Tabelas 7 e 8.

Tabela 7 – Estimativas de $P(\text{Destino} | \text{Origem})$.

| | Destino Local | Destino Remoto |
|---------------|---------------|----------------|
| Origem Local | 0,7247 | 0,2753 |
| Origem Remota | 0,8351 | 0,1649 |

Tabela 8 – Estimativas de $P(\text{Resultado} | \text{Origem}, \text{Destino})$.

| (Origem, Destino) \ Resultado | Sucesso | Fracasso |
|-------------------------------|---------|----------|
| LL | 0,9973 | 0,0023 |
| LR | 0,9852 | 0,0148 |
| RL | 0,9691 | 0,0309 |
| RR | 0,9901 | 0,0099 |

Para compor a Tabela 8 contamos sucessos e fracassos de mensagens que passaram pelo sistema no período especificado. Algumas delas poderiam ter tido mais de um processamento pelo centro de serviço e seriam as mensagens adiadas conforme indicamos na Tabela 2. Dos dados obtidos, percebemos que a distribuição desses adiamentos depende da classificação da mensagem. Destacamos que as mensagens que irão falhar têm comportamento bem diferente das que irão se tornar sucesso. No caso de fracasso, as tentativas são no máximo duas e a

modelagem usou uma distribuição discreta de probabilidade. Para as mensagens bem sucedidas, as distribuições de probabilidade são construídas de forma similar ao que foi feito no desdobramento de mensagens. Isto é, consideramos intervalos com distribuição Uniforme Contínua, ponderados pela frequência relativa de ocorrência. O número de tentativas é o maior inteiro contido no valor aleatório gerado dessa forma. As Tabelas 9a e 9b apresentam, em função da classificação, as escolhas que foram feitas para modelar os adiamentos.

Tabela 9a – Tentativas/ Adiamentos- Origem Local.

| Classificação | Função de Probabilidade | | | | |
|---------------|-------------------------|--------|---------|---------|----------|
| LLS | Tent. | [1, 2) | [2, 3) | [3, 26) | [26, 31) |
| | Prob. | 0,8772 | 0,0944 | 0,0156 | 0,0118 |
| LLF | Tent. | 1 | | | |
| | Prob. | 1 | | | |
| LRS | Tent. | [1, 2) | [2, 14) | | |
| | Prob. | 0,9725 | 0,0275 | | |
| LRF | Tent. | 1 | 2 | | |
| | Prob. | 0,9600 | 0,0400 | | |

Tabela 9b – Tentativas/ Adiamentos- Origem Remota.

| Classificação | Função de Probabilidade | | | | | |
|---------------|-------------------------|--------|--------|---------|----------|----------|
| RLS | Tent. | [1, 2) | [2, 3) | [3, 15) | [15, 25) | [25, 31) |
| | Prob. | 0,9877 | 0,0946 | 0,0007 | 0,0007 | 0,0063 |
| RLF | Tent. | 1 | 2 | | | |
| | Prob. | 0,9900 | 0,0100 | | | |
| RRS | Tent. | [1, 2) | [2, 3) | [3, 16) | | |
| | Prob. | 0,9603 | 0,0116 | 0,0281 | | |
| RRF | Tent. | 1 | | | | |
| | Prob. | 1 | | | | |

O tempo de serviço de cada mensagem é seu processamento nos centros de Recepção e de Serviço, buscando a entrega ou o retorno ao usuário. Assumimos que um tempo fixo é gasto pela mensagem na sua passagem pela Recepção. Apesar de poucas diferenças nesses tempos, distinguimos os vários tipos e as mensagens com e sem adiamento. As mesmas distinções são feitas nos Centros de Serviço para a caracterização do tempo de serviço. O ajuste, com as técnicas já mencionadas, é apresentado na Tabela 10.

Tabela 10 – Tempos de atendimento (segundos).

| Tipo de Processamento | Centro de Recepção (tempo fixo) | Centro de Serviço (tempo aleatório) |
|------------------------------|--|--|
| LLS | 0,121 | Erlang (0,657;3) |
| LLF | 0,143 | 5,83×Beta (0,588;2,38) |
| LLA | 0,146 | Erlang (0,443;2) |
| LRS | 0,104 | Weibull (16,6;0,851) |
| LRF | 0,115 | 63×Beta (0,576;0,911) |
| LRA | 0,118 | Weibull (21,5;0,429) |
| RLS | 0,124 | Lognormal (0,834;1,08) |
| RLF | 0,130 | 3,75×Beta (0,366;2,44) |
| RLA | 0,120 | Lognormal (0,314;0,365) |
| RRS | 0,112 | Weibull (12,7;0,533) |
| RRF | 0,144 | Unif. Cont. (0,80) |
| RRA | 0,132 | Weibull (40,9;0,411) |

O modelo de filas, descrito na Figura 2, foi simulado nas condições descritas acima, utilizando-se o *software* Arena versão 3.01. Ao leitor interessado em maiores detalhes sobre o programa de simulação construído, solicitamos que entre em contacto com os autores. O tempo simulado correspondeu a um total de 10 horas de funcionamento do sistema (9 às 19 horas) e foi precedido de um aquecimento de 604800 segundos (7 dias). Os resultados foram baseados em 30 réplicas.

Para validar o modelo de simulação, utilizamos um vetor de 16 quantidades referentes ao tempo total no sistema e ao número de mensagens processadas em cada uma das 8 possíveis classificações das mensagens. Construímos Intervalos de Confiança (99%) para essas quantidades e verificamos em quantos deles o verdadeiro valor estava incluído. O único valor fora do intervalo foi o referente ao tempo total de mensagens LLF. Tendo em vista existirem apenas 25 mensagens desse tipo, dentre um total de 16725 processadas num dia típico, consideramos que o modelo pode ser admitido como validado.

4. Alterações no Sistema

O efeito sobre as medidas de desempenho do sistema pode ser avaliado através da perturbação nos parâmetros do modelo simulado. Concentramos nossa avaliação no tempo médio total no sistema para as mensagens classificadas como LLS, LRS e RLS. Esses tipos de mensagens correspondem, em conjunto, à cerca de 94% do tráfego total processado em um dia típico. Destacamos na nossa análise 5 perturbações: chegada, falha-manhã, falha-tarde, adiamento e lista.

A intensidade de chegadas foi modificada, mantendo-se a mesma distribuição Exponencial, mas com a média aumentada. O volume de chegadas locais e remotas foi aumentado e os resultados sobre o tempo médio total representados na Figura 4. Esses tempos permanecem insensíveis até praticamente dobrarmos o volume de chegadas. A partir daí o sistema começa a acusar um certo atraso, sendo que as mensagens do tipo LRS são as mais afetadas. Note que essas mensagens são processadas no Centro de Serviço 2, o qual tem os maiores tempos médios totais. O comportamento observado na Figura 4 parece indicar que o sistema está operando com capacidade ociosa que, entretanto, é preenchida primeiro no Centro 2, que passa então a ser fortemente afetado pelo congestionamento.

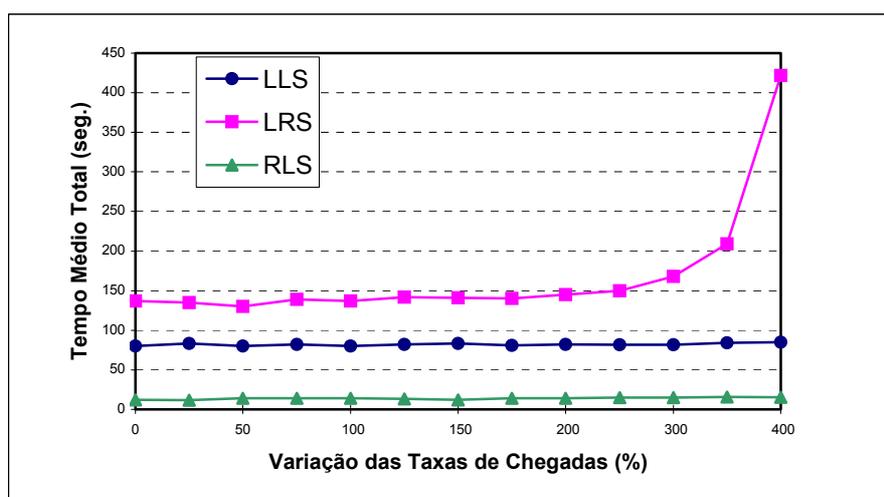


Figura 4 – Efeito do acréscimo nas chegadas.

Avaliamos o efeito de interrupções no sistema. Fruto de algum tipo de falha, consideramos que o sistema suspende suas atividades de atendimento por um certo intervalo de tempo. Chegadas continuam a ocorrer normalmente, mas nenhuma mensagem é processada nos centros de atendimento. Ao retomar as operações, o sistema continua o atendimento a partir da situação em que parou. As falhas foram programadas em um momento fixo da manhã (10 horas) ou da tarde (16 horas) e seu tempo de duração variou de 15 a 120 minutos. Distinguimos os períodos da manhã e da tarde, pois acreditávamos que eles poderiam ocasionar diferentes efeitos propagatórios, tendo em vista que as chegadas ocorrem a partir das 9 horas e se encerram às 19 horas, período esse em que são coletadas as estatísticas das medidas de interesse. Entretanto, dos resultados obtidos observamos um comportamento praticamente idêntico entre as suspensões temporárias matutinas e vespertinas. Como ilustração, apresentamos na Figura 5 o efeito no tempo total de uma paralisação matutina. Aparentemente, o sistema é robusto para paradas de pequena monta, algo em torno de 30 minutos, mas a partir daí o efeito no tempo médio total se agrava e acarreta um aumento considerável nessa medida.

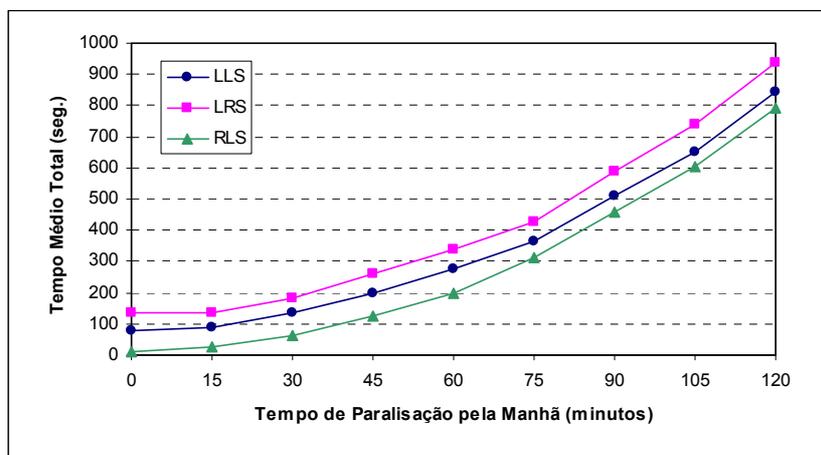


Figura 5 – Efeito das paralisações de atendimento (manhã).

Para completar nosso estudo sobre os efeitos no tempo total médio de permanência no sistema, avaliamos o efeito dos desdobramentos de mensagens e das tentativas. Para as distribuições empíricas obtidas nas Tabelas 5, 6 e 9, modificamos seus valores de modo a obter valores crescentes da média do número de tentativas e do número de mensagens desdobradas.

Na Figura 6, apresentamos os resultados referentes aos acréscimos percentuais no número médio de tentativas. O efeito explosivo é evidente e poderia ser previsível pelo atraso que o sistema impõe a cada mensagem, entre os sucessivos retornos ao centro de serviço.

A Figura 7 mostra o efeito do desdobramento, segundo diversos acréscimos percentuais no tamanho médio das listas. De novo, o sistema parece absorver os aumentos iniciais, indicando uma certa margem para o crescimento das listas. Porém, a partir do acréscimo de 200% no tamanho médio das listas de origem remota, as mensagens do tipo LRS são fortemente afetadas e seu tempo médio total de permanência sofre um aumento significativo.

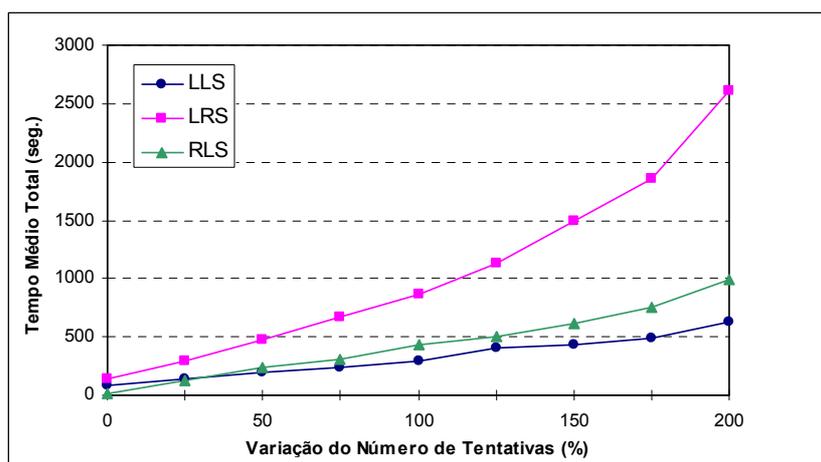


Figura 6 – Efeito do aumento das tentativas.

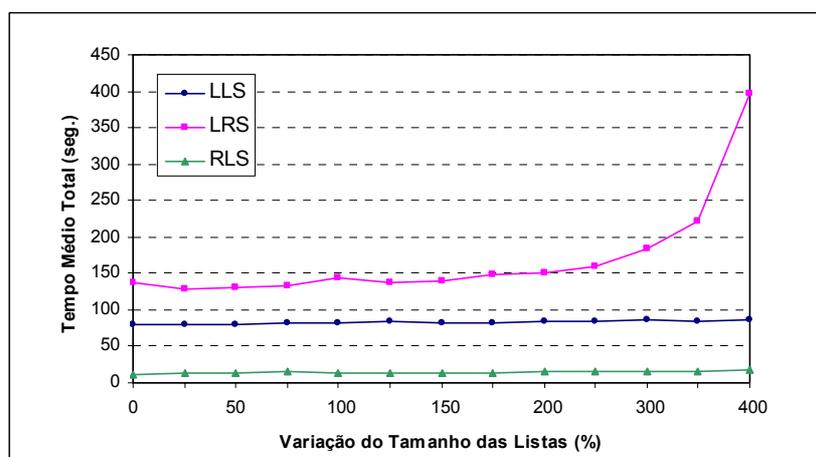


Figura 7 – Efeito do aumento nas listas.

5. Conclusões

Este trabalho analisou o sistema de correio eletrônico do IME-USP. Partindo da avaliação dos dados disponíveis, chegou-se a um modelo de filas que foi simulado para estudar a performance do sistema. Dentre as várias perturbações e diferentes possibilidades de alterações, escolhemos avaliar os efeitos sobre o tempo médio total de permanência no sistema para mensagens do tipo LLS, LRS e RLS.

Pela avaliação realizada o sistema tem um bom espaço para absorver crescimento em várias frentes. Ele se mostrou praticamente insensível à duplicação no número de chegadas de mensagens. Por outro lado, as paradas no sistema têm um efeito importante nas medidas estudadas. O efeito dos desdobramentos foi também avaliado. Eles se originam em listas de interesse e têm um grande potencial multiplicador no tráfego de mensagens pela rede. No sistema estudado, o acréscimo aplicado ao tamanho das listas pode ser em grande parte absorvido pela capacidade existente no sistema.

Mais que uma contribuição ao gerenciamento do sistema que foi estudado, este trabalho ilustra as possibilidades de análise do comportamento de sistemas computacionais, em especial os de correio eletrônico, através de um modelo de filas.

Atualmente, os servidores de *e-mail* têm sido inundados por mensagens, comerciais ou não, na maioria das vezes indesejáveis. Segundo Gomes *et al.* (2004), o chamado “spam” tem número de mensagens ao redor de 50% do volume total e corresponde a uma carga adicional de 10% no tráfego na rede. O efeito desse fenômeno não foi considerado no presente estudo, e parece ser um tópico promissor de pesquisa futura.

Agradecimentos

Agradecemos a colaboração dos administradores da Rede IME-USP e os comentários dos revisores anônimos.

Referências Bibliográficas

- (1) Bertolotti, L. & Calzarossa, M.C. (2001). Models of mail server workloads. *Performance Evaluation*, **46**, 65-76.
- (2) Doy, F.E.; Pereira, G.H.A.; Magalhães, M.N. & Bressan, G. (1999). Características Estatísticas do Serviço de *E-mail* da Rede IME-USP de Computadores. *Anais do XXXI Simpósio Brasileiro de Pesquisa Operacional*, Universidade Federal de Juiz de Fora, 1522-1532.
- (3) Charzinski, J. (2003). Observed performance of elastic Internet applications. *Computer Communications*, **26**, 914-925.
- (4) Garetto, M. & Towsley, D. (2003). Modeling, Simulation and Measurements of Queuing Delay under Long-tail Internet Traffic. *ACM-SIGMETRICS/Performance'03*, 47-57.
- (5) Heyman, D.P. & Lucantoni, D. (2003). Modeling Multiple IP Traffic Streams With Rate Limits. *IEEE/ACM Transactions on Networking*, **11**(6), 948-958.
- (6) Gomes, L.H.; Cazita, C.; Almeida, J.M.; Almeida, V. & Meira Jr., W. (2004). Characterizing a Spam Traffic. *ACM-ICM'04*, 356-369.
- (7) Magalhães, M.N. (1996). *Introdução à Rede de Filas*. 12°. SINAPE, Associação Brasileira de Estatística, São Paulo, SP.
- (8) Menascé, D.A. & Almeida, V.A.F. (2003). *Planejamento de Capacidade para Serviços na Web*. Editora Campus, Rio de Janeiro, RJ.
- (9) Soule, A.; Salamatian K.; Taft N.; Emilion; R. & Papagiannaki, K. (2004). Flow Classification by Histograms. *ACM-SIGMETRICS/Performance'04*, 49-60.