Universidade Federal do Rio de Janeiro

Núcleo de Computação Eletrônica

Hélio José Galvão e Silva

Contingenciamento e Balanceamento de Carga para Sites e Serviços: uma Análise Comparativa entre DNS Adaptativo e BGP

Rio de Janeiro

2010

Hélio José Galvão e Silva

Contingenciamento e Balanceamento de Carga para Sites e Serviços: uma Análise Comparativa entre DNS Adaptativo e BGP

Monografia apresentada para obtenção do título de Especialista em Gerência de Redes de Computadores no Curso de Pós-Graduação Lato Sensu em Gerência de Redes de Computadores e Tecnologia Internet do Núcleo de Computação Eletrônica da Universidade Federal do Rio de Janeiro – NCE/UFRJ.

Orientador:

Prof. Moacyr Henrique Cruz de Azevedo, M.Sc., UFRJ, Brasil

Rio de Janeiro

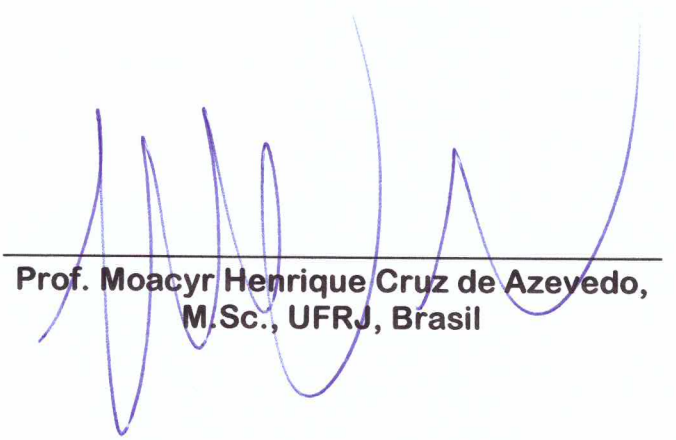
2010

Hélio José Galvão e Silva

Contingenciamento e Balanceamento de Carga para Sites e Serviços: uma Análise Comparativa entre DNS Adaptativo e BGP

Monografia apresentada para obtenção do título de Especialista em Gerência de Redes de Computadores no Curso de Pós-Graduação Lato Sensu em Gerência de Redes de Computadores e Tecnologia Internet do Núcleo de Computação Eletrônica da Universidade Federal do Rio de Janeiro – NCE/UFRJ.

Aprovada em agosto de 2010.



Prof. Moacyr Henrique Cruz de Azevedo, M.Sc., UFRJ, Brasil

“Quanto menos inteligente um homem é,
menos misteriosa lhe parece a existência."
(Schopenhauer)

RESUMO

SILVA, Hélio José Galvão e. **Contingenciamento e Balanceamento de Carga para Sites e Serviços: uma Análise Comparativa entre DNS Adaptativo e BGP**. Monografia (Pós-Graduação em Gerência de Redes e Tecnologia Internet). Núcleo de Computação Eletrônica, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2010.

Com o crescimento da rede mundial de computadores e dos serviços prestados através dela, a redundância dos canais de acesso tem sido um tema bastante discutido. Estes mecanismos possuem alta relevância, afinal incidem diretamente sobre as questões de disponibilidade de aplicações e demais serviços oferecidos pelas diversas organizações presentes na grande rede.

Esta pesquisa avaliará dois métodos propostos, onde entenderemos suas vantagens, desvantagens e cenários de utilização.

ABSTRACT

SILVA, Hélio José Galvão e. **Contingenciamento e Balanceamento de Carga para Sites e Serviços: uma Análise Comparativa entre DNS Adaptativo e BGP**. Monografia (Pós-Graduação em Gerência de Redes e Tecnologia Internet). Núcleo de Computação Eletrônica, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2010.

With the growth of the worldwide computer network and services provided through it, the redundancy of access channels has been a subject widely discussed. These mechanisms are highly relevant, after all directly affect the issues of availability of applications and other services offered by many organizations in the large network.

This research will evaluate two proposed methods, where we will understand its advantages, disadvantages, and usage scenarios.

LISTA DE FIGURAS

LISTA DE ABREVIATURAS E SIGLAS

AS – Autonomous Systems
ASN – Autonomous System Number
BGP – Border Gateway Protocol
CIR – Commited Information Rate
DDR – Dial-on-Demand Routing
DNS – Domain Name System
EIR – Exceeded Information Rate
HA – High Availability
HTTP – Hyper Text Transfer Protocol
ICMP – Internet Control Message Protocol
IP – Internet Protocol
ISP – Internet Service Provider
LACNIC – Latin America and Caribbean Internet Addresses Registry
LP – Linha Privada
RAS – Remote Access Server
RFC – Request for Comments
RTT – Round Trip Time
SLDD – Serviço
TCP – Transmission Control Protocol
TTL – Time-to-Live

SUMÁRIO

1 INTRODUÇÃO

1.1 OBJETIVO

O propósito desta pesquisa é reduzir o esforço administrativo e o investimento financeiro para ambientes cuja alta disponibilidade (*high availability* ou HA) dos sistemas de informação é um fator determinante.

Há diversas abordagens com este objetivo, e podemos incluir a montagem de *clusters* de servidores para os sistemas (*software*), equipamentos redundantes (hardware), fornecimento ininterrupto de energia elétrica, climatização adequada e controles de segurança. Mas a base deste trabalho está definida no caminho percorrido até os serviços, ou seja: questões de conectividade, links de acesso, rede, ou simplesmente *internet*. De fato, mesmo com todos os recursos de redundância citados anteriormente, os sistemas tornar-se-ão inacessíveis caso existam problemas nos canais de acesso.

A resolução de nomes em caráter adaptativo tem por objetivo entregar ao nó solicitante a melhor resposta para o recurso desejado. Em suma, entende-se por melhor resposta fatores como menor número de saltos e com menor tempo de acesso, além do estado e utilização do enlace. Logo, todos estes itens determinarão o resultado em face de uma requisição.

1.2 RELEVÂNCIA

A disponibilidade é uma propriedade que visa garantir o acesso à informação pelos usuários autorizados sempre que necessário, e em termos de negócio, seja público ou privado, a questão permanece inalterada.

2 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados alguns conceitos sobre resolução de nomes, endereçamento e conectividade entre as redes.

2.1 PROCESSOS DE ROTEAMENTO BÁSICO

O acesso aos recursos em uma rede é feito utilizando-se um endereço previamente fornecido, e mecanismos de entrega (definição da rota/caminho) serão acionados para que os datagramas sejam enviados e recebidos. Não serão abordados os métodos de roteamento (direto ou indireto) bem como os algoritmos envolvidos. Apenas faz-se entender a idéia de uma rede de pacotes na qual os dados trafegam por diversos equipamentos e que por muitas vezes podem chegar ao destino fora de ordem. Outro dado importante é que o caminho percorrido nem sempre é o mesmo, o que gera discretas variações nos tempos de acesso para a troca de informações.

2.2 DISTRIBUIÇÃO DOS ENDEREÇOS DE REDE E SISTEMAS AUTÔNOMOS

Segundo a RFC 1918 [1], os endereços lógicos em uma rede de datagramas IP são divididos em classes e distinguem-se entre públicos e privados. Por se tratar de um recurso único e limitado, determinados órgãos regulam a sua distribuição no âmbito público. Tal distribuição é efetuada em blocos, o que dá origem aos Sistemas Autônomos ou simplesmente "AS" como define a RFC 1930 [2]. Um Sistema Autônomo é uma área administrativa, regida por seus equipamentos devidamente endereçados e que oferecem recursos através da rede de dados (fig. 1). Os sistemas autônomos são identificados por números (ASNs), cadastrados nos órgãos de registro de endereços (LACNIC, por exemplo), e o processo exige pré-requisitos técnicos e administrativos

além de tarifas de manutenção. É comum definirmos operadoras de telecomunicações ou provedores de internet (ISP) como sistemas autônomos, porém o mesmo não se aplica às pequenas empresas ou usuários residenciais conectados à rede mundial.

Alguns AS brasileiros:

AS 8167 Brasil Telecom
AS 1916 RNP (Rede Nacional de Pesquisa)
AS 10429 Telefônica
AS 15201 UOL
AS 17222 Mundivox
AS 22055 Banco Central do Brasil

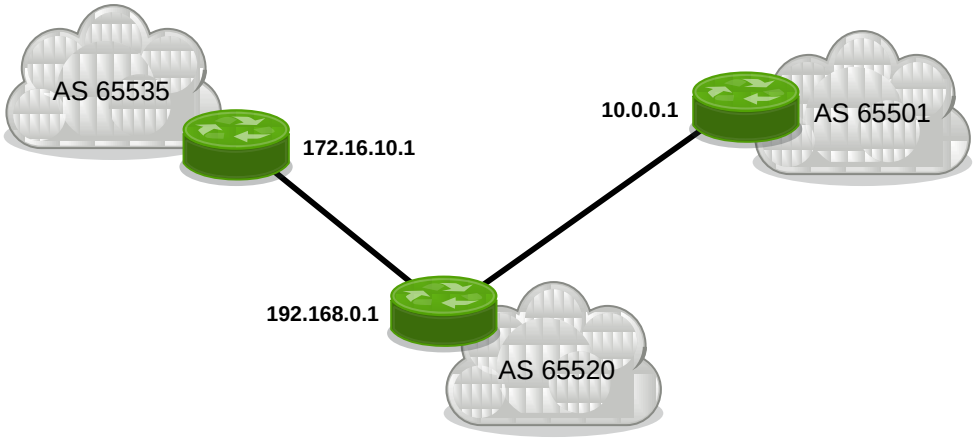


Figura 1 – Esquema de conexão entre ASs

2.3 COMUNICAÇÃO ENTRE OS SISTEMAS AUTÔNOMOS

De fato, para que haja conectividade entre todos os Sistemas Autônomos, é necessário eleger um mecanismo de divulgação dos endereços pertencentes a estes sistemas. Este processo evita que os Sistemas Autônomos se tornem ilhas isoladas com recursos inalcançáveis. Em suma, a divulgação do bloco de endereços públicos tem por finalidade apresentar ao mundo a existência

de um Sistema Autônomo e o caminho (rota) para alcançá-lo. Podemos dizer então que a internet - a rede mundial de computadores - é formada por diversos ASs, interconectados e relacionados. Um protocolo amplamente difundido e que corresponde às expectativas de comunicação entre os sistemas autônomos é o BGP (*Border Gateway Protocol*). O BGP é definido pelas RFCs 1771 [3] e 1772 [4].

> ***"O funcionamento da internet global depende atualmente do uso do protocolo BGP-4, que, por sua vez, requer a sinergia entre todos os ASs para funcionar de forma estável, precisa e segura. Caso ocorram problemas com as informações de roteamento que trafegam nas mensagens do BGP trocadas entre ASs, estas redes podem sair do ar ou ficar inacessíveis temporariamente, causando grandes prejuízos para todos os usuários da internet, principalmente para empresas." (Rede Nacional de Pesquisa, 2001) [5]***

2.4 ARQUITETURA PARA RESOLUÇÃO DE NOMES

O DNS, ou sistema de nome de domínios, é um esquema de nomenclatura hierárquico baseado em uma árvore e é definido pelas RFCs 1034 [6] e 1035 [7]. O DNS opera com um arranjo cliente/servidor e o processo de resolução de nomes em uma rede IP tem por objetivo resgatar o endereço lógico vinculado a um nome de recurso, como, por exemplo, uma página contendo informações (texto, imagens etc.) ou um repositório de dados. O endereço de rede é retornado mediante uma consulta a um tipo de registro, normalmente um registro do tipo *host* (A). Este registro associa um determinado nome no padrão da árvore do sistema de nome de domínios a um endereço de rede. Por exemplo, o valor "www.empresa.com.br" é um nome de recurso baseado nas especificações do DNS (fig. 2).

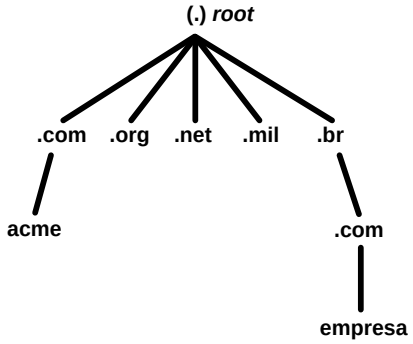


Figura 2 – Árvore DNS

O cliente inicia o processo através de um *DNS resolver*, que é basicamente um software e/ou serviço embutido no sistema operacional do nó em questão. A consulta atinge então um determinado servidor DNS na rede, cuja função é atender tais requisições (fig. 3).

Um dado relevante sobre a resolução de nomes é que o resultado da consulta fica armazenado pelo nó solicitante por um tempo determinado. Para este tempo - computado em segundos (s) - é dado o nome de *time-to-live*, ou simplesmente TTL. Findado este prazo o nó fará uma nova consulta para o mesmo recurso. Isto garante que possíveis mudanças no endereço lógico do destinatário sejam transparentes para os usuários do serviço. É possível ainda associar diversos endereços de rede a um nome DNS. Neste caso, um único nome de recurso estará vinculado a dois ou mais endereços lógicos, que serão entregues ao solicitante:

www.empresa.com.br. A 172.16.10.55 TTL=60
www.empresa.com.br. A 172.16.10.60 TTL=60

Em uma implementação tradicional, um mecanismo de *round robin* é utilizado e, portanto, os endereços serão entregues ao solicitante de forma cíclica, sem avaliar critérios como o estado atual do nó (ligado ou desligado) ou o

estado do enlace (tempo de acesso, número de saltos etc.). Logo, não podemos considerar a entrega cíclica de endereços como uma técnica efetiva para balanceamento de carga ou aumento da disponibilidade.

A arquitetura DNS prevê a utilização de pelo menos dois servidores para a resolução de nomes, portanto a disponibilidade destes não é um fator relevante para esta pesquisa e sim a maneira pela qual atendem às requisições.

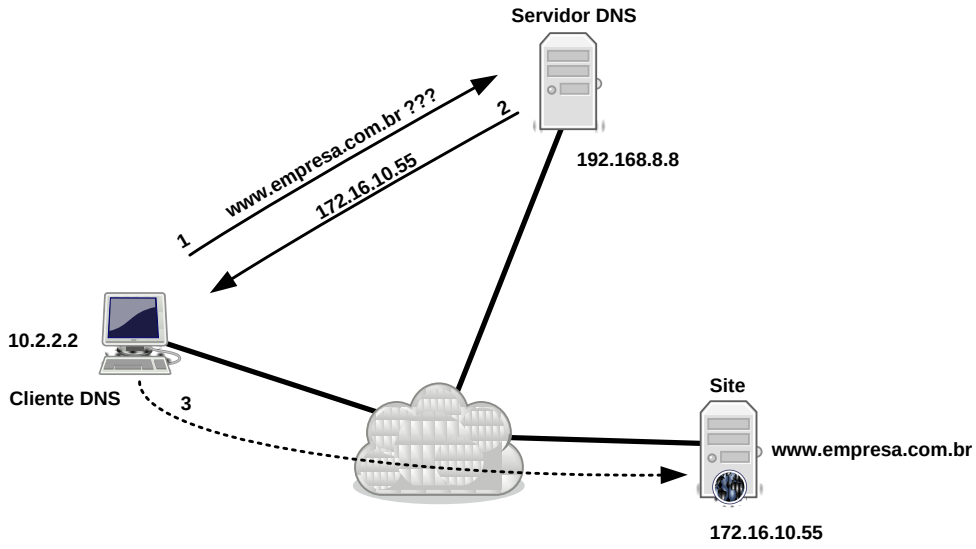


Figura 3 – Processo básico de resolução de nomes

2.5 DESAFIOS DO ACESSO REDUNDANTE

São considerados *multihomed* os ambientes de redes nas corporações que possuem conexões internet redundantes, para um mesmo provedor de acesso (ISP) ou provedor de acesso internet diferentes (dois ou mais ISPs).

Diversas complicações são introduzidas com um projeto de redes que inclui múltiplas conexões para a internet. O primeiro desafio são as complicações

inerentes ao esquema do endereçamento IP que são resultantes de dois planos empregados em sites *multi-homed*:

a) uma única rede IP é atribuída para a rede interna. Isto exigirá comunicação e cooperação entre os ISPs participantes do cenário *multi-homed* de forma a anunciar as devidas rotas desta rede IP para o resto da internet.

b) cada ISP atribui à rede interna uma faixa de endereços IP distinta. Sendo assim, múltiplas faixas de endereços IP estarão ativas simultaneamente para a rede interna.

Cada um dos esquemas apresenta o seu próprio conjunto de desafios, e no primeiro caso, quando um único endereço IP é utilizado, os ISPs deverão coordenar e trabalhar juntos para anunciar a devida rota desta única sub-rede para o restante da internet. Além disto, deverão garantir que todos os enlaces poderão ser utilizados para o tráfego de entrada.

Se somente um ISP estiver sendo utilizado para encaminhar o tráfego de entrada para a rede interna, desta forma, parte das motivações e benefícios de um ambiente *multihoming* não serão realizados.

No segundo caso, quando faixas distintas de endereços IP vindas de cada ISP são atribuídas para a rede interna, haverá um problema sobre qual faixa será utilizada para o tráfego de saída. Se uma faixa de rede interna atribuída por um ISP estiver sendo utilizada e este enlace falhar, não ocorrerá a resposta do tráfego de retorno para a rede, uma vez que o mundo externo somente conhece a faixa acessível pelo ISP cujo enlace acabou de falhar.

Logo, há também o problema sobre quais endereços IP deverão ser anunciados para a internet visando o tráfego de entrada. Por exemplo, se a

rede interna possui um servidor de aplicação que precisa ser acessado por usuários da internet, que endereço IP deverá ser atribuído para o nó em questão? Se for atribuído de um dos ISPs, o servidor ficará inacessível caso o ISP responsável por aquela faixa ficar indisponível seja qual for o motivo. Mas, se um endereço IP de ambas as faixas for anunciado, então mecanismos de redundância precisarão ser analisados e implementados de maneira eficiente.

Além dos problemas que são resultantes dos cenários expostos, uma rede *multihomed* poderá não usufruir de alguns benefícios importantes, mesmo que possua diversos enlaces de saída. Como exemplo, podemos citar o seguinte caso: mesmo com o mais sofisticado dos protocolos de roteamento, o verdadeiro balanceamento de carga jamais será alcançado pelos múltiplos enlaces de saída. Os protocolos de roteamento desconhecem as condições de carga de tráfego dos enlaces e as sessões ativas em cada um deles. Qualquer decisão para o balanceamento de carga que um protocolo de roteamento fizer será considerada, no máximo, como "compartilhamento de carga". Neste caso, foge-se um pouco do objetivo que é o balanceamento de carga e a alta disponibilidade do ambiente.

Ainda retratando o tráfego de saída, podemos afirmar que alguns nós na internet são acessados de maneira mais eficiente através de um ISP do que outro(s). Logo, mesmo que os protocolos de roteamento estejam configurados de maneira apropriada visando desempenho e boas práticas, poderão até conhecer algumas informações sobre o alcance ou proximidade, mas jamais poderão combinar vários fatores como número de saltos até o

destino, latência e carga atual no roteador, para então tomar decisões de roteamento.

Em contrapartida, para o tráfego de entrada um ISP poderá oferecer um caminho melhor para a rede interna do que outro(s) ISP(s). Neste caso não há um mecanismo inteligente presente em uma rede *multihomed* que combine informações de proximidade, DNS e carga no roteador em relação à decisão sobre qual ISP proverá o melhor acesso à rede externa.

De fato, os cenários *multihoming* internet tornam-se projetos complexos tendo em vista não só os esquemas de endereçamento IP, protocolos de roteamento e DNS, mas também pela questão da subutilização de todos os recursos envolvidos.

3 SOLUÇÕES DE CONTINGÊNCIA PARA WAN

O planejamento da contingência dos serviços de uma organização deve levar em consideração a comunicação dos seus circuitos WAN, os quais são geralmente utilizados para interligação de filiais ou da própria organização com outras organizações para troca de informações de serviços dependentes. A arquitetura de uma rede WAN resiliente deve prever redundância para garantir continuidade dos serviços, agregar balanceamento de tráfego, e uma rápida recuperação automática da comunicação para redução dos tempos de indisponibilidade.

Mas, existem alguns fatores nas redes WAN que mostram-se desafiadores no âmbito do contingenciamento dos serviços. O principal deles é que a maior parte do provimento de acesso através de redes WAN é feita pelas operadoras de telecomunicações. Estes serviços são taxados, e quanto maior a necessidade por altas taxas de transmissão e número de circuitos, maior o valor gasto pelas organizações com estas redes. Aliado a isto, as redes WAN, por serem de propriedade das operadoras, possuem as características de disponibilidade de acordo com o seu ambiente interno. Este ambiente não é controlado pela organização e está sujeito às suas próprias falhas, gerando conseqüências ao cliente final.

Outro fator a citar é a largura de banda disponível em circuitos de longa distância, se comparada à largura de banda disponível nas redes locais. Enquanto as redes locais possuem altas taxas de transmissão, as WANs ainda estão um pouco limitadas e possuem um alto custo mensal. Para minimizar este problema, podem ser empregadas técnicas ou diversas

tecnologias para prover Qualidade de Serviço (QoS) nas redes WAN com o objetivo de regular o tráfego e reduzir a necessidade do aumento da largura de banda, baixando o custo dos circuitos. No contexto desta pesquisa, os links WAN são de extrema importância para conexões entre sites remotos, provendo a redundância de acessos e fornecendo continuidade do negócio.

Salientada a importância da rede WAN para a contingência, a seguir são descritas algumas abordagens de contingência para conexões de rede em longas distâncias.

3.1 LINKS DE COMUNICAÇÃO REDUNDANTES

Entende-se por links redundantes o acesso físico e lógico duplicado para a interligação de redes locais distantes entre si, empregando tecnologias de rede WAN. Os links redundantes podem ser compostos por combinações de elementos distintos, como a utilização de links dedicados e com taxa de transmissão elevadas para o ambiente de TI principal da organização, como por exemplo um site principal, e links dedicados, ou dinâmicos, com taxas de transmissão baixas para o site contingência. Porém, é importante que estas limitações estejam bem definidas, documentando no plano de contingência que somente serviços importantes trafegarão pelos links alternativos, ou que o serviço será provido de forma não tão eficiente por estar em estado de contingência.

Devemos levar em consideração que o ambiente interno da rede WAN da organização, que inclui circuitos e equipamentos, deve ser redundante tanto no acesso físico, incluindo-se roteadores, cabos e até tubulações por onde transitam, como no acesso lógico, que está baseado na tecnologia

empregada. Tal consideração se deve ao fato de que um acidente pode comprometer ambos os enlaces redundantes, eliminando a continuidade da comunicação.

3.2 DIAL-BACKUP

Conexões de longa distância podem ser realizadas através da rede pública de telefonia, utilizando-se um acesso discado (dial-up). A forma de contingenciar enlaces WAN através de conexões discadas chama-se *dial-backup*.

A implantação de *dial-backup* é uma solução de baixo custo para prover a alta disponibilidade de uma conexão WAN. Com esta solução não é necessário contratar um enlace redundante para a contingência, o que aumentaria os custos de manutenção. Com esta alternativa, só haveria gastos com o tempo de utilização do enlace redundante, ou seja, o tempo de duração da chamada telefônica. Para alcançar este objetivo é necessário possuir uma interface conectada a um modem, e este à rede pública de telefonia. A segunda etapa seria a configuração de um roteador para que a conexão seja estabelecida no caso da queda de um dos enlaces.

Há duas abordagens relevantes para o recurso de *dial-backup: backup interface* e *dial-on-demand routing* (DDR).

A primeira abordagem, *backup interface*, aciona a conexão discada sempre que é percebido o estado de queda do enlace que está sendo contingenciado. Sendo assim, é necessário que se perceba corretamente o estado do enlace WAN, pois se o protocolo está ativo mas por algum motivo a comunicação não pode ser estabelecida, a conexão discada não será acionada. Também é

possível acionar a conexão discada quando o enlace primário está sobrecarregado e se deseja balancear a carga pela conexão discada.

***Dial-on-demand routing*, ou DDR, estabelece a conexão discada somente quando necessário, através da categorização de tráfego e/ou necessidade de utilização do enlace WAN que está sendo contingenciado. Esta é uma maneira de se reduzir custos, pois a conexão só será acionada se houver necessidade da utilização do link através da definição de que tipo de tráfego é considerado importante e que deve trafegar pela rede. O DDR é utilizado geralmente em organizações que possuem pouco tráfego no enlace WAN e que não possuem necessidade de conexão permanente com o site remoto.**

A estratégia de implementação de *dial-backup* para contingenciar seus links principais pode ser melhorada pelas organizações com a utilização de um elemento concentrador de conexões discadas, como um *Remote Access Server* (RAS). Esse tipo de equipamento possui a característica de poder receber conexões de nós remotos e conectá-los a uma LAN, por exemplo, tornando-se o ponto concentrador de conexões temporárias.

3.3 UTILIZAÇÃO DE REDES DETERMINÍSTICAS

Segundo BRUSCATTO [8], as redes WAN determinísticas são formadas por circuitos dedicados, especializados e exclusivos, ponto-a-ponto ou ponto-multiponto, transmitindo dados entre equipamentos pré-estabelecidos e conhecidos. Este serviço é conhecido no ambiente das operadoras de telecomunicações pela sigla SLDD (Serviço por Linha Dedicada Digital, ou LPs - Linhas Privadas).

As redes determinísticas possuem baixa latência entre os dados trafegados, pois os equipamentos das operadoras por onde passam os circuitos encaminham os dados para pontos definidos. Tal característica compatibiliza estas redes com aplicações sensíveis a baixo retardo, como aplicações multimídia (voz e vídeo).

Por não compartilhar recursos com nenhuma outra rede de outra organização, e somente com a operadora, estas redes possuem latência conhecida e segurança implícita, exatamente por serem dedicadas. A utilização destas redes como uma solução de enlaces redundantes implica em estabelecer dois ou mais pontos dedicados e permanentes entre duas redes.

A grande vantagem de tal solução é o enlace permanente estabelecido entre as redes, com taxas de transmissão previamente estabelecidas e constantes. Outras vantagens incluem as baixas latências dos circuitos e a segurança dos dados trafegados, pois as conexões não são compartilhadas por nenhuma outra organização.

3.4 UTILIZAÇÃO DE REDES ESTATÍSTICAS

Segundo BRUSCATTO [8] as redes estatísticas baseiam-se na multiplexação estatística, associada a uma garantia mínima de tráfego, onde diferentes canais de comunicação compartilham os recursos físicos e a capacidade de transmissão, ocupando-os dinamicamente em função da demanda e da garantia mínima.

Este tipo de rede geralmente tem custo inferior às redes determinísticas pois as operadoras de telecomunicações compartilham os recursos destas redes

com vários clientes. A tecnologia de rede que serve como o maior exemplo deste tipo de rede é o *Frame Relay*. Os circuitos virtuais permanentes, oriundos desta tecnologia, possuem um parâmetro chamado CIR (*Commited Information Rate*), que é o volume mínimo de informação que a rede da operadora se compromete a transmitir, independente do tráfego total gerado na rede. Já o parâmetro EIR (*Excess Information Rate*), é a quantidade de informação que excede o CIR contratado e que a rede é capaz de transmitir, e que depende diretamente da ocupação estatística dos canais virtuais que compartilham o meio.

As redes estatísticas são boas soluções para conexão de pontos remotos, já que possuem baixo custo e a comunicação pode ser alterada para alcançar outros pontos terminais em caso de falhas na rede.

3.5 UTILIZAÇAO DO DNS ADAPTATIVO PARA REDUNDÂNCIA

Esta técnica utiliza os conceitos básicos da resolução de nomes, porém agrega redundância e balanceamento de carga. O objetivo é alcançado empregando-se um serviço adaptativo de resolução de nomes ao invés do DNS tradicional (*bind* para ambientes UNIX ou DNS Microsoft, por exemplo). Assim como o DNS tradicional, este serviço permite a utilização de endereços lógicos distintos para os serviços da rede, bem como operar com enlaces de diferentes provedores de acesso. De fato, a utilização de ISPs distintos já possibilita redundância, mas o diferencial deste método é avaliar qual acesso está em melhores condições em termos de disponibilidade e tráfego. Ao entender a atual conjuntura dos pontos de acesso, o servidor DNS poderá então entregar a melhor resposta ao cliente. E, esta resposta é

um endereço lógico (IP) resolvido e entregue não de forma cíclica, mas de forma dinâmica após análise dos enlaces. O parâmetro mais comum para avaliação dos enlaces é o envio de um "sinal de vida" (*keep-alive*) para determinados endereços - públicos ou não (fig. 4). O interessante neste ponto é que a utilização de endereços públicos para verificação dos enlaces permite que o serviço de DNS Adaptativo teste a conectividade através de um ponto distante de sua rede local. Se este teste fosse restrito à rede local ou à nós muito próximos, não haveria como garantir se há acesso à internet ou não, por exemplo. Estes *keep-alives* podem ser feitos com a ferramenta *ping* (ICMP *echo/reply*) e comandos HTTP entre outros, e um intervalo de repetição é definido para que o teste seja efetuado algumas vezes com segurança antes de desativar um determinado enlace (figs. 4 e 5).

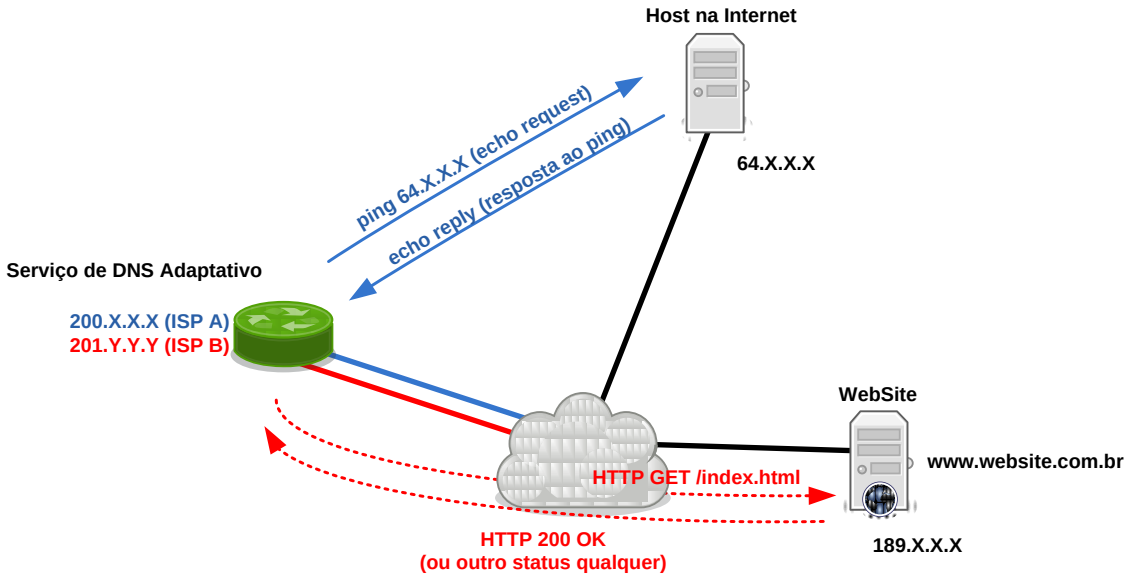


Figura 4 – Execução de *keep-alives* para verificação dos enlaces

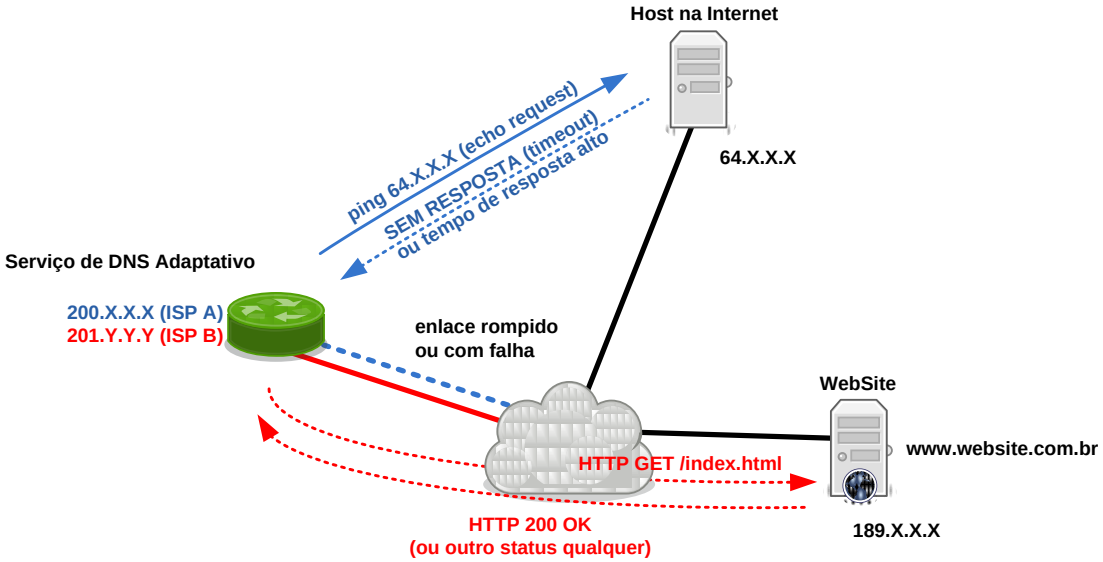


Figura 5 – Um enlace é desativado ao se constatar uma queda

Baseado nos tempos, tipos de resposta e disponibilidade dos enlaces, o servidor avaliará se um determinado endereço de rede associado a um recurso será entregue ou não a um cliente. É neste ponto que se determina a questão adaptativa da resolução de nomes (fig. 6).

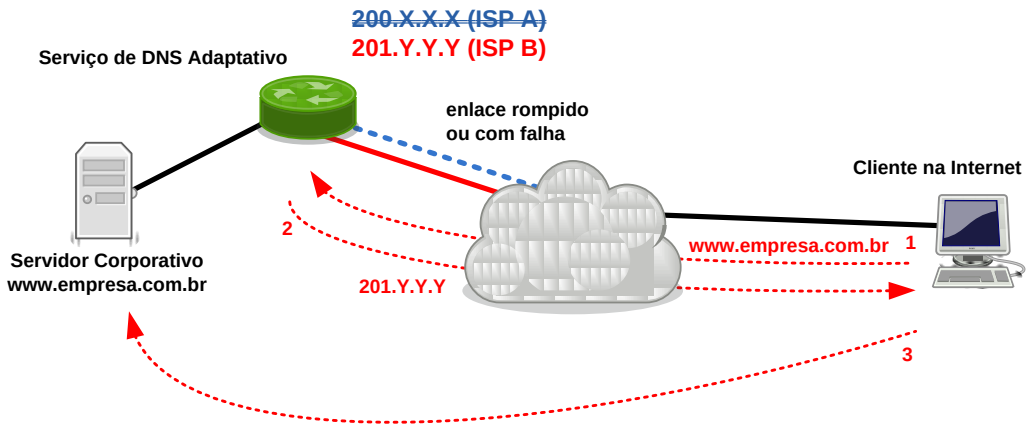


Figura 6 – O cliente não receberá o endereço IP do enlace que está com falha

Supondo que os dois enlaces estivessem operacionais, o endereço IP entregue ao cliente poderia pertencer tanto ao provedor A (200.X.X.X - ISP A) quanto ao provedor B (201.Y.Y.Y - ISP B).

Outro dado importante é que o tempo de vida para os registros retornados aos clientes é sempre zero (TTL=0), o que os obriga a realizar novas consultas para o mesmo registro já que não o armazenam em *cache*. Isto garante que o cliente sempre tenha em mãos uma resposta atualizada. Supondo que em um determinado momento o cliente obteve como resposta de uma consulta o IP 200.X.X.X, e, após alguns minutos o enlace detentor deste endereço não estiver mais operacional, numa arquitetura tradicional o recurso estaria inacessível (o Servidor Corporativo é um exemplo, como na fig. 6). Mesmo com um mecanismo de *round robin* este endereço continuaria a ser disponibilizado para os clientes, ocasionando falha no acesso. No caso do DNS Adaptativo este endereço IP não seria entregue pois a falha foi detectada e o cliente retornará sempre que necessário para realizar uma nova consulta.

A questão da redundância dos servidores DNS permanece inalterada, visto que na implementação tradicional já é prevista e exigida. Para todo e qualquer registro é necessário que informemos pelo menos dois servidores de nomes (NS) distintos, e de preferência geograficamente dispersos.

3.6 UTILIZAÇAO DO DNS ADAPTATIVO E ALGORITMOS DE PROXIMIDADE

Um roteamento eficiente resulta em baixa variação no atraso para a entrega de pacotes, assim, o algoritmo de controle de fluxo poderá receber um tráfego maior em suas interfaces. Por outro lado, um mecanismo eficiente de

controle de fluxo tende a rejeitar um tráfego excessivo que porventura gere atrasos na entrega dos pacotes, saturando os recursos da rede. Fica claro então que os mecanismos de roteamento e controle de fluxo estão inter relacionados. Os problemas nos modelos clássicos de roteamento estático e dinâmico foram estudados por Gallagher [9], Bertsekas [10], Segall [11], entre outros. Estes modelos consideram principalmente o número mínimo de saltos que um pacote precisa seguir entre uma origem e um destino para minimizar o atraso. Mas o congestionamento e a perda de pacotes não são avaliados, assim como o RTT.

Logo, além dos benefícios descritos anteriormente com a utilização do DNS Adaptativo, podemos agregar outra funcionalidade que está associada a um Algoritmo de Proximidade. A proximidade em redes visa alcançar um determinado recurso utilizando-se o menor *round trip time* (RTT), ou seja, o menor tempo de viagem entre uma origem e um destino. Este recurso é interessante pois o pacote poderá ser endereçado às redes mais próximas que possuam o conteúdo solicitado sem precisar viajar longas distâncias (redes intercontinentais, por exemplo). Do ponto de vista dos provedores (ISPs) pode representar economia e, para o cliente, eficiência nos tempos de resposta.

Basicamente, o nó que executa este algoritmo fará uma verificação do endereço de destino em busca do melhor caminho avaliando não só o estado do enlace, mas também o *round trip time* dos pacotes. Esta verificação será executada durante a primeira conexão a um *host* remoto, visto que uma análise prévia poderia resultar em *timeouts* para o cliente. Os resultados

serão armazenados em uma tabela e constantemente atualizados para que as próximas conexões utilizem este recurso. No exemplo da figura 7, o equipamento avalia ambas as saídas para a internet e determina que o enlace do provedor “A” (ISP A) possui o melhor caminho para o endereço de destino solicitado pelo cliente da rede interna.

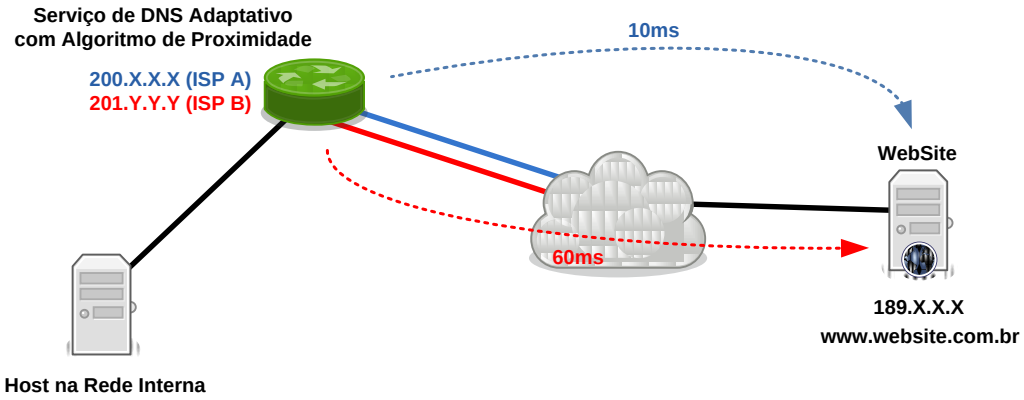


Figura 7 – Decisão do melhor caminho para entrega do pacote

O mesmo mecanismo será empregado nas conexões originadas na rede externa, ou seja, como o cliente externo fará a consulta ao servidor DNS adaptativo e este executa também um algoritmo de proximidade. A resposta (endereço IP) fornecida levará em conta diversos fatores incluindo o RTT entre o cliente remoto e o recurso local que se deseja acessar. Vale lembrar que o algoritmo deve calcular o RTT baseado no endereço IP do cliente e não do servidor DNS recursivo que efetuou a busca para este nó. Se isto não for verificado o resultado do cálculo será comprometido. Este problema pode ser resolvido através de sinalizações TCP, como análise de pacotes SYN, por exemplo.

3.7 AUTONOMIA DOS SISTEMAS

Em se tratando de um provedor de internet (ISP) ou grande instituição, há interesse pela autonomia de seus sistemas visto que esse processo resultará na alocação dedicada de um bloco de endereços de rede. Tais endereços serão especificados e entregues pelo órgão regulador local, como por exemplo, a LACNIC que detém o controle destes recursos na América Latina e Região do Caribe. Após a definição de um Sistema Autônomo, os roteadores de borda serão configurados de forma que o mundo externo conheça os endereços e o caminho (rota) até os recursos oferecidos por esta localidade. A interconexão dos Sistemas Autônomos se dá através de canais físicos (fibra óptica, condutores de cobre, etc.), mas a singularidade deste canal de acesso não define um mecanismo de redundância. Desastres naturais ou falhas nos equipamentos comprometeriam a estrutura, isolando o Sistema Autônomo. Nesta etapa do projeto, vale lembrar que o estabelecimento da autonomia exige um processo formal junto à autoridade reguladora destes recursos.

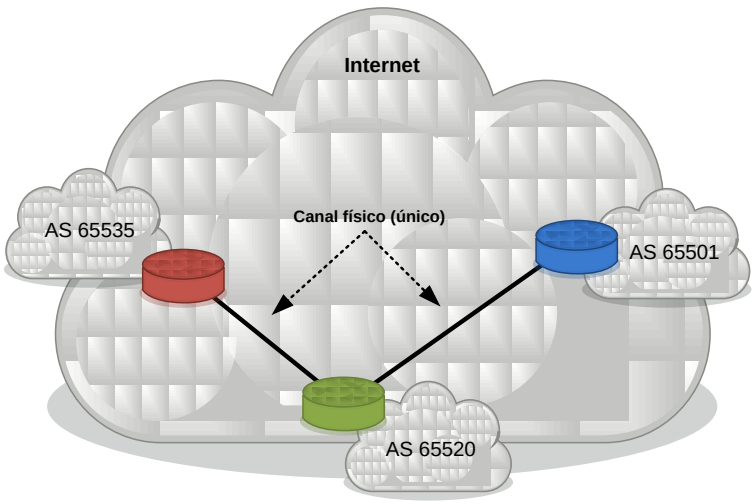


Figura 8 – Interconexão de ASs através de canais únicos

3.8 SISTEMAS AUTONÔMOS, BGP E CAMINHOS REDUNDANTES

Uma vez estabelecida a Área ou Sistema Autônomo com seus roteadores de borda devidamente configurados e anunciando seus endereços, é possível definir caminhos redundantes tanto para o acesso interno quanto externo. Graças ao BGP este anúncio de endereços é efetuado e permite a interconexão entre os diversos Sistemas Autônomos. Em suma, este protocolo divulga o bloco de endereços IP detidos pela entidade, que circulam na rede entre os nós vizinhos e permite que outros Sistemas Autônomos conheçam as rotas para alcançá-lo.

Para que este processo ocorra em sua normalidade, é necessária a viabilização de um canal de acesso dedicado para interconexão a outros Sistemas Autônomos (fig.8). Mas, em se tratando de acesso redundante, serão necessários dois ou mais enlaces dedicados. Desta maneira o bloco de endereços será divulgado por todos os enlaces contratados, oferecendo dois ou mais pontos de entrada e saída para a localidade em questão. O BGP oferece mecanismos de balanceamento de carga através de avaliações sobre o estado do enlace (tráfego, banda disponível, e número de saltos), e todos estes itens serão avaliados para a composição da métrica de sua tabela de roteamento.

A partir deste ponto, nota-se a elevação do custo e complexidade do projeto, visto que a contratação de enlaces dedicados entre ASs requer alocação de recursos financeiros e administrativos. A grande vantagem é que durante a queda de um dos pontos de conexão os endereços continuarão a ser anunciados normalmente pelo canal ativo, garantindo a redundância do

acesso. Mas, para que isto ocorra, a rede necessita de um tempo para que os nós vizinhos entendam o novo arranjo (topologia). Este período é chamado de tempo de convergência. Fatores como a utilização de equipamentos em regime de alta disponibilidade (HA) estão suprimidos nesta pesquisa, porém fazem parte da política de contingenciamento e balanceamento de carga. Encontra-se então, por exemplo, dois ou mais roteadores de borda, assim como outros dispositivos relativos à conectividade (*switches, hubs* etc.) duplicados.

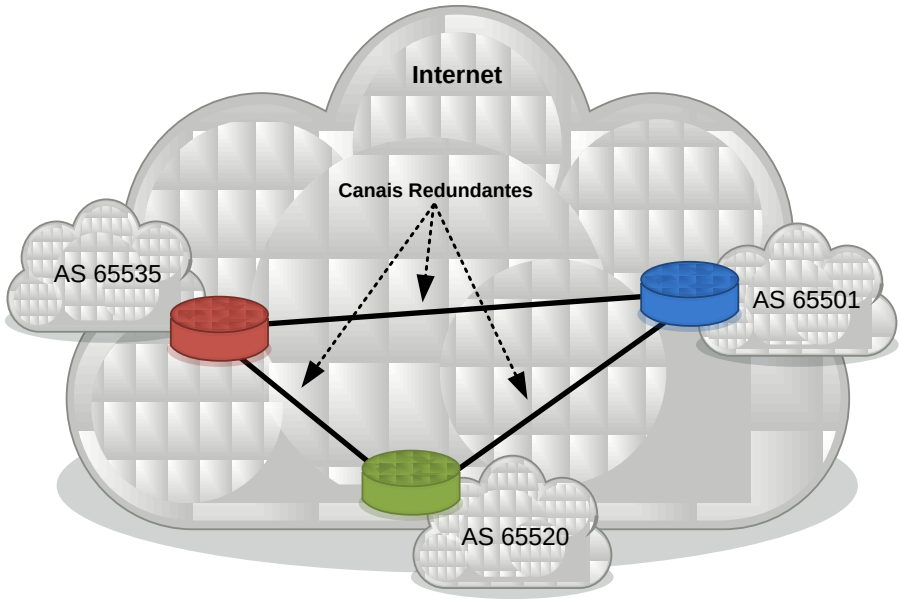


Figura 9 – Interconexão de ASs através de canais redundantes

4 METODOLOGIA DE PESQUISA

Quanto aos objetivos e meios, esta pesquisa é:

a) ***Exploratória,*** **pois tem por objetivo facilitar a familiaridade com o problema objeto da pesquisa de forma a permitir a construção de hipóteses ou tornar a questão mais clara;**
b) ***Bibliográfica,*** **através da utilização de variadas fontes de dados e bibliotecas online.**

5 CONCLUSÃO E ANÁLISE COMPARATIVA

5.1 UTILIZAÇÃO DO BGP

No arranjo tradicional das redes com seus Sistemas Autônomos, concluiu-se que a divulgação dos blocos de endereços detidos por estas áreas é imprescindível para seu funcionamento. Em conjunto com o protocolo BGP, os roteadores de borda serão responsáveis por manter a estrutura em constante operação, afinal é um processo dinâmico.

Percebe-se neste momento a necessidade de um administrador ou engenheiro de redes com mínimo conhecimento nos equipamentos de borda e protocolos para o anúncio das redes do Sistema Autônomo. Além dos custos envolvidos com recursos humanos, há ainda o custo para o estabelecimento de um AS, contratação de canais redundantes e blocos de endereços dedicados.

5.2 UTILIZAÇÃO DO DNS ADAPTATIVO

Em ambientes onde não há viabilidade financeira para a manutenção de sistemas autônomos e especialistas de redes, a utilização de servidores para resolução adaptativa se torna uma saída com excelente custo x beneficio.

Existem soluções de código aberto, como o Ibnamed (Unix) ou equipamentos dedicados, com custo total inferior à manutenção de um sistema autônomo. O custo estaria direcionado principalmente para os enlaces contratados, e por boa prática de operadoras distintas para garantir alta disponibilidade.

BIBLIOGRAFIA

[1] RFC 1918 (http://datatracker.ietf.org/doc/rfc1918) acessado em 08/05/2009

[2] RFC 1930 (http://datatracker.ietf.org/doc/rfc1930) acessado em 20/05/2009

[3] RFC 1771 (http://datatracker.ietf.org/doc/rfc1771) acessado em 02/10/2009

[4] RFC 1772 (http://datatracker.ietf.org/doc/rfc1772) acessado em 02/10/2009

[5] RNP (http://rnp.br/newsgen/0101/bgp4-dicas.html) acessado em 20/01/2010

[6] RFC 1034 (http://datatracker.ietf.org/doc/rfc1034) acessado em 19/06/2009

[7] RFC 1035 (http://datatracker.ietf.org/doc/rfc1035) acessado em 19/06/2009

[8] BRUSCATTO, Alexandre Carlos; MUNCINELLI, Gianfranco; PINHO, Antônio Carlos; RATTMANN, Amilton Carlos. Tutoriais Banda Larga e VoIP, 2005 (http://www.teleco.com.br/tutoriais/tutorialwan/default.asp) acessado em maio de 2010.

[9] R.G Gallagher and D.P.Bertsekas, Data Networks, PHI 1992.

[10] D.P.Bertsekas – "Dynamic models of shortest routing algorithms for communication networks with multiple destinations" IEEE Trans Automat Contr, Fev 1982 AC-27 p. 60-74.

[11] A.Segall, "The modeling of adaptive routing in data communication networks", IEEE Trans commun, vol.COM-25, p 85-95, Jan 1977.