

Caracterização, modelagem e simulação de enlace congestionado de uma universidade

Description, modeling and simulation of the congested link of a university

Mariana Vazquez¹
Reinaldo Morabito²
Cesar Marcondes³

Resumo: Este trabalho apresenta um estudo de avaliação de desempenho de uma rede de comunicação congestionada de uma Universidade. Nesta rede, é comum a formação de filas, devido à alta demanda e ao descarte de requisições em horários de pico, resultando diretamente na queda da qualidade de serviço aos usuários, que têm suas conexões demoradas ou interrompidas e não conseguem acessar páginas na web. O objetivo deste trabalho foi caracterizar o problema de congestionamento empiricamente, capturando rastros de pacotes em logs e estudando atrasos e perdas, e modelar e analisar esse sistema, com base nas teorias de filas e simulação discreta, para avaliar seu desempenho, identificar gargalos e propor soluções para a melhoria da qualidade de serviço. O estudo mostrou que em situações de pico de tráfego no sistema, a abordagem quantitativa que produziu resultados mais próximos do empírico foi a simulação, pois os modelos analíticos de filas estudados não forneceram boas aproximações, principalmente por não refletirem o controle exercido pelo protocolo TCP.

Palavras-chave: Análise de desempenho; Simulação; Teoria de filas; Redes de Internet; Controle de congestionamento.

Abstract: *This paper presents a performance evaluation study of a congested communication network in a university. Queuing often occurs in this network due to high demand and disposal of requests at peak times, resulting directly in decreased service quality for its users, who have their connections delayed or interrupted and cannot access the web pages. The goal of the present work was to characterize this congestion problem empirically, capturing packet traces in logs and studying delays and losses, and to model and analyze this system based on the queuing theory and discrete event simulation to evaluate its performance, identify bottlenecks and proposing solutions to improve the quality of service. The study showed that, in situations of peak traffic, simulation was the quantitative approach that produced results closer to the empirical ones, as the analytical queuing models studied did not provide good approximations, mostly by not reflecting the control exercised by the Transmission Control Protocol (TCP).*

Keywords: *Performance evaluation; Simulation; Queuing theory; Internet network; Congestion control.*

1 Introdução

Com o crescimento massivo da Internet, cresceu também a demanda por diferentes tipos de serviços nas redes de comutação por pacotes. Essas redes deixam de transmitir exclusivamente dados para possibilitar a transmissão de áudios e vídeos com qualidade próxima a das redes de comutação por circuito. Entretanto, para que isso seja possível, a rede deve ser capaz de oferecer serviços dentro de certos padrões de qualidade em termos de largura de banda, atraso, variação do atraso e perda de pacotes.

A evolução da tecnologia de transmissão óptica tem feito a velocidade das redes aumentarem rapidamente. Entretanto, o problema de congestionamento não tem sido resolvido de forma satisfatória quando a demanda é superior à transmissão, e essa tem se tornado uma das principais barreiras para se alcançar a desejada qualidade requisitada pelos serviços mencionados. A utilização da Internet como uma rede multisserviços tem influenciado a proposição de mecanismos de QoS (Qualidade de Serviço), uma vez que a Internet

¹ Departamento de Análise e Desenvolvimento de Sistemas, Centro Estadual de Educação Tecnológica Paula Souza – CEETEPS, Faculdade de Tecnologia de Americana – FATEC, Rua Emílio de Menezes, s/n, CEP 13469-111, Americana, SP, Brasil, e-mail: vazquez.prof@gmail.com

² Departamento de Engenharia de Produção, Universidade Federal de São Carlos – UFSCar, Rodovia Washington Luís, Km 235, CEP 13565-905, São Carlos, SP, Brasil, e-mail: morabito@ufscar.br

³ Departamento de Computação, Universidade Federal de São Carlos – UFSCar, Rodovia Washington Luís, Km 235, CEP 13565-905, São Carlos, SP, Brasil, e-mail: marcondes@dc.ufscar.br

foi concebida para prover serviços que inicialmente não exigiam garantia de banda, limites de atraso e variação do atraso (*jitter*) (Pinto et al., 2003), ou seja, seguia o serviço de melhor esforço.

No Brasil, o provimento de Internet para instituições de ensino e pesquisa em nível federal é feito pela Rede Nacional de Pesquisa, que é responsável pela alocação de enlaces (*links*) para as Universidades. A Universidade Federal de São Carlos (denominada neste trabalho simplesmente como “Universidade”) é uma das beneficiadas, porém a atualização dos enlaces leva tempo e pode gerar situações de congestionamento, como reportadas neste artigo. Essas situações foram objeto de estudo desta pesquisa. No período de avaliação empírica reportada neste trabalho, o enlace de saída para a Internet desta Universidade estava saturado nos horários de pico, com taxas próximas a 155Mbps, que era a velocidade máxima do enlace.

Essa condição de congestionamento é rara de ser descrita empiricamente na literatura, com medidas quantitativas, como as apresentadas neste trabalho. Existem poucos trabalhos como o de (Lee et al., 2011), que apresenta uma visão não convencional baseada em medições quantitativas, em que a degradação de um enlace operando a 100% de sua capacidade (em horários de pico) não corresponde diretamente à taxa de perda de pacotes, que em geral revela-se pequena, mas influencia diretamente a QoS oferecida ao usuário, pois este percebe o atraso adicionado nas transmissões, que têm seu RTT (*Round-Trip Time*) aumentado.

Avaliar o desempenho da rede também pode ser importante quando se deseja instalar um novo serviço que resultará em um tráfego pesado e intenso, permitindo verificar se a rede comportará este novo serviço e, se não, auxiliar no redimensionamento de uma nova topologia (Brito, 2012). Deste modo, faz-se necessária a caracterização e posteriormente, a modelagem de tráfego e análise de desempenho do sistema, com determinadas métricas a serem estabelecidas, por exemplo: tempo de resposta (total, por componente); taxa de transmissão (*throughput*); utilização; escalabilidade; disponibilidade; confiabilidade; entre outros.

O objetivo deste trabalho é caracterizar empiricamente o problema de congestionamento da Universidade estudada, capturando rastros de pacotes em *logs* (processo de registro de eventos) e estudar atrasos e perdas, assim como modelar e analisar esse sistema, com base nas teorias de filas e simulação discreta, para avaliar seu desempenho, identificar gargalos e propor soluções para a melhoria da QoS. Na maioria das situações, os roteadores possuem uma fila única do tipo FIFO (*First In First Out*), que coleta pacotes de ingresso e os desenfileira tão rápido quanto o *hardware* permitir. Essas filas em geral têm tamanho proporcional à capacidade do enlace multiplicado por

um tempo de ida e volta de 200ms (Srikant, 2004). Como os enlaces de rede Internet transportam dados de uma forma serializada, a fila é exigida para gerenciar os pacotes na saída.

A ideia básica subjacente à análise é que, para um determinado enlace, se a capacidade e o fluxo médio são conhecidos, é possível computar o atraso médio de um pacote nesse enlace por meio da aplicação de teoria de filas. Entretanto, devido à complexidade de tráfego de chegada, muitas vezes não é possível determinar com precisão usando modelos Markovianos mais simples, tais como o modelo $M/M/1/k$ (processo de chegada de Poisson, tempos de serviço exponencialmente distribuídos, um servidor e capacidade da fila igual a k). Se esse modelo representar bem o problema, é possível calcular de forma fácil e precisa, por exemplo, o atraso médio fim a fim pela convolução de filas de todos os enlaces (Kleinrock, 1975). No entanto, como também é mostrado neste trabalho com base no estudo de caso da Universidade, o modelo $M/M/1/k$ não é adequado para a análise deste sistema.

Os pacotes que chegam ao enlace são armazenados no *buffer* do roteador e ficam aguardando transmissão; porém, se a taxa de chegada é excessiva, ocorre um congestionamento; em seguida, o roteador inicia o processo de descarte. Para garantir que os dados se comportarão corretamente na presença de congestionamento, os protocolos utilizados para a transmissão de dados na Internet incluem mecanismos de controle do congestionamento de ponta a ponta, que diminuem automaticamente a taxa com que os dados são transmitidos quando é detectada a congestão. Um dos protocolos mais utilizados é o TCP (*Transmission Control Protocol*), que faz o envio em rajadas (milhares de arquivos transmitidos em poucos milissegundos), com o intuito de enviar o maior número possível suportado pelo enlace.

Essa dinâmica entre roteador e TCP poderia ser bem descrita pela técnica de controle de fluidos, entretanto, esta técnica tem limitações por não capturar outros protocolos, como: UDP, DNS, protocolos de imagens em geral. Por outro lado, a teoria de filas é mais genérica, podendo-se utilizar qualquer protocolo (Bu & Towsley, 2001). A busca por soluções para problemas de congestionamento motiva estudos nas áreas de planejamento, gerenciamento e dimensionamento de redes. Todas essas áreas têm como suporte a modelagem e a caracterização do tráfego (Van Woensel et al., 2010). O processo de caracterização e modelagem do tráfego são pontos preponderantes na evolução das redes de telecomunicações. Uma modelagem simples e precisa do tráfego pode permitir o entendimento de um problema físico da rede como um problema matemático/estatístico cuja solução é mais simples. A modelagem também permite a realização de simulações de desempenho de uma rede antes de sua

implantação, e a correção e adequação de parâmetros durante sua existência.

Modelos de avaliação de desempenho de sistemas de filas têm sido aplicados utilizando métodos analíticos (exatos e aproximados), simulação e técnicas relacionadas. Em alguns casos, esses modelos são utilizados de maneira integrada. É comum a aplicação de um modelo analítico para reduzir o conjunto de possíveis configurações do sistema e identificar os parâmetros que mais afetam seu desempenho. Na sequência, utilizando um modelo experimental de simulação, determina-se a melhor entre as configurações avaliadas (Leung & Suri, 1990). Assim, a simulação pode ter um papel importante de investigar como o sistema se comporta, confrontando com a experimentação, medição e análise. Evidentemente, as quatro abordagens (simulação, experimentação, medição e análise) são necessárias, cada uma desempenhando uma função fundamental. A medição é necessária para a verificação da realidade e para desafiar suposições implícitas. Experimentos são cruciais para lidar com questões de implementação, que a princípio podem parecer triviais, mas geralmente introduzem complexidades inesperadas. A experimentação também desempenha papel fundamental na exploração de novos ambientes e como os protocolos da Internet devem operar neles (Banks et al., 2010).

Este trabalho está organizado da seguinte maneira: na seção 2, é apresentado o estudo de caracterização de tráfego real da rede da Universidade e as medições realizadas com a utilização de *softwares* específicos; na seção 3, a análise estatística dos dados coletados neste estudo; na seção 4, o desenvolvimento dos modelos de simulação para aplicação na rede da Universidade; na seção 5, os resultados das medidas de desempenho avaliadas; na seção 6, as conclusões

desta pesquisa, juntamente com as perspectivas futuras do trabalho.

2 Metodologia de coleta e processamento de dados

Para este estudo empírico do congestionamento de uma rede acadêmica houve o acesso aos dados da Universidade, comunidade acadêmica de aproximadamente 12 mil 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 Universidade. Do ponto de vista topológico, a rede universitária está fisicamente organizada como uma estrela estendida. No núcleo da estrela está a Secretaria Geral de Informática (SIn), que abriga os equipamentos do núcleo da rede e o roteador de borda.

Destaca-se que o equipamento usado para o roteamento de borda da Universidade, na época (maio/2012) era um CISCO modelo 7200 com interface SDH (Synchronous Digital Hierarchy) de 155 Mbps com a operadora e, por conseguinte, com a Internet. Depois do estudo do mapeamento da rede, verificou-se que esse roteador era o gargalo da rede de Internet, uma vez que o enlace principal de Internet chegava na SIn com a capacidade de 155 Mbps e saía para os departamentos (ligados diretamente no roteador) com capacidade de 1 Gbps. A topologia desta época com os principais departamentos ligados diretamente ao roteador pode ser visualizada na Figura 1, com destaque para a saída da Internet representada como uma nuvem. Dentro da topologia do campus, destaca-se que os enlaces principais da estrela são conexões de 1Gbps e as ramificações internas (parcialmente descritas na Figura 1) apresentam conexões a 100Mbps.

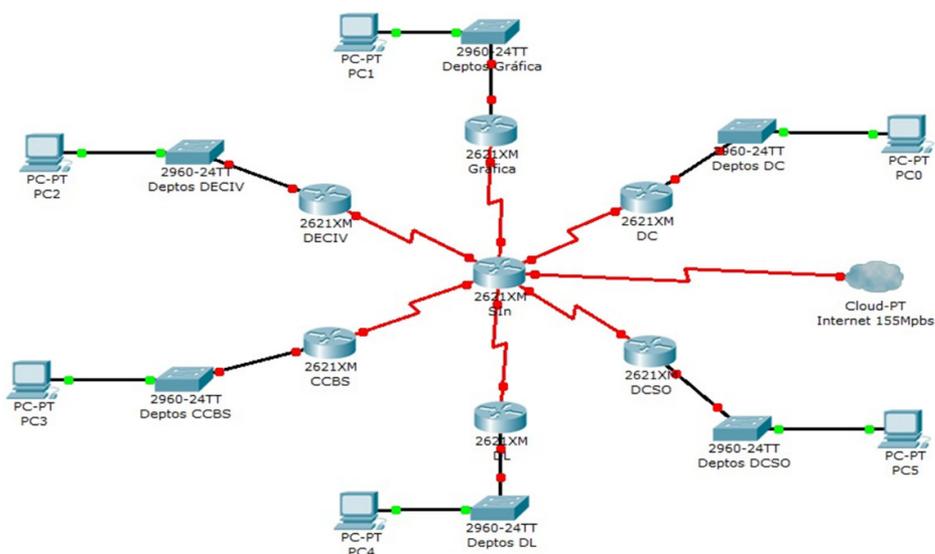


Figura 1. Topologia da rede da Universidade, com os departamentos que possuem roteadores.

Do ponto de vista do enlace de 155Mbps com a operadora, a saída para a Internet acontece através de uma NREN (*National Research and Education Network*) brasileira chamada Rede Nacional de Pesquisa (RNP). O *backbone* (rede de transporte) da RNP, não ilustrado na Figura 1, possui enlaces de 10Gbps entre os Estados e, no caso da Universidade, destaca-se o enlace de 155 Mbps como sendo o gargalo. Foram utilizadas diversas ferramentas de monitoração que se complementam, observando-se vários aspectos simultaneamente. Como exemplo, o atraso de um único fluxo foi capturado e processado pelos *softwares* *TCPDump* (utilizado para coletas de arquivos na rede) e *Wireshark* (utilizado para análise forense), enquanto que informações agregadas (exemplo, utilização total) dos enlaces foram medidas pelos *softwares* de monitoramento de rede Nagios e MRTG.

2.1 Descrição das ferramentas e aplicações

Análise de um único fluxo: A seguir são descritas em mais detalhes as ferramentas e suas aplicações.

TCPDump: utilizou-se o *TCPDump* com configurações de filtro para capturar a transmissão de um único fluxo por meio de segmentação por endereços IP origem e destino e endereços de porta TCP origem e destino. O *TCPDump* tem se provado útil para examinar e avaliar a retransmissão e as operações de gerenciamento de janelas de implementações TCP (Ostermann, 2012). **Wireshark:** em complemento ao *TCPDump*, utilizou-se a ferramenta gráfica de captura *Wireshark* para análises mais complexas, pois este *software* consegue capturar muitos detalhes dos pacotes, como endereçamento (IPs de origem e destino), tamanho dos arquivos, tempo de transmissão, reenvio, perda e, num nível mais elevado, até mesmo acessar o conteúdo dos arquivos (Orebaugh et al., 2006).

Análise de informações agregadas do enlace: De modo a compartilhar a experiência observada por um único fluxo em uma fila compartilhada do enlace de saída, também foram utilizadas ferramentas para análise macroscópica, para possibilitar uma relação com a análise anterior, microscópica. A Universidade contava com um esquema de monitoramento baseado em Nagios (Harlan, 2003) e MRTG (Oetiker, 1998).

Nagios: foi utilizado para monitorar serviços de diferentes protocolos (como SMTP, POP3, entre outros) e também os recursos computacionais de equipamentos de rede (como a carga do processador, a temperatura, a taxa de descarte por interface).

MRTG (*Multi Router Traffic Grapher*): do ponto de vista da monitoração do *link* agregado de saída para a Internet, fez-se o uso do MRTG que monitora dados coletando estatísticas consolidadas pelo equipamento a cada 5 minutos. Foi então acionado um programa *SNMP* que extrai essas informações e apresenta gráficos da utilização da última hora, 24 horas, 1 mês e dados anuais. Para o ambiente de coleta, foram utilizados períodos de grande congestionamento, conforme apresentado pelo MRTG.

Com o uso dos *softwares* *TCPDump* e *Wireshark*, foi coletado todo o tráfego da rede em intervalos de 5 minutos, enquanto que os *softwares* Nagios e MRTG forneceram informações sobre as condições de largura de banda durante estas coletas, como mostram os gráficos das Figuras 2, 3 e 4. Os gráficos apresentam: banda lado a lado com perdas, erros de conexões e utilização do enlace.

Na Figura 2, pode-se notar que o tráfego de chegada (*Inbound*), sentido em que a Internet chega para o campus, está com 127.33 Mbps, ou seja, bastante saturado no instante de uma das coletas. O enlace de 155 Mbps, do ponto de vista de carga útil, pode transmitir dados a pouco mais de 140 Mbps, por meio do cálculo de exclusão de cabeçalhos. Assim sendo, no instante da coleta a utilização do enlace estava a 90,95%. As Figuras 3 e 4 apresentam os resultados de perdas de pacotes em diferentes escalas. Na Figura 3, tem-se as perdas medidas em intervalos de 2 horas com média de 3 erros/segundo. Também se pode notar que o padrão diurno gera mais tráfego e, conseqüentemente, mais erros entre 10 e 18 h. A Figura 4 apresenta um detalhamento dos erros, em intervalos de tempo de 20 min, entre 10 e 13 h.

2.2 Metodologia para a coleta

Para a coleta de dados foi estipulado o seguinte cenário: (a) coletas de arquivos da rede a cada 5 minutos até totalizar 30 minutos, com o uso do *software* *TCPDump*; (b) injeção de tráfego na rede

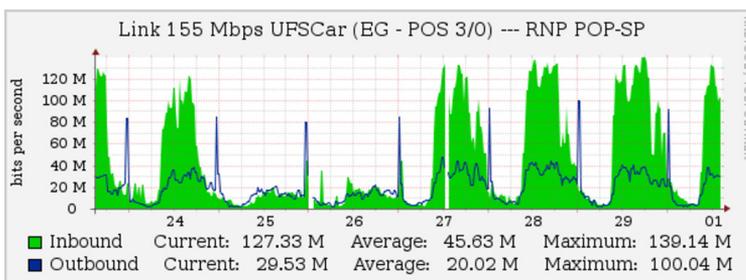


Figura 2. Utilização da rede da Universidade nos dias do mês.

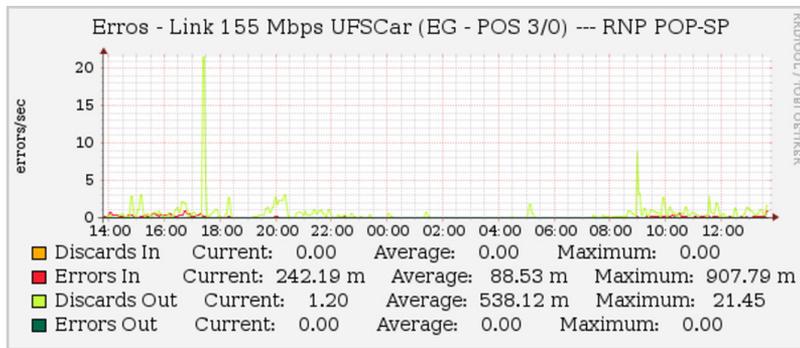


Figura 3. Erros nas conexões em períodos de 2 horas.

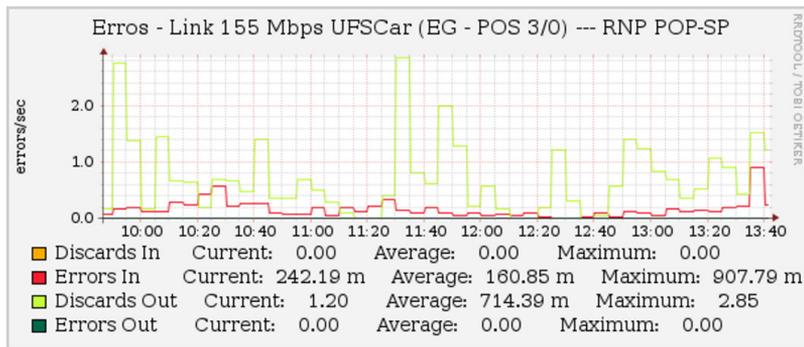


Figura 4. Erros nas conexões em períodos de 20 minutos.

(medição ativa), efetuando *download* dentro e fora da rede (medição passiva); (c) injeção de pacotes ICMP (*Internet Control Message Protocol*) para medir o desempenho dentro e fora da rede; (d) geração de gráficos entre os nós avaliados com o Nagios e o MRTG (*Multi Router Traffic Grapher*).

Medição Passiva: O método de medição passiva analisa o desempenho de redes por meio do uso de dispositivos passivos, que são assim chamados por não interferirem no tráfego da rede quando realizam suas medições. Ou seja, esses dispositivos apenas observam o tráfego corrente que passa pelo ponto de observação que são acionados periodicamente para que as informações sejam coletadas. É desta forma que o desempenho e o estado da rede são analisados. Medir o desempenho utilizando medição passiva não aumenta o tráfego da rede no momento da medição, já que esta técnica utiliza o tráfego real. Por outro lado, é preciso acessar o meio para coletar os dados e os alarmes geram tráfego, que em alguns casos pode ser substancial. Além disto, a quantidade de dados coletados pode ser vigorosa, especialmente se a análise de fluxo ou a coleta de informações requerer a observação de todos os pacotes de dados trafegando na rede.

As medições passivas são valiosas para avaliar o desempenho quando se quer detectar o problema da rede, porém existe a limitação em emular quadros de erros ou isolar o local exato do problema. Um outro

problema é a segurança. Este tipo de solução para medir o desempenho precisa acessar informações de todos os pacotes para caracterizar o estado da rede, comprometendo assim a privacidade dos usuários. A garantia da segurança dos dados torna-se um importante ponto a ser considerado.

Durante o período de coleta, realizaram-se medições passivas em momentos de pouco tráfego (3 A.M.) e de muito tráfego, ou seja, onde congestionamentos tinham maior probabilidade de ocorrer. Os registros nos *logs* (processo de registro de eventos) de milhares de fluxos mostraram que, em média, fluxos diurnos sofriam um acréscimo substancial do tempo de ida e volta (RTT) quando comparados aos fluxos noturnos. Entretanto, devido à natureza complexa de relacionar o RTT e a localização do congestionamento, optou-se pela realização de medições ativas, apresentadas a seguir.

Medição Ativa: A medição ativa é um método para analisar o desempenho de redes e tem como propósito injetar pacotes de teste na rede ou enviar pacotes a servidores e aplicações para, na sequência, medir o desempenho da rede, avaliando como os pacotes de teste se comportam durante o tráfego. No entanto, este método tem a desvantagem de acrescentar tráfego extra, sendo que o tráfego adicional não faz parte do comportamento normal da rede, ou seja, é considerado tráfego artificial que poderia emular uma transação de um usuário. O volume e outros parâmetros do

tráfego adicional são inteiramente ajustáveis, sendo o tráfego adicional pequeno, quando comparado com o tráfego total, e suficiente para realizar medições significativas.

Medir o desempenho da rede utilizando medições ativas fornece o controle explícito na geração de pacotes para a realização das medidas. Esse controle inclui a natureza do tráfego gerado, técnicas de amostragem, temporização, frequência, agendamento, tamanho e tipo de pacotes (que podem variar a fim de simular vários tipos de aplicações), qualidade estatística, caminho e funções escolhidas para monitoramento. O estado de ser ativo implica testar o que se deseja quando é necessário. A simulação e a verificação do tráfego se tornam uma tarefa simples, se a QoS e os acordos de nível de serviço forem atendidos.

Como a rede da Universidade 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 14 h e 15 h. Depois do mapeamento de utilização da rede nos horários de pico, também se observou quais eram os departamentos que possuíam tráfegos maiores, consequentemente, maior utilização da rede, conforme a análise passiva detalhada anteriormente. O *software* Nagios fornece as porcentagens de utilização dinamicamente.

Portanto, de forma amostral, as medidas indiretas foram realizadas no Departamento de Computação (departamento com altas taxas de utilização na Universidade), em 6 coletas de 5 minutos de captura de tráfego em horário de pico (com o TCPDump), em outubro/2012. Na sequência, com a utilização do *software* Wireshark, foi possível identificar e separar os fluxos por IPs de origem e destino, identificando inclusive as perdas de pacotes. As medidas foram realizadas indiretamente com a utilização dos *softwares* Wireshark e TCPDump para coletar o fluxo de pacotes da rede e o MRTG (*Multi Router Traffic Grapher*), para gerar os gráficos de interfaces. O *software* de gerenciamento Nagios juntamente com o MRTG geram graficamente estatísticas com o uso do protocolo SNMP (*Simple Network Management Protocol*), que funciona como um cliente-servidor. Deste modo, o que se faz é monitorar o desempenho da rede e observar como uma interface se comporta por meio da análise dos arquivos *dump*. Nenhum tipo de controle é exercido sobre o roteador.

3 Análise estatística

Nesta seção, são apresentados os procedimentos utilizados para o tratamento e a análise dos dados coletados na Universidade. O maior RTT (*Round-Trip Time*) na rede analisada é de 8 ms (tempo máximo que um pacote leva para ir e voltar, determinado pelo comando

Ping – utilitário de sistemas operacionais que usa o protocolo ICMP para testar a conectividade entre equipamentos). O comando Ping informa o tempo de ida e volta dos pacotes, a quantidade de pacotes transmitidos, recebidos e a porcentagem de perda.

Os dados obtidos do *Wireshark* foram analisados estatisticamente por dois *softwares* comerciais: o *InputAnalyzer* do ARENA e o Bestfit. As distribuições *heavy-tailed* (i.e., distribuições que quando comparadas com distribuições normais ou Gaussianas, apresentam uma quantidade muito maior de dados ao longo de uma “cauda longa”) são comumente encontradas quando se faz a análise dos dados de entrada de um sistema de Filas (Rodríguez-Dagnino, 2004). De acordo com (Crovella et al., 1998), caracterizam-se como distribuições *heavy-tailed* as distribuições dos tamanhos dos arquivos que trafegam na Internet, incluindo arquivos requisitados pelos usuários, arquivos transmitidos pela rede, a duração da transmissão dos arquivos e a quantidade de arquivos armazenados nos servidores (filas).

Segundo (Banks et al., 2010), o uso da Internet dando suporte à telefonia, como o uso de VOIP (*voice over IP*), por exemplo, incitou o desenvolvimento de novos modelos que sugerem a utilização de distribuições *heavy-tailed*, como Pareto e Weibull. O uso de fax e as conexões com a Internet transformaram radicalmente o comportamento estatístico do tráfego, e o uso de processo de Poisson passou a não corresponder à realidade. O trabalho de (Willinger & Paxson, 1998) mostra que os processos de chegada de pacotes na Internet não são processos de Poisson. Uma vez que a variância no tamanho dos arquivos transferidos é muito grande, as distribuições *heavy-tailed* mostram-se bastante adequadas a esta representação.

Assim, após as análises acima citadas, confirmou-se que as taxas de chegada e serviço seguem distribuições *heavy-tailed*. A amostra utilizada foi de 2000 valores (intervalos de tempo entre chegadas e tempos de serviço). De acordo com os testes de aderência (Quiquadrado, Kolmogorov-Smirnov, Anderson-Darling) realizados pelos *softwares* estatísticos Bestfit e *Input Analyzer* do ARENA, as distribuições Lognormal, Loglogística, Weibull e Pareto alternaram-se como as distribuições mais aderentes aos dados, tanto para os intervalos de tempo entre chegadas como para os tempos de serviço.

Segundo o método prático de cálculo de tamanhos de amostra apresentado em (Chwif & Medina, 2010), 1600 valores seriam suficientes, com 95% de confiança. Para 99% de confiança, o tamanho necessário da amostra seria de 1838, com 3% de erro amostral. A população era de 574821 pacotes, em 5 minutos de coleta. 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 0,45 ms e desvio padrão 0,90 ms. Para os tempos de serviço,

obteve-se também uma distribuição Lognormal, com os valores de média 0,47 ms e desvio padrão 0,43 ms.

Nas condições de *heavy-traffic*, os valores estimados estatisticamente foram: taxa de entrada $\lambda = 2,222$ (chegadas por milissegundo), com média dos intervalos entre chegadas de $E(x) = 0,45$ ms, em que x é a variável do intervalo de tempo entre chegadas; taxa de serviço $\mu = 2,127$ (requisições processadas por milissegundo), pois o tempo médio de serviço é $E(s) = 0,47$ ms, em que s é a variável do tempo de serviço de cada requisição. Assim, o nível médio de utilização $\rho = \lambda/\mu = 1,047$, superior a 1, o que verifica a condição de *heavy-traffic*. Os valores da variância e desvio padrão da variável x , foram, respectivamente, $V(x) = 0,80$ e $\sigma(x) = 0,89$; para a variável s , tem-se: $V(s) = 0,19$ e $\sigma(s) = 0,43$. Esses valores foram obtidos com os softwares *InputAnalyzer* do ARENA e o *Bestfit*.

3.1 Equilíbrio amostral – Método SPC (Statistical Process Control)

O Método SPC (Mahajan & Ingalls, 2004) é uma técnica gráfica utilizada em ambientes que necessitam de múltiplas replicações, que tem como objetivo avaliar se a amostra coletada para simulação do sistema está em equilíbrio. Pode-se descrever esse método por meio de quatro passos:

- I. Executar n replicações do modelo simulado, em que cada replicação terá o mesmo tamanho m , para definir os valores de Y_{mn} e as médias de \bar{Y}_m ;
- II. Agrupar as médias em x grupos de tamanho k , em que as médias serão representadas por grupo $\bar{Y}(k)$ b para $b=1,2,\dots,x$. O tamanho destes grupos deve ser proposto de forma a garantir que as médias por grupo sejam aceitas no teste de aderência para normalidade de Anderson-Darling e no teste de correlação de Von Neuman (Mahajan & Ingalls, 2004). É aconselhável um mínimo de 20 grupos. Depois da divisão dos grupos, determina-se a série temporal por:

$$Y(k) = [\bar{Y}(k)1, \bar{Y}(k)2, \dots, \bar{Y}(k)x]$$

- III. Gerar os limites de controle para essas séries temporais. Considerando a estimativa de média populacional (μ) e desvio padrão (σ), a partir da última metade da série $Y(k)$. Depois do cálculo do desvio padrão e da média, determina-se o limite de controle (CL) pela fórmula:

$$CL = \frac{\mu \pm z\sigma}{\sqrt{2}}, \text{ em que } z = 1, 2 \text{ e } 3 \quad (1)$$

IV. Construir a carta de controle, utilizando os três limites encontrados. Por meio dos quatro passos seguintes, verifica-se a saída de controle:

- Existência de pontos fora do limite de controle de 3σ ;
- Dois pontos fora do limite de controle de 2σ de três consecutivos;
- Quatro pontos fora do limite de controle de 1σ de cinco consecutivos;
- Existência de oito pontos consecutivos acima da média ou então oito pontos consecutivos abaixo da média.

3.2 Procedimentos realizados com a amostra da Universidade

A amostra de 2000 valores do Intervalo entre chegadas foi dividida em 100 grupos de 20 elementos, resultando numa série de 100 elementos (Y_1, Y_2, \dots, Y_{100}). Foi calculada a média entre a média Y_{50} e a média Y_{100} , que resultou no valor médio $M(k)$ de 0,000427797 (em ms). O desvio padrão ($\sigma(k)$) resultante entre os elementos Y_{1001} a Y_{2000} da amostra foi de 0,000877976 (em ms). Assim, considerando os 3 intervalos (faixas de controle) de z (-1,1), (-2,2) e (-3,3), o valor de $b=100$ (grupos), $k=20$ (quantidade de valores por grupo), $\mu(k) = 0,000427797$ e $\sigma(k)=0,000877976$, estabeleceram-se os seguintes limites de controle por meio da Equação 1: $CL[-1,1] = (0,0003036324, 0,0005519616)$; $CL[-2,2] = (0,0001794679, 0,0006761261)$; e $CL[-3,3] = (0,0000553033, 0,0008002907)$. Na sequência, verificou-se a validade dos quatro passos acima citados, nos dados amostrais, nos três limites de controle e, para todos eles, a amostra foi validada. Deste modo, pode-se aferir que a amostra analisada está em equilíbrio.

4 Modelagem da simulação

4.1 Simulação

Para a construção do modelo de simulação, foi utilizado o software ARENA. Este software (*Rockwell Automation*) é de simulação geral e pode ser usado para simular sistemas discretos e contínuos (Banks et al., 2010). Modelos de simulação são construídos com objetos gráficos denominados módulos (que definem a lógica do sistema e seus componentes físicos, como máquinas, operadores, etc.). Os módulos são representados por ícones associados a dados de entrada em uma janela de diálogo. Esses ícones são conectados para representarem o fluxo das entidades. Os módulos são organizados em coleções denominadas

“templates”, que propiciarão a modelagem de vários tipos de aplicações.

No processo de modelagem também foram adicionados os recursos, as filas, a lógica do processo e os dados do sistema. A linguagem de simulação utilizada pelo ARENA é a SIMAN. O *InputAnalyzer* do ARENA automatiza o processo de seleção da distribuição mais adequada e seus parâmetros para a representação dos dados existentes, tais como processo e tempo entre chegadas. O *OutputAnalyzer* automatiza a comparação de diversas alternativas de projeto e configuração do sistema que se quer modelar para a simulação. Para o desenvolvimento do modelo de simulação, é necessário o cálculo dos seguintes valores: número de replicações, período de *warm-up* (fase transiente ou período de aquecimento do sistema) e o tamanho das replicações (tempo).

Número de replicações: Para que os resultados de uma simulação sejam confiáveis e próximos à realidade, necessita-se que o modelo seja executado mais de uma vez, ou seja, deve-se executar mais de uma replicação do modelo, dependendo do intervalo de confiança que se deseja alcançar (Pegden et al., 1990). Um intervalo de confiança é um intervalo numérico que possui $(1-a)$ chances de conter o valor real da medida a ser analisada, em que $(1-a)$ é o nível de confiança e a o erro permitido na probabilidade do aparecimento do valor real da variável no intervalo de confiança. Por exemplo, se é permitido um erro de 5% do valor real da variável estar no intervalo, $a = 0,05$, ou seja, há 95% de chance de o valor real estar contido no intervalo (Mahajan & Ingalls, 2004). Para se calcular o intervalo de confiança, é necessário obter a média (μ) e o semi-intervalo (h), sendo o intervalo limitado por $[\mu-h, \mu+h]$. Tendo definido o nível de confiança, é possível determinar o valor tabelado da variável normal padronizada (z), que será usada no cálculo do número de replicações. Aplica-se então a fórmula:

$$n = \left(\frac{100z\sigma}{r\mu} \right)^2 \tag{2}$$

sendo n o número de replicações, z o valor da variável padronizada, σ o desvio padrão, r a precisão requerida e μ a média na amostra. Caso o valor de n encontrado seja decimal, arredonda-se para o próximo valor inteiro. No modelo construído, a média (μ) foi 0,45 ms, o desvio padrão (σ) foi 0,90 ms e a precisão requerida (r) foi 5. Empregando o nível de confiança de 95%, o valor de z é 1,96 (valor tabelado) e o número de replicações é 6146,56 obtido pela Equação 2, arredondando-se para 6147 replicações. Já com o nível de confiança de 99%, o valor de z é de 2,58 (valor tabelado) e o número de replicações obtido com (2) passa a ser 10651 replicações.

Warm-up: O período de *Warm-up* foi obtido por meio do *OutputAnalyzer* do ARENA, por meio da observação da simulação com animação gráfica e dos relatórios estatísticos, que indicaram que, após 5200 ms, o processo de chegada atinge o equilíbrio.

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 Figura 5. No módulo “Chegadas” considera-se a Distribuição Lognormal (0,409; 0,839), 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 de *buffer* ($B = rtt * C = 8ms * 1Gbps [10^9 \text{ bits por seg}] = 8000000 \text{ bits}$) do roteador é de até 1297 (8000000 bits/6166 bits) requisições na fila (porta de entrada do roteador), considerando o tamanho médio das requisições de 6166 bits. O valor de 8 ms é oriundo de medição, ou seja, o RTT exato da rede (Srikant, 2004).

O módulo “Decide 1” faz a seguinte avaliação: se já houver 1297 requisições na fila do roteador, a próxima requisição passará por um módulo “contador”, denominado na Modelagem “Record 2” (para que seja apresentado nos relatórios estatísticos do ARENA a quantidade de requisições perdidas). Então, seguirá para o módulo de “Perda”. Se ainda não houver 1297 requisições na fila, a requisição seguirá para o bloco “RoteamentoRequisições” (Processo de serviço) e na sequência, irá para o bloco “Saída”, sendo considerada

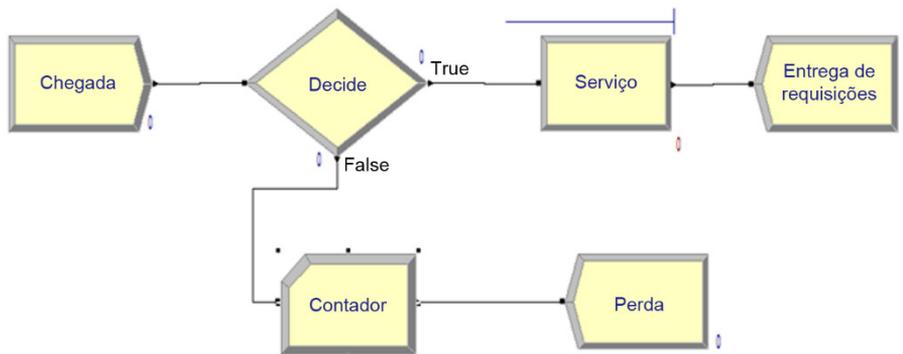


Figura 5. Modelagem no ARENA.

entregue. O bloco RoteamentoRequisições é o bloco do Processo de Serviço. Considerou-se uma distribuição Empírica (uma vez que nenhuma das distribuições teóricas existentes passou nos testes de aderência) com os parâmetros de Probabilidade e Função Distribuição Acumulada fornecidos pelo Input Analyzer do ARENA.

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 para o processo de serviço na rede, pois esse modelo considera que esses processos seguem distribuições genéricas. Embora exista uma limitação do *buffer* do roteador (1297 requisições, em média), por simplicidade considerou-se o modelo G/G/1 sem capacidade de fila, que é bem conhecido da literatura (Whitt, 1983), (Arenales et al., 2007), com fórmulas fechadas facilmente computáveis para algumas medidas de desempenho. Foram consideradas as aproximações de Kraemer e Lagenbach-Belz (Equação 3) e Buzacott e Shanthikumar (Buzacott & Shanthikumar, 1993) (Equações 4 e 5), para o cálculo do número médio de requisições no enlace (essas aproximações consideram disciplina de fila FIFO):

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

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

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

em que 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 (Equações 6 e 7):

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

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

Também analisou-se a aplicação do modelo básico de filas M/M/1/k (Kleinrock, 1975), (Gross & Harris, 1998), (Arenales et al., 2007) (Equação 8):

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

Para os dados amostrais, tem-se que $C_x^2 = 3,99$ e $C_s^2 = 0,83$ 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 (Gross & Harris, 1998), (Arenales et al., 2007).

5 Resultados

A análise dos dados amostrais coletados na Universidade mostra que uma grande parcela das requisições (46,05%) tem protocolo TCP (*Transmission Control Protocol*). Esse protocolo faz várias tentativas de envio de requisições, até que seja atingido o *timeout* (tempo de vida das requisições no enlace, aguardando transmissão) e, conseqüentemente, o descarte da requisição. Desse modo, cada tentativa de reenvio gera duplicações das requisições. Na amostra analisada, todas as tentativas de envio de requisições representam 10,4% da amostra. Entretanto, descontando-se 4,6% de tentativas de reenvio (duplicações), há 5,8% de perda real de requisições (descarte), observadas com o *software* Wireshark.

De acordo com a análise estatística realizada na amostra coletada, a distribuição Lognormal foi aderente para o processo de chegada, com erro quadrático de 0,00612. 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 (Floyd & Paxson, 2001), 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 (Willinger & Paxson, 1998).

O TCP usa o controle de congestionamento fim a fim, o que significa que o remetente limita ou aumenta a taxa de entrega de dados para conexão em função do congestionamento percebido por ele. Por isso, diz-se que o TCP é autorregulado. A conexão TCP é composta de um *buffer* de recepção, um *buffer* de envio e de diversas variáveis. Dentre essas variáveis, tem-se uma janela de congestionamento (“CongWin”), que limita a taxa de envio de pacotes de um remetente TCP. Ao início de cada RTT (tempo de ida e volta), o remetente envia seus pacotes de acordo com o tamanho da CongWin estabelecido, e ao final recebe reconhecimento para os dados, um sinal de que todos os

pacotes foram enviados corretamente. Quando ocorre um evento de perda ou de três ACKs (reconhecimento de pacotes) duplicados (ocasionando desperdício de pacotes), o remetente reduz sua CongWin utilizando a chamada diminuição multiplicativa, reduzindo o valor da CongWin à metade. Porém, existe um limite mínimo do tamanho dessa janela, que é de 1 MSS (*maximum segment size*).

O TCP reconhece que não há congestionamento na rede quando recebe ACKs, então aumenta a CongWin lentamente a cada tempo de ida e volta (aumento aditivo). Esse comportamento do TCP de aumentar a janela de congestionamento lentamente, e depois reduzir à metade bruscamente, gera um comportamento parecido com diagramas do tipo dentes de serra (graficamente). Durante o início de uma conexão TCP tem-se a fase de partida lenta, quando o remetente transmite a uma taxa pequena (normalmente 1 MSS), e depois aumenta sua taxa exponencialmente, duplicando o valor de CongWin a cada tempo de ida e volta até acontecer um evento de perda. O remetente TCP também pode entrar em fase de partida lenta após um evento de esgotamento de temporização, ajustando a janela de congestionamento para 1 MSS e aumentando exponencialmente até que a CongWin alcance metade do valor que tinha antes do evento (*Threshold*) (Kurose & Ross, 2006). Todo esse procedimento de controle do TCP influencia os processos de modelagem e simulação.

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 12,12%, estimativa que pode ser considerada de precisão razoável, uma vez que, na amostra, essa perda sem o desconto dos reenvios e duplicações (no processo de simulação, reenvios e duplicações não podem ser detectados) foi de 10,4%. Vale ressaltar que o controle de congestionamento exercido pelo TCP faz com que uma perda que deveria ser de 10,4%, seja apenas de 5,8% (perda real).

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, uma vez que o $E(L)$ simulado foi de 1280,7, enquanto que no modelo analítico esse valor foi de 238,0 (calculado pela aproximação K-L em (3)), de 239,37 (aproximação B-S1 em (4)) e de 221,10 (aproximação B-S2 em (5)). Isso se deve provavelmente ao fato de essas aproximações não se comportarem bem em sistemas congestionados que têm autoajustes à medida que o sistema tende à saturação (*heavy traffic*), por meio das disciplinas

FIFO baseadas no protocolo TCP (diferente da disciplina básica FIFO considerada nas aproximações do modelo $G/G/1$).

A título de comparação analítica, também calculou-se $E(L)$ usando o modelo básico de fila $M/M/1/k$ (utilizando a expressão $E(L)$ em (8)). Novamente, o resultado $E(L) = 98,99$ ficou muito distante do valor simulado. Ou seja, esse modelo, apesar de considerar a limitação do tamanho da fila (diferentemente do modelo $G/G/1$), também não representa adequadamente o sistema em função das hipóteses assumidas de processos Markovianos de chegadas e serviços e da disciplina FIFO sem autoajuste. Esse resultado vem reforçar relatos da literatura que esse modelo pode não ser adequado para análise destes sistemas com tráfego pesado, apesar de ter sido apresentado em estudos anteriores como alternativa para o cálculo do atraso médio fim a fim pela convolução de filas de todos os enlaces (Kleinrock, 1975).

Com base nos experimentos realizados, como em horários de pico a perda é de aproximadamente 6% (medida amostral), conclui-se que a rede da Universidade estava razoavelmente bem dimensionada para os processos de serviço na época da coleta de dados, assim como para a quantidade de usuários que utilizavam o enlace. Em termos de QoS na Rede, embora na média a perda de pacotes seja relativamente pequena (5,8%) nos departamentos, nos horários de pico, essa porcentagem é muito maior, mais que o dobro. Isso pode ser constatado pela injeção de tráfego e medição dentro de cada departamento e entre departamentos. Esses experimentos foram realizados no Departamento de Computação e no Departamento de Engenharia de Produção da Universidade, e entre os dois. Esse problema pode ser justificado pelo fato de muitas máquinas de acesso ainda possuírem placas de rede de Internet de 100Mbps, ou seja, de baixas capacidades.

6 Considerações finais

Neste trabalho, estudou-se o desempenho de uma rede de comunicação de uma Universidade 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, foram utilizados 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, a abordagem que melhor caracterizou o tráfego de dados e produziu boas estimativas para as medidas de desempenho do sistema foi a simulação. Em termos de perda de requisições, sem descontar os reenvios de dados, o modelo de simulação implementado

no *software* ARENA obteve valores próximos aos valores amostrados no sistema. Como esse modelo não considera explicitamente o controle exercido pelo protocolo TCP, uma diferença entre esses resultados era esperada, além de desvios amostrais aleatórios. Apresentou-se uma prospecção da QoS na rede, assim como a proposição de soluções para sua melhoria.

Quanto à modelagem analítica, as aproximações usadas para o modelo analítico $G/G/1$ não se mostraram adequadas, uma vez que o modelo não considera limitações no tamanho do *buffer* (na rede estudada a capacidade de fila é limitada) e também não considera o controle do protocolo TCP, que faz reenvios e descartes após determinado tempo de espera das requisições na fila. Do mesmo modo, o bem conhecido modelo Markoviano de fila $M/M/1/k$ também não se mostrou adequado, pois os processos de chegada e serviço no sistema estudado não seguem distribuições exponenciais, além deste modelo também não considerar o controle exercido pelo TCP.

Esta pesquisa teve um caráter exploratório e outras investigações ainda deverão ser realizadas, envolvendo outras redes de comunicação em universidades e organizações, para melhor se avaliar o desempenho e eficácia dos métodos aqui utilizados. Estudos específicos para adicionar a dependência dos processos de serviço em relação ao processo de chegada dos pacotes de dados, devido ao controle do protocolo TCP, seriam fundamentais para melhor caracterização do tráfego de dados nas redes de comunicação. Outras pesquisas futuras interessantes seriam o estudo e o desenvolvimento de modelos analíticos de teoria de filas mais elaborados e adequados para representar essas redes, considerando explicitamente o controle do protocolo TCP e utilizando, por exemplo, métodos de estágios e distribuições do tipo fase (Kleinrock, 1975), (Neuts, 1989) para representar melhor as variabilidades dos processos de chegada e processamento dos pacotes de dados e as limitações de *buffers* envolvidos.

Agradecimentos

Os autores agradecem aos dois revisores anônimos os úteis comentários; à CAPES, o apoio financeiro; ao Departamento de Computação e à Secretaria Geral de Informática da Universidade, em especial aos analistas Marcelo Duarte (SIn), que forneceu dados relevantes sobre a estrutura lógica e física da Rede de Internet da Universidade, juntamente com a coleta de Dados na SIn, e Gleise Segatto a coleta de dados realizada no Departamento de Computação.

Referências

- Arenales, M., Armentano, V., Morabito, R., & Yanasse, H. (2007). *Pesquisa operacional*. Rio de Janeiro: Elsevier.
- Banks, J., Carson, J. S., Nelson, B. L., & Nicol, D. M. (2010). *Discrete-event system simulation*. (5th ed.). New Jersey: Prentice-Hall.
- Brito, S. H. B. (2012). *Laboratório de tecnologia Cisco em infraestrutura de redes*. São Paulo: Novatec.
- Bu, T., & Towsley, D. F. (2001). Fixed point approximations for TCP behavior in an AQM Network. In *Proceedings of the 2001 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems - SIGMETRICS'01* (pp. 216-225). New York: ACM.
- Buzacott, J., & Shanthikumar, J. (1993). *Stochastic models of manufacturing systems*. Nova Jersey: Prentice Hall.
- Chwif, L., & Medina, A. C. (2010). *Modelagem e simulação de eventos discretos*. 3. ed. São Paulo: Bravarte.
- Crovella, M. E., Taqqu, M. S., & Bestavros, A. (1998). Heavy-tailed probability distributions in the World Wide Web. In R. J. Adler, R. E. Feldman & M. S. Taqqu, (Eds.), *Practical guide to heavy tails: statistical techniques and applications* (pp. 3-25). Boston: Birkhäuser.
- Floyd, S., Paxson, V. (2001). Difficulties in simulating the internet. *IEEE-ACM Transactions on Networking*, 9(4), 392-403.
- Gross, D., & Harris, C. (1998). *Fundamentals of queueing theory*. 3. ed. New York: John Wiley.
- Harlan, R. C. (2003). Network management with Nagios. *Linux Journal*, 2003(111), 3.
- Kleinrock, L. (1975). *Queueing systems: theory* (Vol. 1). New York: John Wiley & Sons.
- Kurose, J. F., & Ross, K. W. (2006). *Redes de computadores e Internet*. São Paulo: Pearson Addison-Wesley.
- Lee, C., Lee, D. K., Yi, Y., & Moon, S. (2011). Operating a network link at 100%. In 12th Passive Active Measurement Conference (PAM). Atlanta.
- Leung, Y., & Suri, R. (1990). Performance evaluation of discrete manufacturing systems. *IEEE Control Systems Magazine*, 10(4), 77-86.
- Mahajan, P. S., & Ingalls, R. G. (2004). Evaluation of methods used to detect warm-up period in steady state simulation. In Ingalls, R. G., Rossetti, M. D., Smith, J. S., & Peters, B. A. (Orgs.), *Proceedings of the 2004 Winter Simulation Conference* (pp. 663-671). Los Alamitos: IEEE Computer Society.
- Neuts, M. F. (1989). *Structured stochastic matrices of M/G/1 type and their applications*. New York: Marcel Dekker.
- Oetiker, T. (1998). MRTG: The Multi Router Traffic Grapher. In *Proceedings of the 12th Conference on Systems Administration* (pp. 141-148). Berkeley: USENIX Association.
- Orebaugh, A., Ramirez, G., & Beale, J. (2006). *Wireshark & ethereal network protocol analyzer toolkit*. New York: Elsevier Science. 448 p.

- Ostermann, S. *tcptrace TCP dump-file analysis tool*. 2012. Recuperado em 7 de outubro de 2013, de <http://www.tcptrace.org/>.
- Pegden, C. D., Shannon, R. E., & Sadowski, R. P. (1990). *Introduction to simulation using SIMAN*. (2nd ed.). New York: McGraw-Hill.
- Pinto, R. P., Guimarães, E. G., Cardozo, E., & Magalhães, M. F. (2003). Incorporação de qualidade de serviço em aplicações telemáticas. In *21º Simpósio Brasileiro de Redes de Computadores* (pp. 331-346). Natal.
- Rodríguez-Dagnino, R. M. (2004). On the Pareto/M/c and Pareto/M/1/K Queues. In *Proceedings of the 2nd SPIE 5598. Performance, Quality of Service, and Control of Next-Generation Communication Networks* (pp. 183-193). Philadelphia. <http://dx.doi.org/10.1117/12.570535>.
- Srikant, R. (2004). *The mathematics of internet congestion control*. Boston: Birkhäuser. <http://dx.doi.org/10.1007/978-0-8176-8216-3>.
- Van Woensel, T., Andriansyah, R., Cruz, F. R. B., Smith, J. M. G., & Kerbache, L. (2010). Buffer and server allocation in general multi-server queueing networks. *International Transactions in Operational Research*, 17(2), 257-286. <http://dx.doi.org/10.1111/j.1475-3995.2009.00746.x>.
- Whitt, W. (1983). The queueing network analyzer. *The Bell System Technical Journal*, 62(9), 2779-2815. <http://dx.doi.org/10.1002/j.1538-7305.1983.tb03204.x>.
- Willinger, W., & Paxson, V. (1998). Where mathematics meets the internet. *Notices of the AMS*, 45(8), 961-970.