Simulação e distribuição de demandas aleatórias em grupos de tráfego com limites inferior e superior de capacidade

Rafael Castro de Andrade
Universidade Federal do Ceará - UFC
Departamento de Estatística e Matemática Aplicada - DEMA
Campus do Pici, Bloco 910. 60455-760 - Fortaleza, CE - Brasil
rca@lia.ufc.br

Arthur Rodrigues Araruna
Universidade Federal do Ceará - UFC
Departamento de Computação - DC
Campus do Pici, Bloco 910. 60455-760 - Fortaleza, CE - Brasil
araruna@lia.ufc.br

Resumo

Apresentamos uma nova estratégia para a otimização do esforço computacional na simulação e distribuição de demandas aleatórias em grupos de tráfego com restrição de capacidade mínima e máxima. Esse problema surge na simulação de tráfego para o planejamento da expansão de redes dorsais de telecomunicações. Deseja-se prever de forma fidedigna as interações de troca de fluxo de dados entre seus nós a partir de informações probabilísticas sobre o comportamento dos diferentes tipos de dados que trafegam na rede. Esse processo consiste em simular a quantidade e os valores dos fluxos de tráfego, atribuindo-os a pares origem-destino de nós que são subdivididos em grupos de acordo com sua capacidade de gerar/receber tráfego e determinar o volume mínimo e máximo de tráfego inter ou intra esses grupos. Os experimentos mostram um ganho significativo na redução do esforço computacional para obter amostras válidas para o problema.
PALAVRAS-CHAVE: simulação de tráfego, redes de telecomunicações, alocação de dados aleatórios ponderados em grupos. ÁREA: Simulação.

Abstract

We present a new strategy for optimizing the computational effort in simulating and distributing weighted random demands into traffic groups with minimum and maximum capacity constraints. This problem arises in traffic simulation for planning the expansion of backbone telecommunications networks. Traffic forecasts should approximate as real as possible that we observe in practice concerning node to node flow exchange interactions originated from many probabilistic distributions of different types of data. This process consists in simulating the number and the bandwidth of the demands and assigning them to pairs of origin-destination nodes that are subdivided into groups according to their potential in generating/receiving traffic, and determining the minimum and maximum volume of traffic inter or intra these groups. Numerical experiments show a significant reduction in the computational effort in obtaining valid scenarios of traffic.
KEYWORDS: traffic simulation, telecommunications networks, group allocation of weighted random data. AREA: Simulation.

1 Introdução

Neste trabalho apresentamos estratégias para o problema de otimização do esforço computacional na simulação e distribuição de demandas aleatórias de uma rede de telecomunicações em *grupos de tráfego* [Andrade, 2002] com restrição mínima e máxima de capacidade. Nos deparamos com esse problema na simulação de tráfego para o planejamento de expansão de redes dorsais de telecomunicações [Andrade, 2002, Andrade et al., 2004]. Deseja-se prever as interações de troca de fluxo de dados entre nós da rede a partir de informações probabilísticas sobre o comportamentos dos diferentes tipos de dados que nela trafegam [Casilari et al., 2001, Karagiannis et al., 2004], simuladas segundo várias funções distintas de distribuição de probabilidade. Esse processo consiste em simular a quantidade (número) e os valores (largura de banda) dos fluxos de tráfego (demandas) e atribuí-los a pares origem-destino de nós da rede sem violar restrições de capacidade dos grupos de tráfego. Tendo-se em vista a dificuldade de se manter um histórico do fluxo de dados nó a nó da rede, procura-se alternativamente subdividir os nós em grupos de acordo com sua capacidade de gerar/receber tráfego e determina-se um histórico da troca de tráfego inter ou intra esses grupos de nós (de pequeno, médio ou grande porte em gerar/receber dados), dando origem a diferentes grupos de tráfego. Por questões estratégicas de qualidade de serviço, deseja-se que a cada grupo de tráfego seja atribuído, sob certas condições, um percentual mínimo e máximo do volume total de dados que irão transitar na rede, segundo o que se observa na prática, de forma que uma expansão futura da capacidade da rede reduza os riscos de não atendimento das futuras demandas previstas para aquele grupo. Essas condições impõem que a distribuição deva concentrar o mínimo possível demandas oriundas de uma mesma distribuição de probabilidade. Essa abordagem é pouco explorada na literatura e nosso objetivo é mostrar novos resultados em que reduzimos o esforço computacional para obter cenários com grupos de tráfego não violados. Um grupo de tráfego é dito violado quando a soma das larguras de banda das demandas a ele alocadas fica aquém de sua capacidade mínima ou além de sua capacidade máxima.

Esse problema é NP-difícil, pois se trata de uma generalização do problema *Number Partitioning* [Mertens, 2003]. Procuramos resolver uma instância desse problema quando o número de grupos de tráfego for dois e suas capacidades mínima e máxima forem tais que a distribuição dos valores nesses grupos minimize a diferença entre as somas de valores de ambos os grupos.

Na literatura existem duas estratégias para a resolução desse problema [Araruna, 2010]. Ambas utilizam formas diferentes de distribuição das demandas nos grupos de tráfego e seguem estratégias diferentes na tentativa de reduzir o esforço computacional na obtenção de cenários válidos, como veremos mais adiante. Neste trabalho fazemos uma análise comparativa entre os resultados alcançados por essas duas estratégias e pela estratégia aqui proposta.

Nosso propósito é mostrar que a nova estratégia comporta-se como um meio-termo entre as duas anteriores, agregando fatores positivos e reduzindo fatores negativos de ambas, distribuindo as demandas de forma otimizada a um custo computacional razoável. Como aplicação prática, a nova estratégia foi incorporada a um simulador de tráfego para o projeto de expansão de redes dorsais de telecomunicações, constituindo uma importante ferramenta para a solução desse problema [Andrade et al., 2010].

O restante do artigo é organizado como segue. Na seção 2 apresentamos detalhes sobre como são gerados os dados aleatórios. Na seção 3 descrevemos algoritmos existentes para a distribuição de dados em grupos e fazemos uma breve comparação entre eles. Resultados computacionais são apresentados na seção 4. Na última seção apresentamos uma breve conclusão e direções para trabalhos futuros.

2 Simulação e validação de dados

Em nosso trabalho, assumimos que o número de demandas segue distribuição de Poisson, enquanto suas larguras de banda seguem distribuições normal, lognormal ou de Pareto, como descrito na Tabela 1. Usamos quatro instâncias (redes) para os testes e a análise de resultados, compostas de 8, 10, 15 e 21 nós. Cada instância possui três classes de tráfego.

		Classe	Média de demandas	Parâmetro I	Parâmetro II
Instância	8	Normal	40	20,0	1,0
		Lognormal	50	5,0	1,5
		Pareto	30	75,0	1,4
	10	Normal	35	100,0	2,0
		Lognormal	60	7,0	1,3
		Pareto	55	80,0	1,8
	15	Normal	100	75,0	1,3
		Lognormal	40	8,0	2,0
		Pareto	60	115,0	1,5
	21	Normal	90	120,0	1,8
		Lognormal	65	2,0	0,8
		Pareto	35	45,0	2,0

Tabela 1: Instâncias de teste

Na Tabela 1, *Parâmetro I* e *Parâmetro II* representam, respectivamente, média μ e variância σ^2 para a distribuição Normal, μ e σ^2 dos logaritmos para a Lognormal e localização x_m e forma k para a de Pareto. *Média de demandas* é o parâmetro λ para a distribuição de Poisson para cada classe de tráfego de cada instância.

Os grupos de tráfego das redes acima, com suas respectivas restrições de limite de volume de dados, estão descritos na Tabela 2. Cada grupo g_i representa uma estimativa percentual do fluxo de tráfego de dados de uma determinada região da rede.

Instâncias	Grupos	Porcentagem do tráfego total	Variação da margem de tráfego
8	g_1	50	5
	g_2	30	6
	g_3	20	8
10	g_1	50	3,5
	g_2	30	3
	g_3	20	4
15	g_1	35	1,75
	g_2	20	1,4
	g_3	20	2
	g_4	10	1,5
	g_5	7	1,4
	g_6	8	4
21	g_1	35	2,1
	g_2	20	2
	g_3	20	2
	g_4	10	2
	g_5	7	1,75
	g_6	8	2,4

Tabela 2: Limites de tráfego de cada grupo das instâncias da Tabela 1.

Na Tabela 2, os valores apresentados estão em percentuais com relação ao fluxo de dados total simulado. A última coluna determina os limites superior e inferior de volume de tráfego para cada grupo em questão. Somando-se ou subtraindo-se o valor dessa coluna ao valor da terceira coluna, temos o intervalo de percentual de tráfego que pode ser atribuído aos grupos ao final da distribuição das demandas.

3 Distribuição de dados aleatórios em grupos capacitados

A seguir, apresentamos três diferentes estratégias de distribuição das demandas entre os grupos de tráfego, propostas por: [Andrade, 2002], Rayee [Gregory Rayee, Estágio de Mestrado em Informática, Université Libre de Bruxelles (superv.: M. Labbé e R. Andrade), 2007] e por [Araruna, 2010].

Cada uma das estratégias tenta, à sua maneira, aleatorizar a distribuição dessas demandas, e cada uma usa validações distintas para decidir se o cenário em simulação deve ser aceito ou abandonado e ressimulado.

Explicamos a seguir como atua cada uma dessas estratégias.

3.1 Estratégia de ANDRADE

Intuitivamente, imagine que você possui um saco cheio de bolas de várias cores e tamanhos, e algumas caixas onde essas bolas devem ser distribuídas. Pense em cada bola como sendo uma demanda, o saco de bolas como sendo o conjunto de todas as demandas simuladas em um cenário para serem distribuídas nas caixas e cada caixa como sendo um grupo de tráfego com limite máximo e mínimo de volume (tamanho). A cor das bolas representa a classe de tráfego à qual cada demanda pertence, e seu tamanho, sua largura de banda associada.

Nessa estratégia, retiramos uma a uma as bolas do saco sem nos importarmos com tamanho ou cor e reordenamos aleatoriamente as caixas. Depois de reordenadas, tentamos colocar a bola na caixa que ficou na primeira posição. Se a bola não couber nessa caixa, tentamos, em ordem, as próximas caixas até que ela caiba em alguma. Se ela violar a capacidade de todas as caixas, escolhemos ao acaso uma das caixas para alocar essa bola. Caso contrário, ela é alocada na primeira caixa cuja capacidade não seja violada com sua inclusão. Repetimos esse processo até distribuirmos todas as bolas.

Após a última bola ter seu destino escolhido, passamos à etapa em que olhamos para cada caixa e verificamos se sua capacidade mínima foi atingida ou sua capacidade máxima não foi excedida. Se *mais de uma* caixa tiver uma capacidade violada, jogamos as bolas fora e recomeçamos todo o procedimento.

Como podemos perceber, esta estratégia tem grande aleatoriedade, pois todas as demandas têm a mesma probabilidade de pertencer a qualquer grupo de tráfego. Isso não ocorre nas redes reais, onde demandas de grande volume têm grande probabilidade de pertencer a grupos de maior capacidade e pequenas demandas a grupos de menor capacidade. Além disso, dada a forma como as demandas são distribuídas, mostrou-se comum que este procedimento execute várias vezes antes de um cenário de tráfego ser aceito, exigindo um esforço computacional muito grande por cenário válido. Na tentativa de diminuir esse número de execuções, é permitida a violação de um grupo, o que não é desejável na prática.

3.2 Estratégia de RAYEE

Utilizando a mesma intuição da estratégia anterior, imagine agora que sejam retiradas todas as bolas do saco e ordenadas de maneira não-crescente pelo tamanho, sem se importar com a cor. Imagine que as caixas também sejam ordenadas de maneira não-crescente com relação às suas capacidades máximas.

Primeiramente, verificamos se a maior bola cabe na caixa de maior capacidade. Se sim, a deixamos lá e passamos à próxima bola. Caso contrário, simulamos um novo conjunto (saco) de bolas e o procedimento é reiniciado.

Para a distribuição das bolas restantes, adotamos a seguinte estratégia: em uma primeira fase, percorremos as caixas na ordem obtida, colocando bolas até que sua capacidade mínima seja atingida, passando para a próxima caixa nessa ordem. Se uma bola não couber na caixa em que estamos tentando colocá-la, tentamos a próxima. Se não conseguirmos distribuí-la em nenhuma caixa, podemos reiniciar o processo com outro saco de bolas. Caso contrário, passamos à bola seguinte e voltamos a analisar as caixas do início da ordem. Ao terminarmos a primeira fase, percorremos novamente as caixas colocando bolas enquanto a capacidade máxima não for atingida, passando para a próxima caixa quando for o caso.

Se, ao terminarmos esse processo, ainda restarem bolas a serem distribuídas, iniciamos o processo novamente simulando outro saco de bolas.

Apesar de que, como veremos posteriormente, a quantidade de cenários rejeitados cai nessa estratégia em relação à anterior, há uma situação que não ocorre nas redes reais: a probabilidade de uma demanda com maior largura de banda pertencer a um grupo com pequena capacidade é nula, pois ela sempre será destinada ao grupo de maior capacidade. O que deve ocorrer é que essa probabilidade seja pequena, assim como a de uma demanda pequena ocorrer em grupos de grande capacidade. Além disso, a aleatoriedade da distribuição é pequena, já que é imposto que as demandas de grande largura de banda entrem em grupos de grande capacidade, enquanto as de pequena largura têm que se adequar ao restante dos grupos (geralmente de pequena capacidade).

3.3 Estratégia de ARARUNA

A estratégia a seguir é a atualmente utilizada pelo simulador de tráfego que desenvolvemos. Ela é dividida em duas fases: distribuição e validação/redistribuição. Na primeira fase (distribuição) realizamos as seguintes operações: (i) as bolas são agrupadas por cor; (ii) cada agrupamento de cor C é subdividido em dois subconjuntos C_1 e C_2 de forma que, se C tem n bolas, as primeiras $\lfloor \frac{n}{2} \rfloor$ fiquem em C_1 e as restantes em C_2; e (iii) ordenamos as bolas de cada subconjunto em ordem não-crescente de tamanho. Com a ordem obtida, para cada subconjunto de cada cor, as bolas são colocadas na caixa com maior espaço disponível [diferença entre os valores do volume máximo permitido em um grupo de tráfego e do volume atualmente associado a este grupo] naquele momento, sem nos importarmos com os limites da caixa.

Note que, pelo comportamento da estratégia, se uma bola não puder entrar na caixa escolhida, então ela não poderá entrar em nenhuma outra caixa. Isso se dá porque a caixa escolhida para conter a bola é aquela que possui a maior folga entre o que tem e o que pode suportar. Portanto, qualquer outra caixa estará mais "lotada" que ela, impedindo tal bola de entrar em qualquer uma delas também. A atribuição que causaria, digamos, o "menor prejuízo" seria a que já pretendíamos realizar. Por isso não nos preocupamos com outras verificações.

Após todas as bolas terem sido destinadas a uma caixa, inicia-se a segunda fase da estratégia. Na segunda fase (validação/redistribuição), verificamos se o volume de bolas nas caixas está dentro dos limites mínimo e máximo de cada uma. Se alguma caixa tiver algum limite violado, fazemos uma redistribuição das demandas por entre os grupos a fim de melhor acomodá-las. Essa estratégia aplicada após a redistribuição visa reduzir ainda mais o número de simulações rejeitadas.

Perceba que essa estratégia importa-se mais com o comportamento real das redes. Aqui, a maior demanda gerada poderá ser alocada em qualquer grupo, até mesmo em um grupo de pequena capacidade. Isso se dá principalmente pelo fato de as demandas de uma classe serem divididas em dois subconjuntos, podendo a maior demanda pertencer ao conjunto que será analisado depois. Daí a importância dessa subdivisão. Nessa situação, os grupos de maior capacidade podem já estar suficientemente cheios para que um grupo menor seja o escolhido no momento da alocação dessa demanda.

Podemos perceber também que demandas de menor largura de banda tem uma possibilidade maior de serem associadas a grupos de grande capacidade. A estratégia de verificação/redistribuição auxilia na aleatoriedade do processo, pois o destino (grupo) final das demandas ainda pode ser mudado nessa fase.

Uma vez feita a alocação das demandas nos grupos, a atribuição do par (*origem*, *destino*) de cada demanda é feita de forma aleatória a partir dos conjuntos de nós a que pertencem a origem e o destino da demanda que definem o grupo de tráfego. Os algoritmos dessas estratégias podem ser encontrados em [Araruna, 2010].

4 Resultados computacionais

Abaixo mostramos os testes de estimativa de parâmetros realizados para validar o conjunto de dados gerados pelas estratégias de amostragem de variáveis aleatórias.

Na Tabela 3 indicamos, individualmente para cada distribuição, valores teóricos esperados e amostrais de média e variância, obtidos a partir de simulações com algoritmos tradicionais de geração de variáveis aleatórias [Andrade, 2002] (denominada estratégia anterior) e valores obtidos com algoritmos mais eficientes para a geração dessas variáveis aleatórias [Araruna, 2010] (denominada estratégia atual). No caso dos dados referentes à distribuição de Pareto, como a estratégia é a mesma de [Andrade, 2002], tabelamos os dados apenas nas colunas referentes à estratégia atual. Uma observação é que, para a distribuição Lognormal na Tabela 3, colocamos tanto os valores referentes aos dados da amostra quanto os referentes aos logaritmos desses dados (linha "Logaritmos" dessas tabelas, o que representa as variáveis Normais associadas a esses dados). O símbolo '—' indica que a estratégia atual é a mesma de [Andrade, 2002].

Instância		Teórico		Est. Anterior		Est. Atual	
	Classe	Média	Variância	Média	Variância	Média	Variância
8	Normal	20,000	1,000	19,997	1,982	20,007	0,995
	Lognormal	314,191	343697,621	3258,860	1,768e+9	311,276	288802,000
	Logaritmos	5,000	1,500	4,990	4,345	5,003	1,488
	Pareto	262,500	$+\infty$	—	—	231,382	254477,000
10	Normal	100,000	2,000	100,002	7,863	99,993	2,005
	Lognormal	2100,646	1,178e+7	10342,300	9,000e+9	2074,880	1,004e+7
	Logaritmos	7,000	1,300	7,002	3,253	6,999	1,283
	Pareto	180,000	$+\infty$	—	—	173,568	44382,900
15	Normal	75,000	1,300	75,001	3,321	75,003	1,294
	Lognormal	8103,084	4,195e+8	136857,000	2,313e+12	8067,920	3,094e+8
	Logaritmos	8,000	2,000	7,964	7,333	8,005	2,001
	Pareto	345,000	$+\infty$	—	—	315,686	340937,000
21	Normal	120,000	1,800	120,003	6,389	120,003	1,788
	Lognormal	11,023	148,916	16,172	3034,540	10,979	141,147
	Logaritmos	2,000	0,800	2,005	1,245	1,999	0,796
	Pareto	90,000	$+\infty$	—	—	88,224	7608,730

Tabela 3: Resultados dos testes de média e variância entre as estratégias implementadas.

O valor $+\infty$ da variância teórica da distribuição de Pareto se dá pelo fato de o parâmetro k utilizado em todas as instâncias ser menor ou igual a dois. A consequência desse valor de parâmetro para as nossas simulações é que o valor de variância amostral obtido dificilmente será "bem comportado" nesses casos, podendo ser ignorado para fins de validação da amostra gerada.

Observamos nessa tabela que as estatísticas obtidas com a estratégia anterior são muito distantes do esperado em teoria. À exceção da média das Normais (como também dos logaritmos da Lognormal), os outros dados mostram uma dispersão significativa dos valores simulados. Com a melhor performance do algoritmo de geração de Normais com base em coordenadas polares [Araruna, 2010], a estratégia atual melhorou as estimativas de média e variância para as distribuições Normal e Lognormal. Porém, ainda verificamos uma pequena diferença na média das Lognormais quanto ao esperado. Quanto à distribuição de Pareto com variância infinita, observamos que a média obtida é a que mais se distancia do valor teórico esperado. Isso mostra o quanto essa distribuição é dependente de valores localizados em suas caudas, principalmente de valores muito distantes da média, para que as estatísticas amostrais aproximem-se dos valores teóricos. Como valores muito grandes para demandas tipo Pareto são pouco prováveis de ocorrer na prática em uma rede real, para o nosso propósito essa "deficiência" não é importante.

Na Tabela 4 estimamos os parâmetros de média e variância usados para simular as amostras Normais e os logaritmos das amostras Lognormais e os parâmetros de localização e forma usados para simular as amostras de Pareto. Essas estimativas devem ser comparadas com os parâmetros fornecidos na Tabela 1. Também comparamos os dados obtidos tanto com a estratégia anterior [Andrade, 2002] quanto com a em [Araruna, 2010].

Observamos, para a distribuição Normal e de Pareto, que as estimativas com a estratégia atual são muito próximas dos valores usados para simular essas amostras. Para a distribuição Lognormal, observamos um pequeno desvio para a média da instância 15 (a estimativa foi de uma média igual a $8,121$, enquanto o valor teórico deveria ser $8,000$) e uma diferença um pouco mais notável para o parâmetro variância nas três primeiras instâncias. Diferenças como essas, apesar de pequenas, quando levadas ao expoente na geração de Lognormais, podem levar a um grande aumento na dispersão da amostra.

Instância		Est. Anterior		Est. Atual	
	Classe	Par. Média	Par. Var.	Par. Média	Par. Var.
8	Normal	19,997	1,983	20,007	0,995
	Lognormal	5,529	5,121	5,050	1,381
	Pareto	—	—	75,054	1,410
10	Normal	100,002	7,863	99,993	2,005
	Lognormal	7,02185	4,444	7,036	1,204
	Pareto	—	—	80,045	1,811
15	Normal	75,001	3,322	75,003	1,294
	Lognormal	9,415	4,824	8,121	1,750
	Pareto	—	—	115,077	1,517
21	Normal	120,003	6,389	120,003	1,788
	Lognormal	1,51636	2,534	2,008	0,775
	Pareto	—	—	45,023	2,002

Tabela 4: Resultados das estimativas de parâmetros realizadas entre as estratégias implementadas.

Analisando os resultados acima, nota-se que os dados simulados estão de acordo com o esperado em teoria.

4.1 Afetação de demandas ponto a ponto na rede

Uma vez simulados os dados das diversas distribuições, os mesmos devem ser associados às demandas para alimentar a rede. Devemos escolher o par origem-destino de cada demanda de forma que, ao final de todo esse processo, possamos satisfazer algumas restrições de troca de tráfego entre pontos da rede. Daí, para um bom planejamento de expansão, surge a necessidade de se simular as demandas e alocá-las de forma a refletir a troca real de demandas entre diferentes grupos de tráfego, com a condição de que o volume de um dado grupo represente pelo menos um percentual mínimo do volume total de tráfego da rede e não ultrapasse o percentual máximo observado na prática.

Se um cenário de afetação das demandas respeita as condições dos grupos de tráfego pré-definidos, dizemos que o mesmo é válido. Caso contrário, o cenário é dito rejeitado, sendo necessário reiniciar todo o processo a partir da etapa de simulação dos dados até que um cenário válido seja gerado. A ideia é aumentar a taxa de cenários aceitos por rejeitados.

A Tabela 5 mostra a porcentagem de cenários rejeitados com relação ao total de cenários simulados para cada instância utilizada. A coluna *# dem* abstrai o número médio de demandas de cada amostra. Por exemplo, *8N* significa que os parâmetros das instâncias de teste são os mesmos da Tabela 1, exceto o número médio de demandas, que foi multiplicado por 8 para cada classe, e *N* significa que a instância sem modificações foi utilizada (assim como na Tabela 1).

Estratégia	# dem.	Instância			
		8	10	15	21
RAYEE	N	0,000	0,000	22,797	2,570
	2N	0,000	0,000	10,196	0,000
	8N	0,000	0,000	9,447	0,000
ANDRADE	N	0,923	0,056	54,426	5,684
	2N	0,004	0,000	56,151	1,708
	8N	0,000	0,000	51,064	0,559
ARARUNA	N	3,798	0,428	43,322	0,067
	2N	0,974	0,004	16,655	0,000
	8N	0,000	0,000	0,790	0,000

Tabela 5: Porcentagem de cenários rejeitados, por estratégia, de cada instância.

Na Tabela 5 verificamos que o método com menor taxa percentual de rejeição de cenários gerados é, no geral, o de RAYEE, sendo o que melhor se comporta na geração de cenários com pequeno número de demandas. À medida que o número de demandas aumenta, a estratégia de ARARUNA apresenta menores taxas de rejeição que os outros dois métodos. Um dos fatores que contribui para esse comportamento é o fato de essa estratégia tentar redistribuir o tráfego entre os cenários violados após a distribuição.

Ainda sobre a heurística de redistribuição de ARARUNA, verifica-se que para uma pequena granularidade de demandas de uma distribuição de cauda pesada, a estratégia de ARARUNA pode obter um número alto de cenários rejeitados. Mas isso não é um grande empecilho devido ao fato de, em redes de telecomunicações reais, o número de clientes ser bem alto, exatamente a situação onde essa estratégia de distribuição mostrou menor taxa de cenários rejeitados.

O método de ANDRADE, de grande aleatoriedade na distribuição das demandas, apesar da flexibilidade em aceitar violação de até um único grupo de tráfego, mostrou ser o de pior performance nesse quesito. O método de RAYEE mostrou-se concentrar um grande número de demandas de uma mesma classe de tráfego (de maior largura de banda) em um grupo de maior concentração de tráfego total. O método de ARARUNA atinge um meio termo nesse aspecto; a saber, não é tão aleatório quanto o de ANDRADE (o que aumenta a chance de violação dos limites de capacidade dos grupos) nem tão concentrador quanto o de RAYEE.

As Tabelas 6 a 8 mostram exemplos de percentuais de acumulação das demandas e de volume de tráfego em execuções do simulador utilizando cada estratégia para as Instâncias 10, 15 e 21. Essas tabelas relatam o comportamento de cada estratégia ao distribuir o tráfego por entre os grupos de cada instância. Mais resultados sobre esses experimentos (omitidos aqui por razão de espaço) podem ser encontrados em [Araruna, 2010].

Nessas tabelas, a coluna *acumulado* mostra quanto do tráfego total cada classe representa na simulação realizada, a linha homônima representa quanto de tráfego cada grupo acumulou do total gerado após a simulação. A linha *% demandas* informa a porcentagem do número de demandas total do cenário simulado que foi atribuída ao grupo de cada coluna. O cruzamento das demais linhas com as colunas dos grupos mostra o quanto de tráfego de cada classe cada grupo detém.

Estratégia	# demandas	Grupos			acumulado	Classe
		g_1	g_2	g_3		
ANDRADE	N	75,822	18,792	5,384	0,688	Normal
		10,478	72,507	17,014	24,556	Lognormal
		37,108	29,886	33,004	74,754	Pareto
	acumulado	41,136	40,395	18,467		
	% demandas	43,165	30,935	25,899		
	2N	76,728	23,277	0	0,625	Normal
		53,218	27,657	19,124	22,518	Lognormal
		28,567	37,249	34,182	76,856	Pareto
	acumulado	52,838	29,394	17,768		
	% demandas	46,757	27,986	25,256		
	8N	61,328	27,303	11,368	0,602	Normal
		14,814	51,056	34,129	24,516	Lognormal
		39,196	31,663	29,139	74,880	Pareto
	acumulado	38,446	36,674	24,879		
	% demandas	42,222	32,222	25,555		
RAYEE	N	2,510	35,971	61,517	0,652	Normal
		100	0	0	25,584	Lognormal
		63,474	31,635	4,890	73,762	Pareto
	acumulado	21,994	22,535	22,135		
	% demandas	57,746	23,239	19,014		
	2N	50,332	0	49,667	0,635	Normal
		7,400	92,599	0	23,534	Lognormal
		20,045	71,799	8,154	75,829	Pareto
	acumulado	25,926	54,799	19,273		
	% demandas	61,654	21,804	16,541		
	8N	3,778	48,312	47,909	0,616	Normal
		100	0	0	24,954	Lognormal
		74,525	20,776	4,698	74,429	Pareto
	acumulado	26,101	23,029	17,535		
	% demandas	56,444	24,888	18,667		
ARARUNA	N	32,450	35,085	32,464	0,721	Normal
		51,707	29,076	19,215	22,251	Lognormal
		32,077	33,485	34,437	77,026	Pareto
	acumulado	38,745	32,549	28,705		
	% demandas	35,915	32,394	31,690		
	2N	33,811	32,296	33,891	0,631	Normal
		51,286	29,202	19,511	22,157	Lognormal
		33,356	33,610	33,032	77,211	Pareto
	acumulado	39,484	31,703	28,811		
	% demandas	34,962	33,458	31,578		
	8N	33,437	33,441	33,121	0,629	Normal
		51,610	29,132	19,257	23,958	Lognormal
		33,293	33,297	33,409	75,411	Pareto
	acumulado	39,447	31,957	28,595		
	% demandas	36,667	34,250	29,083		

Tabela 6: Acumulação de dados para a Instância 10.

Estratégia	# demandas	Grupos						acumulado	Classe
		g_1	g_2	g_3	g_4	g_5	g_6		
ANDRADE	N	39,857	22,371	29,618	2,031	0	6,121	2,474	Normal
		30,201	3,061	30,341	18,058	0,591	17,745	92,618	Lognormal
		25,206	23,497	11,933	13,805	12,344	13,211	4,906	Pareto
	acumulado	31,755	16,310	23,964	11,298	4,312	12,359		
	% demandas	29,949	20,304	22,842	9,644	6,598	10,659		
	2N	0	0	31,922	16,596	21,499	29,981	2,215	Normal
		57,276	34,484	1,711	0,423	0,224	5,879	93,254	Lognormal
		7,370	1,667	19,134	33,436	20,693	17,697	4,529	Pareto
	acumulado	21,548	12,050	17,589	16,818	14,139	17,852		
	% demandas	6,069	4,046	25,433	20,231	20,809	23,410		
	8N	100	0	0	0	0	0	2,025	Normal
		1,047	39,415	41,773	1,614	5,076	11,073	94,534	Lognormal
		36,985	4,299	2,890	25,898	14,255	15,670	3,439	Pareto
	acumulado	12,677	14,571	14,887	9,171	6,443	8,914		
	% demandas	50,444	8	6,667	13,555	11,777	9,555		
RAYEE	N	23,853	2,164	0	0	0	73,981	2,646	Normal
		0	25,158	0	23,157	0	44,500	89,991	Lognormal
		0	8,331	14,063	30,341	25,709	28,737	7,361	Pareto
	acumulado	7,951	11,884	4,687	17,833	8,569	49,073		
	% demandas	57,277	6,103	9,389	7,981	8,450	10,798		
	2N	100	0	0	0	0	0	2,361	Normal
		76,575	20,440	1,317	0,948	0,718	0	92,568	Lognormal
		2,392	19,022	33,315	17,447	14,439	13,382	5,069	Pareto
	acumulado	26,322	13,154	11,544	6,131	5,052	4,460		
	% demandas	65,289	2,203	6,060	7,162	8,539	10,743		
	8N	100	0	0	0	0	0	1,884	Normal
		47,054	22,738	23,685	4,419	1,361	0,741	94,646	Lognormal
		16,205	0	4,276	25,716	27,698	26,103	3,468	Pareto
	acumulado	21,086	7,579	9,320	10,045	9,686	8,948		
	% demandas	75,777	0,222	0,888	3,111	6,444	13,555		
ARARUNA	N	20,185	18,138	20,224	21,208	0	20,242	2,545	Normal
		35,873	19,542	20,246	9,063	5,714	9,559	90,846	Lognormal
		15,593	16,177	14,468	14,582	23,665	15,512	6,607	Pareto
	acumulado	23,884	17,952	18,313	14,951	9,793	15,105		
	% demandas	22,535	16,431	18,779	17,840	2,816	21,596		
	2N	19,999	1,056	20,000	18,934	20,011	19,996	2,286	Normal
		35,528	20,398	19,976	8,908	5,698	9,490	93,054	Lognormal
		19,910	0	20,452	21,147	19,259	19,229	4,658	Pareto
	acumulado	25,146	7,151	20,143	16,330	14,989	16,238		
	% demandas	19,849	2,763	20,100	19,598	18,341	19,346		
	8N	16,604	16,458	16,733	16,733	16,735	16,734	1,648	Normal
		35,915	19,570	20,208	9,022	5,724	9,558	94,034	Lognormal
		16,739	16,701	16,641	16,686	16,615	16,615	4,316	Pareto
	acumulado	23,086	17,576	17,861	14,147	13,025	14,302		
	% demandas	16,342	17,704	17,250	16,601	15,369	16,731		

Tabela 7: Acumulação de dados para a Instância 15.

Estratégia	# demandas	Grupos						acumulado	Classe
		g_1	g_2	g_3	g_4	g_5	g_6		
ANDRADE	N	31,295	24,210	19,271	10,696	5,789	8,736	74,865	Normal
		11,086	9,312	6,208	13,082	25,277	35,033	4,973	Lognormal
		10,886	9,277	36,475	17,284	15,675	10,400	20,160	Pareto
	acumulado	17,756	14,267	20,651	13,687	15,580	18,056		
	% demandas	24,489	18,877	19,898	12,244	9,693	14,795		
	2N	33,257	21,357	25,411	7,366	4,600	8,007	73,148	Normal
		15,159	24,202	13,829	13,297	19,946	13,563	4,564	Lognormal
		9,650	23,040	10,157	23,112	18,794	15,244	22,286	Pareto
	acumulado	19,355	22,866	16,466	14,592	14,447	12,271		
	% demandas	22,471	20,505	18,539	12,640	12,921	12,921		
	8N	35,305	22,475	27,429	0	6,009	8,780	74,265	Normal
		10,012	9,770	7,478	54,161	10,253	8,323	4,015	Lognormal
		14,503	21,560	15,033	25,132	11,918	11,851	21,718	Pareto
	acumulado	19,940	17,935	16,647	26,431	9,393	9,651		
	% demandas	22,222	19,777	20,222	11,555	12,002	14,222		
RAYEE	N	23,367	21,615	22,370	12,547	10,576	9,522	73,952	Normal
		100	0	0	0	0	0	5,254	Lognormal
		100	0	0	0	0	0	20,792	Pareto
	acumulado	7,789	7,205	7,456	4,182	3,525	3,174		
	% demandas	60,891	10,891	11,386	6,435	5,445	4,950		
	2N	19,221	22,823	23,313	13,844	11,122	9,674	73,868	Normal
		100	0	0	0	0	0	4,625	Lognormal
		97,032	0	1,502	0	0	1,465	21,506	Pareto
	acumulado	38,751	7,607	8,271	4,614	3,707	3,713		
	% demandas	66,304	9,239	9,782	5,706	4,619	4,347		
	8N	25,985	26,102	27,133	15,671	5,106	0	73,584	Normal
		100	0	0	0	0	0	5,001	Lognormal
		70,890	0	1,002	0	10,870	17,236	21,413	Pareto
	acumulado	32,292	8,700	9,378	5,223	5,325	5,745		
	% demandas	66,667	8,667	9,333	5,333	4,444	5,555		
ARARUNA	N	21,416	21,407	21,393	14,506	9,681	11,595	74,335	Normal
		100	0	0	0	0	0	4,865	Lognormal
		63,172	18,881	17,945	0	0	0	20,799	Pareto
	acumulado	28,196	13,429	13,113	4,835	3,227	3,865		
	% demandas	53,465	13,861	14,356	7,425	4,950	5,940		
	2N	22,076	22,097	22,063	13,301	9,217	11,243	74,798	Normal
		100	0	0	0	0	0	4,021	Lognormal
		66,318	16,503	17,177	0	0	0	21,180	Pareto
	acumulado	29,465	12,866	13,080	4,433	3,072	3,747		
	% demandas	56,174	13,801	14,043	6,295	4,358	5,326		
	8N	21,830	21,830	21,824	13,700	9,234	11,580	73,420	Normal
		100	0	0	0	0	0	4,216	Lognormal
		63,454	18,193	18,351	0	0	0	22,363	Pareto
	acumulado	28,428	13,341	13,392	4,566	3,078	3,860		
	% demandas	56,113	14,043	13,983	6,295	4,237	5,326		

Tabela 8: Acumulação de dados para a Instância 21.

Confrontando os resultados das Tabelas 6, 7 e 8 com os da Tabela 5, verificamos que a estratégia de RAYEE mostra uma forte inclinação a concentrar o tráfego de demandas com maior largura de banda nos grupos com maiores porcentagens médias de tráfego, deixando os grupos com menor porcentagem sem nenhuma demanda atribuída. Podemos perceber isso pela quantidade de zeros nas tabelas citadas. A estratégia de ANDRADE não possui essa tendência, mas é bem mais ineficiente quanto ao número de cenários rejeitados gerados como ficou evidenciado na Tabela 5. A estratégia de ARARUNA mostra-se como um meio-termo entre essas estratégias, pois mesmo que algum tráfego ao ser distribuído concentre-se em grupos com maiores porcentagens de tráfego (o que não se mostrou frequente), apresenta um número menor de cenários rejeitados, reduzindo os pontos negativos das outras estratégias.

5 Conclusão

Neste trabalho apresentamos uma nova estratégia para o problema de distribuição de várias classes de demandas aleatórias ponderadas em grupos de tráfego com restrição mínima e máxima de capacidade, de forma a concentrar o mínimo possível demandas oriundas de uma mesma distribuição de probabilidade em algum grupo de tráfego.

Resultados numéricos mostram que a nova estratégia apresenta-se como um meio termo em relação a outras existentes para esse problema. A saber, não é tão aleatória quanto a de ANDRADE (de grande esforço computacional por violar frequentemente os limites de capacidade dos grupos de tráfego) nem tão concentradora de uma classe de tráfego em um único grupo quanto a de RAYEE.

Como contribuição prática, desenvolvemos um simulador de tráfego mais realista e de grande utilidade para o planejamento da expansão de redes dorsais de telecomunicações, que é a próxima fase (em conclusão) deste trabalho.

Agradecimentos

O presente trabalho teve apoio do CNPq (processos 300788/2006-1 e 504244/2007-8).

Referências

[Andrade et al., 2010] Andrade, R., Lisser, A., and Maculan, N. (6-9 June 2010). The design of multi-facility backbone networks under uncertain multi-classes of traffic.

[Andrade et al., 2004] Andrade, R., Lisser, A., Maculan, N., and Plateau, G. (2004). Telecommunication network capacity design for uncertain demand. *Computational Optimization and Applications*, 29(2):127–146.

[Andrade, 2002] Andrade, R. C. (2002). *Synthèse de Réseau à Demande Incertaine*. PhD thesis, Université Paris 13, Paris, France.

[Araruna, 2010] Araruna, A. R. (2010). Simulador de tráfego para a otimização de redes de telecomunicações de alta velocidade com demanda aleatória. Technical report, Universidade Federal do Ceará, Departamento de Estatística e Matemática Aplicada.

[Casilari et al., 2001] Casilari, E., Reyes-Lecuona, A., González, F., Díaz-Estrella, A., and Sandoval, F. (2001). *Characterization of Web Traffic*, pages 1862–1866. Global Communications Conference. IEEE.

[Karagiannis et al., 2004] Karagiannis, T., Molle, M., and Faloutsos, M. (2004). Long-range dependence: Ten years of internet traffic modeling. *IEEE Internet Computing*, 8:57–64.

[Mertens, 2003] Mertens, S. (2003). The easiest hard problem: number partitioning. Technical Report cond-math/0310317.