

# Avaliação de Desempenho da Política EBS para Escalonamento em Aplicações com Restrições de Máximo Tempo Médio de Resposta

Alessandro Nakamuta <sup>1</sup>  
Lourenço Alves Pereira Júnior <sup>1</sup>  
Francisco José Monaco <sup>1</sup>

**Resumo:** Este trabalho apresenta uma avaliação de desempenho do algoritmo EBS como política de escalonamento em aplicações cientes de QoS sob requisitos temporais tipo soft real-time. O estudo analisa parâmetros que descrevem a carga de trabalho e avalia seu impacto na capacidade de atender à especificação de limites superiores de tempo médio de resposta. O algoritmo EBS implementa uma política de escalonamento de prioridade dinâmica para aplicações de tempo-real flexível, a qual tem sido investigada em arquiteturas de escalonamento realimentado e escalonamento dinâmico on-line. São apresentados experimentos e resultados que analisam a influência dos fatores considerados na métrica de SLA.

**Palavras chave:** Avaliação de Desempenho, Tempo-Real, Qualidade de Serviço.

**Abstract:** This paper presents a performance evaluation study on the EBS algorithm as scheduling policy in QoS-aware applications under soft real-time requirements. The study analyses several parameters describing the workload and evaluates their impact on the capability to meet the specification of upper-bounded average response times. The EBS algorithm implements a dynamic-priority scheduling policy for soft real-time applications which has been investigated in feedback scheduling architectures and dynamical on-line scheduling. The paper presents experiments and results which analyze the influence of considered factors on the SLA metric.

**Keywords:** Performance evaluation, Real-Time Systems, Quality of Service.

## 1 Introdução

Um elemento que tem grande influência no desempenho de serviços interativos é o escalonador de processos. O escalonador é responsável por organizar a ordem de atendimento das requisições de modo a atender requisitos de qualidade de serviço (QoS). A fim

---

<sup>1</sup>Instituto de Ciências Matemáticas e de Computação, USP, Caixa Postal 668  
{naka,ljr,monaco}@icmc.usp.br

de satisfazer restrições temporais, em sistemas de tempo real, políticas de escalonamento apropriadas devem ser aplicadas para gerenciar convenientemente a alocação de recursos.

O algoritmo *Exigency Based Scheduler* (EBS) [1] é investigado como política de escalonamento para provisão de garantias de tempo tipo *soft-RT* (soft real-time), proposta em uma linha de pesquisa junto ao GSDPC-ICMC-USP<sup>2</sup>, a qual tem por objetivo o desenvolvimento de abordagens para provisão de garantias de QoS em aplicações cujo SLA determine limites superiores de tempo médio de resposta estabelecidos nos contratos entre o provedor de serviço e o contratante.

Trabalhos anteriores [14, 13, 12, 5, 17, 18, 19, 16] demonstram as propriedades e comparações do algoritmo com outras políticas convencionais no atendimento às especificações estocásticas de responsividade temporal. Esses estudos tem sido feitos mediante de técnicas de simulação utilizando carga de trabalho sintética parametrizada em termos de número clientes do serviço presentes, número de SLAs distintos a que são associados os usuários, taxa de requisições submetidas, taxa de atendimento (tempo de processamento) e outras medidas. Quando em aplicações reais, há interesse em saber-se qual a influência de cada um desses fatores na QoS oferecida, por exemplo, para que o provedor de serviço possa dimensionar eficientemente os recursos e implementar políticas de controle de admissão.

O presente artigo apresenta um experimento de avaliação de desempenho realizado para identificar essas influências e mensurar quantitativamente o impacto de cada um dos fatores analisados na variável de resposta, medida como a taxa de cumprimento do SLA.

## 2 Sistemas de Tempo Real

O requisito-chave na formulação das especificações de tempo-real (RT) é o limite superior para atrasos no tempo de reação do sistema aos estímulos externos. Em sistemas de tempo real, a meta do escalonador corresponde a observar o *deadline* (tempo máximo de resposta) de todas as requisições [2], de onde sua relevância em aplicações de TR. De acordo com o tipo de restrição temporal, sistemas de tempo-real podem ser classificados em três tipos principais: *Hard-RT*, *Firm-RT* e *Soft-RT*. Para sistemas *Hard Real-Time*, a precisão da garantia de tempo de resposta é criticamente importante e não pode ser depreciada em função de outros ganhos como desempenho e eficiência. Nesses sistemas, a falta (não cumprimento) em apenas uma restrição temporal implica em falha do sistema. Para sistemas *Firm Real-Time* a precisão de tempo é importante, porém a falta no seu cumprimento não implica necessariamente em falta, desde que sua ocorrência obedeça limites toleráveis. Finalmente, sistemas *Soft Real-Time*, que correspondem à restrição temporal aplicável ao presente estudo, são semelhantes aos sistemas tipo *Firm-RT*, com particularidade de que, diferente da restrição

---

<sup>2</sup>Grupo de Sistemas Distribuídos e Programação Concorrente, ICMC USP.

*Firm*, as requisições eventualmente atrasadas não são descartadas. Ao contrário, permanecem no sistema e devem ser atendidas, ainda que com *deadline* violado. Nesse cenário, faltas no atendimento das restrições temporais não implicam em falha do sistema, mas na degradação do serviço, quantificável por alguma métrica de QoS. Por exemplo, a política de atendimento de requisições em uma aplicação de comércio eletrônico pode estar sujeita a expectativas de tempo máximo de resposta, porém clientes mal servidos não são descartados.

### 3 EBS: *Exigency Based Scheduler*

A EBS é um algoritmo de escalonamento proposto para sistemas Soft-RT, baseado em prioridade dinâmica e sem suporte a preempção. Sua política de prioridade considera o impacto (exigência) que uma requisição representa ao sistema, que é tanto maior quanto mais próximo o *deadline* e quanto mais longo o tempo esperado de processamento. A prioridade é então dada pela ponderação entre esses dois parâmetros por meio da Eq. 1

$$P = D \cdot C \quad (1)$$

em que  $D$  é o *deadline*,  $C$  é o tempo esperado de execução e  $P$  é a prioridade atribuída a requisição. Em comparação com algoritmos clássicos de escalonamento, a lógica da EBS é, assim como a EDF<sup>3</sup> [10, 15], priorizar requisições com *deadlines* urgentes, porém somente se eles não têm tempo de processamento custoso, analogamente à SJF<sup>4</sup>[3, 8]. Assim, há um balanço entre a otimalidade da política EDF em cumprir *deadlines* e a otimalidade da SJF em reduzir o tempo médio de execução, o que se traduz no correspondente equilíbrio entre os requisitos de tempo-real e potencial *starvation*<sup>5</sup> de requisições pequenas.

O emprego da EBS foi investigado[14] como política de prioridade em uma arquitetura de provisão de QoS baseada em *feedback scheduling* (escalonamento retroalimentado) destinada a garantir o atendimento de limites superiores de tempo médio de resposta. Para isso, o sistema mantém uma lista de SLAs, com os respectivos limites especificados e os tempos médios de resposta efetivamente oferecidos. O *deadline* de cada requisição é calculado como o tempo máximo que aquela pode aguardar em fila até que, considerando ainda seu tempo de processamento, o tempo médio de resposta do cliente que a submeteu não crescerá mais do que o dado pelo SLA. Assim, de posse de todos os *deadlines* recalculados a cada ciclo, o escalonador pondera-os com o tempo esperado de processamento e seleciona para execução o menor produto. O tempo de resposta efetivamente gasto é, então, utilizado para

<sup>3</sup>*Earliest deadline first*: política ótima de prioridade dinâmica não preemptiva para sistemas de tempo-real uniprocessados que seleciona primeiro as requisições com *deadline* mais próximo.

<sup>4</sup>*Shortest job first*: política ótima de prioridade dinâmica não preemptiva que seleciona primeiro as requisições com menor tempo esperado de processamento.

<sup>5</sup>Condição de uma requisição reiteradamente postergada pela chegada de outras requisições de maior prioridade.

recomputar a situação de todos os contratos e fornecer os *deadlines* para o ciclo de escalonamento seguinte. Resultados demonstram a propriedade da EBS em garantir um equilíbrio adequado na alocação de recursos proporcional às demandas impostas por cada classe de serviço (SLA).

## 4 Experimentos

O servidor modelado é do tipo monoprocessado, com uma única fila de espera e com atendimento de requisições do tipo não-preemptivo, onde foi implementada o algoritmo de escalonamento EBS. Para a avaliação do modelo e da política de escalonamento utilizou-se o pacote de simulação Simpack [6, 4], o qual oferece um conjunto próprio de ferramentas para simulação orientadas a eventos. O Simpack foi desenvolvido a partir do modelo SMPL [11] e possui uma biblioteca composta por uma ampla gama de métodos voltados à simulação de redes de filas orientada a eventos discretos[9].

Uma das medidas de interesse em avaliação de desempenho de sistemas computacionais é a taxa de utilização (ou ocupação) imposta ao sistema pela carga de trabalho considerada. Por meio do valor esperado dos tempos de serviço ( $C$ ) e intervalos de chegada ( $A$ ) das requisições dos usuários, é possível descrever a taxa de utilização ( $U$ ) do sistema por meio da Eq. 2.

$$U = \frac{C}{A} \quad (2)$$

A fim de relacionar a ocupação com o nível de QoS oferecido é interessante definir uma métrica de satisfação do usuário. A Eq. 3 indica uma das formas comumente empregadas para essa quantificação em sistemas *Soft-RT*: a taxa de atendimentos bem sucedidos. Segundo essa métrica, o índice de satisfação  $S_i$  do usuário  $i$ , é a razão entre o número de vezes em que a média de latência de sistema dessas requisições ficou abaixo do limiar contratado ( $N_i$ ) e a quantidade total de requisições submetidas por  $i$  ( $R_i$ ). Quanto mais próximo de  $R_i$  for o valor de  $N_i$ , maior será a satisfação proporcionada ao usuário  $i$ . Considera-se que um usuário está satisfeito com o serviço a ele oferecido quando obtiver uma alta porcentagem de requisições atendidas, em média, acima do limiar de qualidade contratado.

$$S_i = \frac{N_i}{R_i} \quad (3)$$

#### 4.1 Planejamento de testes

Como o experimento trata de avaliar um sistema de múltiplos usuários, cada qual podendo submeter uma carga distinta e associado a um SLA distinto, é esperado que o índice de satisfação da QoS recebida por cada usuário possa diferir dos demais. O objetivo do experimento é verificar como essa satisfação é influenciada pelos fatores que caracterizam a demanda e os recursos disponíveis. É interessante, portanto, definir algum parâmetro para comparação entre esses diferentes cenários. Para isso, o experimento considera alguns valores agregados que parametrizam diferentes situações:

- **Média da Satisfação**, como na Eq. 3);
- **Dispersão da Satisfação**: desvio padrão da satisfação dos usuários do sistema;
- **Média de Latência do Sistema**: tempo médio de residência de requisições no sistema.

A média da satisfação é um parâmetro que mensura o quão bem o conjunto de usuários é atendido. Uma vez que dependendo da carga gerada e do SLA estabelecido, diferentes usuários podem ter índices de satisfações bem distintos, computou-se também a dispersão da satisfação para ter-se uma visão da justeza do escalonamento. Foi calculada também a média de latência no sistema, a qual inclui os tempos em fila das requisições.

Para estudar o impacto da demanda no resultado foram analisados quatro fatores de carga no experimento:

- **Taxa de utilização do sistema**: fração de tempo em que o processador fica ocupado;
- **Número de contratos**: quantidade de contratos simultâneos no sistema;
- **Média dos contratos**: máximo tempo médio de resposta a uma requisição do usuário;
- **Dispersão dos contratos**: razão entre o desvio padrão e a média dos contratos.

A taxa de utilização é um valor agregado que representa o quanto o sistema está ocupado. O número de contratos (SLAs) permite avaliar se, para uma mesma ocupação, a existência de muitas ou poucas classes de serviço impactam no resultado. A média dos contratos diz o quão exigente é o conjunto de clientes em termos de *deadlines* e sua dispersão foi calculada para revelar se a disparidade entre os níveis de contratos (apertados e folgados) tem influência na satisfação geral dos usuários. Para realizar o estudo da influência dos fatores de entrada foi utilizado o método *K factorial design*, descrito por Jain [7].

## 5 Resultados

Para cada um dos experimentos realizados neste estudo, foram feitas **100** replicações. Em cada um desses testes são realizadas **100000** requisições no sistema, utilizando-se diferentes sementes<sup>6</sup> disponibilizadas pelo Simpack. Todos os intervalos de confiança foram calculados com grau de confiança de **95%**.

### 5.1 Análise da influência

A análise da influência dos fatores foi realizada com o objetivo de ter uma ideia geral de como esses influenciam, separadamente ou conjuntamente, nas variáveis de saída. Para cada um dos 4 fatores analisados foram ajustados empiricamente 2 níveis de modo a não sobrecarregar o sistema:

- A: taxa média de utilização do sistema: **60%** e **65%**;
- B: número de contratos: **50** e **250**;
- C: tempo médio dos contratos: **150** e **250**;
- D: dispersão dos contratos: **5%** e **25%**.

A Figura 1 é composta por 3 gráficos que representam a influência (eixo das ordenadas), em porcentagem, de cada um dos fatores de entrada (eixo das abscissas). Os Gráficos 1(a), 1(b) e 1(c) mostram as influências no tempo médio de resposta, na satisfação média dos usuários e na dispersão da satisfação, respectivamente. Os gráficos foram alinhados lado-a-lado para melhor comparação. São utilizadas letras de *A* até *D* para representar os fatores de entrada, sendo que os fatores com múltiplas letras representam a interação entre os fatores.

O Gráfico da Figura 1(a) indica que a taxa de utilização é o fator que tem a maior influência (fator *A*, com 95,98%) no tempo de resposta dentre os fatores analisados. Por outro lado, as Figuras 1(b) e 1(c) indicam uma menor influência da taxa de utilização na satisfação e na dispersão da satisfação dos usuários (5,10% e 2,90%, respectivamente). É notada uma dispersão da satisfação com a variação na taxa de utilização (a influência não é estatisticamente nula), porém o efeito é bem menor em comparação com o que acontece com tempo de resposta. O mesmo aplica-se à dispersão da satisfação.

Para verificar se os fatores influem diferentemente em sistemas pouco ou muito carregados, dois novos níveis de taxas de utilização foram fixados (80% e 85%). A Figura 2 mostra os resultados do método fatorial completo aplicado a esse ambiente.

---

<sup>6</sup>Replicações com fluxos de números aleatórios diferentes.

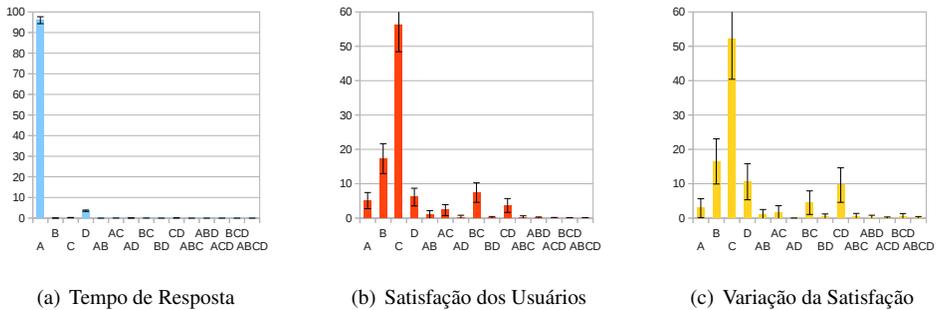


Figura 1. Gráfico da influência dos fatores.

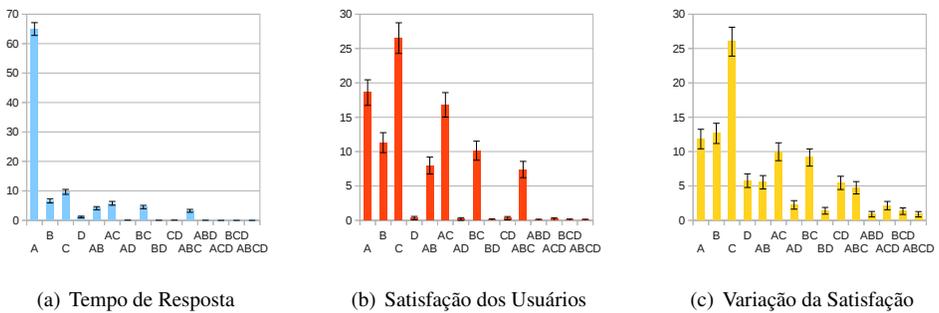


Figura 2. Gráfico da influência dos fatores em um ambiente com maior taxa de utilização

O gráfico da Figura 2(a) indica que a influência da taxa de utilização (fator *A*) no tempo de resposta é 64,94%, que é menor em relação ao que mostrava o Gráfico 1(a) (95,98%). Também podemos notar uma diferença nos outros fatores, indicando que os fatores influem diferentemente dependendo do estado do sistema. Os Gráficos 2(b) e 2(c) também mostram diferenças relação aos Gráficos 1(b) e 1(c). Na seção seguinte (5.2) é feita uma análise completa, fator a fator, de como essas mudanças afetam o sistema.

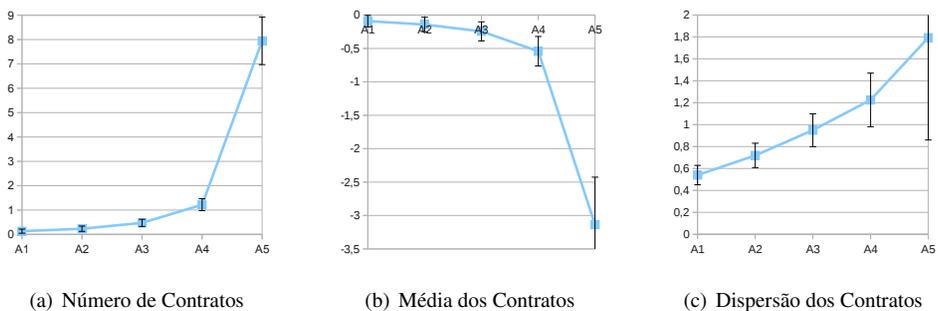
## 5.2 Análise da influência fator a fator

Essa análise tem por objetivo verificar como a influência de um fator reage com a mudança de outro fator. Por exemplo, como a influência do número de contratos muda com

o aumento da taxa de utilização. Com isso podemos verificar se a influência de um fator aumenta ou diminui conforme a taxa de utilização varia e, também, em que taxa ocorre tal mudança. O primeiro experimento consiste em mudar uniformemente a taxa de utilização e verificar como as influências dos outros fatores se comportam.

**5.2.1 Variação da taxa de utilização** Para fazer essa verificação foram feitos experimentos com taxa de utilização de **60%, 65%, 70%, 75%, 80% e 85%**. O fator foi variado uniformemente em passos de 5% e foram feitas 5 análises utilizando o método fatorial completo. Os níveis não mudam entre os experimentos (com exceção da taxa de utilização, que é o fator variante neste caso) e foram escolhidos de modo a representar um sistema variando de um ambiente pouco exigente para um com maior exigência (do experimento A1 para o A5).

A primeira variável de resposta analisada é o tempo médio de resposta. A Figura 3 mostra um gráfico para cada um dos 3 fatores de entrada analisados neste experimento (número de contratos, média dos contratos e dispersão dos contratos, respectivamente). O eixo das abscissas representa cada um dos 5 experimentos enquanto o eixo das ordenadas representa os valores dos  $q_i$ 's com seus respectivos intervalos de confiança.



**Figura 3.** Influência no tempo de resposta com a variação da taxa de utilização

É possível observar a evolução da influência dos fatores de entrada enquanto a taxa de utilização aumenta (de A1 para A5). Quanto mais afastado de 0, mais um fator influi no sistema, independentemente se é negativo ou positivo. Uma influência “negativa” atua como “diminuidor” da variável de resposta, da mesma forma de uma influência “positiva” atua como “aumentador” da variável de resposta. Nesse caso, um aumento no número de contratos (Figura 3(a)) ou na dispersão dos contratos (Figura 3(c)) irá afetar aumentando o tempo médio de resposta (influência positiva). Em contrapartida, o aumento da média dos contratos (Figura 3(b)) irá afetar diminuindo o tempo médio de resposta (influência negativa).

Foi observada uma situação similar nas Figuras 3(a) e 3(b), onde o aumento linear da taxa de utilização (de A1 para A5) é acompanhado com um aumento exponencial da influência desses fatores. Na Figura 3(c) há também um aumento, porém menor se comparado com os outros dois fatores, visualmente mais próximo a um aumento linear. Podemos verificar que o fator que foi mais influenciado é o número de contratos. Já a média dos contratos é o segundo fator que foi mais influenciado e por último a dispersão dos contratos.

Assim, enquanto o sistema está com baixa taxa de utilização há pouca influência do número de contratos e da média dos contratos (chegando a ser menor que a influência da dispersão dos contratos), porém à medida que aumenta a taxa de utilização, mais esses fatores irão afetar o sistema. Em outras palavras, o número de contratos e a média dos contratos pouco necessitam ser levados em consideração em sistemas com pouca taxa de utilização, ao contrário de sistemas com alta ocupação.

A segunda variável de resposta analisada é a satisfação dos usuários. A análise mostrou resultados bastante similares se comparado com a primeira variável de resposta analisada (tempo médio de resposta). Nesse experimento houve também um aumento exponencial com o aumento linear da taxa de utilização. A principal diferença (em relação ao experimento anterior) é a inversão nas influências, ou seja, os gráficos do número de contratos (4(a)) e da dispersão dos contratos (4(c)) mostram influências negativas e o gráfico da média dos contratos (4(b)) mostra uma influência positiva. Esse era um resultado esperado, pois um sistema que tem um aumento no tempo de resposta tende a diminuir a satisfação dos usuários. Assim como na análise anterior, o número de contratos foi o fator mais influenciado com a dispersão da taxa de utilização, seguido da média dos contratos e por último a dispersão dos contratos.

Também é notável que os fatores influenciem mais no tempo de resposta do que na satisfação dos usuários se compararmos a escala dos gráficos da Figura 3 com os da Figura 4. Isso quer dizer que o escalonador consegue manter a satisfação, mesmo com o tempo médio de resposta aumentando.

A terceira variável de resposta analisada é a dispersão da satisfação dos usuários. Nesse caso as influências têm a mesma sinalização dos gráficos mostrados na Figura 3. Isso quer dizer que os 3 fatores influem na dispersão da satisfação dos usuários da mesma forma que o tempo médio de resposta. O número de contratos foi o fator mais influenciado, porém diferente dos experimentos anteriores, a dispersão dos contratos teve uma influência maior do que a média dos contratos, como pode ser visto comparando os Gráficos 5(b) e 5(c). A observação faz sentido se considerado que a dificuldade em garantir justiça no atendimento aos diversos contratos está mais relacionada ao nível de exigência de cada cliente do que à demanda média por eles gerada. A média dos contratos afeta a satisfação como um todo, enquanto a dispersão dos contratos afeta o quão diferente os contratos são e consequentemente o quão diferente o sistema irá tratar os seus usuários.

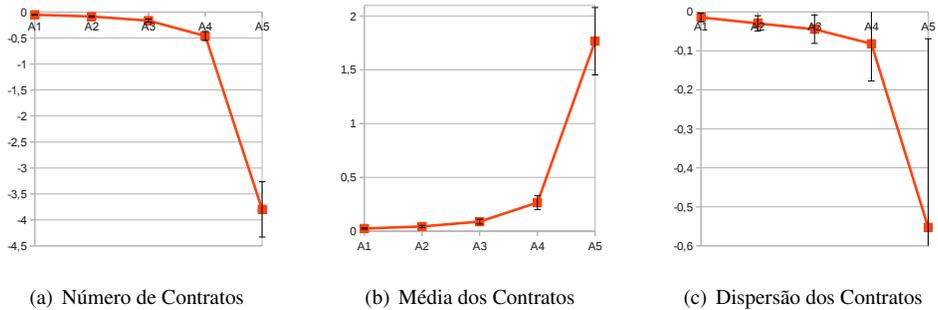


Figura 4. Influência na satisfação dos usuários com a variação da taxa de utilização

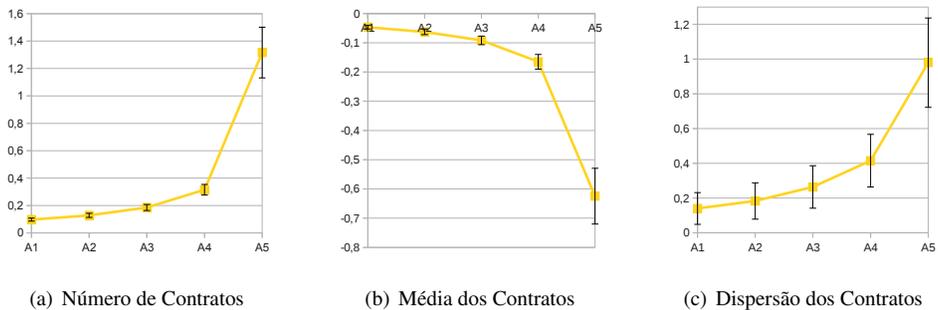


Figura 5. Influência na dispersão da satisfação dos usuários com a variação da taxa de utilização

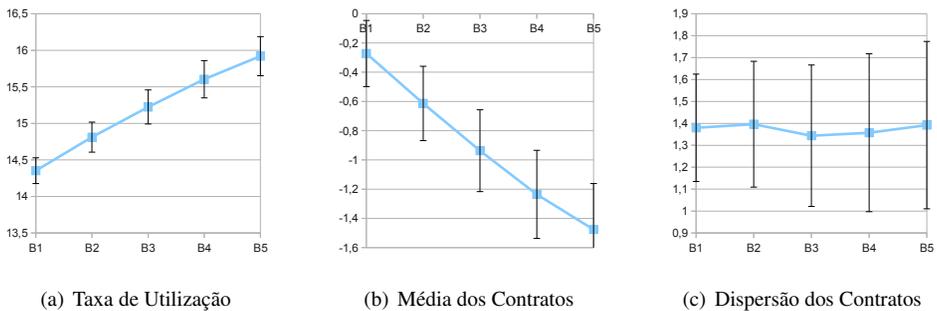
Portanto, o que pode ser observado nos três experimentos, é que a mudança na taxa de utilização muda a influência dos fatores diferentemente em grau e sinal. A influência varia de acordo com os gráficos apresentados nas Figuras 3, 4 e 5. A influência dos 3 fatores analisados é baixa quando o sistema está com pouca taxa de utilização, aumentando exponencialmente quando a taxa de utilização aumenta linearmente. Em relação ao tempo médio de resposta e à satisfação dos usuários, o número de contratos é o fator que foi mais influenciado com o aumento da taxa de utilização, seguido da média dos contratos e por último da dispersão dos contratos. Já em relação à dispersão da satisfação dos usuários, o número de contratos também é o fator mais influenciado, porém é seguido da dispersão dos contratos e por último a média dos contratos. Quanto maior a escala do gráfico, maior será a diferença entre os experimentos, portanto maior será a variação da influência do fator.

A seguir foi computada a interação entre todos os fatores e analisado como se comportam com a mudança no número de contratos, na média dos contratos e na dispersão dos contratos.

**5.2.2 Variação do número de contratos** Nessa seção, o fator variado é o número de contratos. Foram utilizados 6 níveis: **2, 50, 100, 150, 200 e 250**.

A primeira variável de resposta analisada é o tempo médio de resposta. Nos Gráficos 6(a) e 6(b) nota-se um aumento na influência com o aumento do número de contratos (do experimento *B1* para a *B5*). Diferente da variação exponencial visto no experimento anterior, nesse caso essa variação se aproxima bem de uma variação linear. Ou seja, com o aumento linear do número de contratos há um aumento linear na influência da taxa de utilização e da média dos contratos. Já o Gráfico 6(c) mostra que não houve variação na influência da dispersão dos contratos.

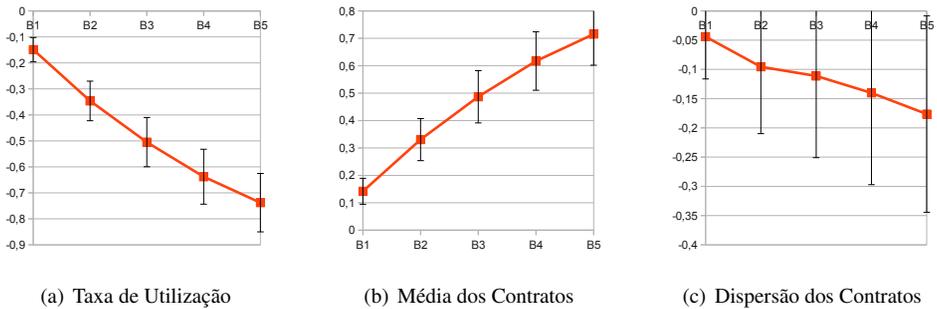
Pelos gráficos é verificado que taxa de utilização é o fator que mais foi influenciado, em seguida a dispersão dos contratos e por último a média dos contratos. Porém há uma tendência da influência da média dos contratos ficar maior que a da dispersão dos contratos à medida que o número de contratos aumenta. Assim a média dos contratos tem pouca influência em sistemas com baixo número de contratos e aumenta linearmente à medida que esse fator aumenta. Já a taxa de utilização (Figura 6(a)) tem uma influência alta, mesmo com baixo número de usuários.



**Figura 6.** Influência no tempo de resposta com a variação do número de contratos

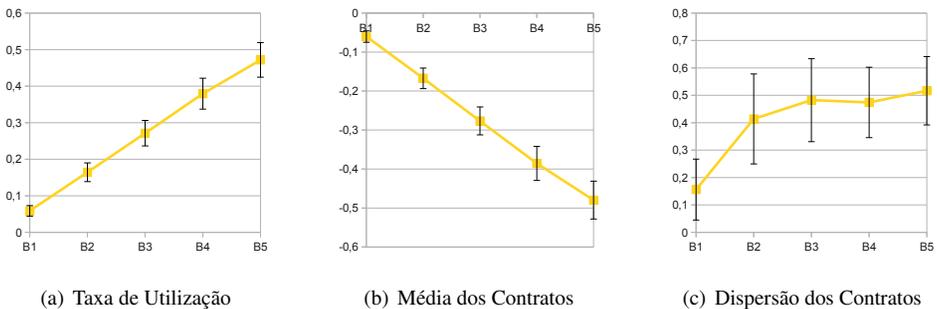
A segunda variável de resposta analisada é a satisfação dos usuários. Aqui também foi observado um aumento linear da influência da taxa de utilização e da média dos contratos e uma estagnação da dispersão dos contratos. Nesse caso, os Gráficos 7(a) e 7(b) se mostraram bem espelhados, com a taxa de utilização e a média dos contratos influenciando da mesma

forma, uma positiva e outra negativamente. O gráfico da Figura 7(c) mostra que o intervalo de confiança engloba o valor 0, o que significa que estatisticamente não há nenhuma influência desse fator na satisfação dos usuários. A taxa de utilização e a média dos contratos foram os fatores que mais influenciaram.



**Figura 7.** Influência na satisfação dos usuários com a variação do número de contratos

A terceira variável de resposta é a dispersão (variação) na satisfação dos usuários. Os gráficos são parecidos com os obtidos nos dois experimentos anteriores, onde os gráficos das Figuras 8(a) e 8(b) são espelhados e o da Figura 8(c) mostra uma estagnação a partir do experimento B1. Nesse caso, a dispersão dos contratos foi o fator que mais influenciou o sistema, porém esse cenário tende a mudar a partir de B5, ou seja, a partir de 250 usuários, com o aumento da influência da taxa de utilização e da média dos contratos.



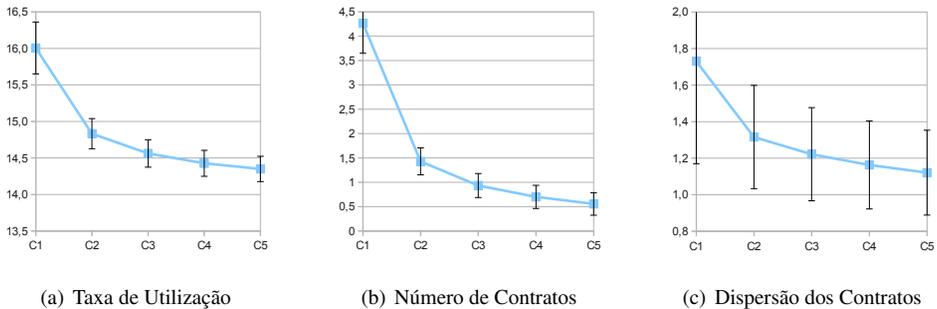
**Figura 8.** Influência na satisfação dos usuários com a variação do número de contratos

Assim a variação do número de usuários interferiu pouco na influência dos outros

fatores. As influências da taxa de utilização e da média dos contratos aumentam linearmente com o aumento do número de contratos. Já a influência da dispersão dos contratos não aumenta ou há uma leve tendência de aumento com o aumento no número de contratos.

**5.2.3 dispersão da média dos contratos** Nessa seção, o fator variado é a média dos contratos. Foram utilizados os seguintes níveis: **125, 150, 175, 200, 225 e 250**.

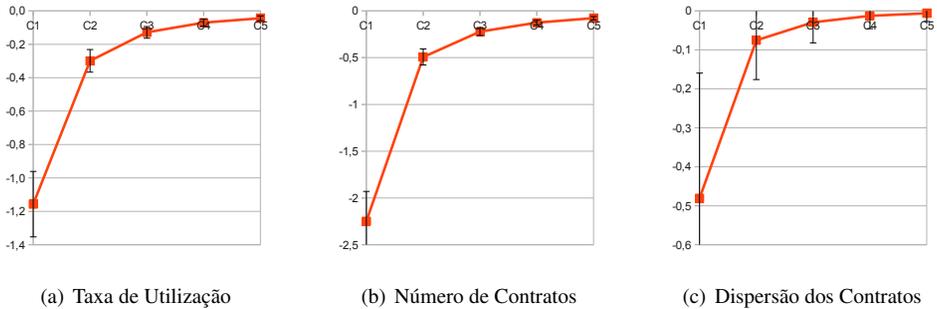
A primeira variável de resposta analisada é o tempo médio de resposta. Foi observado um comportamento inverso aos experimentos até então analisados. Nesse caso, o aumento da média dos contratos (de  $C1$  para  $C5$ ) diminui a influência dos outros fatores (as influências se aproximam de 0). Isso acontece, pois ao contrário dos outros fatores, um aumento na dispersão dos contratos faz com que o sistema fique menos carregado. É observada uma diminuição logarítmica da influência dos fatores com o aumento linear da dispersão dos contratos. O fator que foi mais influenciado é a taxa de utilização, seguido do número de contratos e por último da dispersão dos contratos. A Figura 9(c) mostra intervalos de confiança altos, dificultando conclusões estatísticas.



**Figura 9.** Influência no tempo de resposta com a dispersão da média dos contratos

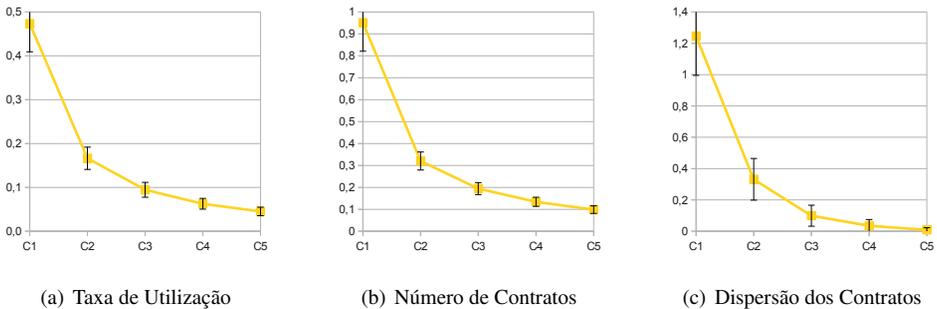
A segunda variável de resposta é a satisfação dos usuários. Como no experimento anterior, verificamos uma diminuição logarítmica da influência dos fatores. Nesse caso, o número de contratos foi o fator que mais foi influenciado, seguido da taxa de utilização e por último a dispersão dos contratos. O Gráfico 10(c) mostra que a partir de  $B2$  a dispersão dos contratos não tem mais influência na satisfação dos usuários do sistema. Ou seja, com contratos altos a sua dispersão não tem influência na satisfação.

A terceira variável de resposta é a dispersão da satisfação. Aqui também as influências dos fatores diminuíram logaritmicamente com o aumento dos contratos. A dispersão dos contratos (Figura 11(c)) é o fator que mais perde influência, inclusive ficando estatisticamente



**Figura 10.** Influência na satisfação dos usuários com a dispersão da média dos contratos

nula a partir de  $C4$ .

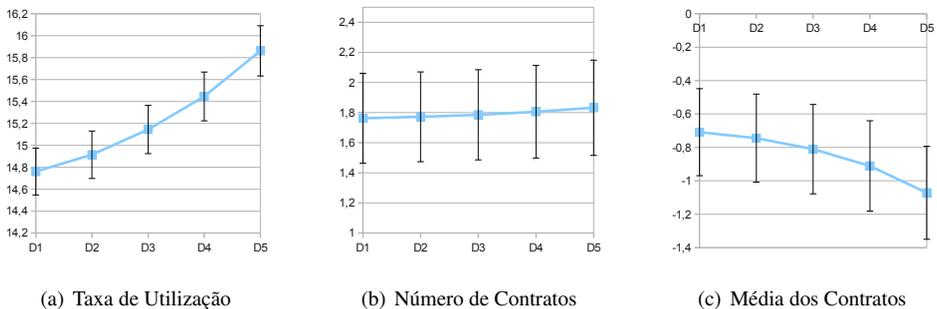


**Figura 11.** Influência na dispersão da satisfação dos usuários com a dispersão da média dos contratos

As influências dos fatores diminuem logaritmicamente com o aumento da média dos contratos. Em cada uma das variáveis de resposta, um fator de entrada foi o que teve mais variação na influência. No caso do tempo de resposta foi a taxa de utilização, na satisfação dos usuários foi o número de contratos e na variação da satisfação foi a dispersão dos contratos. Essa diminuição logarítmica indica que as influências tendem a estabilizar conforme a média dos contratos aumenta, ou seja, há um ponto aonde o aumento dos contratos não irá mais afetar a influência dos outros fatores.

**5.2.4 Variação da dispersão dos contratos** Por último, nessa seção o fator variado é a dispersão dos contratos. Foram utilizados os seguintes níveis: **5%, 10%, 15%, 20%, 25% e 30%**.

A primeira variável de resposta analisada é o tempo médio de resposta. Quanto à taxa de utilização (Figura 12(a)), há uma leve tendência de aumento na sua influência. Já quanto ao número de contratos (12(b)) e à média dos contratos (6(c)) não ocorre um aumento significativo de suas influências. A taxa de utilização foi o fator que mais influenciou o tempo médio de resposta.

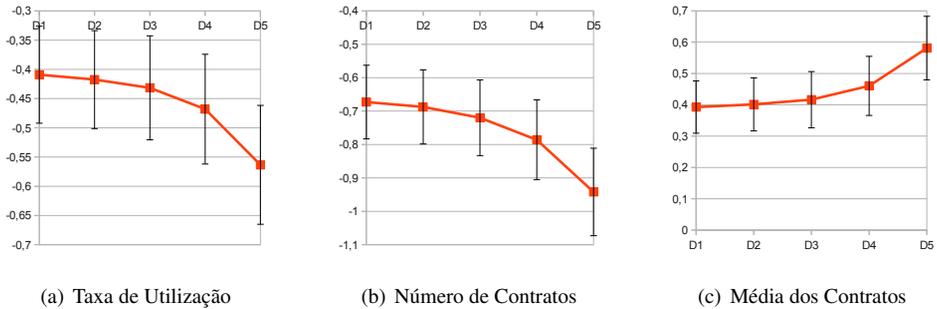


**Figura 12.** Influência no tempo de resposta com a variação da dispersão dos contratos

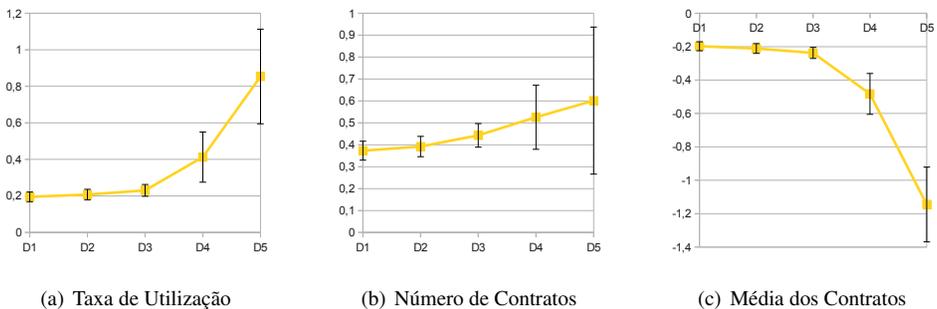
A segunda variável de resposta é a satisfação dos usuários. Como no experimento anterior, os gráficos mostraram um intervalo de confiança bem alto. Em todos os casos ocorreu uma leve tendência de aumento, porém bem pouco se comparado com a variação dos outros fatores.

A terceira variável de resposta é a dispersão da satisfação. Esse experimento também mostra que há pouca mudança na influência dos fatores, indicando que o aumento da dispersão dos usuários não aumenta, ou aumenta pouco a influência dos outros fatores. No caso da influência na dispersão da satisfação, a média dos contratos (Figura 14(c)) é uma exceção, pois apresenta intervalos de confiança menores, indicando que a influência da média dos contratos aumenta com o aumento da variação dos contratos.

As influências dos fatores variam pouco com o aumento da dispersão dos contratos se comparadas com a variação dos outros fatores. Isso indica que a dispersão dos contratos é o fator que menos interage com os outros fatores.



**Figura 13.** Influência na satisfação dos usuários com a variação da dispersão dos contratos



**Figura 14.** Influência na dispersão da satisfação com a variação da dispersão dos contratos

## 6 Conclusão

Este trabalho faz uma análise das influências dos fatores em um ambiente com variações controladas, com o objetivo de possibilitar a avaliação do comportamento de um sistema de tempo-real com algoritmo de escalonamento EBS, assim sendo possível a preparação do sistema para futuras modificações que venham a ocorrer no sistema ou no ambiente. Primeiramente foi constatado que a influência obtida de cada um dos fatores muda, conforme o ambiente (Seção 5.1). Com isso foi necessário uma variação constante do ambiente para que fosse possível uma melhor análise do comportamento desse sistema. Foram analisadas as influências de cada um dos 4 fatores escolhidos, uma a uma, e foram traçados gráficos com o comportamento obtido no decorrer de uma variação (Seção 5.2).

Três dos quatro fatores analisados atuam no sistema da mesma forma. A taxa de utilização, o número de contratos e dispersão dos contratos aumentam o tempo de resposta e a dispersão da satisfação (influência positiva) e diminuem a satisfação dos usuários (influência negativa). Já a média dos contratos atua de forma oposta, ou seja, diminuindo o tempo de resposta e a dispersão da satisfação e aumentando a satisfação dos usuários. Esse comportamento é o esperado, pois quanto maior a taxa de utilização, o número ou a dispersão dos contratos, teoricamente maior será a dificuldade do sistema, portanto maior será o tempo de resposta e menor será a satisfação dos usuários. Por outro lado, quanto maior a média dos contratos, menor a dificuldade do sistema, portanto menor será o tempo de resposta e maior será a satisfação dos usuários.

Todos os fatores aumentaram sua influência, em menor ou maior grau, com o aumento da taxa de utilização, número ou dispersão dos contratos. Porém esses aumentos diferem um dos outros, com alguns fatores tendo variação mais próxima do exponencial, outras mais próximas do linear e outros não tendo variação. Por outro lado todos os fatores diminuíram com o aumento da média dos contratos. Assim a influência do fator  $A$  mostrado no Gráfico 1(a) não diminuiu em relação ao Gráfico 1(b), e sim aumentou menos que os outros fatores.

A variação da taxa de utilização resultou em gráficos com aumentos exponenciais na maioria dos casos, mostrando sua grande influência nos outros fatores. Também mostrou que o número de contratos e a média dos contratos não tem muita importância em sistemas com baixa taxa de utilização, porém tem muita importância em sistemas com alta taxa de utilização. A variação do número de contratos por sua vez gerou gráficos com aumentos lineares na taxa de utilização e na média dos contratos e nenhum aumento substantivo na dispersão dos contratos. Isso quer dizer que não importa o número de contratos, a influência da sua dispersão é sempre a mesma. A variação da média dos contratos mostrou diminuições logarítmicas na influência dos fatores analisados à medida que a média dos contratos aumenta. Essas diminuições tendem a um valor, que mostra que a partir de um determinado ponto, o constante aumento da média dos contratos não irá mais beneficiar o sistema. E por último a variação da dispersão dos contratos teve poucas mudanças nas influências na maioria dos casos, indicando que esse fator não interage fortemente com os outros fatores.

## Referências

- [1] Lucas S. Casagrande, Rodrigo F. de Mello, Ricardo Bertagna, Jose A. Andrade Filho, and Francisco J. Monaco. Exigency-based real-time scheduling policy to provide absolute QoS for web services. *19th International Symposium on Computer Architecture and High Performance Computing - SBAC-PAD*, 0:255–262, Oct. 2007.
- [2] A. M. K. Cheng. *Real-Time Systems: Scheduling, Analysis, and Verification*. Wiley, 1 edition, 2002.

- [3] Alan Cobham. Priority assignment in waiting line problems. *Journal of the Operations Research Society of America*, 2(1):70–76, 1954.
- [4] R. M. Cubert and P. Fishwick. *Sim++*, version 1.0. Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, 1995.
- [5] J. C. Estrella, M. J. Santana, R. H. C. Santana, and F. J. Monaco. Real-time compression of soap messages in a soa environment. In *26th ACM International Conference on Design of Communication*, pages 163–168, Lisbon, Portugal, September 22-24 2008.
- [6] Paul A. Fishwick. Simpack: getting started with simulation programming in C and C++. In *Proceedings of the 24th conference on Winter simulation - WSC*, pages 154–162, New York, NY, USA, 1992. ACM.
- [7] R. Jain. *The art of computer systems performance analysis : techniques for experimental design, measurement, simulation, and modeling*. New York, NY, USA, Wiley, 1991.
- [8] Thomas E. Phipps Jr. and W. R. Van Voorhis. Machine repair as a priority waiting-line problem. *Operations Research*, 4(1):pp. 76–86, 1956.
- [9] K. H. Kumar and S. Majhi. Queuing theory based open loop control of web server. In *Proceedings of the 2004 American Control Conference*, 3:2314–2315, 2004.
- [10] Jane W. S. W. Liu. *Real-Time Systems*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition, 2000.
- [11] M. H. MacDougall. *Simulating computer systems: techniques and tools*. MIT Press Series in Computer Systems, Cambridge, MA, USA, 2 edition, 1989.
- [12] F. J. Monaco, Nery M., and M. L. Peixoto. An orthogonal multi-resource real-time scheduling architecture for responsiveness qos requirements in soa environments. In *24th Annual ACM Symposium on Applied Computing - ACM SAC*, pages 1990–1995, Honolulu, USA, March 8-12 2009.
- [13] F. J. Monaco, M. Nery, and M. M. L. Peixoto. An orthogonal real-time scheduling architecture for responsiveness QoS requirements in SOA environments. In *Proceedings of the 2009 ACM symposium on Applied computing*, pages 1–6, New York, NY, USA, 2009. ACM. to be published.
- [14] Francisco José Monaco and Pedro Northon Nobile. Feedback-based adaptive resource control in qos-aware soa systems with soft real-time requirements. *Quality of Service in Heterogeneous Networks*, 22(1):799–810, 2009.
- [15] N. Nisanke. *Real Time Systems*. Prentice Hall series in computer science. Prentice Hall PTR, 1997.

- [16] Pedro Northon Nobile. *Projeto de um broker de gerenciamento adaptativo de recursos em computação em nuvem baseado em técnicas de controle realimentado*. PhD thesis, Universidade de São Paulo — USP, 2013.
- [17] M. L. M. Peixoto, R. F. Tott, and F. J. Monaco. Política de escalonamento de tempo-real para garantia de qos absoluta em cluster de servidores web heterogêneos. In *WebMedia '07: Proceedings of the XIII Brazilian Symposium on Multimedia and the Web*, Gramado, RS, Brazil, 2007. ACM Digital Library.
- [18] M. M. L. Peixoto and Monaco. F. J. Arquitetura de escalonamentos ortogonal de tempo-real para garantias de qos em servidores web. In *Workshop em Desempenho de Sistemas Computacionais e de Comunicação - Wperformance, junto ao XXVIII Congresso da Sociedade Brasileira de Computação - SBC*, pages 18–37, Belém do Pará, Brasil, Jul. 2008.
- [19] P. T. M. Saito, P. N. Nobile, and F. J. Monaco. Escalonamento realimentado para diferenciação de serviços e garantia de desempenho em ambientes soa com requisitos soft-rt. In *X Simpósio de Sistemas Computacionais — WSCAD-SSC*, São Paulo, SP, Brazil, 2009. Sociedade Brasileira de Computação — SP.