Desempenho de Redes com Comutação de Circuitos

Desempenho e Dimensionamento de Redes

Prof. Amaro de Sousa (asou@ua.pt)

DETI-UA, 2017/2018

Encaminhamento em redes com comutação de circuitos

1. Encaminhamento *fixo*: considera um único percurso para cada fluxo de chamadas suportado pela rede.

2. Encaminhamento *alternativo*: considera uma sequência ordenada de percursos alternativos para cada fluxo (estabelece no *n*-ésimo percurso se nenhum dos percursos até ao (*n*-1)-ésimo tiver recursos).

3. Encaminhamento *dinâmico*: considera uma sequência ordenada de percursos alternativos para cada fluxo e essa sequência varia ao longo do tempo.

    - Numa rede em malha completa, o percurso *direto* é o percurso com uma ligação, que interliga o nó origem e o nó destino.

    - É preferível encaminhar uma chamada pelo percurso direto (os percursos alternativos consomem mais recursos).

    - *Trunk reservation*: Reserva de um conjunto de circuitos em cada ligação para chamadas no percurso direto (limita o excesso de encaminhamento alternativo).

Encaminhamento fixo

Vamos abordar 3 métodos para o cálculo das probabilidades de bloqueio de cada fluxo de chamadas

- Método exato
  - Computacionalmente pesado
  - Exige identificação de todos os estados possíveis da rede
- Teorema do limite do produto
  - Computacionalmente leve
  - É um majorante da probabilidade de bloqueio exata
- Aproximação de carga reduzida
  - Uma aproximação (normalmente boa) dos valores exatos
  - Matematicamente complexo
  - Existem algoritmos iterativos de cálculo

Encaminhamento fixo – Método Exato

- Considere-se uma rede com J ligações que serve K fluxos de chamadas.
- A ligação $j = 1, \ldots, J$ tem capacidade C_j (em número de circuitos).
- Ao fluxo $k = 1, \ldots, K$ está associada uma taxa de chegada, λ_k, um tempo médio de serviço $1/\mu_k$ (a intensidade de tráfego é $\rho_k = \lambda_k / \mu_k$), uma largura de banda b_k (em número de circuitos) e um percurso de encaminhamento fixo $R_k \subseteq \{1,2,\ldots, J\}$:

  (1) As chamadas do fluxo k chegam de acordo com um processo de Poisson à taxa λ_k.

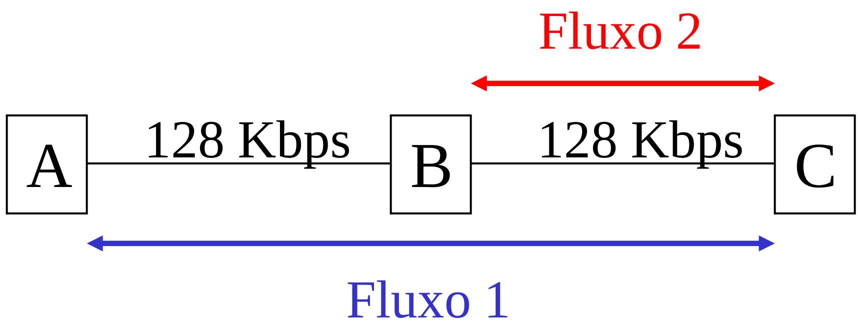
  (2) As chamadas do fluxo k admitidas pelo sistema ocupam b_k circuitos e têm uma duração exponencialmente distribuída com média $1/\mu_k$.

  (3) A duração das chamadas é independente entre chamadas e independente dos instantes de chegada para todos os fluxos.

- O conjunto dos fluxos que atravessam a ligação j é dado por K_j .

Encaminhamento fixo (Exemplo 1)

Considere a rede de figura que suporta 2 fluxos de chamadas: fluxo 1 com taxa de chegada λ_1 = 3 chamadas/hora, duração média das chamadas $1/\mu_1$ = 2 minutos e cada chamada ocupa b_1 = 64 Kb/s; fluxo 2 com taxa de chagada λ_2 = 4 chamadas/hora, duração média das chamadas $1/\mu_2$ = 3 minutos e cada chamada ocupa b_2 = 128 Kb/s.



K = 2 fluxos:

1: $\rho_1 = \lambda_1/\mu_1 = 3/60\times2 = 0.1$ Erlangs, b_1 = 1 circuito, R_1 = {AB,BC}

2: $\rho_2 = \lambda_2/\mu_2 = 4/60\times3 = 0.2$ Erlangs, b_2 = 2 circuitos, R_2 = {BC}

J = 2 ligações:

AB: C_{AB} = 2 circuitos, K_{AB} = {1}

BC: C_{BC} = 2 circuitos, K_{BC} = {1,2}

Encaminhamento fixo – Método Exato

Seja n_k o número de chamadas do fluxo k no sistema, $\mathbf{n} = (n_1, \ldots, n_k)$

Uma chamada do fluxo k não é aceite pela rede se em pelo menos uma das ligações pertencentes ao percurso de k:

$$b_k + \sum_{l \in K_j} b_l n_l > C_j$$

O espaço de estados do processo de nascimento e morte multidimensional é

$$S = \left\{ \mathbf{n} \in I^K : \sum_{k \in K_j} b_k n_k \leq C_j, \quad j = 1, \ldots, J \right\}$$

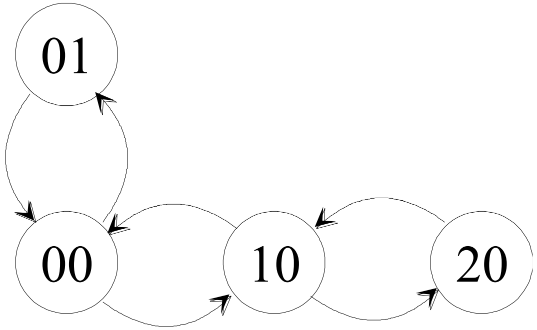
onde I é o conjunto dos inteiros não negativos.

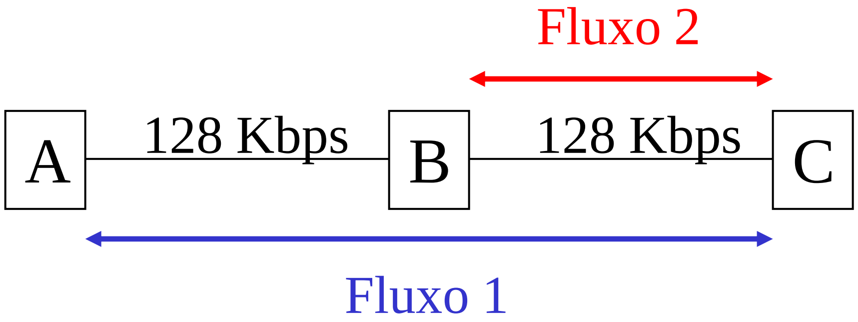
Seja S_k o subconjunto dos estados nos quais uma chamada do fluxo k é admitida quando chega à rede, isto é,

$$S_k = \left\{ \mathbf{n} \in S : \sum_{l \in K_j} b_l n_l \leq C_j - b_k, \quad j \in R_k \right\}$$

Encaminhamento fixo (Exemplo 1)

Considere a rede de figura que suporta 2 fluxos de chamadas: fluxo 1 com taxa de chegada λ_1 = 3 chamadas/hora, duração média das chamadas $1/\mu_1$ = 2 minutos e cada chamada ocupa b_1 = 64 Kb/s; fluxo 2 com taxa de chagada λ_2 = 4 chamadas/hora, duração média das chamadas $1/\mu_2$ = 3 minutos e cada chamada ocupa b_2 = 128 Kb/s.





Espaço de estados: S = {[0,0], [1,0], [2,0], [0,1]}

Estados nos quais uma chamada do fluxo 1 é admitida:

S_1 = {[0,0], [1,0]}

Estados nos quais uma chamada do fluxo 2 é admitida:

S_2 = {[0,0]}

Encaminhamento fixo – Método Exato

A probabilidade limite de cada estado é dada por

$$P(\mathbf{n}) = \frac{1}{G}\prod_{k=1}^{K}\frac{\rho_k^{n_k}}{n_k!} \qquad \mathbf{n} \in S$$

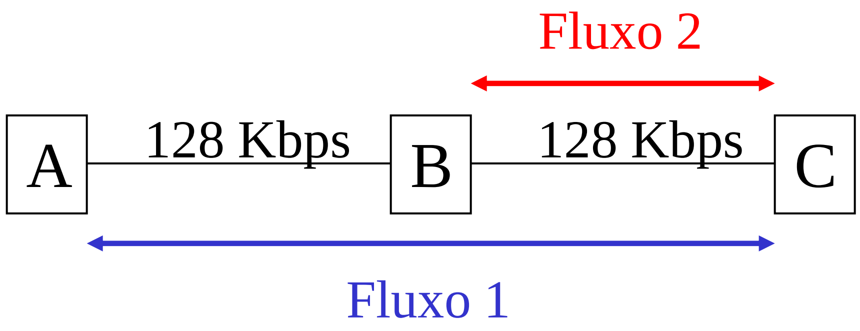
onde $$G = \sum_{\mathbf{n}\in S}\prod_{k=1}^{K}\frac{\rho_k^{n_k}}{n_k!}$$

e a probabilidade de bloqueio da classe *k* é dada por

$$B_k = 1 - \frac{\sum_{\mathbf{n}\in S_k}\prod_{l=1}^{K}\frac{\rho_l^{n_l}}{n_l!}}{\sum_{\mathbf{n}\in S}\prod_{l=1}^{K}\frac{\rho_l^{n_l}}{n_l!}} \quad \Leftrightarrow \quad B_k = \frac{\sum_{\mathbf{n}\in S\setminus S_k}\prod_{l=1}^{K}\frac{\rho_l^{n_l}}{n_l!}}{\sum_{\mathbf{n}\in S}\prod_{l=1}^{K}\frac{\rho_l^{n_l}}{n_l!}}$$

Encaminhamento fixo (Exemplo 1)

Considere a rede de figura que suporta 2 fluxos de chamadas: fluxo 1 com taxa de chegada λ_1 = 3 chamadas/hora, duração média das chamadas $1/\mu_1$ = 2 minutos e cada chamada ocupa b_1 = 64 Kb/s; fluxo 2 com taxa de chagada λ_2 = 4 chamadas/hora, duração média das chamadas $1/\mu_2$ = 3 minutos e cada chamada ocupa b_2 = 128 Kb/s.



1: ρ_1 = 0.1 Erlangs

2: ρ_2 = 0.2 Erlangs

S = {[0,0], [1,0], [2,0], [0,1]}

S_1 = {[0,0], [1,0]}

S_2 = {[0,0]}

$$B_k = 1 - \frac{\sum_{\mathbf{n}\in S_k} \prod_{l=1}^{K} \frac{\rho_l^{n_l}}{n_l!}}{\sum_{\mathbf{n}\in S} \prod_{l=1}^{K} \frac{\rho_l^{n_l}}{n_l!}}$$

$$B_1 = 1 - \frac{\frac{0.1^0 0.2^0}{0!\,0!} + \frac{0.1^1 0.2^0}{1!\,0!}}{\frac{0.1^0 0.2^0}{0!\,0!} + \frac{0.1^1 0.2^0}{1!\,0!} + \frac{0.1^2 0.2^0}{2!\,0!} + \frac{0.1^0 0.2^1}{0!\,1!}}$$

$$= 1 - \frac{1 + 0.1}{1 + 0.1 + \frac{0.01}{2} + 0.2} = 0.157 = 15.7\%$$

Encaminhamento fixo (Exemplo 1)

Considere a rede de figura que suporta 2 fluxos de chamadas: fluxo 1 com taxa de chegada λ_1 = 3 chamadas/hora, duração média das chamadas $1/\mu_1$ = 2 minutos e cada chamada ocupa b_1 = 64 Kb/s; fluxo 2 com taxa de chagada λ_2 = 4 chamadas/hora, duração média das chamadas $1/\mu_2$ = 3 minutos e cada chamada ocupa b_2 = 128 Kb/s.

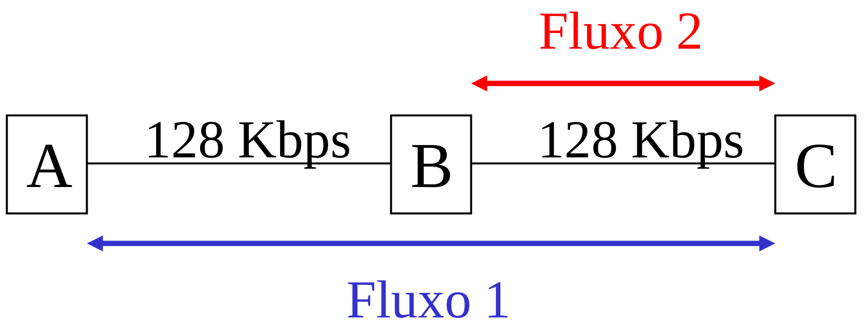
1: ρ_1 = 0.1 Erlangs

2: ρ_2 = 0.2 Erlangs

S = {[0,0], [1,0], [2,0], [0,1]}

S_1 = {[0,0], [1,0]}

S_2 = {[0,0]}



$$B_k = 1 - \frac{\sum_{\mathbf{n}\in S_k} \prod_{l=1}^{K} \frac{\rho_l^{n_l}}{n_l!}}{\sum_{\mathbf{n}\in S} \prod_{l=1}^{K} \frac{\rho_l^{n_l}}{n_l!}}$$

$$B_2 = 1 - \frac{\frac{0.1^0 0.2^0}{0!\,0!}}{\frac{0.1^0 0.2^0}{0!\,0!} + \frac{0.1^1 0.2^0}{1!\,0!} + \frac{0.1^2 0.2^0}{2!\,0!} + \frac{0.1^0 0.2^1}{0!\,1!}}$$

$$= 1 - \frac{1}{1 + 0.1 + \frac{0.01}{2} + 0.2} = 0.234 = 23.4\%$$

Teorema do Limite do Produto

- Aplica-se apenas quando as chamadas de todos os fluxos *k* requerem a mesma largura de banda b_k (em número de circuitos).
- Seja a intensidade de tráfego suportada pela ligação *j* dada por:

$$\bar{\rho}_j = \sum_{k \in K_j} \rho_k$$

- O *teorema do limite do produto* declara que

$$B_k \leq 1 - \prod_{j \in R_k} \left(1 - ER\left[\bar{\rho}_j, C_j\right]\right)$$

C_j – capacidade da ligação *j* (em número de circuitos)

em que $ER[\rho, C]$ representa a fórmula de ErlangB.

- Prova-se matematicamente que este valor é um majorante das probabilidades de bloqueio exatas. É uma boa aproximação quando:

(1) os fluxos atravessam poucas ligações

(2) as probabilidades de bloqueio são pequenas (menores que 1%)

Aproximação de carga reduzida

- Uma possibilidade para melhorar a aproximação associada ao teorema do limite do produto é reduzir o tráfego oferecido à ligação j, tomando em linha de conta o bloqueio nas restantes ligações do percurso de cada fluxo.
- O teorema do limite do produto implica que a probabilidade de uma ligação j estar totalmente ocupada é majorada por

$$ER\left[\sum_{k \in K_j} \rho_k, C_j\right]$$

- Fazendo a substituição de ρ_k por $\rho_k t_k(j)$, em que $t_k(j)$ corresponde à probabilidade de existir pelo menos uma unidade de capacidade disponível em cada ligação pertencente a $R_k - \{j\}$ temos que a probabilidade de bloqueio (aproximada) da ligação j vem dada por

$$L_j = ER\left[\sum_{k \in K_j} \rho_k t_k(j), C_j\right]$$

Aproximação de carga reduzida

- Tomando como aproximação adicional que o bloqueio é independente de ligação para ligação, resulta

$$t_k(j) = \prod_{i \in R_k - \{j\}} (1 - L_i)$$

e, finalmente, combinando as duas aproximações anteriores resultam as seguintes <u>equações de ponto fixo</u> (uma por cada ligação da rede)

$$L_j = ER\left[\sum_{k \in K_j} \rho_k \prod_{i \in R_k - \{j\}} (1 - L_i), C_j\right] \quad , j = 1, 2, ..., J$$

- Admitindo novamente que o bloqueio é independente de ligação para ligação resulta a probabilidade de bloqueio das chamadas do fluxo *k* é

$$B_k \approx 1 - \prod_{j \in R_k} (1 - L_j) \quad k = 1,\ 2,\ ...,\ K$$

Algoritmo iterativo de cálculo da aproximação de carga reduzida

Seja **L** = (L_1, L_2, …, L_J) e o operador **T**(**L**) = (T_1(**L**), T_2(**L**), …, T_J(**L**)) onde

$$T_j(\mathbf{L}) = ER\left[\sum_{k\in\mathcal{K}_j}\rho_k \prod_{i\in R_k-\{j\}}(1-L_i), C_j\right]$$

As equações de ponto fixo podem ser expressas na forma **L** = **T**(**L**).

Método iterativo – partindo de um vetor inicial $\mathbf{L}\in[0,1]^J$ aplica-se sucessivamente o operador **T**:

$$\mathbf{L}^0 = \mathbf{L}$$

$$\mathbf{L}^m = \mathbf{T}\left(\mathbf{L}^{m-1}\right), \qquad m = 1,\ldots,n$$

Partindo de $\mathbf{L}^0$ = (1, 1, …, 1), o método dá origem a $\mathbf{L}^1$ = (0, 0, …, 0), $\mathbf{L}^{2n}$ converge para um $\mathbf{L}^+$ e $\mathbf{L}^{2n+1}$ converge para um $\mathbf{L}^-$ tal que $\mathbf{L}^- \leq \mathbf{L}^* \leq \mathbf{L}^+$.

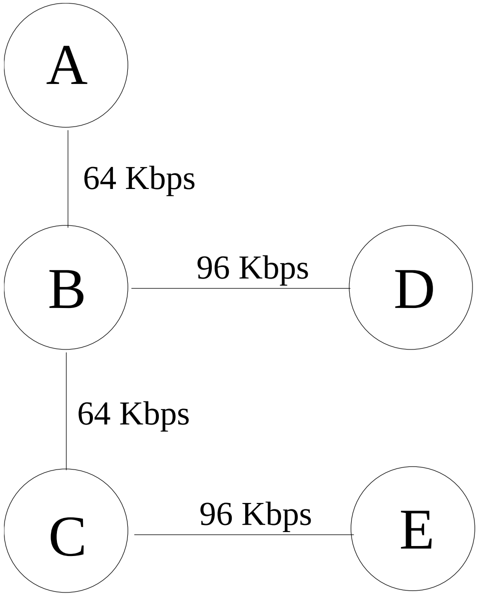
As sucessivas iterações *m* determinam majorantes da solução **L*** quando *m* é par e minorantes da solução **L*** quando *m* é impar. Termina-se o algoritmo quando os dois limites estão suficientemente próximos.

Exemplo 2

Considere a rede da figura. A rede suporta 3 fluxos de chamadas: *fluxo 1* entre A e D, *fluxo 2* entre C e D e *fluxo 3* entre E e B. As chamadas chegam de acordo com processos de Poisson com taxa λ_1 = 20 chamadas/hora, λ_2 = 60 chamadas/hora e λ_3 = 20 chamadas/hora. Em todos os fluxos, a duração de cada chamada é exponencialmente distribuída com média $1/\mu$ = 3 minutos e cada chamada requer uma largura de banda de 32 Kbps.

(1) Calcule um limite superior para a probabilidade de bloqueio de cada fluxo usando o teorema do limite do produto.

(2) Escreva as equações que permitem calcular as probabilidades de bloqueio de cada fluxo através da aproximação de carga reduzida.



Encaminhamento dinâmico da rede telefónica

Os métodos de encaminhamento dinâmico são usados nas redes de transporte dos operadores telefónicos.

Estas redes têm conectividade total, ou seja, incluem uma ligação direta entre todos os pares de nós de acesso.

Um nó de acesso é uma central que liga uma rede de acesso à rede de transporte.

Assim, numa rede com N nós:

- existem $N\,(N-1)\,/\,2$ ligações
- o número de percursos com apenas duas ligações entre quaisquer duas centrais é $N-2$.

Encaminhamento dinâmico da rede telefónica

Os métodos de encaminhamento dinâmico têm as seguintes características em comum:

1. quando é pedido o estabelecimento de uma chamada entre duas centrais, a chamada é encaminhada no percurso direto, se houver pelo menos um circuito disponível;
2. quando o percurso direto estiver indisponível, a chamada pode ser encaminhada num dos percursos alternativos permitidos;
3. os percursos alternativos têm sempre apenas duas ligações, ou seja, as chamadas nunca são estabelecidas em percursos com três ou mais ligações.

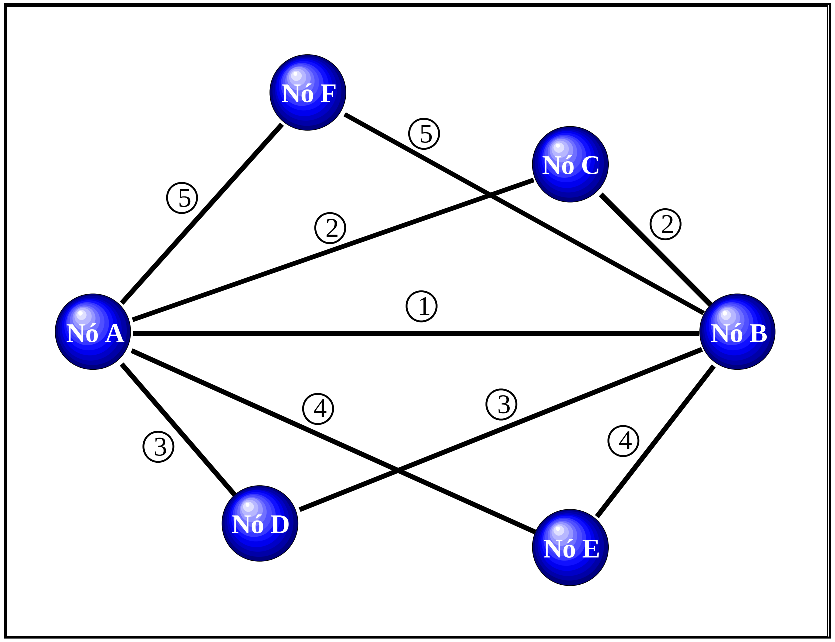
Os diferentes métodos de encaminhamento diferem na forma como é definido, em cada instante, o conjunto de percursos alternativos.

Encaminhamento sequencial

- É a base do método introduzido pela AT&T nos anos 80 designado por DNHR - *Dynamic Non-Hierarchical Routing*.
- A cada par de centrais origem-destino associa-se uma lista ordenada de percursos alternativos. Se uma chamada que chega encontra o percurso direto indisponível:
  1) a chamada é estabelecida no primeiro percurso alternativo disponível da lista ordenada;
  2) se todos os percursos alternativos estiverem indisponíveis a chamada é bloqueada.
- Na implementação do DNHR, os diferentes parâmetros da rede (a lista de percursos alternativos, a ordenação dos percursos na lista, a percentagem de circuitos reservados em cada ligação, etc...) variam no tempo, sendo considerados até 10 períodos distintos em cada dia.
- Quando a segunda ligação de um percurso alternativo estiver indisponível, a central intermédia tem de sinalizar a central origem dessa ocorrência para que esta possa tentar outro percurso alternativo (função designada por *crankback*).

DNHR

Nó A → Nó B



Lista #N	Percursos recomendados presentes na lista	Período de tempo
Lista #1	1→3→2→4	10:00– 12:00
Lista #2	1→4→2	12:00 – 17:00
Lista #3	1→3→5→2	17:00– 19:00
Lista #4	1→5→4	19:00– 23:00
Lista #5	1→2→3	23:00 – 10:00

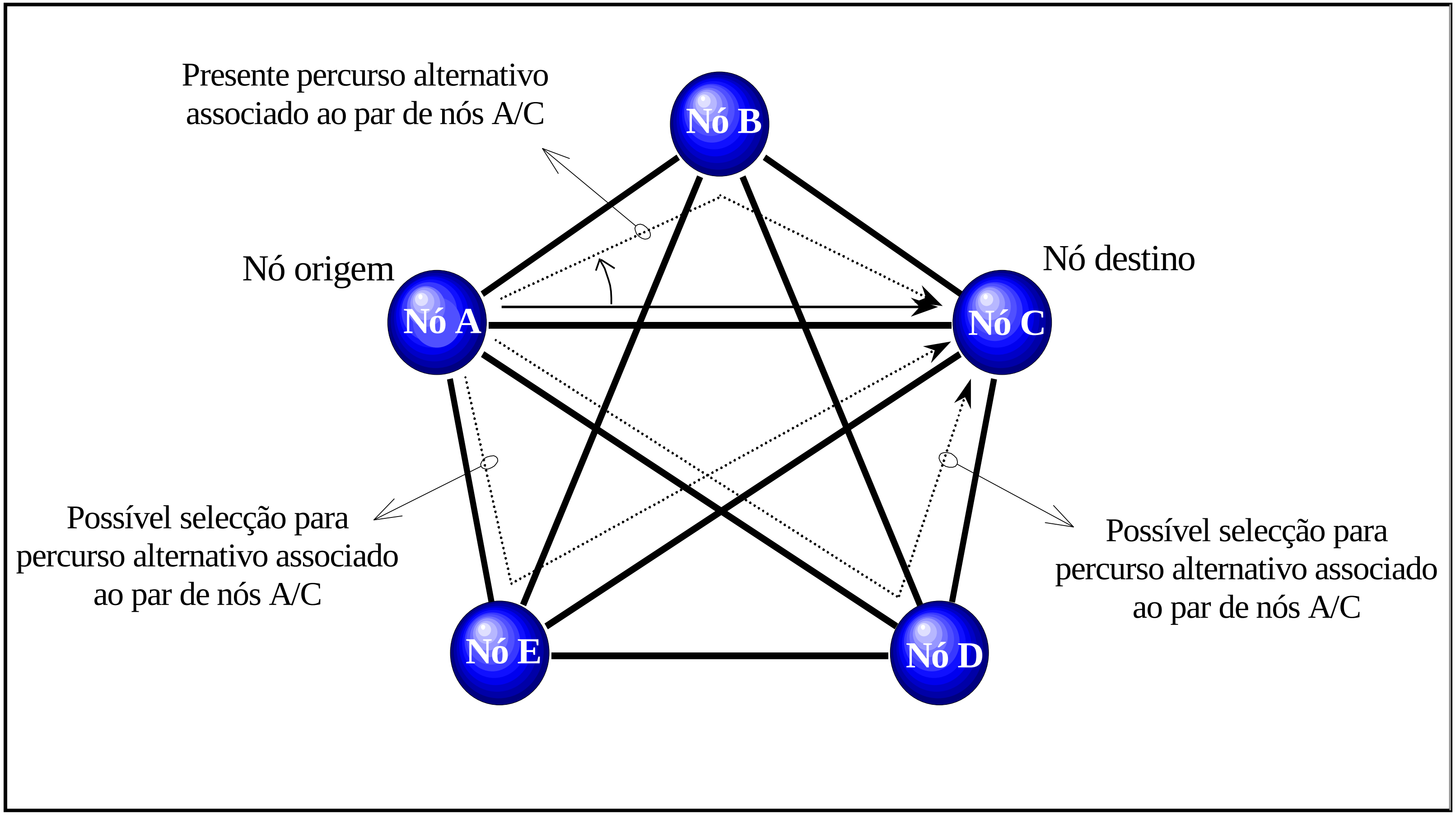
Encaminhamento aleatório retardado

- Este método é a base do protocolo DAR - *Dynamic Alternative Routing* introduzido pelos British Telecom nos anos 90.
- Associa-se a cada par de centrais origem-destino um percurso alternativo (e apenas um):
  1) se quando uma chamada chega o percurso direto está indisponível e o percurso alternativo está disponível a chamada é estabelecida no percurso alternativo;
  2) caso contrário, a chamada é bloqueada e é escolhido um novo percurso alternativo para as chamadas subsequentes;
  3) o novo percurso alternativo é escolhido aleatoriamente de entre os N – 3 percursos alternativos restantes.

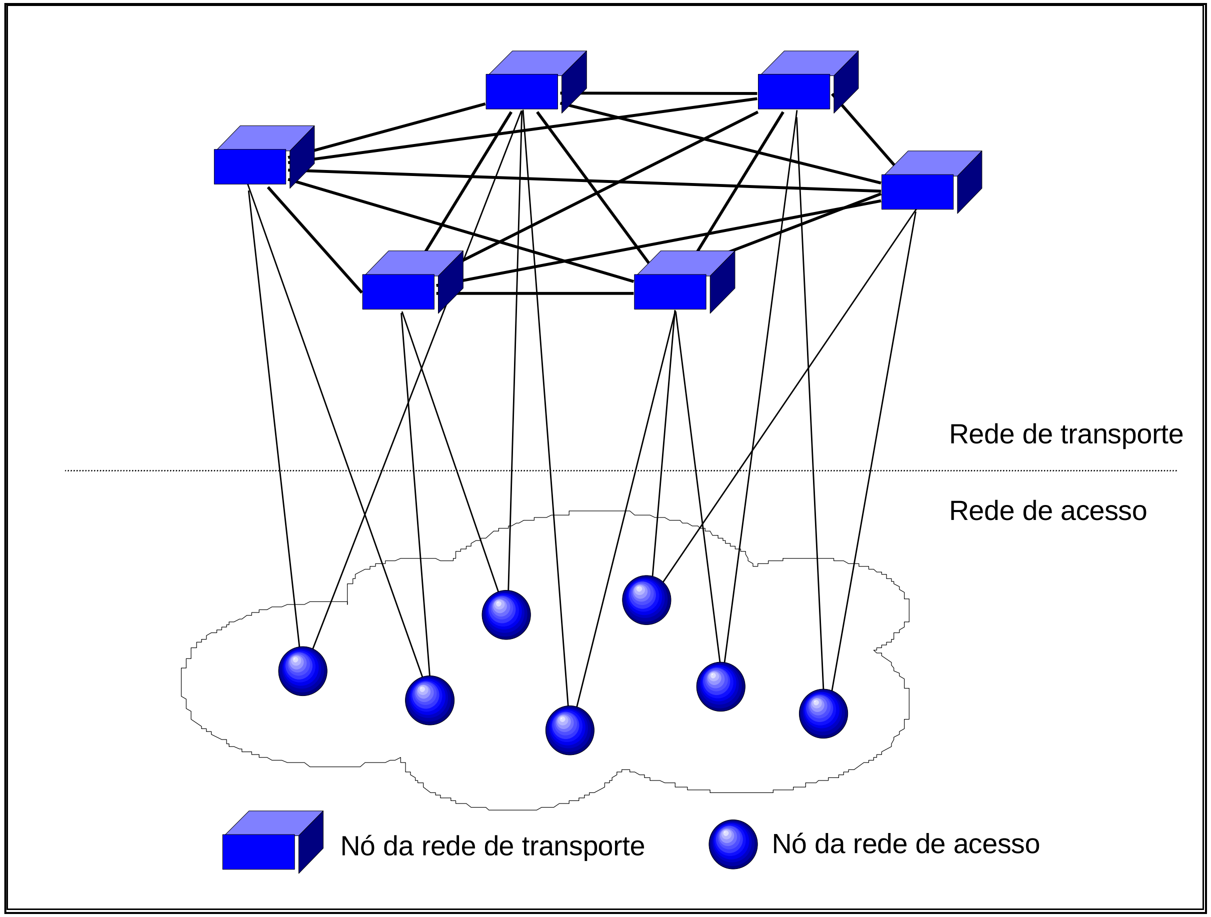
Relativamente ao DHNR:

- não é necessário implementar a função de *cranckback*, porque apenas é tentado um percurso alternativo (os mecanismos de sinalização são mais simples)
- adapta-se dinamicamente ao tráfego sem necessidade de gestão centralizada dos percursos alternativos

Operação do protocolo DAR

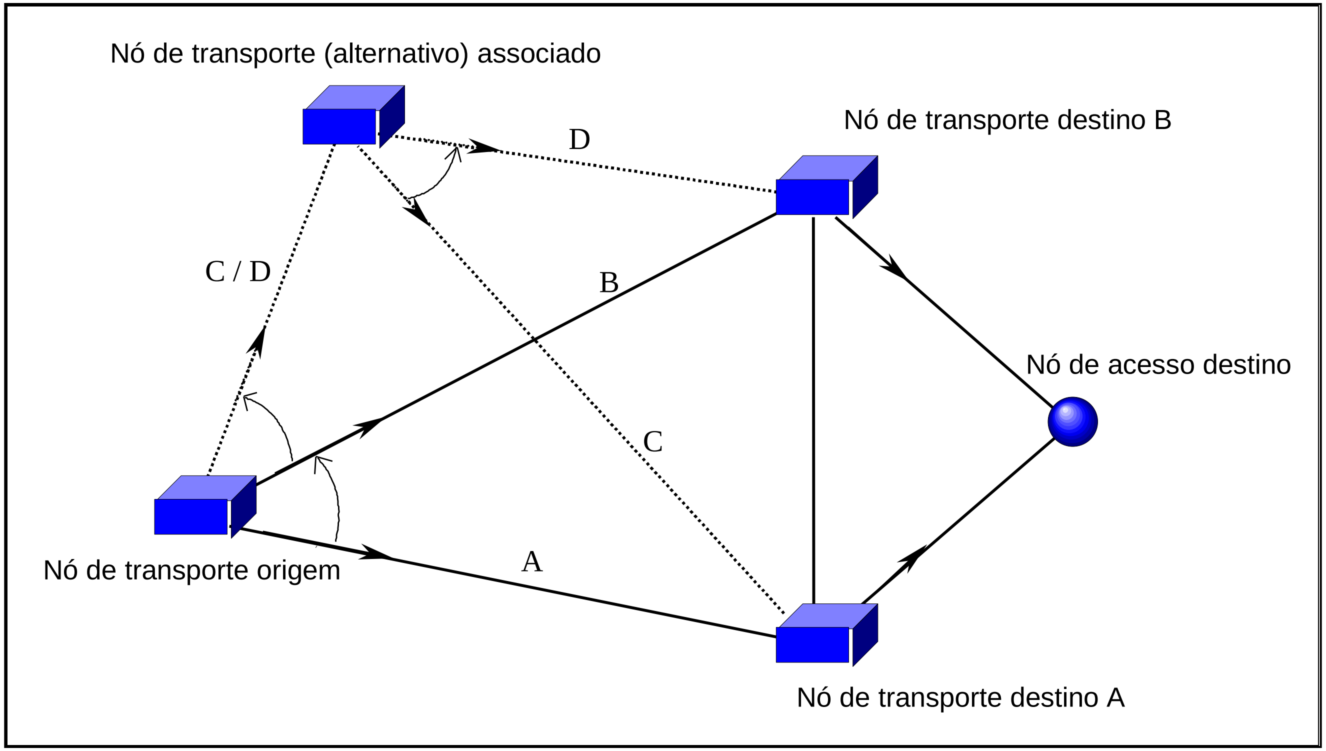


DAR na rede da British Telecom



- Rede organizada em rede de transporte e rede de acesso.
- Rede de transporte com conectividade física total.
- Cada nó de acesso com ligação física a dois nós de transporte (para proteção de falha individual de ligação).

DAR na rede da British Telecom



- No nó de transporte onde chega o pedido de chamada existem 2 percursos diretos e um nó de transporte alternativo.
- Primeiro são tentados os dois percursos diretos.
- Depois são tentados os dois percursos via nó alternativo.
- Se a chamada bloquear, outro nó alternativo é escolhido para o próximo pedido de chamada para o mesmo destino.

Encaminhamento de menor carga

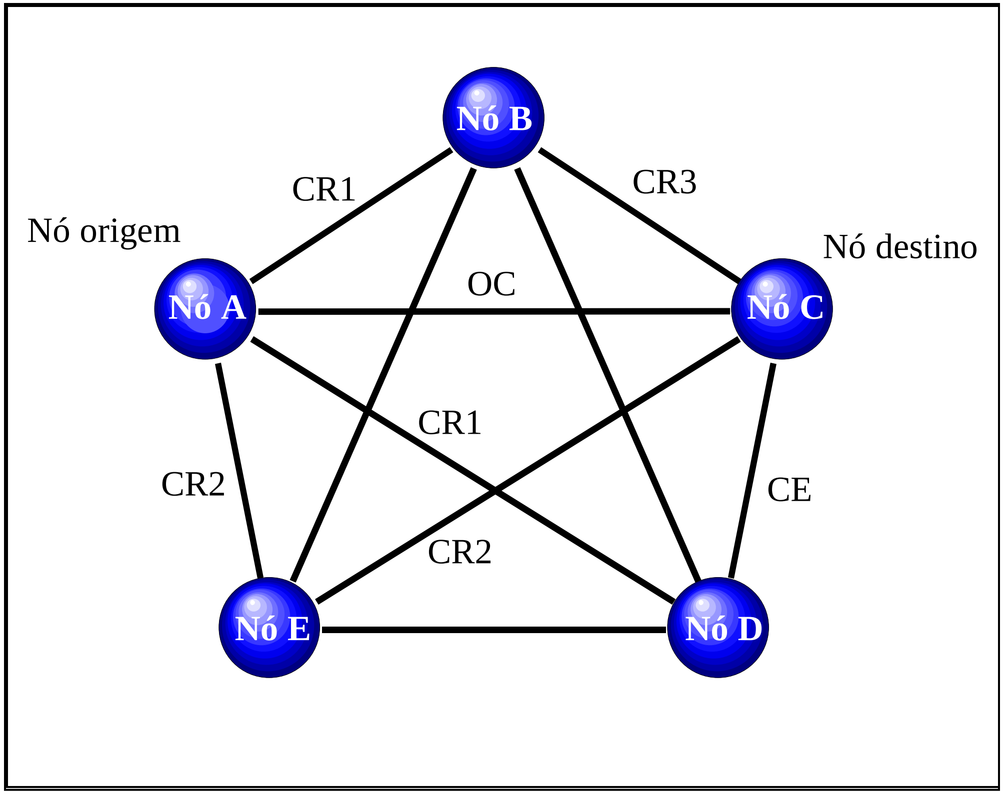
- Este método mantém um registo da capacidade não utilizada em cada percurso alternativo (corresponde ao número de circuitos não ocupados, para lá dos circuitos reservados).
- Quando o percurso direto está indisponível, escolhe o percurso alternativo com maior capacidade não utilizada, de entre o conjunto de percursos alternativos permitidos; se todos os percursos alternativos estiverem indisponíveis a chamada é bloqueada.

Este método está na base do método RTNR - *Real-Time Network Routing* introduzido pela AT&T no início dos anos 90:

1) quando o percurso direto está indisponível, a central origem pergunta à central destino qual a capacidade não utilizada de todas as suas ligações;
2) após receber essa informação, a central origem determina qual o percurso alternativo com menor capacidade utilizada, com base na informação de que dispõe sobre a capacidade não utilizada das suas próprias ligações.

- Este método é mais eficiente que o DAR mas exibe uma maior complexidade nos mecanismos de sinalização.

RTNR



Níveis de carga considerados:

CR1 - Carga Reduzida de nível 1

CR2 - Carga Reduzida de nível 2

CR3 - Carga Reduzida de nível 3

CE - Carga Elevada

RS – Reservado

OC - Ocupado

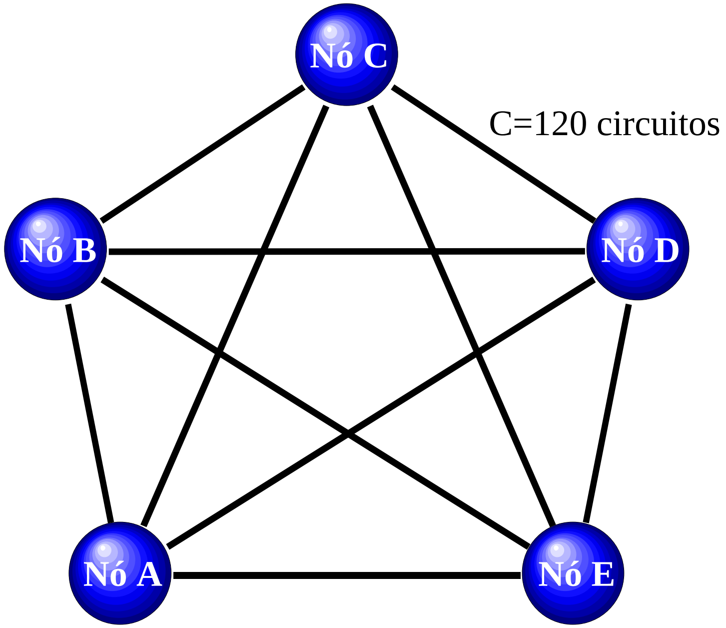
Nó A → Nó C

A chamada é estabelecida pelo nó intermédio E

Metaestabilidade e reserva de recursos

Exemplo:

- rede com conectividade total;
- encaminhamento aleatório global;
- tráfego oferecido igual para todos os pares origem-destino;
- mesma reserva de recursos **r** para todos os percursos diretos.

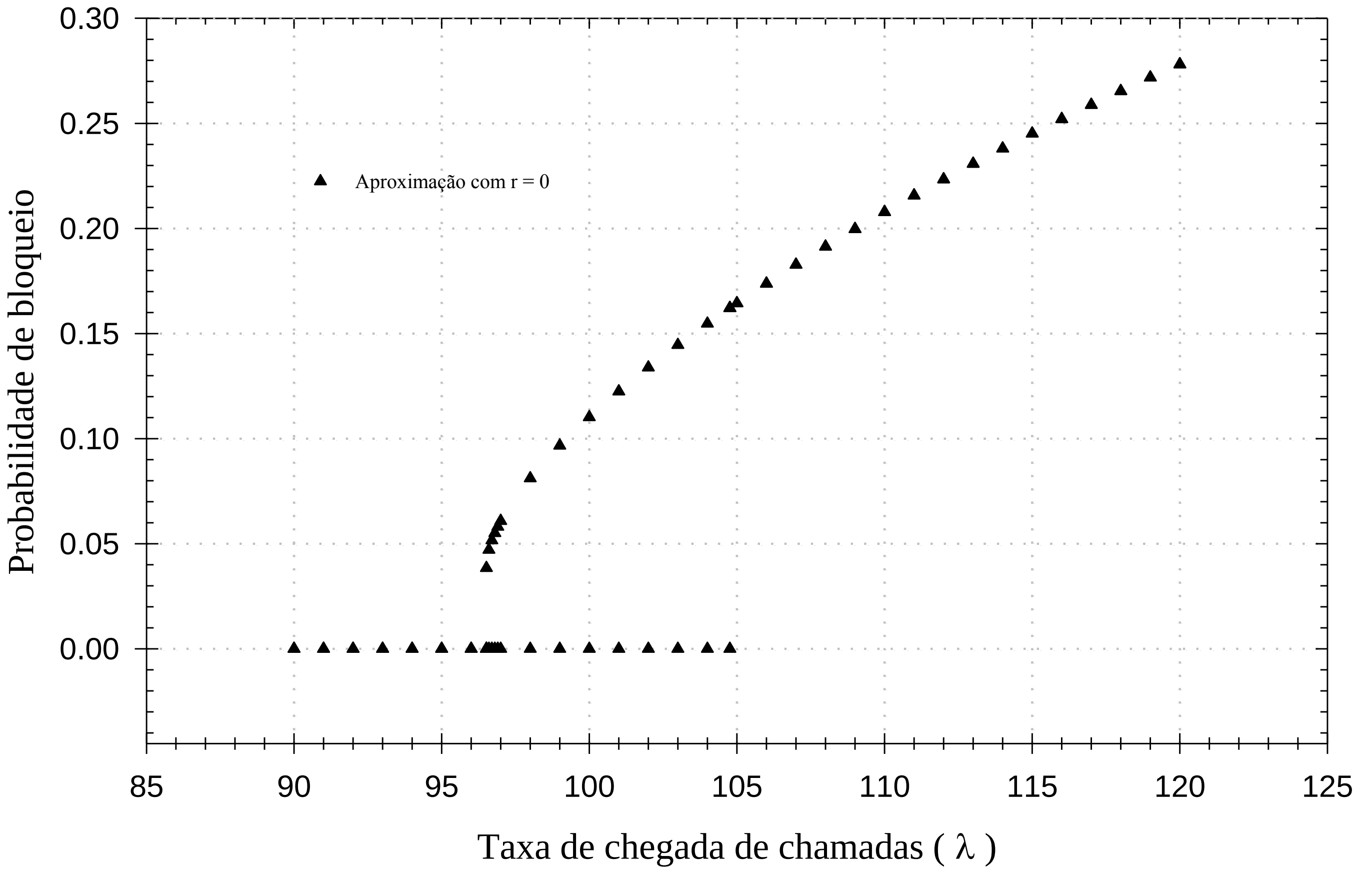


O desempenho desta rede pode ser calculado de forma aproximada por processos analíticos.

Exemplo sem reserva de recursos (r=0)

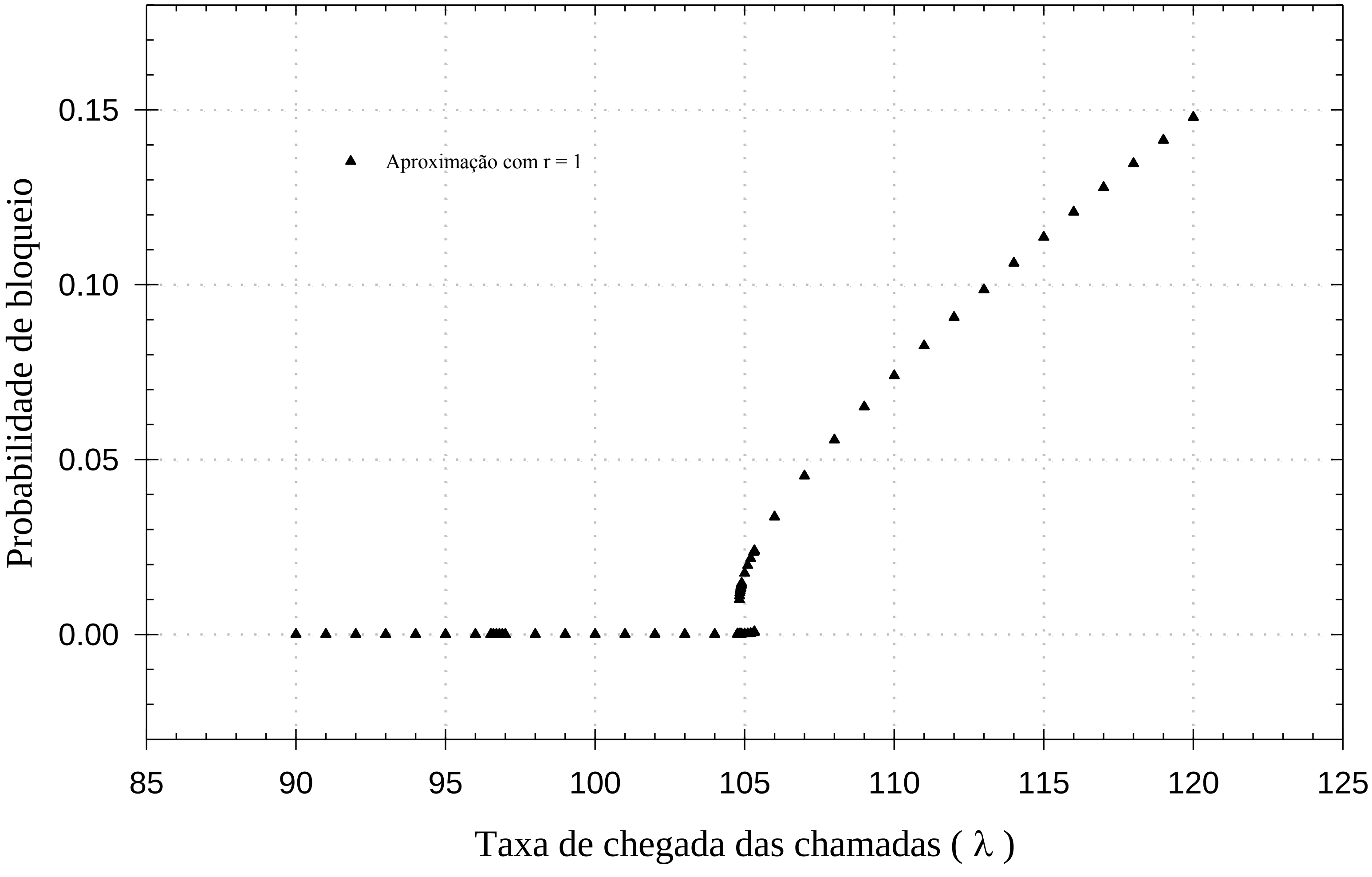
Instabilidade para valores de λ entre 96 e 105:

- Existem intervalos de tempo em que a maior parte do tráfego é encaminhada pelo percurso direto (probabilidade de bloqueio desprezável).
- Existem intervalos de tempo em que a maior parte do tráfego é encaminhada pelos percursos alternativos (probabilidade de bloqueio elevada).



Exemplo com reserva de um circuito (r=1)

A instabilidade é reduzida com uma reserva de 1 em 120 circuitos



Exemplo com reserva de 2 ou mais circuitos

Valores crescentes de reserva anulam a instabilidade e aumentam o desempenho da rede (diminuem a percentagem de tráfego que usa percursos alternativos)