



Engenharia de Tráfego de Serviços *Anycast*

Modelação e Desempenho de Redes e Serviços

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

DETI-UA, 2022/2023

Serviços *Unicast* vs. *Anycast*

Serviço *Unicast*:

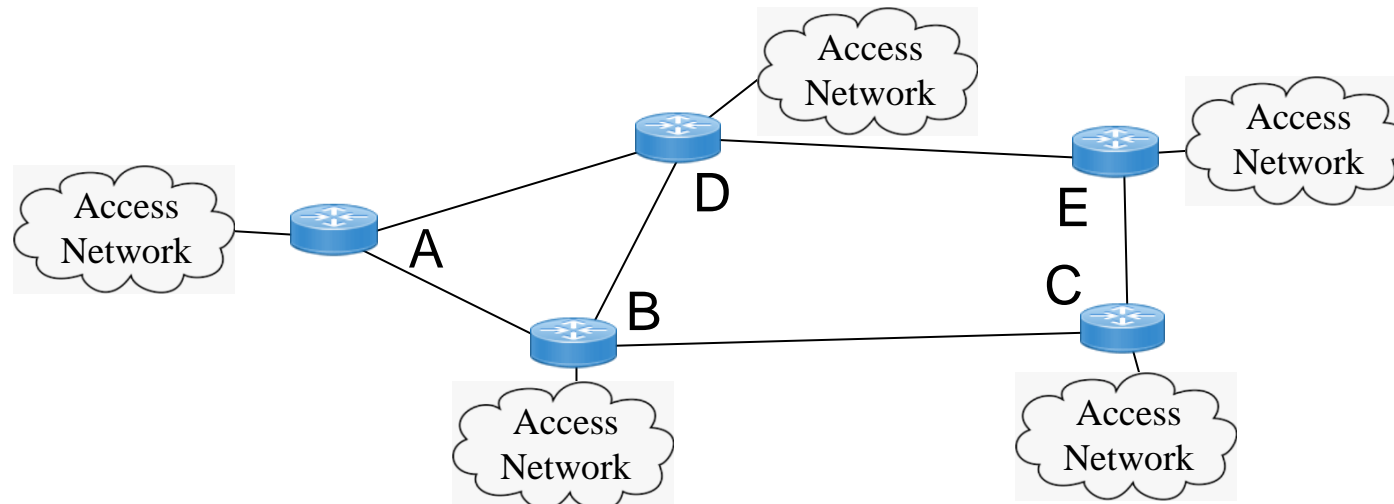
- O tráfego de um serviço *unicast* é definido por um conjunto de fluxos de tráfego de cada nó origem para cada nó destino.

Serviço *Anycast*:

- Exemplos: serviços de *streaming* de filmes e/ou de música (tais como o Netflix, Youtube, Amazon Prime Video, Spotify)
- Num serviço *anycast*, existe um conjunto de nós (designados por nós *anycast*) associados ao serviço, i.e., os nós da rede ao qual estão ligados os servidores do serviço, tipicamente hospedados em *Data Centers* (DCs).
- O tráfego de um serviço *anycast* é definido por um conjunto de fluxos de tráfego de cada nó origem.
- O destino de cada fluxo de tráfego pode ser qualquer um dos nós *anycast*, i.e., depende do encaminhamento dos fluxos na rede que liga os diferentes nós.

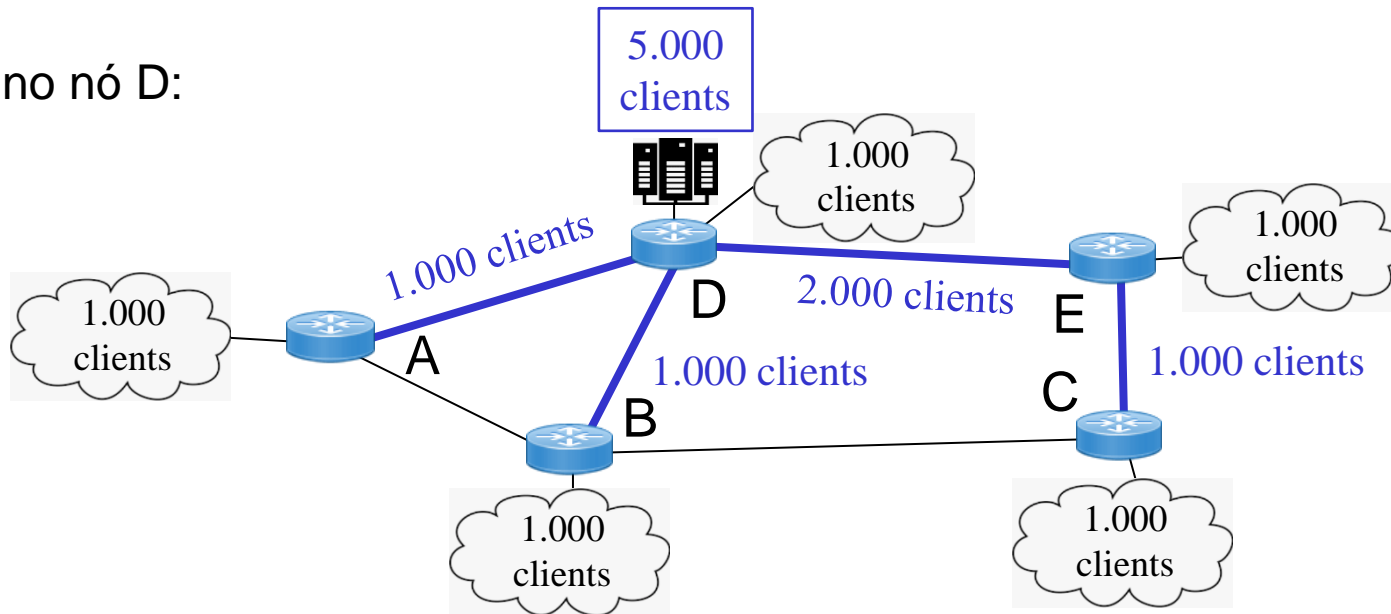
Serviços *anycast*

- Quando um cliente se liga a um serviço *anycast*, por norma, a rede encaminha a comunicação para o nó *anycast* mais próximo (em número de ligações ou em termos de atraso).
- O número de nós *anycast* e a sua localização na rede influenciam:
 - os recursos necessários na rede para suportar o serviço;
 - o desempenho da rede tanto em termos de atraso como de disponibilidade do serviço.
- Considere-se a seguinte rede a suportar um serviço *anycast* com 1000 clientes ligados em cada rede de acesso:



Serviços *anycast*

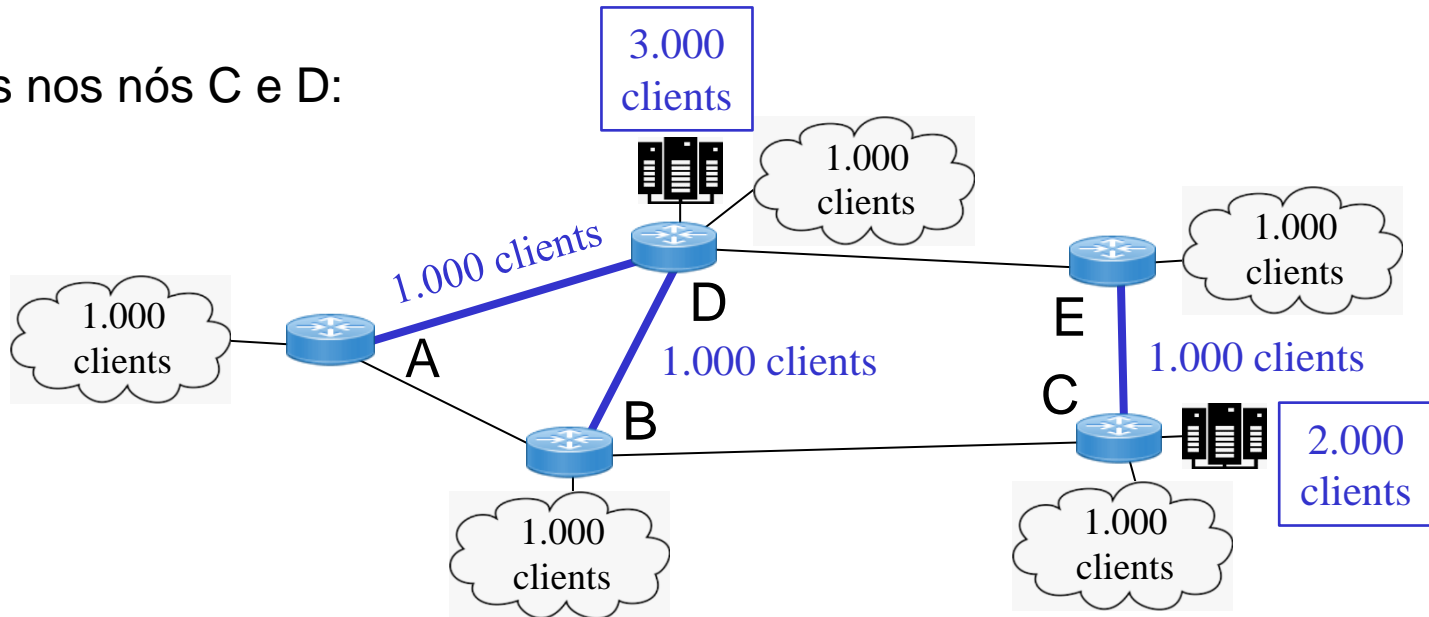
1 DC no nó D:



- 3 ligações têm de ter capacidade suficiente para suportar o tráfego de 1000 clientes e 1 ligação tem de ter capacidade suficiente para suportar o tráfego de 2000 clientes.
- O DC tem de ter a capacidade para suportar o tráfego de 5000 clientes.
- Se o DC falhar, o serviço *anycast* falha completamente.

Serviços *anycast*

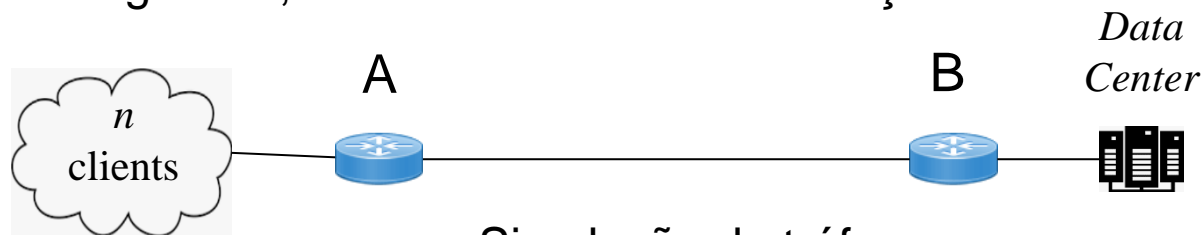
2 DCs nos nós C e D:



- 3 ligações têm de ter capacidade suficiente para suportar o tráfego de 1000 clientes (ocupa menos recursos de rede que no caso anterior).
- O DC no nó D tem de ter a capacidade para suportar o tráfego de 3000 clientes e o DC no nó C tem de ter a capacidade para suportar o tráfego de 2000 clientes.
- Se um dos DCs falhar, o serviço *anycast* sofre uma degradação mas não falha completamente.

Agregação de tráfego

- Considere-se a simulação do tráfego *downstream* de n clientes numa rede de acesso a aceder a um servidor de *streaming* de filmes (hospedado num DC remoto) caracterizada por:
 - dependendo do terminal, o filme é transmitido num *stream* de 6, 12 ou 24 Mbps (todos os formatos igualmente prováveis)
 - duração do acesso de cada cliente ao serviço (em contínuo) entre 0.5 e 2 horas (segundo uma distribuição uniforme)
 - intervalo de tempo entre acessos ao serviço de cada cliente exponencialmente distribuído com média de 2 horas
 - em cada acesso, uma pausa entre 2 e 18 minutos (segundo uma distribuição uniforme) com probabilidade p .
- Nos slides seguintes, são visualizadas 10 simulações.



Simulação do tráfego
agregado na ligação de
B para A

Agregação de tráfego sem pausas



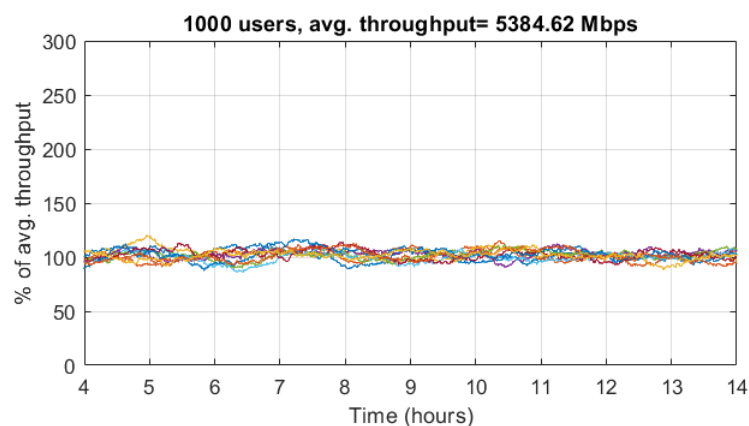
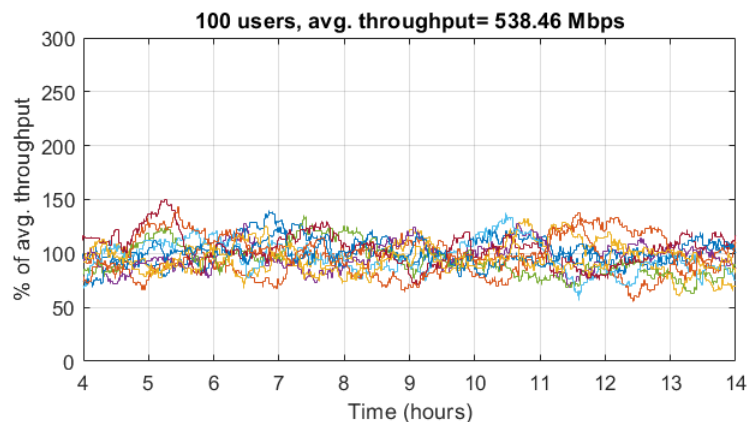
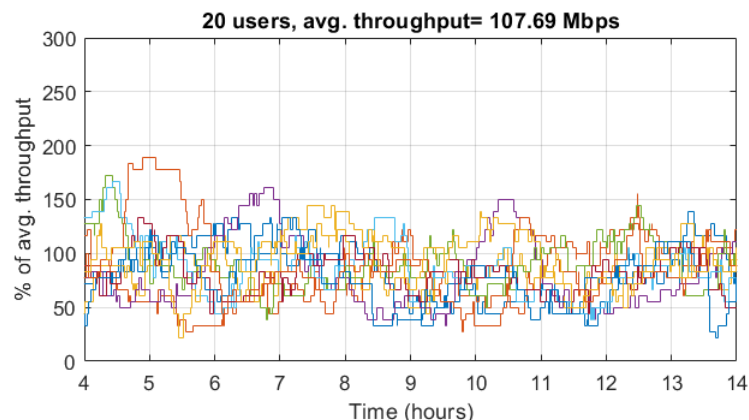
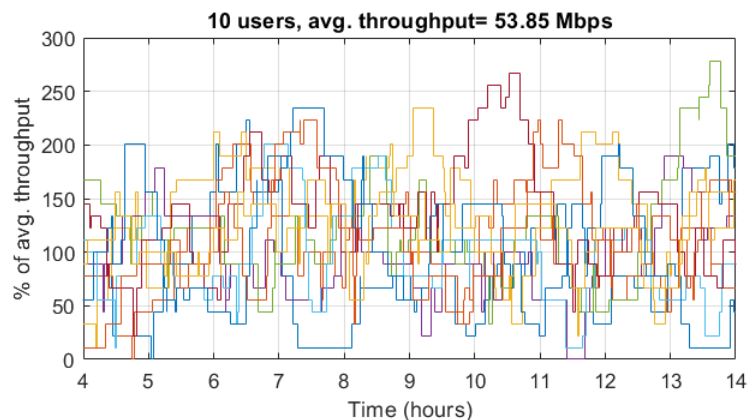
A



B



Data Center



Para um número de clientes grande, o tráfego agregado varia pouco em torno do tráfego médio.

Agregação de tráfego



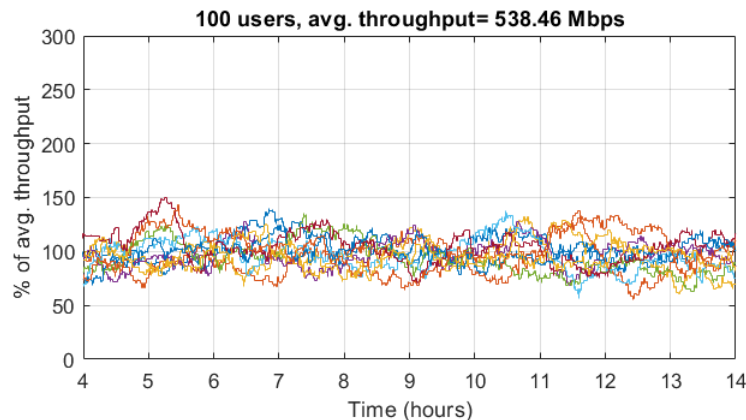
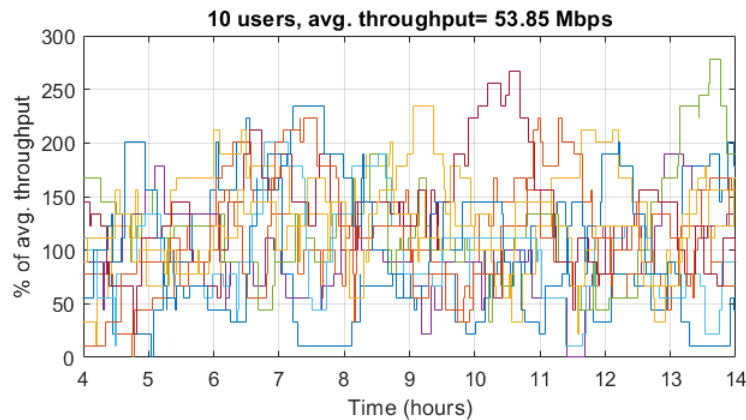
A



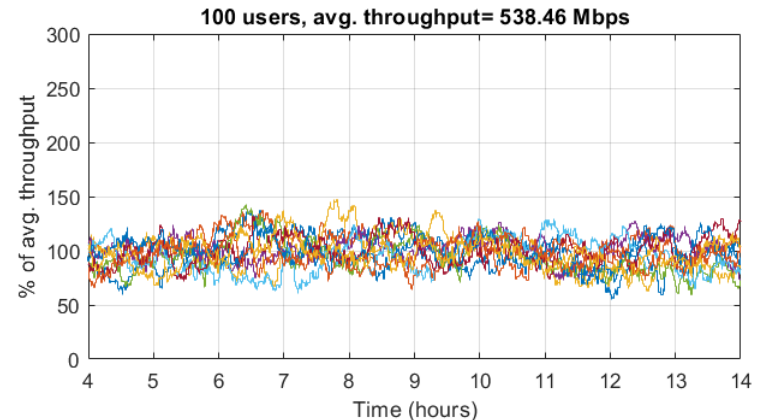
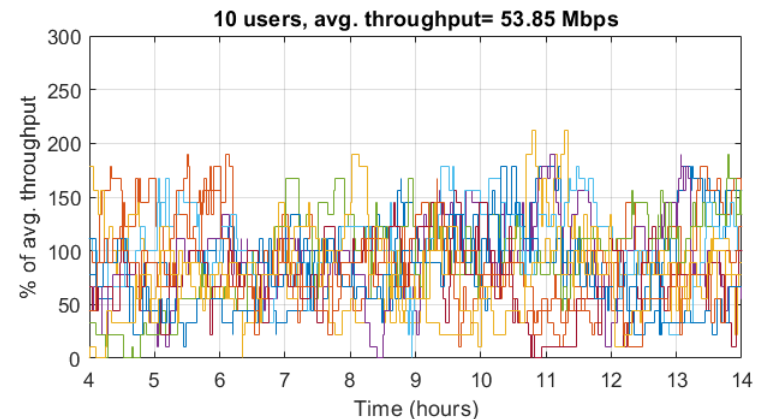
B



Sem Pausas ($p = 0.0$)



Com Pausas ($p = 0.99$)



As pausas reduzem pouco a variação do tráfego agregado em torno do tráfego médio e a redução é observável apenas para um número de clientes pequeno.

Tráfego médio de um agregado de fluxos

As simulações anteriores mostram que o tráfego suportado pela rede em cada fluxo é aproximadamente igual ao tráfego médio para serviços com muitos clientes.

Considere-se o tráfego *downstream* de n clientes (numa rede de acesso) a aceder a um servidor de *streaming* de filmes.

Os filmes têm uma duração média de 90 minutos e são disponibilizados em 3 formatos (por exemplo, HD, FHD ou 4K) cujo ritmo de transmissão é respetivamente de 6, 12 ou 24 Mbps.

Cada cliente vê em média 2 filmes por dia cujo formato é em HD com 20% de probabilidade, em FHD com 30% de probabilidade e em 4K com 50% de probabilidade.

1. Determine o ritmo médio r de transmissão *downstream* por cliente.
2. Determine o ritmo médio R de transmissão *downstream* para um agregado de $n = 1000$ clientes.



Tráfego médio de um agregado de fluxos



1. Determine o ritmo médio r de transmissão *downstream* por cliente.

O ritmo médio de transmissão de um filme por cliente é:

$$0.2 \times 6 + 0.3 \times 12 + 0.5 \times 24 = 16.8 \text{ Mbps}$$

A fração média de tempo em que o cliente está a ver filmes por dia é:

$$2 \times 1.5 / 24 = 0.125 (=12.5\%)$$

Finalmente: $r = 16.8 \times 0.125 = 2.1 \text{ Mbps}$

2. Determine o ritmo médio R de transmissão *downstream* para um agregado de $n = 1000$ clientes.

$$R = n \times r = 1000 \times 2.1 = 2100 \text{ Mbps} = 2.1 \text{ Gbps}$$

Probabilidade de bloqueio de um servidor

- Considere-se um servidor com capacidade para atender m clientes. Qual a probabilidade de bloqueio do servidor, i.e., a probabilidade de um pedido ser recusado porque o servidor está a atender m clientes?
- Se os pedidos de serviço forem um processo de Poisson com taxa λ (pedidos por unidade de tempo) e o tempo de serviço de cada cliente for exponencialmente distribuído com média $1/\mu$ (unidades de tempo), o desempenho do servidor é caracterizado por um sistema $M/M/m/m$.
- Probabilidade do servidor estar a atender n clientes:

$$P_n = \frac{(\lambda/\mu)^n / n!}{\sum_{i=0}^m (\lambda/\mu)^i / i!} \quad n = 0, 1, \dots, m$$

- Probabilidade de bloqueio (fórmula de ErlangB):
$$P_m = \frac{(\lambda/\mu)^m / m!}{\sum_{i=0}^m (\lambda/\mu)^i / i!}$$
- É possível demonstrar que as fórmulas são válidas para qualquer estatística do tempo de serviço desde que o tempo de serviço seja estatisticamente independente dos instantes de pedido de serviço.

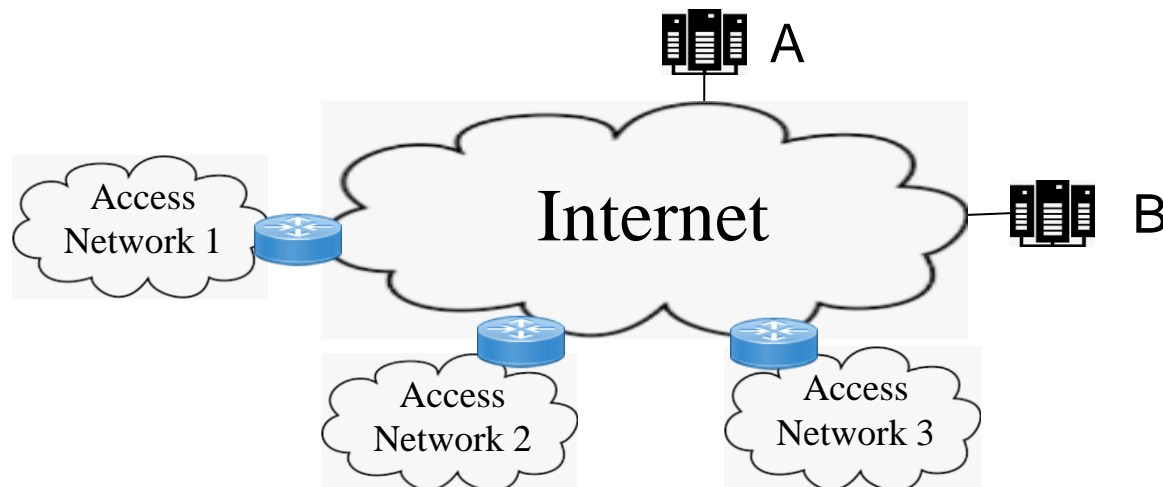
Probabilidade de bloqueio e disponibilidade de um serviço *anycast*

- Considere-se um serviço em que quando todos os servidores estão operacionais, um cliente pode ser servido por qualquer servidor.
- Assim, se o servidor mais próximo estiver totalmente ocupado, o pedido de serviço é encaminhado para o servidor disponível mais próximo
 - Geralmente, existe um módulo do serviço para balanceamento de carga entre servidores.
- Então, em termos de probabilidade de bloqueio, é conceptualmente equivalente a considerar um único servidor cuja capacidade é a soma das capacidades de todos os servidores.
- Se cada servidor tiver um valor de disponibilidade associado, então a probabilidade de bloqueio é a soma pesada da probabilidade de bloqueio da capacidade disponível em que o peso é a probabilidade de cada caso de falha.
- Exemplo no próximo slide.

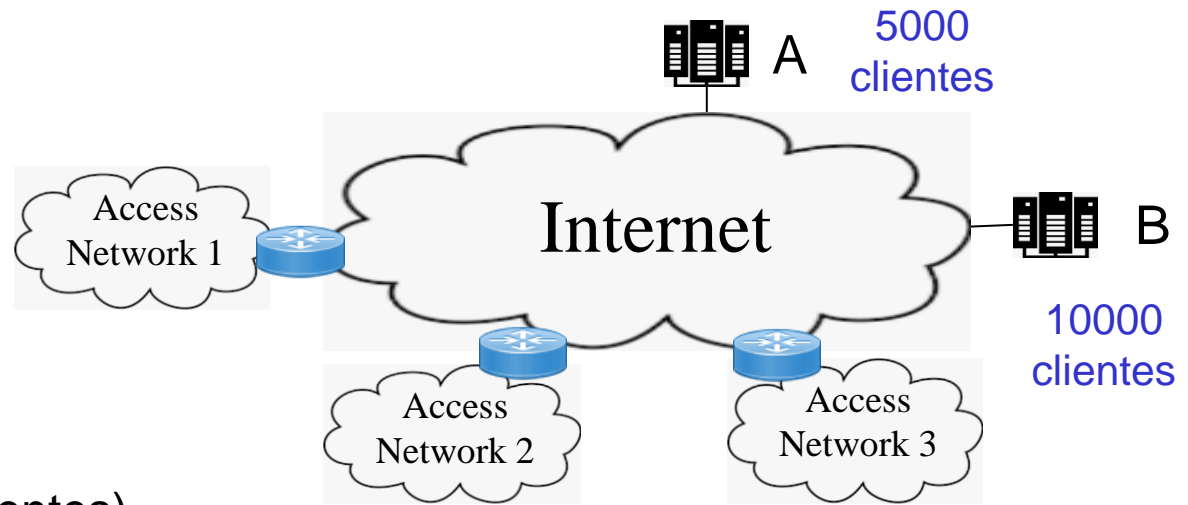
Exemplo

Considere-se um serviço de *streaming* de filmes com 2 servidores ilustrado na figura: um servidor hospedado no *Data Center A* com uma disponibilidade de 0.99 e uma capacidade de 5000 clientes e outro servidor hospedado no *Data Center B* com uma disponibilidade de 0.999 e uma capacidade de 10000 clientes.

1. Determine a capacidade média C do serviço (em número de clientes).
2. Determine a probabilidade de bloqueio P do serviço considerando que os clientes geram 5000, 3000 e 2000 pedidos por hora (nas redes de acesso 1, 2 e 3, respetivamente) e que os filmes têm uma duração média de 90 minutos.



Exemplo



1. Capacidade média C do serviço (em número de clientes).

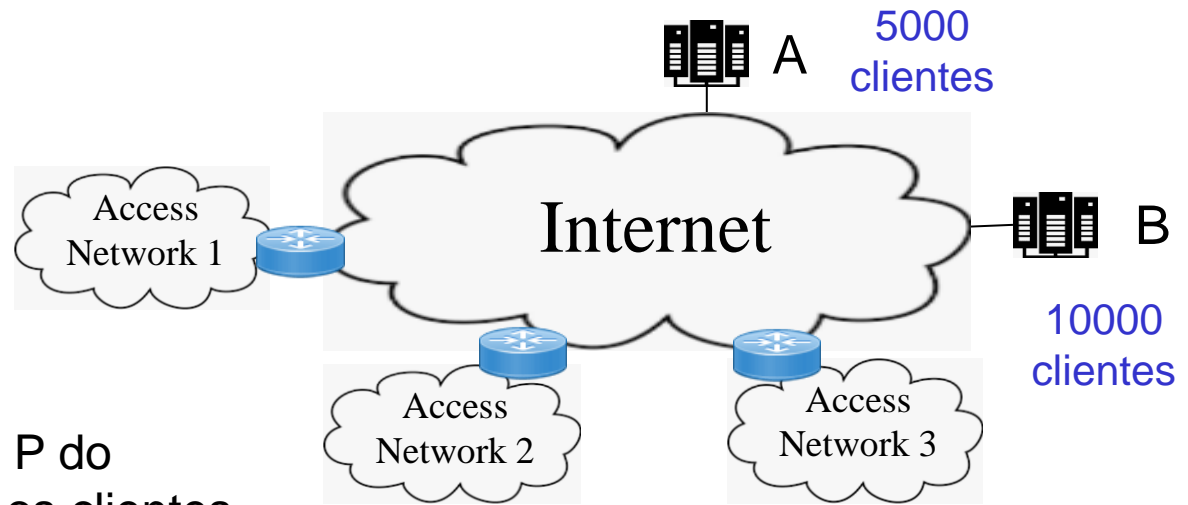
Probabilidade dos 2 servidores estarem operacionais

Probabilidade do servidor A estar operacional e o servidor B estar inoperacional

Probabilidade do servidor A estar inoperacional e o servidor B estar operacional

$$\begin{aligned} C &= (5000+10000) \times 0.99 \times 0.999 + 5000 \times 0.99 \times (1-0.999) + 10000 \times (1-0.99) \times 0.999 = \\ &= 15000 \times 0.989 + 5000 \times 0.00099 + 10000 \times 0.00999 = 14939.85 \text{ clientes} \end{aligned}$$

Exemplo



2. Probabilidade de bloqueio P do serviço considerando que os clientes geram 5000, 3000 e 2000 pedidos por hora (nas redes de acesso 1, 2 e 3, respetivamente) e que os filmes têm uma duração média de 90 minutos.

$$\lambda = 5000 + 3000 + 2000 = 10000 \text{ pedidos/hora}$$

$$1/\mu = 1.5 \text{ horas}$$

$$\lambda/\mu = 10000 \times 1.5 = 15000$$

$$E(\lambda/\mu, m) = \frac{(\lambda/\mu)^m / m!}{\sum_{i=0}^m (\lambda/\mu)^i / i!}$$

$$P = E(\lambda/\mu, 15000) \times 0.989 + E(\lambda/\mu, 5000) \times 0.00999 + E(\lambda/\mu, 10000) \times 0.00099 =$$

$$= 0.0065 \times 0.989 + 0.6667 \times 0.00099 + 0.3335 \times 0.00999 = 0.0104 = 1.04\%$$

↑
Probabilidade de bloqueio
para capacidade de
15000 clientes

↖
Probabilidade de bloqueio
para capacidade de 5000
clientes

↖
Probabilidade de bloqueio
para capacidade de
10000 clientes