

Metodologia baseada em inteligência computacional para otimização de Qualidade de Serviço em *links* Internet

Ulisses C. Cavalca¹, Caio Mesquita¹, Adriano C. M. Pereira², Eduardo G. Carrano³

¹Departamento de Computação
Centro Federal de Educação Tecnológica de Minas Gerais – Belo Horizonte, MG – Brasil

²Departamento de Ciência da Computação – Universidade Federal de Minas Gerais
Belo Horizonte, MG – Brasil

³Departamento de Engenharia Elétrica – Universidade Federal de Minas Gerais
Belo Horizonte, MG – Brasil

ulisses@ccc.cefetmg.br, caioboninho@gmail.com

adrianoc@dcc.ufmg.br, egcarrano@ufmg.br

Abstract. *The classification and treatment of different kind of traffic on Internet links have become a strategy on Quality of Service (QoS) policies implementation. However, the correct configuration of traffic shaping mechanisms is a challenge in terms of performance and robustness. This work proposes a methodology for optimizing operational parameters of traffic shaping in network environments, using evolutionary algorithms and artificial neural networks. For method validation, we use an actual data from a large network infrastructure of a Federal educational institution. Besides the applicability of the work, the results show improvements in the QoS of the classes with higher priorities.*

Resumo. *A classificação e tratamento dos diferentes tipos de tráfego de links Internet têm se tornado estratégias na implementação de políticas de Qualidade de Serviço (QoS). Entretanto, a correta configuração de mecanismos de traffic shaping é um desafio em termos de desempenho e robustez. O presente trabalho propõe uma metodologia para otimização dos parâmetros operacionais de traffic shaping em ambientes de rede, através de algoritmos evolucionários e redes neurais artificiais. Para validação, o método utiliza dados reais de uma grande infraestrutura de rede pertencente a uma instituição federal de ensino. Os resultados mostraram a melhoria na Qualidade de Serviço das classes de tráfego com maior prioridade.*

1. Introdução

O conceito *Next Generation Network* (NGN) foi proposto em 2004 pela *International Telecommunication Union* (ITU-T), ao qual oferece uma nova abordagem no suporte a diversas tecnologias, expansão, e gerência de serviços e demandas em redes baseadas em pacotes [ITU-T 2004]. Adicionalmente, o termo NGN abrange o tratamento a diversos tipos de serviços, o que inclui serviços multimídias, *streaming* e tráfego em tempo real e não-real. Sendo a Qualidade de Serviço (QoS) um princípio da NGN, a sua aplicação visa garantir robustez, escalabilidade, confiança e disponibilidade dos recursos da rede. A regulamentação da ITU-T prevê ainda a aplicação das arquiteturas de QoS *IntServ* [Braden et al. 1994] e *DiffServ* [Blake et al. 1998], usadas, respectivamente, na garantia de recursos fim-a-fim e entre domínios de rede.

Aliado as políticas de Qualidade de Serviço, o comportamento de usuários tende a mudar à medida que novas tecnologias e recursos são ofertados. No contexto da Internet, a oferta de largura de banda pode representar, a longo prazo, aumento significativo

da demanda de tráfego de rede. Em diversos casos, a capacidade de transmissão é expressivamente elevada em relação a real carga de trabalho do ambiente, com reduzido percentual de utilização relativa. Como consequência, embora esse recurso possa parecer infinito, a capacidade de transmissão deve ser otimizada para minimizar seu custo de implantação e operação, com a maximização do seu desempenho. Para os demais casos, onde a utilização relativa do *link* de Internet é próxima ou igual à 100%, é natural a implementação de *traffic shaping* como solução de QoS, na minimização das perdas de desempenho.

Sob a perspectiva da modelagem, métodos de otimização têm sido amplamente utilizados na alocação eficiente de recursos e tomada de decisão. Técnicas de inteligência artificial, tais como Redes Neurais, Lógica *Fuzzy* e Computação Evolucionária, representam métodos para maximizar o desempenho e minimizar os custos. Naturalmente, a combinação desses métodos pode ser aplicada no contexto de gerenciamento do desempenho dos recursos da rede.

Este artigo aborda a necessidade de gerenciar e otimizar o desempenho de recursos de rede, especificamente *links* de Internet, dos quais suportam serviços como *streaming* de vídeo e áudio, P2P e aplicações típicas da Web 2.0. Propõe-se uma metodologia para configuração de *traffic shaping* através de um processo de otimização, cujas variáveis de decisão são *throughput* (vazão de dados) ofertado to_i e perda de pacotes pp_i , sob uma demanda de tráfego td_i . As variáveis do problema contemplam os parâmetros de *traffic shaping*, tais como taxa de transmissão garantida ar_i , taxa máxima de transmissão c_i e prioridade p_i para cada classe de tráfego i .

A principal contribuição deste trabalho é a metodologia para otimizar a configuração de *traffic shaping*, composta pelas etapas descritas a seguir: *Caracterização de tráfego*, a qual identifica perfis de utilização e demanda de rede; *Emulação de tráfego*, que reproduz maior parte das características técnicas do ambiente de operação, e gera conjunto de dados para treinamento das redes neurais que modelam o mecanismo de *traffic shaping*; *Identificação de modelo*, que obtém os modelos de redes neurais artificiais a serem usados como função de aproximação do mecanismo de *traffic shaping*; *Otimização*, etapa onde as configurações ótimas são obtidas através da computação evolucionária, e redes neurais na avaliação de soluções; e *Análise*, etapa final que consiste na validação das soluções ótimas obtidas, antes da configuração do *traffic shaping* no ambiente de rede em operação. A metodologia pode ser aplicada no planejamento e configuração de *links* Internet, e adequação do plano de capacidade de Provedores de Serviços Internet (ISPs) ou ambiente de *cloud computing*.

Além disso, a metodologia pode ser aplicada em qualquer tipo de implementação de QoS, onde o mecanismo de *traffic shaping* considera para cada classe de tráfego: taxa mínima garantida, taxa máxima e prioridade. Como estudo de caso, o presente trabalho considera ambientes com servidores GNU/Linux, e o *Hierarchical Token Bucket* (HTB) [Devera 2002] como solução de *traffic shaping*. Dados reais de rede e configurações manuais do HTB do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) são utilizados como referência para validação da metodologia proposta. Essa validação, sob o ponto de vista de desempenho, ressalta o processo de otimização de parâmetros de *traffic shaping* frente à configuração manual usualmente realizada em *links* Internet.

As demais Seções deste artigo estão organizadas da seguinte forma: a Seção 2 apresenta trabalhos relacionados a otimização de QoS. A Seção 3 apresenta a metodologia proposta, com foco na descrição das etapas de otimização. Em seguida, na Seção 4 é descrito o estudo de caso. Por fim, a Seção 5 apresenta a conclusão e trabalhos futuros.

2. Trabalhos relacionados

Além dos tópicos apresentados na Seção 1 sobre *Next Generation Network* (NGN), a ITU-T considera como características em redes baseadas em pacotes a convergência de redes. Adicionalmente, [Gupta 2006] aponta como futuras práticas nas redes da próxima geração o cenário de gerenciamento de rede com equipamentos auto gerenciáveis, desempenho dinâmico em função de esquemas de utilização, e soluções de monitoramento e configuração baseadas em sistemas bioinspirados.

Na otimização de atributos de QoS, [Canfora et al. 2005] introduzem esse tipo de problema como \mathcal{NP} -difícil, cuja solução é encontrada a partir do emprego de Programação Inteira e algoritmos genéticos. Nesse trabalho, os autores consideram como função de avaliação a combinação entre custo, tempo de resposta, disponibilidade e confiabilidade. A otimização de atributos de QoS também é abordada por [Yao and Chen 2009], onde os autores consideram preço, tempo de resposta, disponibilidade, rendimento e reputação de serviços.

Em trabalhos recentes, [Rosenberg et al. 2010] reintroduzem a otimização de atributos como um problema \mathcal{NP} -difícil. Entretanto, os autores apresentam uma otimização meta-heurística ao qual pode ser implementada com algoritmos genéticos, busca tabu, e *simulated annealing*. [Pop et al. 2011] contribuem na problemática de otimização de atributos de QoS com o uso do algoritmo *Differential Evolution* na abordagem de tempo de resposta, confiança, disponibilidade e custo. [Mikoshi et al. 2009] incrementam a otimização em ambientes com *DiffServ*, ao aplicar Programação Inteira para otimização do método de controle de admissão proposto no trabalho. [Qaraawy et al. 2012] implementam o algoritmo evolucionário *Particle Swarm Optimization* (PSO), na otimização do gerenciador de fila ativo proposto no trabalho, em cenários de redes congestionadas.

Na aplicação de *traffic shaping*, [Marcon et al. 2011] discutem a problemática dos picos de tráfego no desempenho de um provedor de serviço de Internet. O trabalho traz uma análise de diferente tipos de *traffic shaping*, cujo objetivo é minimizar os picos de tráfego, número de fluxos modelados, e respectivos atrasos. [Kanuparth and Dovrolis 2011] propõem método para detecção de mecanismos de *traffic shaping*, eventualmente aplicados em ISPs, cuja validação está baseada em dados obtidos dos principais servidores de Internet do mundo. Este abordagem abre discussão sobre a neutralidade na Web, tema fortemente debatido no âmbito da Qualidade de Serviço, onde conexões de redes não devem ser limitadas em função de origem ou destino das requisições, tipo de serviço requisitado, meio de acesso, dentre outros critérios.

A breve discussão desses trabalhos estreita a relação entre técnicas de otimização, em especial inteligência computacional e computação evolucionária, e o emprego de *traffic shaping* em rede de computadores. Após a apresentação das referências relacionados a metodologia proposta neste artigo, não foram encontrados trabalhos que abordam a otimização dos parâmetros de *traffic shaping*. A metodologia proposta neste artigo tem grande contribuição na otimização dos parâmetros de *traffic shaping*, e na implementação

combinada de redes neurais artificiais e algoritmos genéticos.

3. Metodologia

Essa Seção discute a classificação das classes de tráfego, modelo de carga de trabalho em redes de computadores, e limitações de *traffic shaping*. Além disso, apresenta o problema de otimização de *traffic shaping* e os passos da metodologia proposta.

3.1. Discussão inicial

Essa metodologia requer a classificação da demanda do tráfego de entrada e saída td_i , ordenada por serviço ou localização, tais como serviços Web, DNS, SMTP, P2P, fluxos por IP de origem e destino, dentre outros critérios. Cada classificação corresponde a uma classe de tráfego i , que possui taxa de transmissão garantida ar_i , taxa máxima de transmissão c_i e prioridade p_i como atributos associados.

Em ambientes de redes, requisições feitas no intervalo de tempo δ_{k-1} refletem em δ_k , da mesma forma que requisições feitas em δ_k influenciam na demanda de rede em δ_{k+1} . A metodologia de otimização dos parâmetros de configuração de *traffic shaping* (AR_k^*) considera a demanda de rede registrada no intervalo de tempo δ_k , caracterizada por tráfego entrante (*download*) e saínte (*upload*). Para períodos δ_l onde $l \neq k$, os parâmetros ótimos AR_k^* devem ser obtidos através de uma nova execução da metodologia. Esse requisito caracteriza uma limitação da versão inicial da metodologia ora proposta neste artigo.

Esse trabalho aborda a otimização dos parâmetros de QoS como um problema empírico, e considera o mecanismo de *traffic shaping* como modelo caixa preta. Para qualquer tipo de controle de tráfego que tem associado taxa de transmissão garantida ar_i , taxa de transmissão máxima c_i e prioridade p_i como parâmetros de QoS, essa metodologia é robusta o bastante para ser aplicada. Vale ressaltar que esse trabalho adota o HTB (*Hierarchical Token Bucket*) como solução de *traffic shaping* em servidores GNU/Linux que atuam como *gateway* ou *proxy*.

A principal limitação da metodologia considera no seu processo de modelagem matemática, a análise do tráfego em intervalos de tempo discretos com a demanda de rede representada por valores constantes. Como a metodologia otimiza o *traffic shaping* para apenas uma única janela de tempo δ_k , oscilações da demanda de rede não são tratadas. Esse comportamento implica, conforme discussão na subseção 3.3.2, na emulação de tráfego a partir do modelo de carga *Constant Bit Rate* (CBR). Consequentemente, é sabido que o desempenho da rede não é totalmente maximizado quando, modelos de carga do tipo *Variable Bit Rate* (VBR) e replicação de tráfego a partir de *traces*, são descartados na etapa de emulação.

Como alternativa, para trabalhos futuros, a otimização de parâmetros de *traffic shaping* pode considerar a modelagem do comportamento dos protocolos TCP e UDP a partir de sistemas dinâmicos, conforme apresentado por [Hollot et al. 2001] e [Li et al. 2005]. A partir do momento que o proposto método seja executado em um ambiente de emulação, a avaliação das soluções candidatas apresentam elevado custo computacional. Dessa forma, o uso de modelos analíticos torna exequível a metodologia para a otimização de parâmetros de *traffic shaping*, quando modelado dinamicamente ao longo do tempo. Para grandes intervalos de tempo e de interesse ao administrador da rede (janela de 18 horas por exemplo), espera-se que a simulação de um ambiente de rede

seja executada na grandeza de segundos. Esse mesmo tempo também será aceitável nas implementações de algoritmos genéticos na avaliação de soluções candidatas.

3.2. Modelagem do problema

Seja um volume de tráfego T em um ambiente com a largura de banda L , classificado em classes de tráfego $i = 1, 2, \dots, n$, no intervalo de tempo δ_k . A taxa garantida ar_i , limite de taxa c_i e prioridade p_i caracterizam cada classe de tráfego i . Em cada intervalo de tempo δ_k , considera-se *throughput* demandado td_i como demanda de rede para cada classe de tráfego i , em que $\sum td_i = T$ representa o total da carga de trabalho. O mecanismo de *traffic shaping* oferece um *throughput* to_i , e em cenários congestionados, observa-se a perda de pacote assumidas por pp_i . A Figura 1 ilustra o modelo proposto para otimização dos parâmetros operacionais de *traffic shaping*.

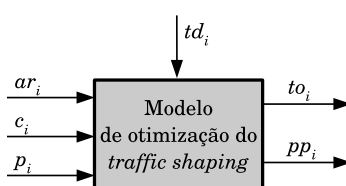


Figura 1. Modelagem de otimização do *traffic shaping*

O HTB admite as seguintes restrições na sua configuração:

$$c_i = L \quad (1)$$

$$0 < ar_i < L \quad (2)$$

$$\sum_{i=1}^n ar_i = L \quad (3)$$

A taxa de transferência máxima (c_i) para todas as classes de tráfego possui o mesmo valor de largura de banda L (Equação 1). Vale ressaltar que essa restrição apenas simplifica a implementação da metodologia, de forma a obter o máximo desempenho para cada classe de tráfego i . As taxas garantidas estão contidas no intervalo $(0, L)$ (Equação 2), cuja soma equivale à largura de banda L (Equação 3).

3.3. Etapas

Essa Subseção representa as 5 etapas para a metodologia proposta, conforme apresentado.

3.3.1. Caracterização do tráfego

Consiste na coleta dos dados e na descrição de algumas variáveis como *throughput*, pacotes enviados, pacotes com erros, latência, dentre outras métricas, através de ferramentas como Cacti¹, Zabbix² ou protocolo SNMP. A metodologia proposta requer somente o *throughput* demandado td_i para cada classe de tráfego i , em um intervalo de tempo δ_k pré-estabelecido. É responsabilidade do administrador da rede a escolha do intervalo δ_k representativo ao ambiente a ser otimizado, bem como a granularidade da caracterização do

¹Cacti: <http://www.cacti.net/>, acessado em 08.dez.2012

²Zabbix: <http://www.zabbix.com/>, acessado em 08.dez.2012

tráfego (5, 10, 30 ou 60 minutos). A metodologia não interfere no processo de definição das classes de tráfego e prioridades da solução de *traffic shaping* do ambiente de rede manipulado.

3.3.2. Emulação de tráfego

Reproduz as mesmas características do ambiente de rede analisado, com os objetivos de realizar testes com o mecanismo de *traffic shaping*, e obter o conjunto de dados para treinamento das redes neurais da etapa subsequente de identificação de modelo. A Figura 2 ilustra o ambiente de teste implementado com máquinas virtuais. A VM-1 representa a demanda da rede local, caracterizando a entrada de tráfego (*download*). Analogamente, a VM-3 representa a demanda Web, representando a saída de tráfego (*upload*). O volume de tráfego demandado (td_i) será implementado pelo modelo CBR (*Constant Bit Rate*) em VM-1 e VM-3, usando *iperf*³.

A VM-2 implementa o mecanismo (criação de classes de tráfego, classificação e marcação de pacotes) de *traffic shaping* HTB, através da ferramenta *traffic control*⁴. As ferramentas de QoS elaboradas nesse trabalho geram, por meio de simulação de Monte Carlo, valores de taxa garantida ar_i . Junto com a prioridade p_i e taxa máxima c_i , as ferramentas ainda fornecem as métricas *throughput* ofertado to_i e perda de pacote pp_i . Essas variáveis compõem o conjunto de dados para treinamento das redes neurais. O conjunto de *scripts* para geração de tráfego e manipulação de regras de *traffic shaping* (*qostools*) está disponível em <http://sourceforge.net/projects/qostools/>. Cada execução leva 2 minutos, onde 1 minuto é dedicado para configuração do experimento. Uma janela de tempo a ser otimizada (instância) requer, no mínimo, 2000 amostras para a etapa de Identificação do Modelo.

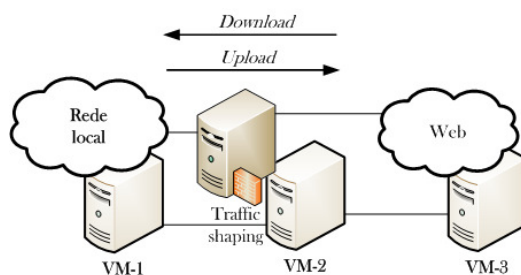


Figura 2. Ambiente de emulação de tráfego

3.3.3. Identificação do modelo

Obtém um modelo empírico e representativo do funcionamento de *traffic shaping* (HTB), por meio de uma função aproximada por Rede Neural Artificial. O modelo será construído baseado em redes neurais *Multilayer Perceptron* (MLP) [Haykin 1998], ou ANFIS [Jang 1993], ou *Radial Basis Function* [Haykin 1998]. Cada modelo de rede neural terá n

³**iperf**: <http://sourceforge.net/projects/iperf/>, acessado em 08.dez.2012

⁴**traffic control**: <http://tldp.org/HOWTO/Traffic-Control-HOWTO/>, acessado em 08.dez.2012

entradas, referentes as taxas garantidas ar_1, \dots, ar_n , e uma saída pertinente ao *throughput* ofertado ou perda de pacote de cada classe de tráfego i . Para implementações com MLP, a metodologia irá avaliar modelos com 3, 5, 10, 12, 15, 20, 25 e 30 neurônios na camada escondida. A escolha da melhor rede neural de aproximação será feita em função do menor erro de validação. Esse modelo obtido é essencial para o processo de otimização, considerando que a avaliação de soluções candidatas requer um tempo de simulação praticamente inaceitável.

3.3.4. Otimização

Nessa etapa são implementadas variações de algoritmo genético para otimização mono objetivo de *throughput* e perda de pacotes, conforme apresentados na Tabela 1.

Tabela 1. Variações dos algoritmos genéticos propostos

Variação	Seleção	Cruzamento	Mutação
AG-1	<i>Stochastic Universal Sampling</i> (SUS)	Interpolação real	Gaussiana
AG-2	<i>Stochastic Universal Sampling</i> (SUS)	Interpolação real	Polinomial
AG-3	<i>Stochastic Universal Sampling</i> (SUS)	<i>Simulated Binary Crossover</i> (SBX)	Gaussiana
AG-4	<i>Stochastic Universal Sampling</i> (SUS)	<i>Simulated Binary Crossover</i> (SBX)	Polinomial
AG-5	Torneio binário	Interpolação real	Gaussiana
AG-6	Torneio binário	Interpolação real	Polinomial
AG-7	Torneio binário	<i>Simulated Binary Crossover</i> (SBX)	Gaussiana
AG-8	Torneio binário	<i>Simulated Binary Crossover</i> (SBX)	Polinomial

A otimização mono objetivo do *throughput* ofertado $TO = (to_1, to_2, \dots, to_n)$ possui a função de avaliação (*fitness*) descrita na Equação 4, em função das variáveis $AR = (ar_1, ar_2, \dots, ar_n)$, $P = (p_1, p_2, \dots, p_n)$ e $TD = (td_1, td_2, \dots, td_n)$. A relação to_i/td_i indica o desempenho do *throughput* ofertado pelo *throughput* demandado para a classe de tráfego i . Valores de to_i/td_i próximos de 1 implicam no atendimento da classe de tráfego i , assim como to_i/td_i menores que 1 representam ambientes super demandados. Cada relação possui sua relevância associada aos pesos p_i normalizados por ω_i . Como a otimização dessa variável contempla a maximização do *throughput* ofertado F^{TO} , na implementação dos algoritmos genéticos, otimizar *throughput* ofertado equivale minimizar $-F^{TO}$.

$$F^{TO}(AR, P, TD) = \sum_{i=1}^n \omega_i(p_i) \frac{to_i(ar_i, p_i, td_i)}{td_i} \quad (4)$$

Analogamente a Equação 5 apresenta a função de avaliação F^{PP} para minimização do percentual de perda de pacote $PP = (pp_1, pp_2, \dots, pp_n)$, também em função de AR, P e TD . Cada PP_i de cada classe de tráfego i possui relevância associada aos pesos ω_i normalizados.

$$F^{PP}(AR, P, TD) = \sum_{i=1}^n \omega_i(p_i) pp_i(ar_i, p_i, td_i) \quad (5)$$

A execução dos algoritmos genéticos prevê uma população de 100 indivíduos, avaliados ao longo de 200 gerações. Cada variação de algoritmos será executada 33 vezes, para que seja analisado a média das melhores soluções. A probabilidade de cruzamento será de 80% e a probabilidade de mutação 0,01% por variável.

3.3.5. Análise

Nessa etapa, os parâmetros obtidos através dos algoritmos genéticos são avaliados, em uma nova execução, no ambiente de emulação conforme proposto na seção 3.3.2. Para uma infraestrutura de *traffic shaping* já existente, a análise também contempla a verificação entre os parâmetros operacionais definidos manualmente e as configurações ótimas encontradas na etapa de otimização. Essa etapa de análise e validação de soluções ótimas precede a configuração do *traffic shaping* no ambiente de rede em operação.

4. Estudo de caso

Esta Seção descreve o cenário real no qual foi aplicado e validado o processo de otimização proposto. Adicionalmente são apresentados os resultados da otimização mono objetivo do *throughput* ofertado e percentual de perda de pacotes. Por fim, é feita a análise da aplicação da otimização para diferentes valores de largura de banda e validação dos resultados ótimos com resultados obtidos a partir da configuração manual.

4.1. Cenário real

A metodologia proposta foi aplicada em um caso de estudo real, observado na rede local do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)⁵ - Campus II. A infraestrutura dessa rede inclui serviços Web, DNS e banco de dados, além de conexão a Internet para aproximadamente 1.200 *hosts*. A ferramenta Cacti é usada para executar a gerência da rede, e o *tc (traffic control)* aplicado na garantia do *traffic shaping* através do HTB.

Atualmente o CEFET-MG possui uma largura de banda de 1Gbps. Entretanto, o ambiente já operou com largura de banda de 6Mbps até fevereiro de 2012, para a mesma demanda de tráfego. Para uma carga de tráfego de 29,640 Mbps, conforme Tabela 2, esse cenário resultava em um ambiente com expressivo congestionamento, cuja solução de *traffic shaping* não era suficiente para assegurar Qualidade de Serviço. Esse cenário motivou este trabalho, na busca de valores ótimos para configuração do HTB. Entretanto, é sabido que tanto a demanda como o comportamento dos usuários serão diferentes a partir do momento que não há congestionamento no *link* de dados. Esse trabalho assume essa demanda como simplificação da aplicação da metodologia, excluindo o esforço no provisionamento de carga de trabalho em ambientes congestionados.

A demanda total de rede de dados do CEFET-MG foi dividida em 6 classes de tráfego, conforme apresentado na Tabela 2, que mostra os valores de *throughput* demandado na rede de dados para o dia 20 de setembro de 2012, de 12h às 13h, que serão utilizados como referência na aplicação da metodologia de otimização neste trabalho. A escolha dessa janela de tempo foi uma escolha aleatória para aplicação da metodologia, onde o maior valor de demanda representaria cenário mais complexo para otimização.

Neste trabalho, a implementação da metodologia para otimização dos parâmetros de *traffic shaping* considera 6 instâncias, diferenciadas na largura de banda da infraestrutura de rede: INST-1 - 36Mbps; INST-2 - 30Mbps; INST-3 - 24Mbps; INST-4 - 18Mbps; INST-5 - 12Mbps; INST-6 - 6Mbps. A proposição dessas 6 instâncias visa avaliar a alocação de recursos de transmissão de dados, de maneira que seja possível obter níveis

⁵CEFET-MG: <http://www.cefetmg.br/>, acessado em 3.dez.2012

Tabela 2. Tráfego em 20.set - 12h às 13h (granularidade 1 hora)

Classes de tráfego	Prioridade	Download [Kbps]		Upload [Kbps]	
10: Servidores locais	3	114	0,38%	199	7,22%
20: Web: Tráfego HTTP	4	15.160	51,15%	1.897	68,83%
30: Web: Redes sociais	6	13.357	45,06%	543	19,70%
40: CEFET-MG Web	1	159	0,54%	36	1,31%
50: CEFET-MG sistemas	2	801	2,70%	71	2,58%
60: Tráfego não classificado	5	49	0,17%	10	0,36%
Total		29.640	100,00%	2.756	100,00%

de serviço aceitáveis com a implantação ótima de *traffic shaping*. Como a carga de tráfego de entrada e de saída é respectivamente 29,640 Mbps e 2,756 Mbps, o volume de tráfego de entrada torna-se objeto de otimização desta metodologia por representar um cenário congestionado e passível de aplicação de Qualidade de Serviço.

A instância INST-6 com largura de banda de 6Mbps será analisada mais detalhadamente, por representar o cenário de rede do CEFET-MG até fevereiro de 2012. Como a demanda de tráfego de saída é menor que a largura de banda considerada, não será abordada a otimização para *upload*. Cabe ressaltar que tanto o *throughput* ofertado como a perda de pacotes serão gerados a partir das estatísticas geradas pela ferramenta *traffic control*, que implementa *traffic shaping* em ambiente GNU/Linux.

4.2. Otimização

Nesta Seção são apresentados os resultados da identificação do modelo e otimização, conforme apresentados nas seções 3.3.3 e 3.3.4. Como a implementação da otimização de *traffic shaping* considera 6 classes de tráfego serão necessários 12 modelos de redes neurais para aproximação de função de *throughput* ofertado (to_i) e perda de pacotes (pp_i). A partir da Tabela 2 verifica-se que a arquitetura *Multilayer Perceptron* (MLP), com reduzido número de neurônios na camada escondida, foi o modelo que melhor representa o funcionamento do HTB para *throughput* ofertado e perda de pacotes. Além disso, as redes neurais para as variáveis pp_1 , pp_4 , pp_5 e pp_6 obtiveram erro de treinamento e validação igual a 0, devido as classes de tráfego 1, 4, 5, e 6 não apresentarem perda de pacotes.

Tabela 3. Métricas dos modelos de redes neurais para *throughput* ofertado e perda de pacotes (INST-6, *download*)

Variável	Modelo (RNA)	Erro treinamento	Erro validação	Épocas
to_1	MLP(6,3,1)	3,54E-03	2,53E-03	8
to_2	MLP(6,3,1)	6,60E-04	3,77E-04	34
to_3	MLP(6,5,1)	1,00E-07	1,00E-07	656
to_4	MLP(6,5,1)	4,44E-03	2,52E-03	10
to_5	MLP(6,3,1)	2,67E-03	2,50E-03	10
to_6	MLP(6,3,1)	4,17E-03	2,43E-03	10
pp_1	MLP(6,3,1)	0,00E+00	0,00E+00	0
pp_2	MLP(6,3,1)	2,15E-03	1,46E-03	430
pp_3	MLP(6,3,1)	1,57E-04	1,14E-04	60
pp_4	MLP(6,3,1)	0,00E+00	0,00E+00	0
pp_5	MLP(6,3,1)	0,00E+00	0,00E+00	0
pp_6	MLP(6,3,1)	0,00E+00	0,00E+00	0

Após a obtenção dos modelos de aproximação de função, a partir da rede neural com menor erro de validação, as variações de algoritmos genéticos são aplicados para busca dos parâmetros ótimos de configuração. Conforme a função de avaliação do *throughput* ofertado (Equação 4), dado que todas as relações to_i/td_i sejam atendidas, o valor ótimo de cada solução será $-\sum \omega_i$. Como ω é o valor normalizado das prioridades

das classes de tráfego, $\omega = [0, 66\ 0, 50\ 0, 16\ 1\ 0, 83\ 0, 33]$, logo o valor de avaliação ótimo será $-3,5$. A Tabela 4 apresenta as melhores soluções encontradas para cada variação de algoritmo genético, e os valores de referência provenientes da configuração manual do CEFET-MG. Nessa implementação, a variação com seleção por torneio binário, cruzamento por interpolação real e mutação gaussiana (AG-5) obteve o melhor desempenho.

Tabela 4. AR^* para *throughput* ofertado (INST-6, download)

AG	Função de avaliação (<i>Fitness</i>)	$AR = [ar_1\ ar_2\ \dots\ ar_n]$
CEFET-MG	-2,9744	[500 3400 1000 500 500 100]
AG-1	-2,995444	[226,5 133,9 0,1 231,3 5311,3 96,9]
AG-2	-2,994846	[255,3 28,5 28,1 29,9 5630,0 28,1]
AG-3	-2,994900	[46,9 26,8 26,8 26,8 5846,1 26,8]
AG-4	-2,994847	[339,4 27,7 27,7 27,7 5549,6 27,7]
AG-5	-2,995500	[93,4 80,5 0,1 112,2 5700,4 13,5]
AG-6	-2,994847	[380,9 27,9 27,5 27,5 5508,5 27,5]
AG-7	-2,994902	[157,8 46,0 26,1 27,6 5716,3 26,1]
AG-8	-2,994847	[339,0 27,7 27,7 27,7 5550,0 27,7]

Analogamente, as variações de algoritmos genéticos foram implementadas para otimização da perda de pacote. A função de avaliação da perda de pacotes (Equação 5) possui valor ótimo igual a 0. A Tabela 5 apresenta os resultados obtidos na otimização dessa variável e os valores de configuração manual do CEFET-MG, onde o algoritmo genético 5 (AG-5) novamente obteve melhor desempenho.

Tabela 5. AR^* para perda de pacotes (INST-6, download)

AG	Função de avaliação (<i>Fitness</i>)	$AR = [ar_1\ ar_2\ \dots\ ar_n]$
CEFET-MG	0,5248	[500 3400 1000 500 500 100]
AG-1	0,502303	[231,9 180,9 1,3 254,6 112,6 5218,7]
AG-2	0,502764	[29,3 29,3 29,3 29,3 29,3 5853,7]
AG-3	0,502706	[26,5 74,0 26,5 27,9 26,5 5818,4]
AG-4	0,502764	[29,3 29,3 29,3 29,3 29,3 5853,7]
AG-5	0,502160	[170,7 199,8 0,1 11,8 48,2 5569,4]
AG-6	0,502764	[29,3 29,3 29,3 29,3 29,3 5853,7]
AG-7	0,502712	[277,0 63,1 25,5 36,4 25,5 5572,5]
AG-8	0,502764	[29,3 29,3 29,3 29,3 29,3 5853,7]

A Figura 3 apresenta a evolução das variações dos algoritmos genéticos ao longo das 200 gerações de simulação, ao ilustrar a média dos melhores indivíduos e o erro padrão em cada geração. Todas as implementações de algoritmos genéticos tiveram melhor desempenho que a configuração manual do CEFET-MG, conforme valores de avaliação apresentados nas Tabelas 4 e 5. AG-1 e AG-5 obtiveram melhor desempenho, o que significa semelhança entre os operadores de seleção. Para a otimização do *throughput* ofertado e perda de pacotes não se pode afirmar sobre o melhor desempenho entre AG-2, AG-4, AG-6 e AG-8. Ainda sobre essas variações, AG-3 e AG-7 apresentam pouco ganho de desempenho para ambas as variáveis.

4.3. Análise dos Resultados

A implementação do método de otimização de *traffic shaping* considerou diferentes valores de largura de banda, conforme apresentação das 6 instâncias de testes na seção 4.1. Essa abordagem permite verificar o comportamento do ambiente de rede para diferentes recursos de transmissão de dados em conjunto com políticas de QoS. Em termos práticos, essa estratégia permite que uma organização adquira um recurso de menor custo com a possibilidade de atendimentos dos níveis de qualidade dos serviços Web.

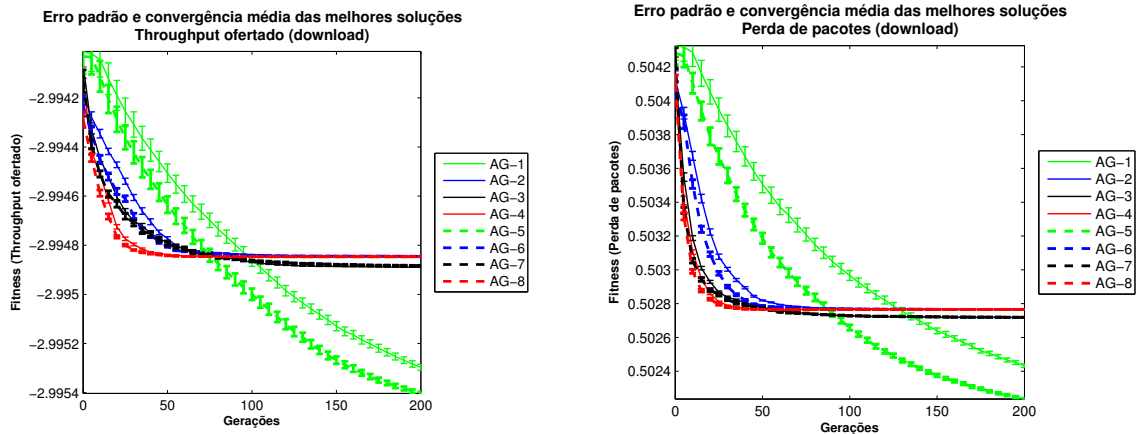


Figura 3. Média e erro padrão dos melhores indivíduos: INST-6 download (throughput e perda de pacotes)

A Figura 4 ilustra o comportamento do processo de otimização para as 6 instâncias de teste, conforme apresentado na seção 4.1. Vale ressaltar que as instâncias INST-1 (L=36Mbps) e INST-B2 (L=30Mbps) foram propostas apenas para validação da metodologia de otimização, visto que os valores de largura de banda para esses cenários não oferecem congestionamento na rede. Os valores de avaliação dessas instâncias para ambas as variáveis estão próximos ao valor ótimo. Para ambas as variáveis, as instâncias INST-5 (L=12Mbps) e INST-4 (L=18Mbps) obtiveram significativa melhora em relação à INST-6 (L=6Mbps) e ao valor de referência do CEFET-MG.

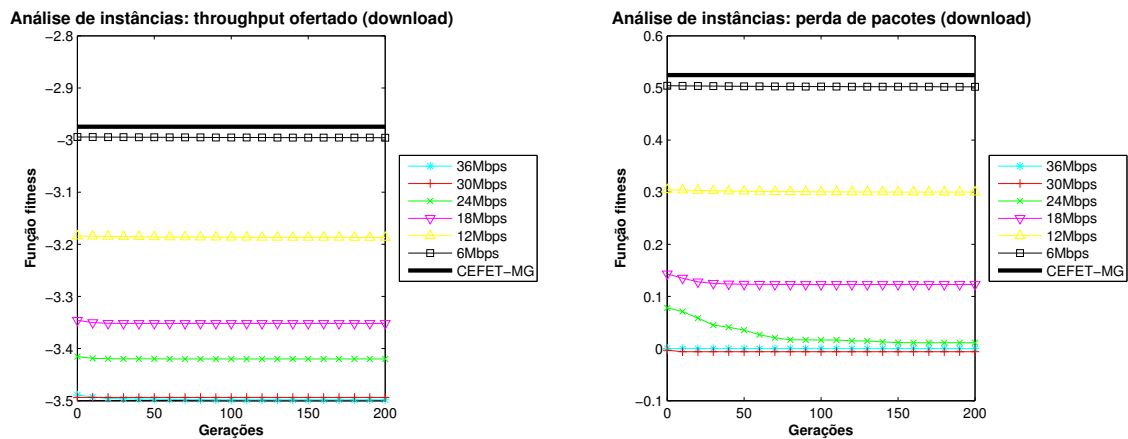


Figura 4. Avaliação de cenários com diferentes larguras de banda (download)

A Tabela 6 apresenta a validação dos valores ótimos de *throughput* ofertado e perda de pacote para a instância INST-6. A partir da configuração manual do CEFET-MG e das soluções ótimas encontradas pelos algoritmos genéticos, o *throughput* ofertado e a perda de pacote podem ser confrontados com a demanda de tráfego apresentada na Tabela 2. A Tabela 6 ainda ilustra o ganho relativo de cada variável para cada classe de tráfego. Percebe-se o ganho para *throughput* ofertado de 25,7% na classe 20, e perda de -99,9% na classe 30. Verifica-se também, para perda de pacotes, melhora de 8,11% para a classe de tráfego 20, porém queda de -8,12% para a classe de tráfego 30.

Tabela 6. Validação dos resultados ótimos (INST-6), ordenados por prioridade

Classes de tráfego	40		50		10		20		60		30	
TO^{CEFET} [Kbps]	165,2	-0,3%	831,5	-0,2%	118,5	-0,4%	3898,5	25,7%	25,9	-0,1%	1009,0	-99,9%
TO^* [Kbps]	164,7		829,5		118,1		4900,6		25,9		1,0	
PP^{CEFET} [%]	0,0	0,0%	0,0	0,0%	0,0	0,0%	0,74	8,11%	0,0	0,0%	0,925	-8,12%
PP^* [%]	0,0		0,0		0,0		0,68		0,0		0,999	

O resultado da otimização de *traffic shaping* reside no pronto atendimento das classes de tráfego com maior prioridade. Embora o HTB possua implicitamente esse comportamento na alocação de banda, o processo de otimização dos parâmetros de configuração forçou a oferta de *throughput* para as classes mais relevantes. A Figura 5 ilustra a alocação acumulada de largura de banda, à medida que cada classe de tráfego é atendida em função da sua prioridade. Pode-se observar que o *throughput* ofertado foi praticamente alocado para todas as classes de tráfego com prioridade menor igual a 4. Analogamente, a perda de pacote foi diminuída a partir da classe com prioridade 4.

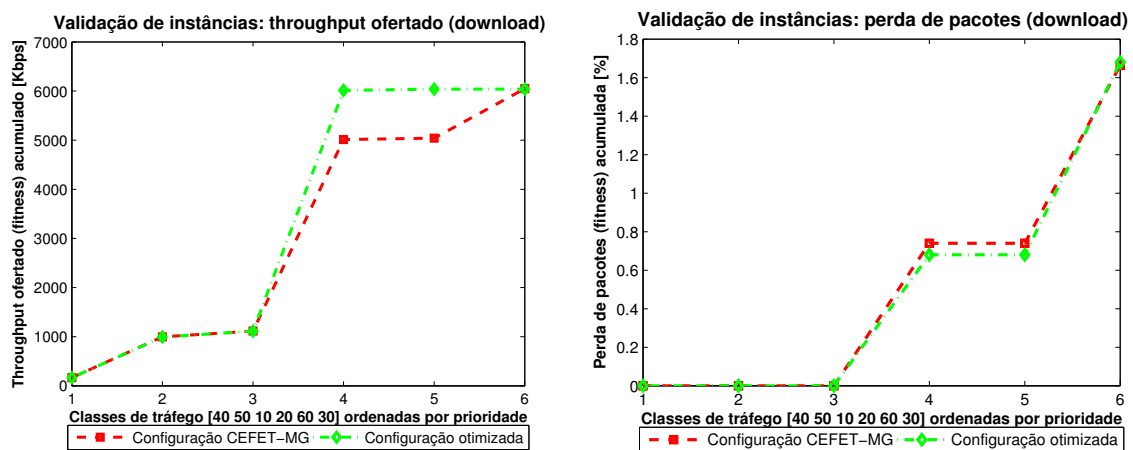


Figura 5. Avaliação de cenários com diferentes larguras de banda (download)

Cabe ressaltar que quanto maior a complexidade da rede, em termos de controle de tráfego e convergência de serviços, maior o valor do trabalho aqui proposto. A aplicação da metodologia concentra-se na garantia de recursos de conectividade para locais que atuam como provedores de acesso à Internet, serviços Web e *cloud computing*. Além disso, a caracterização do tráfego em função de janelas de tempo e heterogeneidade das requisições é um importante aliado na otimização de QoS.

5. Conclusão

Este trabalho propõe uma nova metodologia para otimização de parâmetros de *traffic shaping*, com foco no mecanismo HTB (*Hierarchical Token Bucket*) em ambientes GNU/Linux, devido a sua popularidade e aplicabilidade. A implementação desse processo de otimização, neste trabalho com dados reais do CEFET-MG, contempla o uso de técnicas de inteligência artificial na gerência de redes conforme previsto pelas Redes da Próxima Geração (*NGN*).

A metodologia contempla as etapas de caracterização e emulação de tráfego, identificação de modelo para descrição do funcionamento do mecanismo de *traffic shaping* com redes neurais artificiais, processo de otimização com algoritmos genéticos e

análise dos resultados. A utilização de redes neurais na descrição do funcionamento de *traffic shaping* permite que qualquer modelador de tráfego que possua taxa de transmissão garantida, taxa de transmissão limite e prioridade seja otimizado. Adicionalmente, a abordagem empírica e a modelagem matemática caixa preta do problema permitem a aplicabilidade da metodologia em ambientes reais, cuja carga possui características heterogêneas de serviço, de acesso e operação.

Por outro lado, a metodologia possui a principal limitação na abordagem do tráfego de rede em um único intervalo de tempo. Como consequência, a redução de toda a carga de tráfego nesse intervalo a um modelo CBR (*Constant Bit Rate*) despreza características relevantes de um ambiente de rede, tais como mecanismo de entrega de pacotes e janelas deslizantes do protocolo TCP. Assim, oscilações da demanda de rede não são tratadas pela metodologia.

Essa limitação foi necessária em função do alto custo computacional gasto na etapa de identificação de modelo e avaliação de cada solução candidata na etapa de otimização. A metodologia pode ser aplicada em diferentes intervalos de tempo, à medida que períodos relevantes a operação da rede sejam caracterizados. Para janelas de tempo semelhantes, uma única solução ótima pode ser aplicada para esses períodos. A abordagem dinâmica da metodologia proposta será melhor avaliada em trabalhos futuros.

Os resultados apresentados na Seção 4 ilustram ganho no *throughput* ofertado e perda de pacote para classes de tráfego mais relevantes, em função da carga de tráfego no intervalo de tempo considerado. Deve-se frisar que, no cenário avaliado, o desempenho é pouco aumentado para um ambiente com demanda de 29,6 Mbps em uma infraestrutura com 6 Mbps de largura de banda. Esse caso exige a consideração da operação do ambiente de rede com outras possibilidades de capacidade de transmissão.

Como trabalhos futuros, o presente trabalho abre perspectivas para a otimização multiobjetivo de *throughput* ofertado e perda de pacote, além da consideração da métrica latência na otimização dos parâmetros de QoS. Essa abordagem permite ao administrador da rede a tomada de decisão mais fundamentada sobre qual configuração de *traffic shaping* utilizar. Outro desafio futuro contempla a abordagem dinâmica ao longo do tempo desse método de otimização de *traffic shaping*, cujo objetivo é abordar as variações da demanda de rede.

Referências

- Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and Weiss, W. (1998). An Architecture for Differentiated Services. RFC 2475 (Informational). Updated by RFC 3260.
- Braden, R., Clark, D., and Shenker, S. (1994). Integrated Services in the Internet Architecture: an Overview. RFC 1633 (Informational).
- Canfora, G., Di Penta, M., Esposito, R., and Villani, M. L. (2005). An approach for qos-aware service composition based on genetic algorithms. In *Proceedings of the 2005 conference on Genetic and evolutionary computation, GECCO '05*, pages 1069–1075. ACM, New York, NY, USA.
- Devera, M. (2002). Htb linux queuing discipline manual - user guide. Technical report.

- Gupta, A. (2006). Network management: Current trends and future perspectives. *Journal of Network and Systems Management*, 14:483–491. 10.1007/s10922-006-9044-7.
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition.
- Hollot, C., Misra, V., Towsley, D., and Gong, W.-B. (2001). A control theoretic analysis of red. In *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1510–1519.
- ITU-T (2004). Y.2001: General overview of ngn.
- Jang, J.-S. (1993). Anfis: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(3):665–685.
- Kanuparth, P. and Dovrolis, C. (2011). Shaperprobe: end-to-end detection of isp traffic shaping using active methods. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, IMC '11*, pages 473–482, New York, NY, USA. ACM.
- Li, W., Zeng-zhi, L., and Yan-ping, C. (2005). A control theoretic analysis of mixed tcp and udp traffic under red based on nonlinear dynamic model. In *Information Technology and Applications, 2005. ICITA 2005. Third International Conference on*, volume 2, pages 747–750.
- Marcon, M., Dischinger, M., Gummadi, K., and Vahdat, A. (2011). The local and global effects of traffic shaping in the internet. In *Communication Systems and Networks (COMSNETS), 2011 Third International Conference on*, pages 1–10.
- Mikoshi, T., Takenaka, T., Fujiwara, T., Oki, E., and Shiimoto, K. (2009). Optimization of input admissible traffic flow guaranteeing qos. *Communications Letters, IEEE*, 13(1):49–51.
- Pop, F.-C., Pallez, D., Cremene, M., Tettamanzi, A., Suciu, M., and Vaida, M. (2011). Qos-based service optimization using differential evolution. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation, GECCO '11*, pages 1891–1898, New York, NY, USA. ACM.
- Qaraawy, S., Ali, H., and Mahmood, A. (2012). Particle swarm optimization based robust controller for congestion avoidance in computer networks. In *Future Communication Networks (ICFCN), 2012 International Conference on*, pages 18–22.
- Rosenberg, F., Müller, M. B., Leitner, P., Michlmayr, A., Bouguettaya, A., and Dustdar, S. (2010). Metaheuristic optimization of large-scale qos-aware service compositions. pages 97–104.
- Yao, Y. and Chen, H. (2009). Qos-aware service composition using nsga-ii. In *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, ICIS '09*, pages 358–363. ACM, New York, NY, USA.