

UMA APLICAÇÃO METODOLÓGICA PARA A CARACTERIZAÇÃO DE TRÁFEGO E AVALIAÇÃO DE DESEMPENHO DA REDE WIRELESS DE UMA INSTITUIÇÃO DE ENSINO SUPERIOR¹

Mariana Vazquez Godoy Miano²

RESUMO

O presente artigo apresenta um estudo de caso da aplicação de uma metodologia de coleta e análise de dados para o desenvolvimento de um modelo de simulação, com foco na análise de desempenho via Teoria de Filas, apresentada em [14]. O objetivo desta pesquisa é mostrar a aplicação da metodologia em outra Instituição de Ensino Superior (IES) - conforme sugerido em Vazquez [14], com aspectos topológicos não tão distintos, mas com infraestrutura e capacidades dos enlaces bem diversas. A IES do estudo também possui os enlaces bastante congestionados, principalmente àqueles dedicados à rede Wireless, ocorrendo assim perdas das requisições e conseqüentemente, baixa qualidade de serviço. Esta aplicação mostrou que, em termos de modelos analíticos de filas, o modelo M/M/1/k se mostrou muito mais adequado do que o G/G/1, embora a distribuição de serviço não seja exponencial, significando que a capacidade da fila (tamanho do *buffer*) é um indicador importante para a construção do modelo. A rede analisada tem um *buffer* muito pequeno e, aproximadamente 60% das requisições possuem protocolo *Transmission Control Protocol* (TCP).

Palavras-chave: análise de desempenho, simulação, teoria de filas, redes de Internet, controle de congestionamento.

ABSTRACT

This paper presents a case study of the application of a methodology for collecting and analyzing data to develop a simulation model, with focus on performance analysis via Queuing Theory, presented in [14]. The objective of this research is to show the application of the methodology in other IES (as suggested in Vazquez [14]), with topological aspects not as distinct, but with a more diversified infrastructure and capability. This Higher Education Institution also has rather congested links, particularly those dedicated to the Wireless network, thus causing loss of requests and poor quality of service. This application demonstrated that, in terms of analytical queuing models, the M/M/1/k model was more suitable than G/G/1, even though the service distribution is not exponential, meaning that queuing capability (buffer size) is an important indicator for the construction of the model. The analyzed network has a very small buffer, and approximately 60% of requests have Transmission Control Protocol (TCP) protocol.

Keywords: performance evaluation, simulation, queuing theory, Internet network, congestion control.

1 INTRODUÇÃO

A utilização da *Internet* como uma rede multisserviços tem influenciado a proposição de mecanismos de Qualidade de Serviço (QoS), uma vez que a *Internet* foi concebida para prover serviços que inicialmente não exigiam garantia de banda, limites de atraso e variação do atraso. Entretanto, o problema de congestionamento não tem sido resolvido de forma satisfatória quando a demanda é superior à transmissão, e esse tem se tornado um dos principais problemas para se alcançar a desejada qualidade. Essa condição de congestionamento é rara de ser descrita empiricamente na literatura, com medidas quantitativas, como as apresentadas neste estudo de caso.

Lee [9] apresenta uma visão não convencional baseada em medições quantitativas, com a degradação de um enlace operando a 100% de sua capacidade (em horários de pico). O trabalho de Vazquez [14] propõe uma metodologia que utiliza teoria de filas para o desenvolvimento de um modelo de simulação.

¹ Artigo recebido em 26/06/2014

² Departamento de Análise e Desenvolvimento de Sistemas – Centro Estadual de Educação Tecnológica Paula Souza (CEETEPS) – Faculdade de Tecnologia de Americana ; Contato: vazquez_prof@gmail.com

R.Tec.FatecAM	Americana	v.2	n.2	p.79-85	set.2014 / mar. 2015
---------------	-----------	-----	-----	---------	----------------------

Assim, o objetivo deste estudo de caso é caracterizar empiricamente o problema de congestionamento da IES estudada, capturando rastros de pacotes em *logs* (processo de registro de eventos) e estudando atrasos e perdas. Na sequência, utilizando um modelo experimental de simulação, determina-se a melhor configuração [10]. A experimentação também desempenha papel fundamental na exploração de novos ambientes e como os protocolos da *Internet* devem operar neles [2].

2. METODOLOGIA DE COLETA E PROCESSAMENTO DE DADOS

Para este estudo empírico do congestionamento da rede de comunicação acadêmica houve o acesso aos dados da Faculdade de Tecnologia de Americana, unidade do Centro Estadual de Educação Tecnológica Paula Souza (CEETEPS), comunidade acadêmica de aproximadamente 2400 usuários, que estava com seu enlace de acesso à *Internet* saturado a mais de 90% de utilização em certos períodos de pico do dia. Inicialmente, é descrito o ambiente utilizado pela IES. Do ponto de vista topológico, a rede de comunicação está fisicamente organizada como uma estrela estendida.

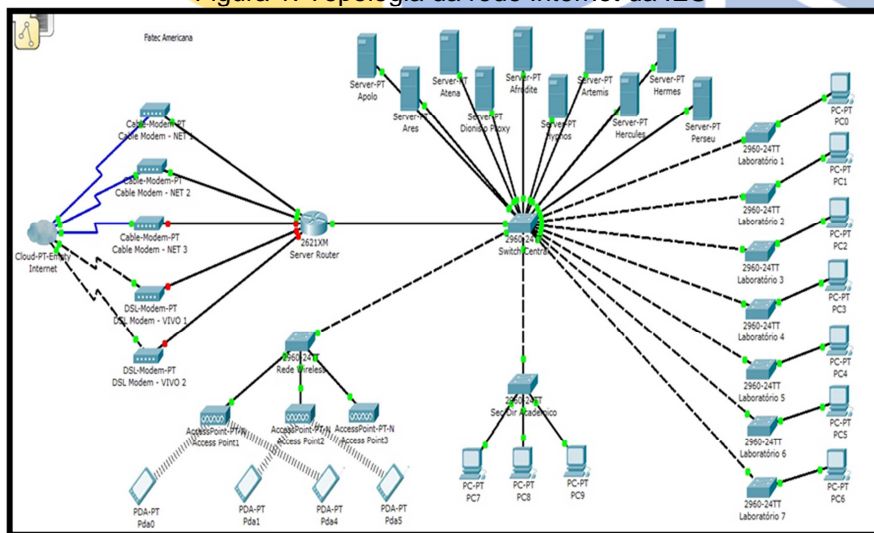
O nó principal é uma máquina com sistema operacional *Linux* que exerce a função de roteador e de balanceador de carga entre os 5 enlaces existentes. A partir desta máquina, está conectado um *switch* sem gerenciamento, que está conectado aos demais *switches* que dão acesso aos laboratórios, à secretaria e ao administrativo; nele também estão conectados todos os servidores presentes na rede.

Após o estudo do mapeamento da rede, verificou-se que esse servidor com função de roteamento e balanceamento de carga era o gargalo da rede de *Internet*, uma vez que os cinco enlaces existentes chegam a ele e depois se transformam em apenas um, que segue para o *switch* central, que distribui as requisições. Dos 5 enlaces existentes, apenas 4 são continuamente operantes (um deles atua como reserva) e juntos, têm uma capacidade nominal de aproximadamente 20 Mbps. Dentre os 5 enlaces, 3 deles possuem características de acesso residencial provido pela operadora *NET* e os outros 2 são providos pela Cia. de Processamento de Dados do Estado de São Paulo (Prodesp).

Com o monitoramento da rede utilizando o *software Zabbix* [13], observou-se que a rede da IES tem como característica principal um consumo alto de dados para acesso à *Internet* proveniente de equipamentos móveis (*LapTop, Tablet e Smartphone*). A rede de cobertura *Wireless* no campus é ampla e por esse motivo o maior fluxo de dados é proveniente deste tipo de acesso. Assim, o foco deste trabalho será a rede *Wireless*.

A figura 1 (topologia da rede de Internet da IES) apresenta a topologia encontrada, onde se pode notar o formato estrela, com destaque para a saída da Internet representada como uma nuvem.

Figura 1: Topologia da rede Internet da IES



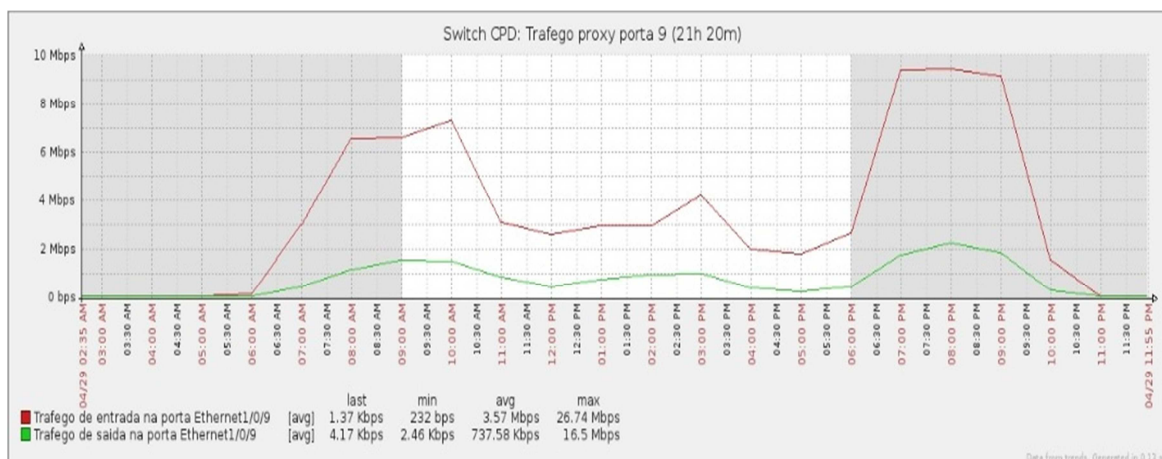
O fluxo de dados do bloco administrativo é separado dos demais e direcionado ao enlace da Prodesp, não sofrendo a interferência dos demais elementos presentes na rede. Foram utilizadas ferramentas de monitoração que se complementam observando vários aspectos simultaneamente, por

R.Tec.FatecAM	Americana	v.2	n.2	p.79-85	set.2014 / mar. 2015
---------------	-----------	-----	-----	---------	----------------------

exemplo, a chegada das requisições na rede foi capturada e processada pelo *software Wireshark* [11], utilizado para análise forense, enquanto que informações agregadas dos enlaces foram medidas pelos softwares de monitoramento de rede *Zabbix*.

Com o uso do *software Wireshark*, foi coletado todo o tráfego da rede em intervalos durante 10 minutos, em 29/04/2013, enquanto que o *software Zabbix* forneceu informações sobre as condições de largura de banda durante esta coleta, e o tráfego de entrada e saída de requisições, como mostra o gráfico da figura 2.

Figura 2: Switch CPD: Tráfego proxy porta 9 (21h20m)



Fonte: Fatec Americana - CPD

Na figura 2 pode-se notar que o tráfego de chegada (*Inbound*), sentido em que a Internet chega para a IES no horário de pico das 8h (p.m.), e está com 9.5 Mbps, ou seja, visivelmente saturado no instante da coleta (realizada entre 8h07m. e 8h17m.). O enlace de 10 Mbps, do ponto de vista de carga útil, pode transmitir dados a pouco mais de 9 Mbps, por meio do cálculo de exclusão de cabeçalhos. Assim, sendo, no instante da coleta a utilização do enlace estava a 100%. Também se pode notar 3 períodos de alta utilização no gráfico: entre 8h (a.m.) e 10h (a.m.) – referente ao período matutino de aulas; entre 2h (p.m.) e 3h (p.m.) – referente ao período vespertino e depois, entre 7h30m. (p.m.) e 8h30m. (p.m.) – referente ao período noturno, no qual o número de alunos da IES é excessivamente maior.

2.1 Metodologia para a coleta

Para a coleta de dados foi estipulado o seguinte cenário: (a) coletas de arquivos da rede durante 10 minutos, com o uso do *software Wireshark*; (b) injeção de tráfego na rede (medição ativa), efetuando *download* dentro e fora da rede (medição passiva); (c) injeção de pacotes *Internet Control Message Protocol (ICMP)* para medir o desempenho dentro e fora da rede.

Como a rede da IES é uma rede monitorada, conforme discutido na seção anterior, foram realizados testes de medição ativa, tanto usando o protocolo *ICMP request/reply*, que fornece relatórios de erros, quanto realizando transferências de arquivos que passavam pelo gargalo da rede em horários característicos de pico, como entre 7h30 min. (p.m.) e 8h30 min. (p.m.).

Portanto, de forma amostral, as medidas indiretas foram realizadas nos laboratórios de aula (medição passiva) e na sala de docentes em Regime de Jornada Integral (RJI). A injeção de tráfego (medição ativa) também ocorreu entre laboratórios de aula e a sala de RJI, em abril de 2013.

Na sequência, com a utilização do *software Wireshark*, foi possível coletar, identificar e separar os fluxos por IPs de origem e destino, identificando inclusive as perdas de pacotes. Nenhum tipo de controle é exercido sobre a máquina servidora que desempenha a função de roteador.

3. ANÁLISE ESTATÍSTICA

Nesta seção são apresentados os procedimentos utilizados para o tratamento e a análise dos dados coletados na IES, conforme metodologia proposta em [14].

R.Tec.FatecAM	Americana	v.2	n.2	p.79-85	set.2014 / mar. 2015
---------------	-----------	-----	-----	---------	----------------------

Segundo o método prático de cálculo de tamanhos de amostra apresentado em [4], para 99% de confiança, o tamanho necessário da amostra seria de 1840, com 3% de erro amostral. A população era de 978.000 pacotes, em 5 minutos da amostra coletada. De acordo com a amostra inserida no software *InputAnalyzer* do Arena, obteve-se como melhor aproximação a distribuição Lognormal para os tempos entre chegadas, com a média de 0,326 ms e desvio padrão 0,795 ms. Para os tempos de serviço, obteve-se uma distribuição empírica, com os valores de média 0,325 ms e desvio padrão 0,377 ms.

Nas condições de *heavy-traffic*, os valores estimados estatisticamente foram: taxa de entrada $\lambda = 3,067$ (chegadas por milissegundo), com média dos intervalos entre chegadas de $E(x) = 0,326$ ms, onde x é a variável do intervalo de tempo entre chegadas; taxa de serviço $\mu = 3,067$ (requisições processadas por milissegundo), pois o tempo médio de serviço é $E(s) = 0,326$ ms, onde s é a variável do tempo de serviço de cada requisição. Assim, o nível médio de utilização $\rho = \lambda/\mu = 1$, o que verifica a condição de *heavy-traffic*. Os valores da variância e desvio padrão da variável x , foram, respectivamente, $V(x) = 0,631$ e $\sigma(x) = 0,795$; para a variável s , tem-se: $V(s) = 0,142$ e $\sigma(s) = 0,377$. Esses valores foram obtidos com os softwares *InputAnalyzer* do Arena e o software comercial *Bestfit*.

4. MODELAGEM DA SIMULAÇÃO

4.1 Simulação

Para a construção do modelo de simulação, foi utilizado o software Arena. Este software é de simulação geral e pode ser usado para simular sistemas discretos e contínuos [2]. Para o desenvolvimento do modelo de simulação, é necessário o cálculo dos seguintes valores: número de replicações (10651), período de *warm-up* (fase transiente ou período de aquecimento do sistema – 290 ms) e o tamanho das replicações (tempo - 1000 ms). Todos os detalhes dos cálculos citados podem ser vistos em [14].

Modelagem no Arena: A modelagem foi elaborada no software Arena com a implementação de 6 módulos das barras de componentes (*Basic Process e Advanced Process*), conforme descrito em [14]. No módulo “Chegadas” considera-se a Distribuição Lognormal (0,426; 2,21), com os respectivos parâmetros fornecidos pelo *Input Analyzer* do Arena. O módulo “Decide 1” avalia o número de requisições que chegam ao roteador. A capacidade do *buffer* ($B = rt^*C = 10\text{ms} \cdot 20\text{M bps}$ [10^6 bits por seg] = 200.000 bits) do roteador é de até 34 (200.000 bits/5823 bits) requisições na fila (porta de entrada do roteador), considerando o tamanho médio das requisições de 5823 bits [8]. O valor de 10 ms é oriundo de medição, ou seja, o RTT exato da rede [12].

4.2 Modelos analíticos M/M/1/k e G/G/1

Um possível modelo analítico de fila considerado para a modelagem do enlace congestionado é o modelo G/G/1 (processo de chegada e serviço genéricos, um servidor). Este modelo foi escolhido devido às análises estatísticas apontarem para distribuições não exponenciais, tanto para o processo de chegada como o processo de serviço na rede, pois esse modelo considera que esses processos seguem distribuições genéricas. Embora exista uma limitação do *buffer* do roteador (34 requisições, em média), por simplicidade considerou-se o modelo G/G/1 sem capacidade de fila, que é bem conhecido da literatura [16, 1], com fórmulas fechadas facilmente computáveis para algumas medidas de desempenho. Foram consideradas as aproximações de Kraemer e Lagenbach-Belz e Buzacott e Shanthikumar [3], para o cálculo do número médio de requisições no enlace (essas aproximações consideram disciplina de fila FIFO) [14]:

$$E(L) = \rho + \frac{(\rho^2(C_x^2 - C_s^2))}{(2(1-\rho))} \quad \text{Kraemer e Lagenbach-Belz (K-L)} \quad (3)$$

$$E(L) = \rho + \frac{(\rho^2(1 + C_s^2)(C_x^2 + \rho^2 C_s^2))}{(2(1-\rho)(1 + \rho^2 C_s^2))} \quad \text{Buzacott e Shantikumar (BS-1)} \quad (4)$$

$$E(L) = \rho + \frac{(\rho^2(1 + C_s^2)(2 - \rho)C_x^2 + (\rho^2 C_s^2))}{(2(1-\rho)(2 - \rho + \rho C_s^2))} \quad \text{Buzacott e Shantikumar (BS-2)} \quad (5)$$

onde C_x^2 e C_s^2 são os coeficientes quadráticos de variação das variáveis aleatórias x e s , sendo definidos a partir de $E(x)$, $E(s)$, $V(x)$ e $V(s)$, como:

$$C_x^2 = \frac{V(x)}{E(x)^2} \quad (6)$$

R.Tec.FatecAM	Americana	v.2	n.2	p.79-85	set.2014 / mar. 2015
---------------	-----------	-----	-----	---------	----------------------

$$C_s^2 = \frac{V(s)}{E(s)^2} \quad (7)$$

Também se analisou a aplicação do modelo básico de filas $M/M/1/k$ [7], [6], [1]

$$E(L) = \frac{\rho}{1-\rho} - \frac{(k+1)\rho^{k+1}}{1-\rho^{k+1}} \quad (8)$$

Para os dados amostrais, tem-se que $C^2_x = 5,937$ e $C^2_s = 1,336$ utilizando (6) e (7), indicando que a distribuição do intervalo entre chegadas tem uma alta variabilidade. Com os valores amostrais de ρ , λ e μ e o valor computado de $E(L)$, calculado pelas aproximações em (3), (4), (5) e (8), pode-se facilmente aplicar a fórmula de *Little* para se obter outras medidas de desempenho do sistema [6, 1].

5. RESULTADOS

A análise dos dados amostrais coletados na IES mostra que uma grande parcela das requisições (61,3%) tem protocolo *Transmission Control Protocol* (TCP). Esse protocolo faz várias tentativas de envio de requisições, até que seja atingido o *timeout* e conseqüentemente, o descarte da requisição. Desse modo, cada tentativa de reenvio gera duplicações das requisições. Na amostra analisada, as perdas de requisições representam, em média, 8,4% da amostra, observadas com o *software Wireshark*. Porém, em alguns momentos houve perda de 20%, considerado esse o limite máximo observado.

De acordo com a análise estatística realizada na amostra coletada, a distribuição Lognormal foi aderente para o processo de chegada. Entretanto, para o processo de serviço, nenhuma das 26 distribuições teóricas avaliadas nos *softwares Bestfit e Input Analyzer* do Arena passou nos testes de aderência. Alguns estudos na literatura reportam esta dificuldade da modelagem da *Internet* [5], principalmente por esse controle de tráfego exercido pela família de protocolos TCP. Entretanto, como a pesquisa refere-se à modelagem de um enlace, e não de toda a *Internet*, houve a necessidade de verificação da extensão desse problema também descrito em [15]. Todo esse procedimento de controle do TCP influencia os processos de modelagem e simulação [14].

O modelo de simulação foi implementado no *software Arena*, conforme descrito anteriormente, considerando-se a distribuição Lognormal no processo de chegada e uma distribuição empírica no processo de serviço. O valor de perda de requisições da simulação foi de 4,9%, estimativa relativamente distante do valor amostrado. Considera-se que o fato de haver um número muito grande de requisições com o protocolo TCP (61,3%) ocasionou tal diferença.

No que se refere à análise do modelo analítico de fila $G/G/1$, comparou-se o valor de $E(L)$ - número médio de requisições no enlace, calculado com fórmulas fechadas aproximadas - com o valor de $E(L)$ simulado. Os resultados desta comparação mostram que as aproximações usadas para o modelo $G/G/1$ não são adequadas.

A título de comparação analítica, também se calculou $E(L)$ usando o modelo básico de fila $M/M/1/k$ (utilizando a expressão $E(L)$ em (8)). Neste caso, o resultado $E(L) = 15,98$ ficou relativamente próximo ao valor simulado, considerando-se a forte atuação do protocolo TCP. Também se observou o tempo médio de espera das requisições na fila $E(Wq) = 5\text{ms}$, enquanto que o valor simulado foi $E(Wq) = 4,7\text{ms}$.

Como em horários de pico a perda média de requisições é de aproximadamente 8% (medida amostral), chegando a um máximo de 20%, conclui-se que a rede da IES não está bem dimensionada para os processos de serviço atuais, assim como para a quantidade de usuários que utilizam o enlace.

Uma remodelagem na rede, em geral, é preponderante, assim como o aumento das capacidades nominais dos enlaces (*links*) providos pela operadora.

6. CONSIDERAÇÕES FINAIS

Neste estudo de caso, analisou-se o desempenho de uma rede de comunicação de uma IES com enlace congestionado em períodos de pico. Identificou-se o gargalo do sistema e caracterizou-se o problema de congestionamento empiricamente por meio de coletas e análises de dados, capturando rastros de pacotes de *logs* e estudando atrasos e perdas. Para a análise de desempenho do sistema, utilizaram-se modelos baseados em simulação discreta e também alguns modelos analíticos de teoria de filas. Nas situações de *heavy traffic* deste sistema, ou seja, quando a utilização do enlace está próxima de 100% da sua capacidade, nenhuma das abordagens (simulação e teoria de filas)

R.Tec.FatecAM	Americana	v.2	n.2	p.79-85	set.2014 / mar. 2015
---------------	-----------	-----	-----	---------	----------------------

Artigo de estudo de caso Uma aplicação metodológica para a caracterização de tráfego e avaliação...

caracterizou adequadamente o tráfego de dados. Entretanto, essa situação pode ser explicada devido ao fato de que mais de 60% das requisições do sistema eram de protocolo TCP, que exerce um controle do envio de requisições que não foi considerado nem no processo de simulação, nem nos dois modelos de filas analisados. Apresentou-se uma prospecção da QoS na rede, assim como a proposição de soluções para a melhoria deste.

AGRADECIMENTOS: A autora agradece ao CEETEPS, pelo apoio financeiro; ao coordenador do Centro de Processamento de Dados da Faculdade de Tecnologia de Americana, prof. Wladimir da Costa, assim como ao funcionário Luis Ferrary Neto que forneceu dados relevantes sobre a estrutura física da rede de Internet da IES, e ao aluno de iniciação científica Antônio José Tiago Vicente, pela ativa participação nas coletas de dados da rede.

REFERÊNCIAS

- [1] ARENALES, M., ARMENTANO, V., MORABITO, R., YANASSE, H. **Pesquisa operacional**. Rio de Janeiro: Elsevier, 2007.
- [2] BANKS, J.; CARSON, J. S.; NELSON, B. L.; NICOL, D. M. **Discrete-event system simulation**. 5.ed. New Jersey: Prentice-Hall, 2010.
- [3] BUZACOTT, J.; SHANTIKUMAR, J. **Stochastic models of manufacturing systems**. New Jersey: Prentice Hall, 1993.
- [4] CHWIF, L.; MEDINA, A. C. **Modelagem e simulação de eventos discretos**. 3.ed. s.l.p.: Leonardo Chwif, 2010.
- [5] FLOYD, S., PAXSON, V. "Difficulties in Simulating the Internet", in **IEEE - ACM Transactions on Networking**, Piscataway, 2001. p392-403. v.99.
- [6] GROSS, D., HARRIS, C. **Fundamentals of queueing theory**. 3rd. New York: John Wiley, 1998.
- [7] KLEINROCK, L. **Queueing systems: theory**. New York: John Wiley & Sons, 1975. v.1
- [8] Kurose, J. F.; Ross, K. W. **Redes de computadores e internet**. São Paulo: Pearson Addison-Wesley, 2006.
- [9] Lee, C.; Lee, D. K.; Yi, Y.; Moon, S. "Operating a Network Link at 100%". In: 12th Passive Active Measurement Conference (PAM), 12, 2011, Atlanta. **Anais...** Atlanta: Georgia Institute of Technology, 2011.
- [10] Leung, Y.; Suri, R. "Performance evaluation of discrete manufacturing systems", in **IEEE Control Systems Magazine**, Piscataway, junho, pages 77-86. 1990.
- [11] Orebaugh A., Ramirez G., Beale J. **Wireshark & ethereal network protocol analyzer toolkit**. New York: Elsevier Science, 2006. 448 p.
- [12] Srikant, R. **The mathematics of internet congestion control**. Boston: Birkhäuser, 2004.
- [13] Tader, Paul. Server monitoring with Zabbix. **Linux Journal**, Chicago v. 2010, n. 195, p. 7, 2010.
- [14] Vazquez, M., Morabito, R., Marcondes, C. Caracterização, modelagem e simulação de enlace congestionado de uma Universidade. **Gestão & Produção**. Artigo submetido.
- [15] Willinger, W., Paxson, V. (1998). Where mathematics meets the Internet. **Notices of the AMS**, St. Louis, p. 961-970, v. 45, n. 8.
- [16] Whitt, W. "The queueing network analyzer". **The Bell System Technical Journal**, Piscataway, n.62, p. 2779-2815, 1983.



R.Tec.FatecAM	Americana	v.2	n.2	p.79-85	set.2014 / mar. 2015
---------------	-----------	-----	-----	---------	----------------------

Mariana Vazquez Godoy Miano

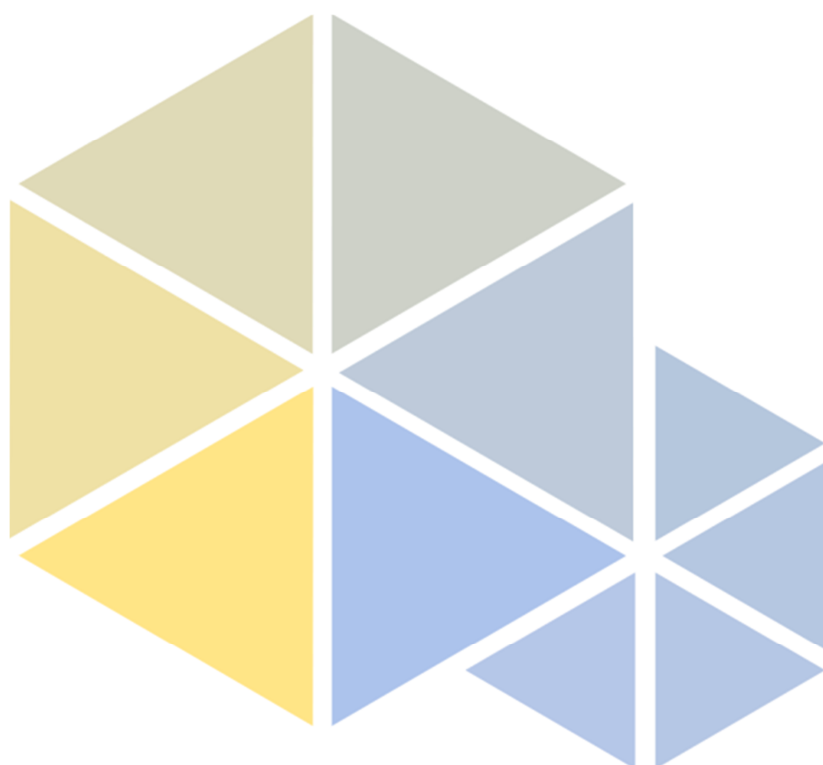
Pós-Doutorado em Engenharia de Produção pela Universidade Federal de São Carlos (UFSCAR - 2013). Doutorado (2009) e Mestrado (2004) em Engenharia Mecânica pela Universidade Estadual de Campinas (UNICAMP). Licenciatura Plena em Matemática pela Universidade Estadual Paulista (UNESP - 2001). Tem experiência nas áreas de Matemática, Engenharia e Computação, com ênfase em Matemática Aplicada à Engenharia e Simulação de Sistemas e Desempenho de Redes de Internet.

Contato: vazquez.prof@gmail.com

Fonte: CNPQ – Currículo Lattes

Justificativa do autor: Artigo referente à minha pesquisa atual de RJI na Fatec-am - finalizando em julho/2014, pesquisa esta vinculada ao meu pós-doutoramento no DEP-UFSCar. Toda a metodologia foi desenvolvida durante o pós-doutoramento, e neste artigo, estou aplicando esta metodologia na Rede de Internet da Fatec-AM, sendo portanto, um estudo de caso.

Como citado no artigo, medições quantitativas são raras de serem encontradas (ou publicadas), assim, trata-se de uma abordagem diferenciada e em especial, uma aplicação na rede de nossa Instituição, com resultados de desempenho e modelo que poderá servir para outras IES.



R.Tec.FatecAM	Americana	v.2	n.2	p.79-85	set.2014 / mar. 2015
---------------	-----------	-----	-----	---------	----------------------