

Estimação do Tempo Médio de Resposta do Servidor em Plataformas de Ensino a Distância do Tipo Cliente-Servidor Utilizando Modelo Baseado em Teoria de Filas

Victor Hugo Teles Costa, Jonatas Abreu Fernandes, Bruno Henrique Pereira Gonçalves, Flávio Henrique Teles Vieira, Leonardo Antônio Alves

Escola de Engenharia Elétrica, Mecânica e de Computação – Universidade Federal de Goiás (UFG)

Av. Universitária, n. 1488 - Quadra 86 - Bloco A - 3º piso 74605-010 - Setor Leste
Universitário - Goiânia – GO – Brazil

victor@vcosta.com.br, jonatasfdv@gmail.com, brunohpg@gmail.com
flavio@eee.ufg.br, leoaalvs@gmail.com.

***Abstract.** In this paper, we present an analytical model to predict the performance of Client-Server based e-Learning Platforms. Our approach is based on the queueing theory and can adequately describe some performance parameters of real e-Learning Platforms such as server response time. We evaluate the modeling capability of the proposed approach by comparing the results provided through the model to the results obtained by other model presented in the literature. However, our approach also takes account of the data loss probability in the system. Most importantly, in this study, we develop a tool for designing and scaling Client-Server based e-Learning Platforms based in client-server architecture.*

***Resumo.** Neste trabalho, apresentamos um modelo para prever o desempenho de plataforma de Ensino a Distância baseadas em cliente-servidor. A nossa abordagem é baseada em Teoria de Filas, focando principalmente na estimação do tempo de resposta do servidor. Avaliamos a eficiência do modelo proposto para a plataforma de Ensino a Distância, comparando os resultados fornecidos deste com os de outro modelo apresentado na literatura. Entretanto, a principal diferença é que nosso modelo leva em consideração a probabilidade de perda de dados. Assim, neste trabalho, desenvolvemos uma ferramenta para a concepção e dimensionamento de plataformas de Ensino a Distância baseadas na arquitetura cliente-servidor.*

1. Introdução

Desde o final do século XX, a TI (Tecnologia da Informação) tem oferecido diferentes paradigmas para o desenvolvimento da Tecnologia Educacional. Hoje, na era da informação, apesar de suas dificuldades e restrições, é importante que as pessoas tenham acesso à educação formal. Neste contexto, o Ensino a Distância (EaD) está emergindo como um novo paradigma, permitindo que a informação seja compartilhada entre todos os públicos, removendo as barreiras físicas e sociais.

Neste novo cenário, os educadores podem criar laboratórios e salas de aula virtuais onde os alunos contam com materiais de ensino de alta qualidade através de ambientes multimídia, incluindo ilustrações tridimensionais, armazenamento de arquivos, videoconferências e reuniões online.

As empresas e instituições que pretendem explorar o Ensino a Distância precisam construir uma plataforma que deve ser suportada por uma infraestrutura de telecomunicações composta principalmente por servidores de rede, permitindo o uso de todo tipo de mídia baseada em computador.

O Ensino a Distância em todo o Brasil e no exterior é basicamente realizado através de sites, um conjunto de páginas, estáticas ou dinâmicas, organizadas numa estrutura conhecida como WWW (*World Wide Web*). Portanto, a *web* tem um papel crucial no mundo da informação, sustentando as chamadas Plataformas de EaD.

O desempenho da Plataforma de EaD do tipo Cliente-Servidor depende diretamente do tempo de resposta do servidor nas conexões, isto é, o atraso para responder à requisições de página [Drakopoulos e Merges 1992] [King 2003]. Por esta razão, focamos no tempo de resposta e nos parâmetros de rede envolvidos na sua estimação. Para calcular o desempenho da Plataforma de EaD, examinamos um modelo Cliente-Servidor baseado na Teoria de Filas [Gross, Shortle, Thompson, e Harris 2009] [Kleinrock 1976][Pujolle e Gelenbe 1998]. Verificamos que o modelo de fila proposto pode ser utilizado para auxiliar no projeto de Plataformas de EaD, assim como para verificar sua escalabilidade.

O restante deste trabalho é organizado da seguinte forma: na seção II, apresentamos uma abordagem para modelagem de Plataformas de EaD baseada na Teoria de Filas, na seção III, avaliamos o desempenho de Plataformas EaD em termos do tempo médio de resposta do servidor de acordo com a equação proposta, finalmente, na seção IV, relatamos nossas conclusões.

2. Tempo Médio de Resposta do Servidor em Plataformas de EaD

Nesta seção, propomos um modelo para estimar o tempo médio de resposta do servidor em plataformas de EaD baseadas na arquitetura cliente-servidor.

No modelo proposto, consideramos que há três fases no processo de atendimento das requisições de usuários na plataforma de EaD: a chegada de uma solicitação, o processamento no servidor e a chegada do processamento no cliente.

A Figura 1 mostra a rede de filas proposta utilizada para a obtenção das equações para o cálculo do tempo de resposta do servidor, onde A é o número de requisições de chegada ao servidor, S é o tamanho da banda (taxa de transmissão) do servidor e C , o tamanho da banda (taxa de transmissão) do cliente e p é a probabilidade ou taxa de perda de pacotes.

Denominamos de S_r , S_s , S_c , as capacidades de atendimento das filas referentes ao servidor Web, à conexão Internet e ao *browser* do usuário, respectivamente. Nesta rede de filas, as requisições de entrada A se juntam na entrada no servidor de entrada de requisições S_r às requisições A' que deverão ser retransmitidas.

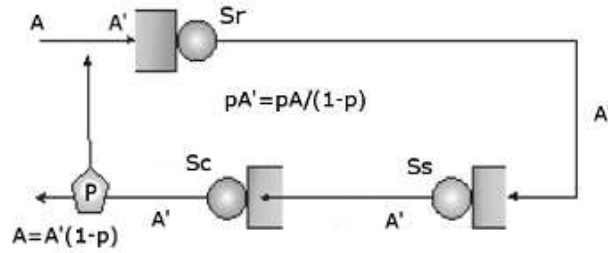


Figura 1. Modelo para Plataforma de EaD.

Essas requisições entram no servidor S_s para serem processadas, e são entregues ao cliente S_c , que poderá solicitar retransmissões ao servidor Web S_r . Denotaremos por C a banda, ou taxa de transmissão do usuário S_s (conexão Internet) e por S a banda do servidor, ou taxa com a qual o servidor responde à Internet. É fácil demonstrar que a capacidade S_r do servidor Web é dada por:

$$S_r = \frac{1}{\left(y + \frac{B}{R}\right)} \quad (1)$$

onde F é o tamanho do arquivo recebido, B é o tamanho do *buffer*, y é o tempo de espera no servidor, R é a taxa dinâmica de atendimento, ou taxa em que o servidor processa o *buffer*.

Podemos enunciar a seguinte proposição referente ao tempo médio de resposta para plataformas de EaD.

Proposição 1: Assumindo que o tráfego em redes de EaD seja Poissoniano, o Teorema de Jackson é aplicável ao modelo de filas apresentado na Figura 1. Assim, o tempo de resposta médio de um servidor EaD utilizando arquitetura cliente-servidor é dado por:

$$E\{T\} = \frac{1}{\frac{1}{y + \frac{B}{R}} - \frac{A}{(1-p)}} + \frac{1}{\frac{C}{F} - \frac{A}{(1-p)}} + \frac{1}{\frac{S}{F} - \frac{A}{(1-p)}} \quad (9)$$

De acordo com Alves, Vieira e Araújo (2009), o modelo de filas que descreve uma Plataforma EaD pode ser considerada uma rede de Jackson e, portanto, o tempo de espera poderia ser calculado através de:

$$E\{T\} = \frac{F}{C} + \frac{I}{1 - AI} + \frac{F}{S - AF} + \frac{F(B + Ry)}{BR - AF(B + Ry)} \quad (10)$$

onde I corresponde ao tempo de inicialização.

Para validar a equação proposta de tempo médio de resposta do servidor em plataformas de EaD, inicialmente implementamos um simulação de fila simples com um servidor com *buffer*. A simulação foi realizada no *software* Rockwell Arena 12, versão estudantil, considerando um *buffer* de 50 Mbits de espaço, uma banda de servidor de 30 Mbps e taxa de requisições Poissonianas.

Na próxima seção, apresentaremos uma comparação dos resultados obtidos com as Equações 9 e 10.

3. Estudos de Otimização de Desempenho de Plataforma de EaD

Analizamos o desempenho da plataforma de EaD em termos do tempo de resposta do servidor, variando parâmetros como a banda do servidor S e o tamanho do arquivo F . Para tal, consideramos como parâmetros para o modelo, valores reais obtidos em uma plataforma real de Ensino a Distância mantida pela empresa NT educação no SENAI-GO.

Calculando os tempos de resposta utilizando a equação proposta temos que, após certo valor de banda (capacidade do servidor), o tempo de resposta do servidor diminui insignificativamente. O que significa que a contratação de mais banda não corresponde necessariamente a uma diminuição no tempo de resposta do servidor. A variação no tempo de resposta foi de menos de 5%. Isso implica que uma plataforma de EaD não depende estritamente da quantidade de banda, mas também da capacidade de atendimento às requisições. A Figura 2 corrobora essa afirmação, onde fizemos $p=0$ e fixamos os valores de $B=2000$ bytes, $R=25547$ bytes/s, $C=38400$ bytes/s (ou 300 kbits/s), $A=3,715$ requisições/s, $F=1500$ bytes, $\gamma=1$ ms e $I=0$ s e variamos a capacidade S do servidor. Note que os valores considerados para B , R , C , A , γ e I são valores que foram encontrados em uma plataforma real de Ensino a Distância mantida pela empresa NT educação no SENAI-GO.

Observamos que para os parâmetros de rede considerados, a variação da capacidade do servidor S a partir de 185855 bytes/s não produz variação significativa no tempo de resposta (variação menor do que 5%). Pode-se então, afirmar que o tempo de resposta independe da capacidade para $S > 185855$ bytes/s, no cenário considerado.

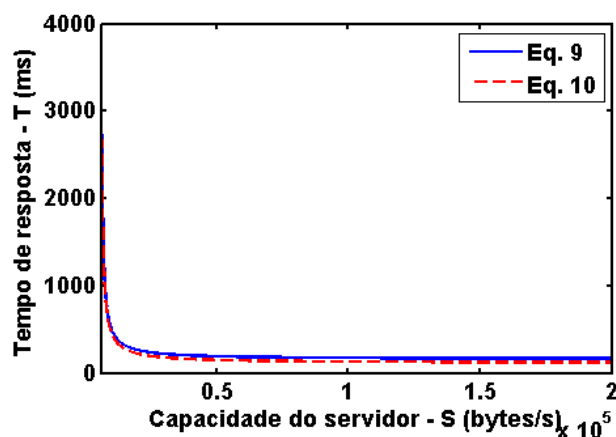


Figura 2. Tempo Médio de Resposta versus Capacidade do Servidor (S).

Quanto maior o tamanho do arquivo (F), maior o tempo de resposta, conforme apresenta a Figura 3 (para o caso de $p=0$). A Figura 3 foi gerada considerando os valores dos parâmetros do sistema fornecidos anteriormente, fixando S em 393216 bytes/s (ou 3 Mbits/s) e variando F .

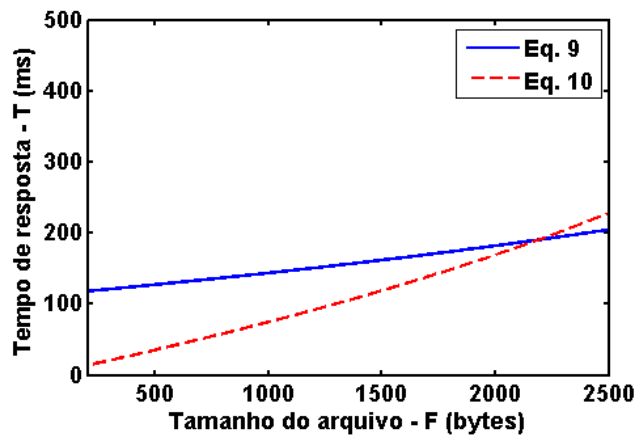


Figura 3. Tempo de Resposta versus Tamanho do Arquivo (F).

A otimização da Plataforma de EaD, com respeito ao tempo de resposta, deve levar em conta a relação entre a variável $E\{T\}$ e as variáveis C , B , S e R . Podemos ajustar os parâmetros C , S e R a fim de obtermos uma melhoria do desempenho do sistema, considerando a disponibilidade de cada recurso. Por exemplo, para obtermos uma redução mais significativa do tempo de resposta no sistema caracterizado pela Figura 3, com $S=185855$ bytes/s, temos de aumentar os parâmetros S ou R dentro dos limites disponíveis.

Ainda considerando $B=2000$ bytes, $R=25547$ bytes/s, $S=393216$ bytes/s, $A=3,715$ requisições/s, $F=1500$ bytes, $\gamma=1$ ms, $I=0$ e $p=0$, na Figura 4, observamos a relação que se desenvolve entre o tempo de resposta e a capacidade C do cliente. Claramente, quanto maior a capacidade C , menor será o tempo de resposta do servidor. No entanto, há um momento em que um aumento na banda pode não significar uma diminuição considerável no tempo de resposta, dependendo da disponibilidade de recursos. Mais especificamente, nas circunstâncias da Figura 4, podemos dizer que para $C > 250709$ bytes/s a variação no tempo de resposta é de menos de 5%, para $S=393216$ bytes/s.

Para os parâmetros considerados anteriormente ($p=0$), e fixando $C=38400$ bytes/s, o comportamento do tempo de resposta em relação à taxa R de atendimento pode ser observado na Figura 5.

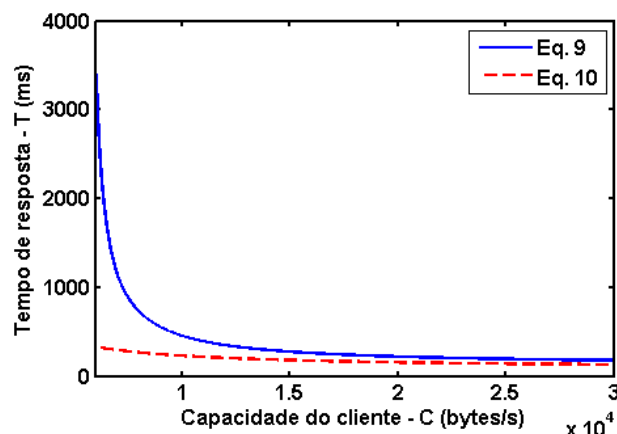


Figura 4. Tempo de Resposta versus Capacidade do Cliente (C).

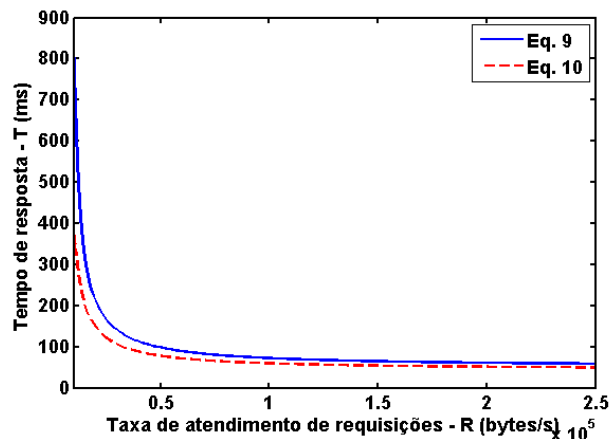


Figura 5. Tempo de Resposta versus Taxa de Atendimento de Requisições (R).

Como na relação entre tempo de resposta e capacidade do cliente, notamos uma diminuição no tempo de resposta com o aumento da capacidade de processamento de requisições. No caso da Figura 5, observamos que, para $R > 764556$ bytes/s, o tempo de resposta não tem diminuição significativa para $p=0$, a variação é inferior a 5%.

Atribuindo alguns valores para p , para que se possa analisar a influência das perdas no sistema, tem-se os resultados de tempo médio de resposta apresentados na Tabela 1. O valor de p corresponde à porcentagem de pacotes que não chegam ao destino devido a erro no pacote ou falha no sistema, sendo que devem ser retransmitidos pelos mecanismos do TCP-IP. Note que nos arquivos de log da plataforma de Ensino a Distância mantida pela empresa NT educação no SENAI-GO encontra-se perdas próximas a 0,1%.

Analisando os resultados de tempo médio de resposta para diferentes valores de p , constatamos que a variação do tempo de resposta $E\{T\}$ em relação a p diferente de zero é pequena. Para $p=0$, notamos uma relativa aproximação do tempo de resposta dado pela equação proposta com o tempo de resposta da Equação 10, como mostra a Tabela 1.

Tabela 1. Tempo de Resposta (em milissegundos) considerando diferentes valores para p

A	E{T} (eq. 9)			E{T} (eq.10)
	$p=0$	$p=0,001$	$p=0,01$	
2,97222	151,7845	151,8224	152,1677	115,1528
3,34375	156,6858	156,7316	157,1492	117,1502
3,71528	161,9595	162,0143	162,5146	119,2606
4,08681	167,6521	167,7174	168,3128	121,4939
4,45833	173,8189	173,8961	174,6016	123,8612
4,82986	180,5252	180,6163	181,4498	126,3748
5,20139	187,8493	187,9568	188,9401	129,0490
5,57292	195,8859	196,0126	197,1726	131,8995
5,94444	204,7505	204,9000	206,2701	134,9444
6,31597	214,5856	214,7623	216,3844	138,2044
6,68750	225,5690	225,7787	227,7060	141,7032

A Figura 6 revela uma influência significativa da perda p no tempo de resposta ao variar a capacidade do cliente (C). Observamos um incremento claro do tempo de resposta com o aumento no valor de p . Como era de se esperar, quanto menor o valor de p , a curva de tempo de resposta se aproxima mais da curva para $p=0$. Para $p=0,1\%$, como observado na Tabela 1, os valores são praticamente os mesmos para $p=0$.

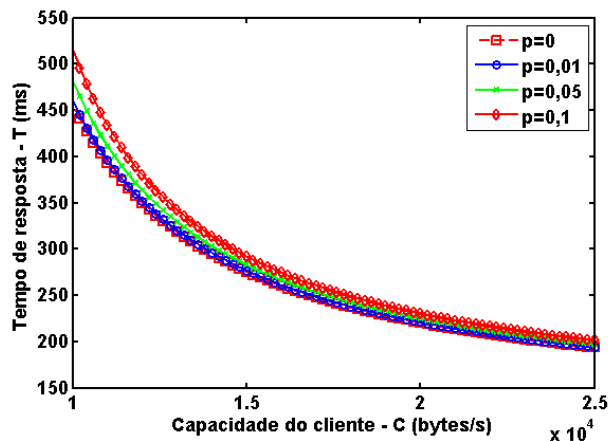


Figura 6. Tempo de Resposta versus Capacidade do Cliente (C).

Na Figura 7, observamos que ao alterar o tamanho do arquivo transmitido, há uma diferenciação considerável dos valores de tempo de resposta para diferentes valores de p quando o arquivo apresenta tamanhos superiores a 4000 bytes. Ou seja, para valores de tamanho de arquivo abaixo de 4000 bytes, o impacto da taxa de perda de dados no tempo de resposta é menor.

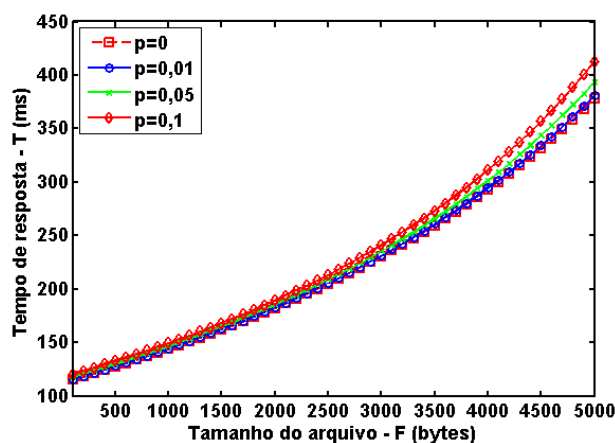


Figura 7. Tempo de Resposta versus Tamanho do Arquivo (F).

4. Conclusões

Neste trabalho, apresentamos um modelo de Plataforma de EaD do tipo Cliente-Servidor baseado em Teoria de Filas. Comparamos o Tempo de Resposta do servidor dado pelo nosso modelo ao obtido com o trabalho desenvolvido em Alves, Vieira e Araújo (2009), verificando a consistência dos valores estimados com o modelo desenvolvido.

Na equação proposta, levamos em conta a influência da taxa de perda de pacotes. Através disso, pudemos observar como a perda pode se relacionar ao aumento do tempo de resposta de uma Plataforma EaD. Também avaliamos a relação do tempo de resposta com vários parâmetros do sistema, mostrando, inclusive, como os parâmetros do sistema poderiam ser escolhidos para otimização de seu desempenho em termos de tempo médio de resposta.

Observe que este tipo de Plataforma de EaD permite que a largura de banda C da rede seja aumentada. Este parâmetro, por exemplo, pode ser usado para maximizar o desempenho da Plataforma de EaD. Ou seja, a partir de nosso modelo, podemos prever o comportamento da Plataforma quando acessada por certa quantidade de usuários. Portanto, a abordagem proposta pode ser vista como uma poderosa ferramenta para projeto e implantação de Plataformas de EaD do tipo Cliente-Servidor em todo o mundo.

5. Referências

- Alves, L. A., Vieira, F. H. T. e Araújo, S. G. An Analytical Performance Model for Client-Server based e-Learning Platforms. ICELW, 10 a 12 de Junho, Nova Iorque, EUA, 2009.
- Drakopoulos, E. and M. J. Merges, "Performance Analysis of Client-Server Storage Systems", IEEE Transactions on Computers, Vol. 41, No. 11, November 1992.
- Gross, D., Shortle, J. F., Thompson, J. M. e Harris, C. M. Fundamentals of Queueing Theory. 4a Edição, John Wiley & Sons, Nova Iorque, 2009.
- King, P. J. B., Computer and Communication Systems Performance Modelling, Prentice Hall International, UK, 2003.
- Kleinrock, L. Queueing Systems, Volume I: Theory. John Wiley & Sons, Nova Iorque, 1976.
- Pujolle, G. e E. Gelenbe. Introduction to Queueing Networks. 2a Edição, John Wiley & Sons, Nova Iorque, 1998.