

Sobre-aprovisionamento de capacidade baseado no modelo de tráfego de Erlang

Helio Waldman¹, Rodrigo Campos Bortoletto², Michel Scassiotte Troyano¹

¹Faculdade de Engenharia Elétrica e de Computação
Universidade Estadual de Campinas (Unicamp) – Campinas – SP – Brazil

²Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP)
07115-000 – Guarulhos – SP – Brazil

waldman@decom.fee.unicamp.br, bortoletto@ifsp.edu.br,

mic0401@decom.fee.unicamp.br

Abstract. *During the last century, the Erlang model was widely accepted to represent the dynamics of resource (channels) assignment in switching networks. However, it is an asymptotic model, whose validity predicates an infinite number of sources, each of which produces an infinitesimal amount of traffic. Although this condition may closely approximate the old public telephone networks, it is questionable in the modern optical networks that form the backbone of the Internet infrastructure. This article seeks to quantify the error caused by the use of the Erlang model in the dimensioning of a single link. For this purpose, it uses the Engset model as a benchmark, with a finite number of sources, each one producing a finite amount of traffic, in a more realistic, albeit more complex setting. The article discusses the need to use recursive formulas found in the literature to express the blocking probabilities, and places these formulas as a function of the same offered traffic in the two models to provide a comparison between the required resources. The existence of overprovisioning caused by the adoption of the Erlang model is then found and assessed as a function of the number of sources.*

Resumo. *Durante o século passado, o modelo de Erlang foi amplamente aceito para representar a dinâmica de alocação de recursos (canais) em redes de comutação. Entretanto, ele é um modelo assintótico, cuja validade pressupõe um número infinito de fontes, cada uma das quais produzindo uma quantidade infinitesimal de tráfego. Embora essa condição fosse bem aproximada nas antigas redes públicas de telefonia, ela é questionável nas modernas redes ópticas que constituem a espinha dorsal da infraestrutura da Internet. Esse artigo procura quantificar o erro causado pelo uso do modelo de Erlang no dimensionamento de um enlace avulso. Para isso, ele usa como referência o modelo de Engset, que é baseado num número finito de fontes, cada uma com tráfego finito, sendo assim mais realista, ainda que mais complexo. O artigo discute a necessidade de usar fórmulas recursivas encontradas na literatura para expressar as probabilidades de bloqueio, e coloca essas fórmulas em função do tráfego oferecido nos dois modelos para propiciar a comparação entre os dimensionamentos preconizados. É então constatada e avaliada a existência do sobredimensionamento causado pela adoção do modelo de Erlang em função do número de fontes.*

1. Introdução

O lançamento de um novo sistema de comunicações envolve diversas considerações, dentre as quais se destaca o dimensionamento do sistema, que deve levar em conta o volume e a possível heterogeneidade do tráfego a ser atendido, bem como o horizonte temporal do esgotamento da sua capacidade face ao crescimento previsto para esse tráfego ao longo dos próximos anos ou décadas. Trata-se de um problema tecno-econômico, que envolve não só questões de engenharia mas também de planejamento e prospecção econômica. Mesmo assim, seu estudo requer a modelagem da interação do sistema com a demanda representada pelo tráfego oferecido à rede de comunicações, que é um problema de engenharia.

Antes do advento dos sistemas de comunicações ópticas no final do século passado, e especialmente da multiplicação em escala global das redes de fibras ópticas ocorrida na virada do século, os sistemas de comunicações eram severamente limitados pela capacidade dos meios físicos disponíveis, que eram basicamente o ar para os sistemas sem fio e os fios de cobre para a telefonia [Agrawal 2021]. Sendo o ar mais transparente que o cobre, posto que menos que o vácuo, ele era reservado para sistemas de radiodifusão e teledifusão, em que poucos emissores geravam conteúdo para ser captado pelas antenas de uma grande massa de ouvintes e telespectadores, ao passo que os sistemas com fio eram destinados às comunicações entre assinantes do serviço de telefonia, cujo número crescia lentamente (para os padrões atuais) a uma taxa em torno dos 10% ao ano, por exigir o lançamento de mais fios na rede [Singal 2016]. A demanda, porém, era muito maior, tendo em vista a crescente necessidade de comunicação entre agentes econômicos numa sociedade capitalista em franco desenvolvimento.

A partir da virada do século, essa situação sofreu uma reviravolta, com a disseminação de um novo meio físico (a fibra óptica) muito mais transparente que o ar, tornando obsoleta a segregação entre os sistemas físicos de suporte à radiodifusão por um lado, e à telefonia pelo outro [Agrawal 2021, Singal 2016]. Porém, como a fibra óptica é um fio, ela não tem a mesma capilaridade que o ar, que está em toda parte. Daí a necessidade de integração entre a fibra, que fornece sua inigualável capacidade, com o ar, que continua fornecendo a sua capilaridade, sempre superior a qualquer sistema com fio.

Nos sistemas com fio, os recursos devem ser dimensionados para atender o tráfego agregado gerado por um certo número de fontes de maneira que a probabilidade de bloqueio de uma chamada seja inferior a um valor especificado no projeto. Quando uma fonte está ocupada por uma chamada, ela não gera novas chamadas, fazendo com que o tráfego gerado por ela não seja Poissoniano. Porém, quando o número de fontes tende a infinito e a taxa de requisições de cada fonte tende a zero gerando uma taxa agregada finita, o tráfego se aproxima assintoticamente do modelo Poissoniano frequentemente adotado na literatura. Essa é a hipótese basilar do modelo de Erlang para a probabilidade de bloqueio dos sistemas de comutação de circuitos [Kumar et al. 2004]. Como essa hipótese era razoavelmente bem aproximada nas redes públicas de telefonia do século passado, a adoção do modelo de Erlang tornou-se então a base das práticas de planejamento e dimensionamento das redes com fio na época.

Em sistemas ópticos de comutação de circuitos, as fontes de tráfego requisitam recursos espectrais das fibras (“*slots*”) para dar suporte a chamadas. Assim, os novos “fios” não são mais fios de cobre, mas sim fatias do espectro óptico, que entretanto de-

semprenham o mesmo papel sistêmico. Mesmo assim, o modelo de Erlang permanece popular, dada a sua simplicidade. Porém, os pressupostos desse modelo, de que há um número muito grande de fontes que permanecem inativas a maior parte do tempo, passaram a ser bastante questionáveis. Assim, o modelo mais realista compatível com a abordagem analítica passa a ser o modelo de Engset, que se baseia num número finito de fontes com atividade frequente. Este trabalho tem a finalidade de avaliar, em função do número de fontes, o sobre-aprovisionamento de recursos espectrais de um enlace avulso (“*single link*”) gerado pelo uso do modelo de Erlang no lugar do modelo de Engset.

O artigo está organizado da seguinte forma. Na Seção 2, são discutidos alguns trabalhos relevantes para a temática abordada. Na Seção 3, são discutidos os modelos analíticos de Erlang e de Engset. Na Seção 4, são apresentados resultados numéricos do cálculo de sobre-aprovisionamento de recursos gerado pela adoção do modelo de Erlang em função do número de fontes.

2. Trabalhos Relacionados

O modelo de Erlang foi proposto inicialmente em 1917 pelo matemático dinamarquês Agner Krarup Erlang (1878-1929) [Erlang 1917], a quem é também atribuída a origem das Teorias das Filas e do Teletráfego. No ano seguinte, o matemático norueguês Tore Olaus Engset (1865-1943) publicou o modelo que leva o seu nome [Engset 1918], que já constava de um relatório apresentado em 1915 à Companhia Telefônica de Copenhague, redescoberto em 1995 [Myskja 1998].

Desde então e até hoje, esses dois modelos vêm sendo extensamente utilizados e refinados sob diferentes pressupostos referentes ao tráfego oferecido e, no caso do modelo de Engset, às funcionalidades das fontes. Uma apresentação bem contextualizada, referenciada e aprofundada dos principais resultados obtidos está disponível em [Zukerman 2021].

O modelo de Erlang pode ser considerado um limite assintótico do modelo de Engset quando o número de fontes é levado ao infinito e a quantidade de tráfego gerada por cada fonte é levada a zero de tal maneira que o tráfego oferecido ao sistema se mantém finito. Por isso, as fórmulas que expressam os cálculos da probabilidade de bloqueio pelos dois modelos guardam alguma semelhança, e compartilham os mesmos problemas numéricos quando utilizadas em cálculos. Esses problemas estão associados à presença de fatoriais de números que podem ser muito grandes, dependendo do tamanho do sistema. Para contornar esses problemas, foram derivadas fórmulas recursivas exatas por alguns autores, que são extensamente discutidas em [Iversen 2015]. Essas fórmulas são usadas nesse artigo e, no caso do modelo de Engset, estendidas para propiciar a comparação entre os dois modelos.

3. Formalização do Problema

O uso do modelo de tráfego de Erlang para dimensionar redes de comunicação foi disseminado durante o século passado em associação com a expansão das redes públicas de telefonia. Os terminais dessas redes eram um grande número de telefones que permaneciam ociosos a maior parte do tempo, conectados a uma central telefônica. Para se comunicarem com telefones ligados a outras centrais, os M telefones servidos por uma central, ao fazer uma chamada, deviam disputar acesso a um dos k canais de um tronco.

Sendo $k < M$, havia a possibilidade de bloqueio da chamada. Assim, havia a necessidade de dimensionar k para que a probabilidade desse bloqueio pela rede fosse suficientemente pequena em comparação com a probabilidade do número chamado estar ocupado.

Na rede telefônica, os telefones atuavam como fontes de tráfego. Quando uma fonte está ativa ou ocupada, ele não gera novas chamadas. Só quando está livre é que ela gera novas chamadas com uma taxa $\hat{\lambda}$, que suporemos igual para todos as fontes para maior facilidade analítica. Assim, quando m fontes estão ativas no tronco, portanto ocupando m dos k canais, o tráfego gerado pelas M fontes é $(M - m) \hat{\lambda}$. A Figura 1 mostra uma cadeia markoviana gerada por esse tipo de tráfego, frequentemente chamado de tráfego liga-desliga (“on-off”) num sistema com $k < M$ canais. Ela representa o chamado modelo de Engset do sistema, no qual a taxa de chegada de novas requisições depende do estado m em que se encontra o sistema markoviano. O estado do sistema é caracterizado apenas pelo número $m \in \{0, 1, 2, \dots, k\}$ de canais ativos.

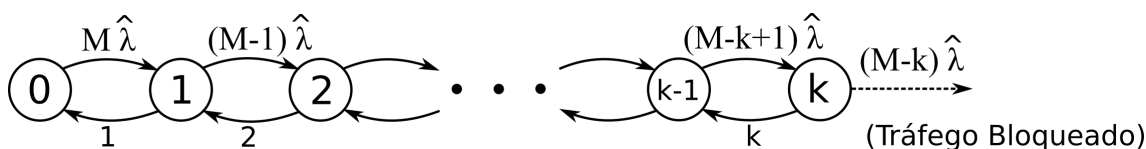


Figura 1. Representação do Modelo de Engset

Caso $M \gg k \geq m$, a expressão $(M - m) \hat{\lambda}$ pode ser aproximada para $M \hat{\lambda}$. A aproximação de $(M - m) \hat{\lambda}$ por $M \hat{\lambda}$ é a base do modelo de Erlang, que descreve um sistema no qual a taxa de chegada de novas chamadas independe do número de canais ativos $m \leq k$, como mostra a Figura 2. A taxa de chegadas de novas requisições de canal é constante, portanto independe do estado do sistema, e é denotada por λ . Observe que o modelo de Erlang ignora o número de fontes M e a taxa de requisições por fonte $\hat{\lambda}$, mas toma por base apenas o produto entre elas, que é aceito como uma taxa λ independente de estado e identificada com o tráfego oferecido ao sistema.

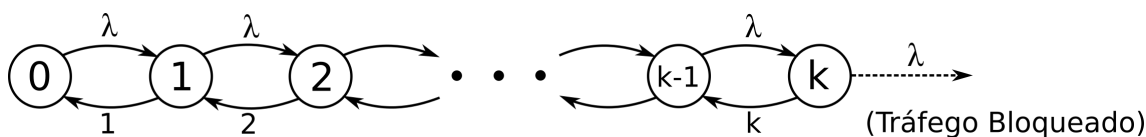


Figura 2. Representação do Modelo de Erlang

A condição $M \gg k \geq m$ era válida nas antigas redes públicas de telefonia, pois o número de fontes era muito maior que o de canais disponíveis para longas distâncias, e o tempo médio entre chamadas por fonte, igual a $\frac{1}{\hat{\lambda}}$, era muito maior que a duração média de um telefonema (da ordem de 3 minutos, pois a tarifa era cobrada por minuto), fazendo com que $\hat{\lambda} \ll 1$ quando expresso em Erlangs [Angus 2001]. Além disso, o modelo de Erlang é mais simples que o de Engset, por não depender da variável M . Essas circunstâncias levaram a uma grande disseminação do uso do modelo de Erlang durante o século XX, em associação com o desenvolvimento da Teoria do Tráfego e em especial com o dimensionamento de sistemas, ou seja, com a determinação da quantidade de recursos físicos (fios, fibras, comprimentos de onda, slots espectrais, etc.) necessários para garantir um certo grau de desempenho.

Quando transportado para a análise de sistemas modernos, especialmente no contexto da espinha dorsal (“*backbone*”) da infraestrutura física da Internet, o uso do modelo de Erlang pode ser inadequado, pois seus pressupostos podem não ser válidos. Para quantificar essa inadequação, vamos comparar os dois modelos com base nas probabilidades de bloqueio P_{br} e $P_{bn}(M)$ previstas respectivamente pelos modelos de Erlang e de Engset a partir do número de canais k e da taxa observada de requisições λ . Dada uma probabilidade de bloqueio P_b considerada aceitável pela rede, obteremos o dimensionamento de Erlang pelo menor número de canais k_r que faz $P_{br} \leq P_b$; e o dimensionamento de Engset, para cada número de fontes M , pelo menor número de canais $k_n(M)$ que faz $P_{bn}(M) \leq P_b$. Em termos percentuais, o sobre-aprovisionamento de recursos espectrais gerado pelo dimensionamento segundo o modelo de Erlang será então dado por:

$$\chi(M) = \left\{ \frac{[k_r - k_n(M)]}{k_n(M)} \right\} \cdot 100\% \quad (1)$$

para cada número M de fontes.

3.1. Dimensionamento segundo Erlang

A Figura 2 mostra um sistema markoviano cujos estados são identificados pelo número m de canais ocupados. Em cada estado m exceto no estado k , há uma transição de m para $(m + 1)$ com taxa λ independente de m . No estado k , essa transição não existe porque não existem $(k + 1)$ canais. Por isso, as chamadas que ocorrem quando o sistema está no estado k são bloqueadas. Porém, as requisições por chamadas continuam chegando com taxa λ , de maneira que a taxa de bloqueios é λ enquanto o sistema está no estado k e zero nos demais estados. Assim sendo, a taxa média de bloqueios ao longo do tempo é $(\lambda \cdot p_k)$, onde p_m é a probabilidade do sistema se encontrar no estado m .

A probabilidade de bloqueio do sistema é a razão entre a taxa média de chamadas bloqueadas e a taxa média de requisições recebidas. A taxa média de chamadas bloqueadas é $(\lambda \cdot p_k)$, como visto acima. A taxa de requisições é λ em todos os estados, pela hipótese básica do modelo. Então, a probabilidade de bloqueio do modelo de Erlang é:

$$P_{br} = \frac{(\lambda \cdot p_k)}{\lambda} = p_k \quad (2)$$

A equação 2 diz que a probabilidade de bloqueio segundo o modelo de Erlang é a probabilidade do sistema estar no estado k . Como as chegadas são Poissonianas, essa conclusão decorre diretamente da chamada propriedade PASTA (**P**oisson **A**rrivals **S**ee **T**ime **A**verages).

Na Figura 2, observa-se que a supressão das transições entre dois estados que se comunicam divide a cadeia em duas partes que não se comunicam, o que caracteriza a cadeia como uma árvore [Kumar et al. 2004]. Se tomarmos dois estados i e j que se comunicam numa árvore com taxas λ_{ij} de i para j e λ_{ji} de j para i , então teremos, em situações de equilíbrio estocástico:

$$p_i \lambda_{ij} = p_j \lambda_{ji} \quad (3)$$

Aplicando essa equação ao caso do modelo de Erlang mostrado na Figura 2, tem-se $p_0\lambda = p_1$, $p_1\lambda = 2p_2$, $p_2\lambda = 3p_3$, \dots , $p_{k-2}\lambda = (k-1)p_{k-1}$, $p_{k-1}\lambda = kp_k$, e portanto:

$$\begin{aligned} p_i &= \left(\frac{\lambda}{i}\right) p_{(i-1)} = \left(\frac{\lambda}{i}\right) \left(\frac{\lambda}{i-1}\right) p_{(i-2)} = \\ &= \left(\frac{\lambda}{i}\right) \left(\frac{\lambda}{i-1}\right) \left(\frac{\lambda}{i-2}\right) p_{(i-3)} = \dots = \left(\frac{\lambda^i}{i!}\right) p_0, \quad \forall i \in \{1, 2, \dots, k\} \end{aligned} \quad (4)$$

Assim sendo, é possível escrever:

$$p_i = \left(\frac{\lambda^i}{i!}\right) p_0, \quad \forall i \in \{0, 1, \dots, k\}, \quad (5)$$

onde o caso $i = 0$ pode ser acrescentado por ser uma identidade.

Como $\sum_{i=0}^k p_i = 1$, pode-se então obter o valor de p_0 a partir da Equação 5 fazendo:

$$\begin{aligned} \sum_{i=0}^k p_i &= p_0 \cdot \sum_{i=0}^k \left(\frac{\lambda^i}{i!}\right) = 1 \\ \therefore p_0 &= \frac{1}{\sum_{i=0}^k \left(\frac{\lambda^i}{i!}\right)} \end{aligned} \quad (6)$$

Inserindo a Equação 6 na Equação 5 e aplicando na Equação 2 para obter a probabilidade de bloqueio para o modelo simples de Erlang, tem-se:

$$P_{br}(k) = \frac{\left(\frac{\lambda^k}{k!}\right)}{\sum_{i=0}^k \left(\frac{\lambda^i}{i!}\right)} \quad (7)$$

Embora simples do ponto de vista analítico, essa expressão pode não ser conveniente do ponto de vista numérico, pois usa operações fatoriais e exponenciações que podem ser muito grandes quando os operandos k e λ atingem valores de apenas algumas dezenas, como pode ocorrer em casos de interesse prático. Porém, esse problema pode ser contornado com o uso de fórmulas recursivas exatas encontradas na literatura, das quais apresenta-se a seguir a que faz a recursão em k .

Para esse fim, aplica-se a Equação 7 para obter $P_{br}(k-1)$:

$$P_{br}(k-1) = \frac{\left(\frac{\lambda^{k-1}}{(k-1)!}\right)}{\sum_{i=0}^{k-1} \left(\frac{\lambda^i}{i!}\right)} \quad (8)$$

Dividindo a Equação 7 pela Equação 8, obtém-se:

$$\frac{P_{br}(k)}{P_{br}(k-1)} = \frac{\left(\frac{\lambda}{k}\right)}{1 + \frac{\frac{\lambda^k}{k!}}{\sum_{i=0}^{k-1} \left(\frac{\lambda^i}{i!}\right)}} = \frac{\lambda}{k + \lambda \left[\frac{\frac{\lambda^{k-1}}{(k-1)!}}{\sum_{i=0}^{k-1} \left(\frac{\lambda^i}{i!}\right)} \right]} \quad (9)$$

Reconhecendo, com base na Equação 8, que o termo entre colchetes é $P_{br}(k-1)$, obtem-se a seguinte expressão recursiva:

$$P_{br}(k) = \frac{\lambda P_{br}(k-1)}{k + \lambda P_{br}(k-1)} \quad (10)$$

Quando $k = 0$, não há possibilidade de transmissão, portanto:

$$P_{br}(0) = 1 \quad (11)$$

Então, a recursão começa com a Equação 11, e em seguida usa-se a Equação 10 para calcular $P_{br}(1)$ a partir de $P_{br}(0)$, $P_{br}(2)$ a partir de $P_{br}(1)$, e assim por diante até $P_{br}(k)$ a partir de $P_{br}(k-1)$ para o valor desejado de k . No caso do dimensionamento, o valor desejado é o argumento do primeiro valor de $P_{br}(k)$ que resulte inferior à probabilidade P_b considerada tolerável pelo projeto.

3.2. Dimensionamento segundo Engset

Quando $M \leq k$, é possível alocar um canal permanentemente para servir cada uma das M fontes, zerando a probabilidade de bloqueio para qualquer taxa $\hat{\lambda}$ de chegada de requisições por fonte. Assim, com cada fonte podendo dispor de um canal privativo para as suas chamadas, não há mais bloqueios.

Quando $M > k$, o dimensionamento baseado no modelo de Engset é obtido a partir do sistema markoviano mostrado na Figura 1, no qual a intensidade do tráfego depende do estado em que o sistema se encontra. No único estado bloqueante, que continua sendo o estado k , a taxa de chegada de requisições é $\hat{\lambda}(M-k)$. Então, a taxa de bloqueios é dada por $\hat{\lambda}(M-k)p_k$. Porém, a taxa média de requisições é dada pela média da taxa de chegada em cada estado ponderada pela probabilidade de cada um:

$$\lambda = \sum_{i=0}^k p_i \cdot \hat{\lambda} \cdot (M-i) \quad (12)$$

Dividindo então a taxa de bloqueios pela taxa de chegadas de requisições, obtem-se a probabilidade de bloqueios no modelo de Engset:

$$P_{bn}(M, k) = \frac{\hat{\lambda}(M-k)p_k}{\sum_{i=0}^k p_i \cdot \hat{\lambda} \cdot (M-i)} = \frac{(M-k)p_k}{\sum_{i=0}^k p_i \cdot (M-i)} \quad (13)$$

Na Figura 1, observa-se que a taxa de transição de um estado i para $(i+1)$ é dada por $\hat{\lambda}(M-i) = \lambda_{i,i+1}$, ao passo que a taxa do estado $(i+1)$ para i é $(i+1) = \lambda_{i+1,i}$. A inserção dessas relações na Equação 3 fornece então a seguinte relação recursiva entre probabilidades de estados sucessivos ao percorrer a cadeia markoviana da Figura 1:

$$p_i \cdot i = p_{i-1} \hat{\lambda} (M-i+1), \quad \forall i \in \{1, 2, \dots, k\}, M > k \quad (14)$$

Aplicando essa equação recursivamente ao caso do modelo de Engset mostrado na Figura 1, tem-se então:

$$\begin{aligned}
p_i &= \left(\frac{\hat{\lambda}(M-i+1)}{i} \right) p_{i-1} = \left(\frac{\hat{\lambda}(M-i+1)}{i} \right) \left(\frac{\hat{\lambda}(M-i+2)}{i-1} \right) p_{i-2} = \\
&= \left(\frac{\hat{\lambda}(M-i+1)}{i} \right) \left(\frac{\hat{\lambda}(M-i+2)}{i-1} \right) \left(\frac{\hat{\lambda}(M-i+3)}{i-2} \right) p_{i-3} = \dots = \\
&= \left(\frac{\hat{\lambda}(M-i+1)}{i} \right) \left(\frac{\hat{\lambda}(M-i+2)}{i-1} \right) \dots \left(\frac{\hat{\lambda}(M-1)}{2} \right) (\hat{\lambda}M) p_0 = \\
&= \frac{\hat{\lambda}^i M!}{i! (M-i)!} p_0 = \binom{M}{i} \hat{\lambda}^i p_0 \quad (15)
\end{aligned}$$

Como $\sum_{i=0}^k p_i = p_0 \sum_{i=0}^k \binom{M}{i} \hat{\lambda}^i = 1$, resulta que:

$$p_0 = \frac{1}{\sum_{i=0}^k \binom{M}{i} \hat{\lambda}^i} \quad (16)$$

Inserindo a Equação 16 em 15, tem-se:

$$p_i = \frac{\binom{M}{i} \hat{\lambda}^i}{\sum_{j=0}^k \binom{M}{j} \hat{\lambda}^j} \quad (17)$$

Finalmente, a Equação 17 pode ser usada na Equação 13 para obter a seguinte fórmula para a probabilidade de bloqueio no modelo de Engset:

$$P_{bn}(M, k) = \frac{(M-k) \binom{M}{k} \hat{\lambda}^k}{\sum_{j=0}^k (M-j) \binom{M}{j} \hat{\lambda}^j}, \quad M > k \quad (18)$$

Por oportuno, observa-se que:

$$(M-j) \binom{M}{j} = \frac{[(M-j) M!]}{(j! (M-j)!)} = M \left[\frac{(M-1)!}{j! (M-j-1)!} \right] = M \binom{M-1}{j} \quad (19)$$

Então, a Equação 18 pode ser simplificada para:

$$P_{bn}(M, k) = \frac{\binom{M-1}{k} \hat{\lambda}^k}{\sum_{j=0}^k \binom{M-1}{j} \hat{\lambda}^j} \quad (20)$$

Embora simples, essa expressão pode não ser adequada para a realização de cálculos numéricos, pois os termos combinatórios envolvem fatoriais de M e expoentes de $\hat{\lambda}$, que podem ser números extremamente grandes quando essas variáveis tomam valores de apenas algumas dezenas, como seria importante considerar em casos

práticos. Por isso, algumas fórmulas recursivas exatas têm sido apresentadas na literatura [Ross and Tsang 1989] para contornar esse problema, sendo possível realizar a recursão em M , em k , ou em M e k [Iversen 2015]. Será discutida aqui apenas a recursão em k . Para isso, inicia-se reescrevendo a Equação 20 para expressar $P_{bn}(M, k - 1)$:

$$P_{bn}(M, k - 1) = \frac{\binom{M-1}{k-1} \hat{\lambda}^{k-1}}{\sum_{j=0}^{k-1} \binom{M-1}{j} \hat{\lambda}^j} \quad (21)$$

A recursão será baseada na divisão da Equação 20 pela Equação 21, onde surge o seguinte termo:

$$\frac{\binom{M-1}{k}}{\binom{M-1}{k-1}} = \frac{(M-1)! (k-1)! (M-k)!}{(M-1)! k! (M-1-k)!} = \frac{M-k}{k} \quad (22)$$

Usando essa identidade na divisão da Equação 20 pela Equação 21, tem-se:

$$\frac{P_{bn}(M, k)}{P_{bn}(M, k-1)} = \frac{\binom{M-k}{k} \cdot \hat{\lambda}}{1 + \binom{M-k}{k} \cdot \hat{\lambda} \cdot \left[\frac{\binom{M-1}{k-1} \hat{\lambda}^{k-1}}{\sum_{j=0}^{k-1} \binom{M-1}{j} \hat{\lambda}^j} \right]} \quad (23)$$

Finalmente, multiplicando numerador e denominador por k e reconhecendo que o termo entre colchetes é $P_{bn}(M, k - 1)$, obtem-se a expressão recursiva desejada:

$$P_{bn}(M, k) = \frac{\hat{\lambda} (M - k) P_{bn}(M, k - 1)}{k + \hat{\lambda} (M - k) P_{bn}(M, k - 1)} \quad (24)$$

A recursão começa com o valor de $P_{bn}(M, 0)$:

$$P_{bn}(M, 0) = 1 \quad (25)$$

que é justificado pelo fato de que, quando não há recursos para transmitir (ou seja, quando $k = 0$), então a probabilidade de bloqueio é 1.

A partir da Equação 25, usa-se a Equação 24 para calcular $P_{bn}(M, 1)$ a partir de $P_{bn}(M, 0)$, $P_{bn}(M, 2)$ a partir de $P_{bn}(M, 1)$, e assim por diante, até o valor de $P_{bn}(M, k)$ para o valor desejado de k . Para o cálculo do dimensionamento, o valor desejado é o primeiro que resultar inferior ao da probabilidade de bloqueio P_b considerada tolerável pelo projeto.

Infelizmente, a Equação 24, que pode ser encontrada na literatura [Zukerman 2021], expressa $P_{bn}(M, k)$ em função da taxa de requisições de cada fonte durante sua inatividade, que não é observável pelo sistema nem é a variável usada na Equação 6 para expressar a probabilidade de bloqueio P_{br} prevista pelo modelo de Erlang. Por isso, para fazer a comparação desejada entre os dois modelos, é necessário obter a relação entre a taxa do tráfego oferecido ao sistema pelas M fontes e a taxa $\hat{\lambda}$ de requisições de cada fonte durante sua inatividade.

Para esse fim, considera-se inicialmente o tráfego oferecido por cada fonte individualmente ao sistema. Cada fonte alterna períodos de atividade e de inatividade ao longo do tempo, conforme mostrado na Figura 3. A duração média dos períodos de atividade é a duração média das chamadas, que é adotada como unidade de tempo, portanto igual a 1. Durante a atividade (chamada), o número de novas chamadas pela fonte ativa é zero. Já durante a inatividade, a taxa de chegada de novas chamadas é $\hat{\lambda}$, de maneira que o tempo médio de espera pela próxima chamada é $\frac{1}{\hat{\lambda}}$. Conclui-se então que o tempo médio entre chamadas sucessivas da mesma fonte é $\left(1 + \frac{1}{\hat{\lambda}}\right)$, e portanto a taxa com que cada fonte emite novas requisições é $\frac{1}{\left(1 + \frac{1}{\hat{\lambda}}\right)} = \frac{\hat{\lambda}}{(1 + \hat{\lambda})}$. Isso significa que o tráfego oferecido por cada fonte ao sistema é sempre menor que 1, e só se aproxima de 1 quando $\hat{\lambda}$ tende ao infinito. Para obter o tráfego λ oferecido pelas M fontes ao sistema, basta então multiplicar por M o tráfego gerado por cada uma:

$$\lambda = \frac{M\hat{\lambda}}{1 + \hat{\lambda}} \quad (26)$$

$$\therefore \hat{\lambda} = \frac{\lambda}{M - \lambda}, \quad 0 \leq \lambda < M \quad (27)$$

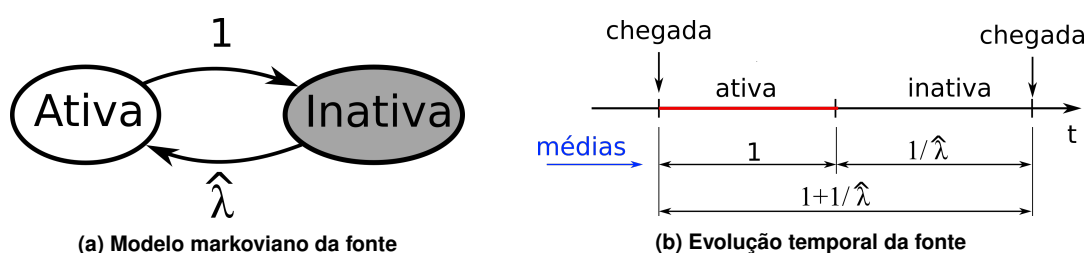


Figura 3. Modelo da fonte de tráfego

É importante observar que a Equação 27 não se aplica para $\lambda > M$, o que é natural, pois se cada fonte contribui com menos que 1 para o tráfego oferecido, então as M fontes só poderão gerar uma taxa de requisições inferior a M . Como o modelo de Erlang não tem essa limitação, pode-se perguntar então como ficaria a comparação entre os dimensionamentos gerados pelos dois modelos para $\lambda > M$. A resposta é que não é possível realizar a comparação, pois não se pode comparar dois modelos para uma situação considerada impossível por um deles. Do ponto de vista do modelo de Engset, se for observado um valor de λ superior à estimativa de M , então essa estimativa deve ser corrigida para algum valor superior a λ . A discrepância naturalmente desaparece quando M tende a infinito, única situação em que os dois modelos convergem.

Inserindo então a Equação 27 na Equação 24, obtém-se uma expressão recursiva baseada na taxa λ observada de requisições feitas ao sistema:

$$P_{bn}(M, k) = \frac{\lambda(M - k) P_{bn}(M, k - 1)}{k(M - \lambda) + \lambda(M - k) P_{bn}(M, k - 1)}, \quad 0 \leq \lambda < M \quad (28)$$

3.2.1. Estimativa do número de fontes

O cálculo do dimensionamento adequado pelo modelo de Engset depende do conhecimento do número de fontes habilitadas para gerar tráfego, que não necessariamente é conhecido. Daí a possível necessidade de estimar M a partir das características do tráfego.

A Figura 3 deixa claro que o intervalo entre chegadas sucessivas do tráfego gerado por uma fonte é dado pela soma de duas variáveis exponenciais independentes, quais sejam:

- a duração da **chamada** cuja distribuição tem média 1, sendo assim com densidade de probabilidade dada por $p_{cham}(t) = e^{-t}$; e
- o tempo de **espera** pela nova chamada após o término da anterior, cuja densidade de probabilidade, para uma fonte markoviana, é dada por $p_{espe}(t) = \hat{\lambda}e^{-\hat{\lambda}t}$.

Como essas duas variáveis são independentes e sua soma é o tempo entre **chegadas** sucessivas, esse tempo tem como densidade de probabilidade a convolução das densidades de probabilidade das duas parcelas:

$$p_{cheg}(t) = \int_0^t p_{cham}(\tau) p_{espe}(t - \tau) d\tau = \int_0^t e^{-\tau} \hat{\lambda} e^{-\hat{\lambda}(t-\tau)} d\tau = \hat{\lambda} e^{-\hat{\lambda}t} \int_0^t e^{(\hat{\lambda}-1)\tau} d\tau \quad (29)$$

Caso $\hat{\lambda}$ seja igual a 1, tem-se então:

$$p_{cheg}(t) = te^{-t}, \quad se \quad \hat{\lambda} = 1, \quad (30)$$

o que caracteriza a densidade de uma distribuição gama mostrada na Figura 4.

Se $\hat{\lambda} > 1$, tem-se da Equação 29 que:

$$p_{cheg}(t) = \hat{\lambda} e^{-\hat{\lambda}t} \left[\frac{e^{(\hat{\lambda}-1)t} - 1}{\hat{\lambda} - 1} \right] = \left(\frac{\hat{\lambda}}{\hat{\lambda} - 1} \right) (e^{-t} - e^{-\hat{\lambda}t}), \quad se \quad \hat{\lambda} > 1 \quad (31)$$

Finalmente, para $0 < \hat{\lambda} < 1$, obtém-se da Equação 29:

$$p_{cheg}(t) = \hat{\lambda} e^{-\hat{\lambda}t} \left[\frac{1 - e^{(\hat{\lambda}-1)t}}{1 - \hat{\lambda}} \right] = \left(\frac{\hat{\lambda}}{1 - \hat{\lambda}} \right) (e^{-\hat{\lambda}t} - e^{-t}), \quad se \quad \hat{\lambda} < 1 \quad (32)$$

A Figura 4 mostra a densidade de probabilidade do tempo entre chegadas sucessivas para $\hat{\lambda} = 1$ com base na Equação 30, para $\hat{\lambda} = 2$ e para $\hat{\lambda} = 5$ com base na Equação 31, e para $\hat{\lambda} = 0.5$ com base na Equação 32. Havendo a possibilidade de identificar a fonte de cada chamada durante a operação, pode-se então analisar a similitude da distribuição média do tempo entre chamadas sucessivas da mesma fonte com a família de distribuições mostrada na Figura para estimar $\hat{\lambda}$, e inseri-lo na Equação 26 para estimar o número de fontes M .

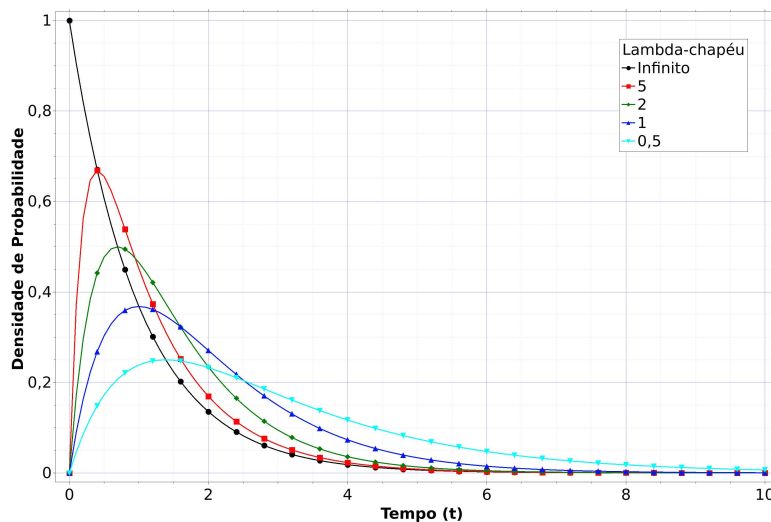


Figura 4. Densidade de probabilidade do tempo entre chegadas sucessivas da mesma fonte.

4. Resultados

A Figura 5 apresenta o percentual de sobre-provisionamento de recursos espectrais em função do número de fontes para uma intensidade de tráfego oferecido $\lambda = 100$ Erlang, para probabilidades de bloqueio de 10^{-2} a 10^{-5} . Como $\lambda < M$ no modelo de Engset, a comparação só faz sentido para $M \geq 101$, sendo que os maiores excessos de provisionamento ocorrem justamente quando M mais se aproxima de λ . Com o crescimento do número de fontes, os dois modelos convergem, de maneira que o sobre-provisionamento tende a zero.

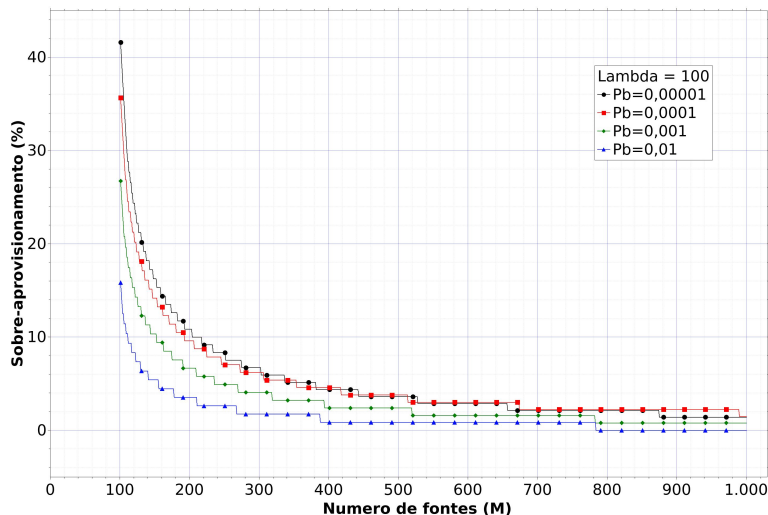


Figura 5. Sobre-provisionamento de recursos espectrais em função do número de fontes (M).

A Figura 6 mostra a evolução das probabilidades de bloqueio sob os dois modelos quando o tráfego oferecido varia de zero a 200 Erlang, para uma probabilidade de bloqueio tolerável de 0,0001. No começo (até $\lambda \cong 70$), as duas probabilidades de bloqueio crescem com λ até atingir a probabilidade de bloqueio de referência 0,0001, quando então elas

caem abruptamente devido ao aumento dos respectivos dimensionamentos k_r e k_n , que são acrescidos de uma unidade a cada descontinuidade. Porém, a partir de $\lambda \cong 90$, a probabilidade de bloqueio deixa de crescer com λ no modelo de Engset, chegando a rigorosamente zero para $\lambda = M = 200$, quando então é possível alocar um canal a cada fonte, sem a necessidade de bloquear nenhuma requisição; ao passo que no modelo de Erlang a probabilidade de bloqueio P_b continua a crescer com λ , decrescendo em pontos discretos com o acréscimo de mais unidades de dimensionamento, por supor a existência de infinitas fontes. Assim, com a continuidade do crescimento de k_r enquanto cessa o crescimento de k_n , há uma aceleração do aumento do sobre-provisionamento com o tráfego oferecido λ , conforme pode ser observado na Figura 7.

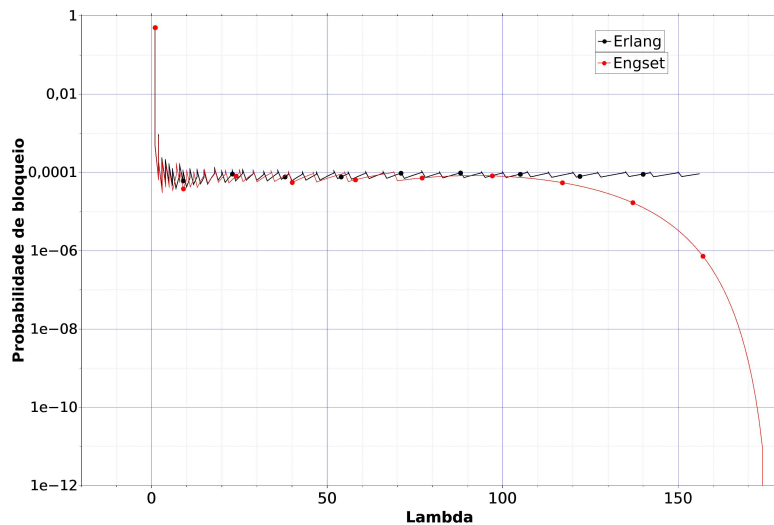


Figura 6. Probabilidades de bloqueio limitadas a 0,0001 pelos dois modelos considerados.

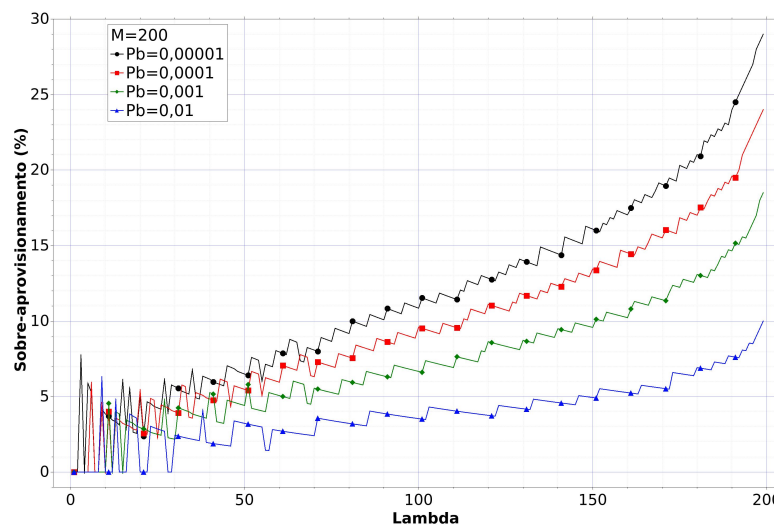


Figura 7. Sobre-provisionamento de recursos espectrais em função do Lambda para $M = 200$.

5. Conclusões

O uso do modelo de Erlang no dimensionamento de redes pode causar um sobre-provisionamento significativo, que pode chegar a dezenas de pontos percentuais. Para

evitar isso, é recomendável utilizar o modelo de Engset, que leva em conta o número finito de fontes. Quando esse número não é conhecido, ele pode ser estimado a partir do tráfego se for possível identificar a fonte de cada chegada, para que se possa estimar o intervalo médio entre chegadas sucessivas da mesma fonte. Além disso, a taxa máxima de chegadas em Erlang ao longo do histórico da rede pode ser considerada um limitante inferior do número de fontes, igualando-o se e quando cada fonte estiver enviando uma nova requisição imediatamente após o término da conexão anterior.

Quando o número de fontes é muito grande, os dois modelos convergem, e por isso o uso do modelo de Erlang no lugar do de Engset se justifica pela sua maior simplicidade. Esse foi o caso das redes públicas de telefonia no século XX. A partir da virada do século, o uso continuado do modelo de Erlang no dimensionamento das redes ópticas passa a se justificar pela sub-utilização do espectro óptico das fibras, gerando tolerância ao sobre-provisionamento de um recurso abundante (“commodity”). Atualmente, porém, há uma clara busca de mais eficiência no uso do espectro óptico, que se traduz no advento de redes ópticas elásticas, de esquemas de sobrevivência com compressão da banda durante eventos emergenciais, etc.. Assim, parece oportuna a adoção do modelo de Engset no planejamento da rede, a fim de evitar o sobre-provisionamento de recursos espectrais escassos.

Agradecimentos

Os autores agradecem o suporte do projeto da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) Proc. 2015/24341-7, e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Proc. 304171/2017-4.

Referências

- Agrawal, G. P. (2021). *Fiber-optic communication systems*. John Wiley & Son, 5 edition.
- Angus, I. (2001). An introduction to erlang b and erlang c. *Telemanagement*, 187:6–8.
- Engset, T. (1918). Die wahrscheinlichkeitsrechnung zur bestimmung der wähleranzahl in automatischen fernsprechamtern. *Elektrotechnische zeitschrift*, 39(31):304–306.
- Erlang, A. K. (1917). Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Post Office Electrical Engineer's Journal*, 10:189–197.
- Iversen, V. B. (2015). *Teletraffic engineering and network planning*. Technical University of Denmark - DTU Fotonik.
- Kumar, A., Manjunath, D., and Kuri, J. (2004). *Communication networking: an analytical approach*. Elsevier.
- Myskja, A. (1998). The engset report of 1915-summary and comments. *TELEKTRO-NIKK*, 94:143–153.
- Ross, K. W. and Tsang, D. H. (1989). The stochastic knapsack problem. *IEEE Transactions on communications*, 37(7):740–747.
- Singal, T. L. (2016). *Optical fiber communications: principles and applications*. Cambridge University Press.
- Zukerman, M. (2021). *Introduction to queueing theory and stochastic teletraffic models*, EE Department. City University of Hong Kong.