Cláudio Fernando Sequeira Assunção
Licenciado em Ciências de Engenharia

Hybrid Link-State Path-Vector Protocol ++

Dissertação para obtenção do Grau de Mestre em Engenharia Electrotécnica e de Computadores
Mestrado Integrado em Engenharia Electrotécnica e de Computadores

Orientador: Luís Bernardo, Professor Auxiliar, FCT-UNL
Co-orientador: Pedro Amaral, Assistente, FCT-UNL

Júri:

Presidente:	Prof. Doutor Paulo da Costa Luís Fonseca Pinto
Arguente:	Prof. Doutor Luís Filipe dos Santos Gomes
Vogais:	Prof. Doutor Luís Filipe Lourenço Bernardo
	Mestre Pedro Miguel Figueiredo Amaral

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Março de 2012

Hybrid Link-State Path-Vector Protocol ++

Aos meus pais

RESUMO

O protocolo usado actualmente na Internet para realizar encaminhamento inter-domínio é o BGP (Border Gateway Protocol). Este protocolo foi desenhado para obter acessibilidade dos domínios e não está preparado para todos os requisitos das redes modernas, sofrendo de problemas graves, como convergência lenta e escalabilidade limitada. O protocolo HLP (Hybrid Link-state Path-vector) foi proposto como uma possível solução para estes problemas.

Nesta dissertação é avaliado o desempenho do HLP face ao BGP para a Internet real. Também é avaliada a sua compatibilidade com o modelo de negócios da Internet. O estudo do HLP revelou que não é compatível com o actual modelo de negócios da Internet, tendo sido desenvolvido o HLP++, um novo protocolo que corrige esta limitação. A implementação do HLP++ foi realizada no simulador de redes 2 (ns-2).

É efectuado um estudo sobre a natureza topológica da Internet, por forma a melhor compreender as relações entre os ASes. Deste estudo resultou uma topologia (um subconjunto da rede Internet) utilizada nas experiências realizadas com os protocolos avaliados, conseguindo assim resultados mais próximos da realidade. Os resultados mostram que o protocolo HLP++ não está adaptado à topologia da Internet, tendo um desempenho inferior quando comparado com o BGP. Apenas tem bom desempenho numa rede hierárquica.

PALAVRAS-CHAVE

Encaminhamento inter-domínio, BGP, escalabilidade, convergência, algoritmos.

ABSTRACT

The protocol currently used in the Internet for inter-domain routing is BGP (Border Gateway Protocol). This protocol was designed for accessibility of domains and is not suitable for all the requirements of modern networks, suffering from serious problems such as limited scalability and slow convergence. The HLP protocol has been proposed as a possible solution to these problems.

The HLP's performance face to BGP in the real Internet is evaluated in this thesis. It is also evaluated its compatibility with the Internet's business model. The study revealed that HLP is not compatible with current business model of the Internet. A new protocol was developed to fix this limitation, HLP++. It was implemented in the network simulator2 (ns-2).

A study of the topological nature of the Internet is also made, in order to better understand the relationships between the ASes. As result it is obtained a topology (a subset of Internet network) that was used in the experiments, thereby achieving results closer to the reality. The results show that the protocol HLP++ protocol is not adapted to the Internet's topology, and it has lower performance as compared with the BGP. It has good performance in a hierarchical network.

KEYWORDS

Inter-domain routing, BGP, scalability, convergence, algorithms.

AGRADECIMENTOS

Na fase final da escrita desta dissertação, gostaria de agradecer às várias pessoas que ajudaram no seu desenvolvimento e realização.

Agradeço ao Prof. Luís Bernardo a amizade, a orientação e apoio constante ao longo da minha dissertação. Agradeço-lhe pela oportunidade de crescimento, aprendizagem, realização profissional e pessoal pela confiança em mim depositada.

Agradeço também ao Prof. Pedro Amaral a co-orientação e apoio, assumida na fase inicial de desenvolvimento da dissertação.

Um especial agradecimento ao Prof. Paulo Pinto, pela sua grande experiência e conhecimento científico. Sempre me incentivou na busca do conhecimento, sendo um exemplo de competência, garra, determinação e disciplina.

A todos os colegas da FCT-UNL, em particular aos colegas da secção de telecomunicações pelo seu apoio e convívio ao longo de todo o curso.

Aos meus amigos, com particular destaque para o Miguel Pereira, Francisco Ganhão, Michael Figueiredo, Hugo Borda D'Água, Filipe Higino, Miguel Silva, Bruno Esperança e Gonçalo Martins, pelas suas sugestões, críticas, apoio e amizade demonstrada.

Ao Eng.º Vasco Aleluia pela preciosa ajuda na implementação desta dissertação, a sua imensa experiência nas redes IP e protocolos, contribuiu para aprofundar o meu conhecimento. Um profissional exemplar e um amigo sincero.

Agradeço também ao projecto MPSat, PTDC/EEA-TEL/099074/2008.

***

Por fim, mas de forma alguma com menos importância, gostaria de mostrar a gratidão mais gentil e especial para com a minha família e amigos por terem compartilhado comigo o outro lado da vida durante toda a realização do curso. Agradeço aos meus pais por todo o apoio e carinho que me deram ao longo da minha vida e que sem eles seria impossível alcançar esta meta.

A todos o meu profundo agradecimento.

Lisboa, Março 2012

Cláudio Fernando Sequeira Assunção

ACRÓNIMOS

ARPA	Advanced Research Project Agency
AS	Autonomous System
ASes	Autonomous Systems
AUP	Acceptable Use Policy
BBN	Bolt Beranek and Newman Corp
BGP	Border Gateway Protocol
C2P	Customer to Provider
CAIDA	Cooperative Association for Internet Data Analysis
CSNET	Computer Science Network
DARPA	Defense Advanced Research Project Agency
DNS	Domain Name System
DV	Distance Vector
eBGP	External Border Gateway Protocol
FCCN	Fundação para Computação Científica Nacional
FCP	Failure Carrying Protocol
FNC	Federal Networking Council
FPV	Fragmented Path-Vector
FTP	File Transport Protocol
GNU	GNU is Not Unix
HLP	Hybrid Link-State Path-Vector Protocol
iBGP	Internal Border Gateway Protocol
ICCC	International Computer Communication Conference
ICMP	Internet Control Message Protocol
IMP	Interface Message Processor
IMPs	Interface Message Processors
IP	Internet Protocol
IPV4	Internet Protocol Version 4
IPV6	Internet Protocol Version 6
IRR	Internet Routing Registry
IS-IS	Intermediate System – Intermediate System

ISP	Internet Service Provider
ISPs	Internet Service Providers
LSA	Link State Advertisements
LAN	Local Area Network
LANs	Local Area Networks
LS	Link State
MED	Multi Exit Discriminator
MILNET	Military Network
MIT	Massachusetts Institute of Technology
MRAI	Minimum Route Advertisement Interval
NASA	National Aeronautics and Space Administration
NCP	Network Control Protocol
NIRA	New Inter-domain Routing Architecture
NPL	National Physical Laboratory
NRLS	Name-to-Route Lookup Service
NSF	National Science Foundation
ns-2	network simulator 2
OSI	Open System Interconnection
OSPF	Open Shortest Path First
P2C	Provider to Customer
PC	Personal Computer
P2P	Peer to Peer
PV	Path Vector
QoS	Quality of Service
RFC	Request For Comments
RIP	Routing Information Protocol
RIPE	Réseaux IP Européens
RPSL	Routing Policy Specification Language
UCLA	University of California Los Angeles
UCSB	University of California Santa Barbara
UDP	User Datagram Protocol
USENET	UNIX User Network
USA	United States of America
UTAH	University of Utah

S2S	Sibling to Sibling
SRI	Stanford Research Institute
TIPP	Topology Information Propagation Protocol
VoIP	Voice over IP
VPNs	Virtual Private Networks
WWW	World Wide Web
XL	Approximate Link-state

ÍNDICE DE MATÉRIAS

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

Capítulo 1.

INTRODUÇÃO

A "Internet" é um termo bastante comum nos dias de hoje. Expressões como "procura na Internet", ou "compra pela Internet" tornaram-se banais entre as pessoas. O que começou por ser um projecto para partilha de ficheiros e recursos entre computadores (ARPANET), cresceu e tornou-se numa das ferramentas mais importantes nos dias actuais. Seja em lazer ou trabalho, a Internet veio alterar a sociedade como nada antes visto. O conjunto de ferramentas actualmente disponíveis na Internet e facilmente acessíveis, permitem aos seus utilizadores a realização de compras *online*, aquisição de informação e interacção social.

O crescimento explosivo da Internet por volta de 1995 levou a imprensa comum a crucificar a Internet e a anunciar que esta se podia desfazer se tal crescimento perdurasse [met]. Hoje, a Internet mantém-se "viva" e continua a crescer. Seja por fibra óptica ou por redes sem fios, a Internet chega a novos utilizadores. O crescimento actual não se reflecte só em número de utilizadores mas também em largura de banda, pois são oferecidos pelos operadores de serviços de Internet (ISPs – *Internet Service Providers*) acessos cada vez mais rápidos. No entanto este aumento obriga a que os ISPs façam planos e estudos cuidados nas suas redes para que desta forma não existam lacunas no fornecimento de largura de banda, e mais importante ainda, não existam falhas de encaminhamento de pacotes.

O protocolo de encaminhamento presente na Internet é o BGP (*Border Gateway Protocol*), actualmente na sua versão 4. Este protocolo sofre de várias restrições [YMBB05], destacando-se a baixa convergência e os problemas de escalabilidade. A principal causa responsável pela baixa convergência resulta da visibilidade global de eventos: uma simples falha de uma ligação pode levar a que os encaminhadores BGP troquem entre si enormes quantidades de mensagens de actualização de rotas por toda a rede. Os problemas de escalabilidade advêm principalmente dos ISPs terem aumentado a sua conectividade por razões de redundância e distribuição de carga, para darem resposta à grande demanda dos seus

utilizadores por recursos. Foi adoptada uma estrutura multi-fornecedor (*multihomed)*, onde os ASes (Autonomous Systems) têm ligações a mais do que um fornecedor. Embora o BGP seja usado na Internet como a solução para o encaminhamento inter-domínio, os ISPs tendem a usá-lo para diferentes propósitos, tais como balanceamento de carga ou para suportar VPNs (Virtual Private Networks). As limitações do BGP reflectem-se nos serviços usados na Internet, como por exemplo voz sobre IP (VoIP – Voice over IP) que é muito sensível a alterações de rotas [KKK07].

O encaminhamento inter-domínio é considerado nos dias actuais uma área de investigação desafiante. A comunidade de investigação apresentou várias soluções académicas: alguns investigadores tentaram atenuar os efeitos das limitações do actual protocolo de encaminhamento, enquanto outros apresentam mesmo novas soluções para substituir o BGP. Nesta dissertação apresenta-se uma extensão para uma das soluções propostas, designada de HLP (Hybrid Link-State Path-Vector Protocol) [SCE+05], desenvolvida por *Subramanian,* onde é proposta uma nova arquitectura para o encaminhamento a nível do Inter-Domínio.

O HLP assume que a topologia rede é uma estrutura hierárquica de relações do tipo fornecedor-cliente entre os sistemas autónomos (ASes – *Autonomous Systems*). O anúncio de rotas é efectuado apenas com base na identificação dos ASes, reduzindo assim a sua tabela de encaminhamento. O HLP classifica-se com um protocolo híbrido a nível do encaminhamento inter-domínio na medida que as alterações topológicas são efectuadas com recurso a uma combinação de dois tipos de protocolos complementares: estado de linha (LS – *Link State*) dentro de uma hierarquia e vectores de caminhos (PV – *Path Vector*) entre hierarquias.

Dada a natureza dos serviços e a evolução natural da Internet, cada AS estabelece relações comerciais diferentes com diferentes ASes, independentemente do nível hierárquico a que pertence, restringindo o tipo de tráfego que se pode transportar em cada AS. Desta forma, fica definido um modelo de negócios da Internet, caracterizado por um conjunto de políticas. O HLP falha na capacidade para suportar o modelo de negócios. Dadas as propriedades dos protocolos LS e PV, as políticas entre ASes não estão asseguradas e poderão ser violadas. Esta dissertação propõem-se modificar o HLP para que este se torne compatível com o modelo de negócios da Internet. As modificações designadas genericamente de HLP++, efectuam-se a nível das regras de exportação de caminhos e dos dois protocolos utilizados de forma híbrida pelo HLP.

1.1. Enquadramento

Na Internet, para que o cliente final tenha acesso a dados globalmente é necessário que os operadores disponham de rotas para qualquer destino (destino IP). Como um operador local não consegue por si próprio endereçar todas as redes na Internet, ele recorre a outros operadores que fornecem as rotas necessárias. Este tipo de operações leva a que os ASes estabeleçam entre si contractos monetários para transporte de tráfego, que se reflectem posteriormente sob a forma de políticas nas configurações dos encaminhadores de fronteira nos ASes.

O trabalho de Gao [Gao01] identificou que a maioria destas políticas são comuns e podem classificar-se como pertencendo a três tipos: fornecedor – cliente, par – par e irmão – irmão. Como é perceptível as entidades envolvidas não estão interessadas em que os contractos sejam violados; ou dito de outra forma, se um AS não pagou por uma ligação então não a poderá usar.

Os protocolos de encaminhamento têm de ser compatíveis com as políticas que definem o modelo de negócios da Internet. Para tal é necessário restringir o anúncio de rotas realizado pelo protocolo, de maneira a não se anunciar rotas inválidas.

1.2. Hipótese

A partir do contexto actual da Internet, apresentado na secção 1.1, é formulada a seguinte hipótese: Se o modelo de negócios da Internet assenta essencialmente em relações do tipo fornecedor – cliente, cliente – fornecedor, par – a – par, e irmão – irmão, então é possível realizar a arquitectura HLP de forma a ela ser compatível com o actual modelo de negócios e assegurar e respeitar as políticas entre ASes.

1.3. Objectivos e Contribuições

O objectivo principal desta dissertação é implementar e modificar o protocolo HLP de forma a estar em conformidade com a hipótese da secção 1.2.

As principais contribuições desta dissertação são o desenho da modificação do protocolo HLP, designado de HLP++, e a sua implementação integral no simulador ns-2.

No entanto, para uma correcta análise de desempenho do protocolo, é necessário usar uma topologia de teste próxima da realidade. Nesta dissertação também é feito um estudo sobre a natureza topológica da Internet com recurso aos resultados do projecto que estudou as relações entre ASes da CAIDA (Cooperative Association for Internet Data Analysis) [cai09]. Nesta dissertação foi definida uma topologia de 54 ASes (subconjunto da rede Internet) para realizar a análise do HLP++ e do BGP++[bg09]. Parte dos resultados apresentados nesta dissertação foram publicados na conferência da *IEEE Globecom'09* [AGA+09]

1.4. Estrutura da Dissertação

A dissertação encontra-se organizada em cinco capítulos, conforme se resume em seguida.

No Capítulo 2 ("Trabalho Relacionado") é apresentada uma breve introdução da história da Internet e a sua evolução. Introduz-se o modelo de negócios da Internet, é feita a comparação entre os protocolos utilizados actualmente e as soluções propostas pelo mundo académico na área do encaminhamento inter-domínio.

No Capítulo 3 ("Arquitectura do protocolo"), sumariza-se o protocolo HLP e são mostradas as suas limitações. De seguida, é apresentado o protocolo HLP++, que o estende tornando-o compatível com o modelo de negócios da Internet. No fim, são apresentadas e explicadas as modificações necessárias, designadas de HLP++.

O Capítulo 4 ("Análise do Desempenho"), inclui uma breve introdução ao simulador ns-2 numa primeira parte. De seguida são apresentadas as características da topologia usada nas simulações. Por último, é feita a avaliação do desempenho do protocolo proposto nesta dissertação.

No Capítulo 5 ("Conclusões"), é feita uma análise global do trabalho realizado, tendo por base a hipótese originalmente estabelecida. Por fim identificam-se as questões em aberto e consequentes trabalhos futuros.

Capítulo 2.

TRABALHO RELACIONADO

Nos dias actuais o encaminhamento inter-domínio (*interdomain routing*) é considerado uma área de investigação desafiante. Este desafio deve-se principalmente a dois factores: às limitações do protocolo de encaminhamento e à natureza distribuída e descentralizada da Internet.

O actual protocolo de encaminhamento inter-domínio utilizado na Internet tem diversas limitações, no entanto, a sua substituição não é trivial devido à sua implantação a nível mundial. As suas limitações tornaram-se especialmente notórias devido ao grande crescimento da Internet nos últimos anos. O crescimento não se refere apenas ao tamanho da rede, mas também à quantidade e variedade de aplicações actualmente disponíveis na Internet. Esta tendência de crescimento tem vindo a expor as limitações dos protocolos de encaminhamento inter-domínio, nomeadamente com um aumento da instabilidade das rotas na rede [SKM09].

Como o próprio nome indica, encaminhamento inter-domínio significa encaminhamento entre domínios ou redes distintas. Nestes domínios compostos por entidades completamente autónomas, a gestão do seu encaminhamento local é baseada em políticas que normalmente têm apenas significado local. Neste tipo de cenário, com concorrência entre domínios e entre empresas e, com total independência na gestão do encaminhamento usando políticas potencialmente conflituosas, o problema do encaminhamento inter-domínio torna-se ainda mais difícil.

Em paralelo com os desafios inerentes ao encaminhamento inter-domínio, o crescimento da Internet suscitou também o interesse da comunidade de investigação em descobrir e analisar a topologia da Internet. Alguns investigadores desenvolveram ferramentas para obter dados referentes à topologia enquanto outros tentaram entender as propriedades Internet.

2.1. A Internet

A internet revolucionou o mundo dos computadores e os meios de comunicação como nada antes visto, desde a invenção da máquina de impressão há mais de 500 anos atrás a Internet é ao mesmo tempo um meio de radiodifusão, um mecanismo de disseminação de informação, e um meio de colaboração e integração de pessoas com computadores independentemente da sua localização geográfica

A internet hoje é uma infra-estrutura ampla de informação, no entanto a sua origem é complexa e envolve aspectos de carácter tecnológico, organizacional e social. A sua influência não afecta apenas os aspectos técnicos das comunicações por computador, mas também toda a sociedade dado o crescimento da utilização de ferramentas *online* para completar tarefas comércio electrónico, aquisição de informação e interacção social.

A sua história gira em torno de quatro aspectos distintos: a evolução tecnológica, a operação e gestão da infra-estrutura, o aspecto social e a comercialização. A evolução tecnológica começou com as primeiras pesquisas sobre a comutação de pacotes na ARPANET, e ainda hoje a pesquisa continua a expandir os horizontes da infra-estrutura em várias dimensões, tais como a escalabilidade, desempenho e funcionalidades de alto nível. Com o crescimento a complexidade da operação e gestão da infra-estrutura da Internet a nível global cresceu. A nível social, a Internet resultou numa ampla comunidade de internautas trabalhando juntos para criar e fazer evoluir a tecnologia. Por fim há o aspecto da comercialização, resultado de uma extrema e eficaz disponibilização dos resultados de pesquisas e investigações numa rede de computadores acessível aos seus utilizadores.

2.1.1. As Origens

Em 1957 preocupado com o facto de os Estados Unidos da América (USA – United States of America) estarem a ser ultrapassados em matéria de conhecimento científico devido ao lançamento do Sputnik[1] pela União Soviética, o presidente Dwigth D. Eisenhower aprovou a criação da ARPA (Advanced Research Project Agency) mais tarde renomeada para DARPA (Defence Advanced Research Project Agency) com a missão de manter os USA na vanguarda da tecnologia em termos de defesa militar.

O primeiro conceito de Internet aparece em 1962 por J.C.R. Licklider com a sua "Rede Galáctica". A ideia era ter um conjunto de computadores globalmente interligados, através

[1] O Sputnik foi o primeiro satélite artificial da terra em órbita, foi lançado numa órbita elíptica de baixa altitude pela União Soviética a 4 de Outubro de 1957.

dos quais todas as pessoas poderiam ter acesso a dados e programas a partir de qualquer lugar (o espírito da Internet de hoje não é muito diferente) [BLM09].

Desde o início da ARPA, Leonard Kleinrock vinha desenvolvendo uma teoria sobre o envio de informação utilizando comutação de pacotes. A ideia era partir a mensagem em fragmentos e enviar separadamente cada um para o seu destino. Do outro lado, fazia-se o processo inverso e reconstruía-se a mensagem original, obtendo-se assim mais flexibilidade e segurança dado que os pacotes não seriam obrigados a seguir todos o mesmo percurso. Convencido por esta teoria, Roberts decidiu explorar a ideia em 1965 ligando dois computadores através de uma linha telefónica de baixa velocidade entre Massachusetts e Califórnia. O resultado desta experiência revelou que a partilha de tempo entre computadores para executar programas e aceder a dados em máquinas remotas era viável. No entanto a linha telefónica não era adequada, o que confirmou a necessidade de utilizar comutação de pacotes em vez de circuitos telefónicos.

Um ano mais tarde Roberts ingressa na ARPA para desenvolver o conceito rede de computadores. Rapidamente coloca em prática um plano para a ARPANET, publicando-o nesse mesmo ano. Nos finais de 1969 as Universidades, UCLA (University of California, Los Angeles), UCSB (University of California, Santa Barbara), UTAH (University of Utah) e o SRI (Stanford Research Institute), são ligados por IMP's (*Interface Message Processor*) os primeiros comutadores de pacotes (*packet-swicthers*) desenvolvidos pela BBN (*Bolt, Beranek and Newman Corp*.). Estes IMPs não eram mais que uma versão primitiva dos actuais comutadores (*switches*), computadores programados para receber e enviar pacotes para um dado destino. A figura 2.1 mostra a título de curiosidade a ARPANET no ano do seu "nascimento", os círculos representam os IMP's e os rectângulos os respectivas máquinas (*hosts*) a eles ligados. A Internet começava a "gatinhar" e o sonho de Licklider tornava-se realidade.

No início os sistemas da ARPANET só permitiam aplicações cliente-servidor como o Telnet e o FTP (*File Transfer Protocol*), e não conseguia lidar com as relações máquina-a-máquina. Os investigadores rapidamente perceberam que o funcionamento da rede estava limitada, tendo desenvolvido um novo protocolo de nome NCP (*Network Control Protocol*) para colmatar este problema, que se tornou assim no primeiro protocolo a nível de transporte da ARPANET e também das redes de computadores [SR95].

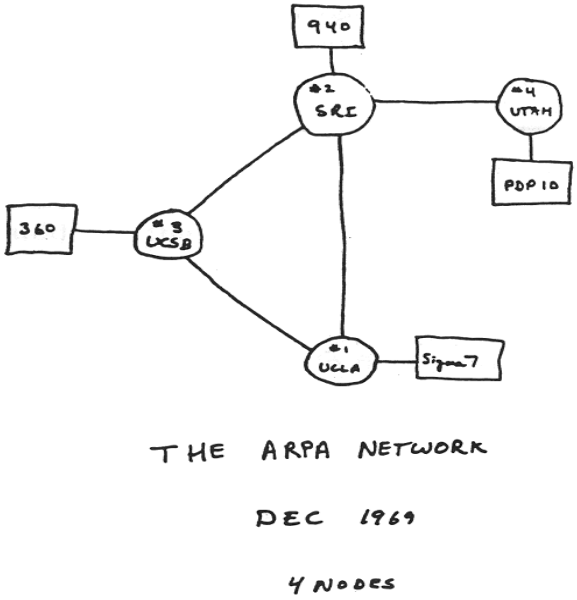

Figura 2.1 – Arpanet em Dezembro de 1969. Disponível em http://www.leidenuniv.nl/letteren/internethistory/arpanet.gif

2.1.2. Da Arpanet para a Internet

A ARPANET original cresceu e tornou-se na Internet. Novos computadores foram adicionados rapidamente e, em finais de 1971 a ARPANET, representada na figura 2.2, tinha crescido de 4 para 23 máquinas e o número não parava de aumentar.

Inúmeras foram as datas que marcaram a história da Internet e podem ser consultadas em [LK08] no entanto, 1972 torna-se especial dado que em Outubro desse ano a ARPANET é tornada pública na International Computer Communication Conference (ICCC) em Washington D.C. Representantes de outras redes de comutação de pacotes experimentais como a NPL (National Physical Laboratory) do Reino Unido e a Cyclades de França estavam presentes nessa conferência. É também em Março do mesmo ano que aparece a primeira aplicação "quente" de rede, o correio electrónico, motivado pela necessidade dos programadores da ARPANET necessitarem de um mecanismo de fácil comunicação, Ray Tomlinson da BBN escreveu o simples programa de envio e recepção de e-mail, que a partir desta data foi a aplicação mais usada na rede durante décadas.

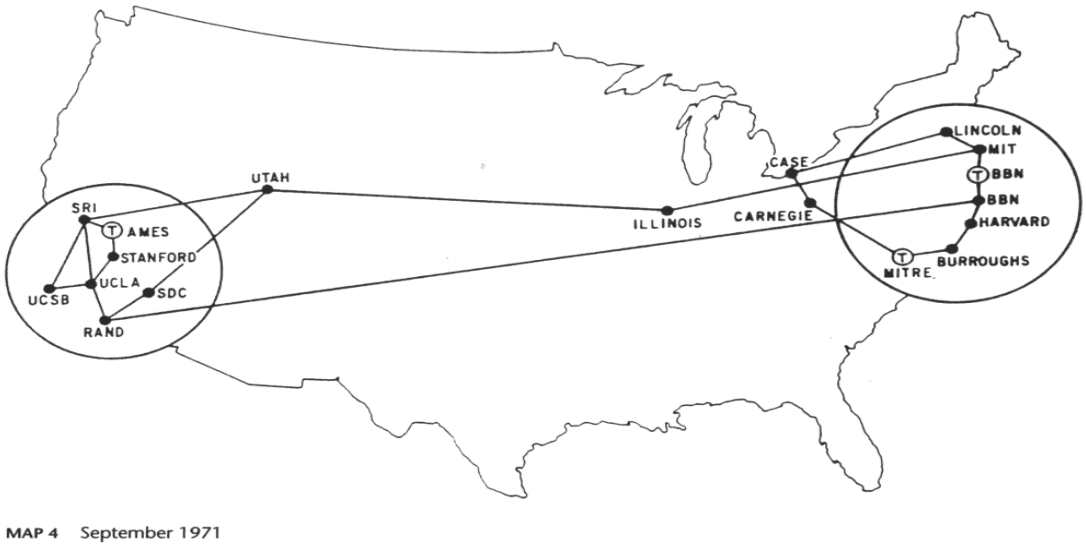

Figura 2.2 – Arpanet em Setembro de 1971. Disponível em http://www.leidenuniv.nl/letteren/internethistory/arpanet1.gif

A Internet foi baseada na ideia que haveria múltiplas redes independentes de projectos arbitrários. A ARPANET foi a pioneira nas redes de comutação de pacotes, mas em breve a Internet incluiria redes de satélites, redes de rádio terrestre e outras redes. A internet de hoje, tal como a conhecemos, incorpora uma técnica chave chamada arquitectura de rede aberta. Neste tipo de abordagem, a escolha da tecnologia para uma rede individual não está dedicada a uma rede em particular, mas pode ser livremente escolhida pelo operador e feita para interagir com outras redes.

No entanto, o protocolo de rede nessa altura era o NCP. Este protocolo não fora desenhado para endereçar redes ou máquinas, o que era claramente um obstáculo para interligar as várias redes. O NCP foi desenvolvido para permitir comunicação ponto a ponto. Se algum pacote se perdesse, as aplicações e o próprio protocolo iriam falhar. Neste modelo o NCP não tinha controlo de erros; este tipo de controlo era feito pelas máquinas.

Dadas estas dificuldades foi decidido criar um novo protocolo que satisfizesse as necessidades de um ambiente em arquitectura de rede aberta. Este protocolo seria mais típico de um protocolo de comunicações enquanto o NCP seria mais utilizado como um driver de dispositivo. Quatro regras determinaram o aparecimento deste novo protocolo de comunicações: Primeira, cada rede teria de ser distinta e nenhuma alteração poderia ser exigida do lado da rede de forma a se conectar a Internet; Segunda, as comunicações teriam por base o melhor esforço, se um pacote não chega-se ao seu destino, devia ser retransmitido

a partir da fonte; Terceira, as caixas negras, que mais tarde viriam a ser chamadas de encaminhadores (*gateways* ou *routers*), não retêm nenhuma informação sobre o fluxo individual dos pacotes, simplificando a sua estrutura e evitando complicados modos de adaptação e recuperação dos vários modos de falha; por último, não haveria controlo global no nível operacional.

O novo protocolo de comunicações viria a ser chamado de TCP/IP (*Transmission Control Protocol / Internet Protocol*) dado que a sua primeira implementação foi apenas o TCP, mas mais tarde verificou-se com a perda de pacotes (em particular pacotes de voz), que por vezes o controlo deveria passar para o lado da aplicação. Desta forma, o protocolo TCP inicial foi separado em dois, o protocolo IP que se destina ao endereçamento e encaminhamento dos pacotes individuais e o protocolo TCP que se preocupa com as características do serviço, tais como o controlo de fluxo e a recuperação de pacotes perdidos. Para as aplicações que não necessitavam dos serviços do TCP, um protocolo alternativo chamado de UDP (*User Datagram Protocol*) foi adicionado, por forma a fornecer acesso directo ao serviço de melhor esforço fornecido pelo IP.

A maior motivação inicial tanto da ARPANET como da Internet foi a partilha de recursos. Por exemplo, se utilizadores de uma certa rede necessitassem de utilizar super-computadores para fazerem cálculos das suas experiências, era muito mais simples interligar as redes e dar acesso a esses mesmos computadores já disponíveis na ARPANET do que duplicar os mesmos. Optimizava-se, desta forma, os recursos e economia das instituições. Enquanto o acesso remoto e a partilha de ficheiros eram as aplicações importantes, o correio electrónico teve provavelmente o impacto mais significativo. O correio electrónico veio possibilitar um novo modelo de comunicação entre as pessoas, e veio modificar a natureza da colaboração entre indivíduos, ainda bem presente nos dias actuais. A utilização do correio electrónico foi massificada como meio de comunicação entre pessoas, empresas, instituições académicas e comércio electrónico. As primeiras implementações do TCP/IP foram feitas para super-computadores (com sistemas operativos de partilha de tempo). Quando os primeiros computadores pessoais começaram a aparecer, David Clark do MIT e a sua equipa de investigação mostraram que era possível uma realização simples e compacta do TCP, disponibilizando-o para o primeiro computador pessoal desenvolvido pela Xerox PARC e mais tarde para o IBM PC. Desta forma, tanto os super-computadores como os computadores pessoais podiam coexistir na mesma rede e ambos fazer parte da Internet.

O desenvolvimento generalizado de LANs (Redes de Área Local), computadores pessoais e servidores por volta dos anos 80, permitiu que a ainda incipiente Internet florescesse. A

mudança de um ambiente com um modesto número de máquinas para partilha de tempo de processamento e ficheiros para um onde existem várias redes, originou uma série de novos conceitos e modificações na tecnologia subjacente. Primeiro, resultou na definição de três classes de rede (A, B e C) para acomodar a gama de redes. A classe A representa as grandes redes em escala nacional (pequeno número de redes com grande número de máquinas). A classe B caracteriza as redes de escala regional. Por fim, as redes locais (grande número de redes com relativamente poucas máquinas) são representadas pela classe C.

Uma outra grande mudança ocorreu como resultado do aumento em escala da Internet e dos seus problemas de gestão associados. Para facilitar a utilização da rede pelas pessoas, às máquinas foram atribuídos nomes de forma que não era necessário lembrarem-se dos endereços numéricos. Originalmente, havia um número bastante limitado de computadores, por isso foi possível manter uma tabela única de todos os servidores com os seus nomes e endereços associados. A mudança para uma grande quantidade de redes geridas de forma independente (por exemplo as LANs) revelou que ter uma única tabela de máquinas e endereços não era mais viável. Desta forma o sistema de domínios de nome (DNS, *Domain Name System*) foi inventado por Paul Mockapetris. O DNS veio permitir um mecanismo escalável e distribuído de resolução nomes de servidores hierárquicos em endereços de rede.

O TCP/IP foi adoptado pelo Ministério da Defesa Americano como protocolo padrão em 1980. Três anos mais tarde a ARPANET já estava a ser usada não só por um vasto número de investigadores da defesa Americana, mas também por uma significativa quantidade de organizações. A transição do seu protocolo inicial NCP para o TCP/IP veio permitir que a ARPANET se dividisse em duas: a MILNET, para suportar os requisitos operacionais a nível militar; e a ARPANET que continuaria a suportar as necessidades de investigação. Assim, em 1985 a Internet não só estava bem estabelecida como a tecnologia de suporte a uma ampla comunidade de pesquisa e desenvolvimento, como também já fazia parte de outras comunidades que a utilizavam como tecnologia de comunicação diária entre computadores. O sucesso da utilização do correio electrónico e partilha de computadores demonstrado pela ARPA, motivou que outras organizações criassem a sua própria rede como foi o caso da SPAN por parte da NASA, o Departamento de Energia cria a MFENet, CSNET para a comunidade de ciência e computadores, a USENET por parte de AT&T entre outras.

2.1.3. A Comercialização da Tecnologia

O que começou com a experiência académica de partilha de tempo e ficheiros entre computadores, tinha-se tornado no que hoje conhecemos como a Internet. Inicialmente todas as redes eram construídas para servir as comunidades académicas (maioritariamente universidades) e eram restritas a outros utilizadores. No entanto, a implementação destas redes tinham custos associados. Por volta do ano 1986, a Fundação Nacional para a Ciência do governo Americano (NSF, *National Science Foundation*) desempenhou um papel importante na privatização da tecnologia com a implementação de uma rede chamada NSFNET e elegendo como protocolo obrigatório o TCP/IP. Steve Wolf director do projecto NSFNET, reconheceu a necessidade de criar uma infra-estrutura de rede para suportar as comunidades de pesquisa e académicas em geral, juntamente com a necessidade de desenvolver uma estratégia para que tal infra-estrutura e outras que viessem mais tarde não dependessem directamente dos recursos federais.

Para este efeito, foram realizadas políticas e estratégias por diversas agências federais que definiram a Internet de hoje. Como exemplos destas políticas e decisões temos o Conselho Federal de Redes (FNC, *Federal Networking Council*) que veio coordenar a partilha dos custos comuns suportados pelas agências federais relativos às infra-estruturas como circuitos transoceânicos entre redes.

O estímulo por parte da NSF em que as suas redes regionais (inicialmente académicas) procurassem utilizadores comerciais e não académicos, levou à implementação de outra política e talvez a mais importante adoptada por esta agência, a aplicação da política aceitável de utilização (AUP, *Acceptable Use Policy*), que proibiu o uso do *Backbone* (o segmento à escala nacional) da rede NSFNET para fins que não fossem de suporte a educação e investigação. Teve resultado previsível e premeditado de encorajar o tráfego comercial a nível regional e local, enquanto intencionalmente o proibia a nível Nacional. O propósito era estimular o surgimento ou crescimento de redes de longa distância competitivas e privadas tais como a PSI, UUNET entre outras que surgiram mais tarde.

A comercialização da Internet envolveu não apenas o desenvolvimento de mercados competitivos e serviços privados, mas também o crescimento de produtos comerciais para implementação da Internet. Se olharmos para o passado a estratégia por parte dos vendedores incorporarem protocolos de comunicação nomeadamente o TCP/IP nos sistemas operativos, foi um dos elementos chave no sucesso da Internet.

Nos últimos anos temos assistido a uma nova fase de comercialização. Inicialmente os esforços por parte das entidades comerciais foram baseados no fornecimento de produtos básicos de rede, com os operadores a oferecerem conectividade e apenas os serviços principais de Internet. Actualmente a Internet tornou-se quase num serviço de "comodidade", e muita da atenção passa agora pela utilização da infra-estrutura global para suportar cada vez mais serviços. Isto tem sido extremamente acelerado pela adopção rápida e generalizada de navegadores de Internet (*browsers*) e da tecnologia WWW (*World Wide Web)* por parte das várias faixas etárias da sociedade. Os operadores oferecem cada vez mais e melhores acessos à Internet, nomeadamente fibra óptica até a casa do cliente, que introduzem requisitos crescentes de largura de banda no núcleo da rede da Internet. Os produtos estão cada vez mais disponíveis e sofisticados o que facilita o fluxo de informação em "cima" da camada de dados da Internet.

No entanto todo este crescimento tem um custo, e não estamos a falar apenas da parte monetária mas sim da tecnologia que suporta todo este desenvolvimento. Para que a informação e serviços cheguem ao utilizador final, existe toda uma tecnologia instalada, essencialmente equipamentos e protocolos (e em especial protocolos de encaminhamento de pacotes), cuja evolução tem de ser gradual. É sobre estes protocolos que assenta esta dissertação, em particular os protocolos de encaminhamento inter-domínio, que vamos falar na próxima secção e nos próximos capítulos.

2.2. Características da Internet

Antes de se entrar em detalhe nos protocolos de encaminhamento, é necessário compreender como é que a Internet actual se caracteriza e onde é que estes protocolos se inserem. Como se viu na secção anterior, o que começou com a ARPANET, uma rede puramente académica para partilha de tempo e ficheiros entre computadores, evoluiu e chegou ao sector privado como consequência de um crescente número de clientes exigentes e de serviços. Com este crescimento surgiu uma necessidade incontornável por parte dos operadores locais para dar cumprimento à presente demanda, e para servir um público cada vez mais amplo para além das empresas e instituições académicas. Pode-se dizer que são duas as características principais que definem a Internet presente: o seu modelo de negócios e a sua topologia.

2.2.1. Modelo de Negócios da Internet

Dada a natureza dos serviços e a evolução natural da Internet, os operadores locais não conseguem oferecer aos seus clientes uma solução completa. Eles necessitam de recorrer a diferentes operadores nacionais e estes, por sua vez, de recorrer a outros internacionais, para que o utilizador final tenha conectividade e possa aceder a dados globalmente. Desta forma são estabelecidos contratos entre estas entidades para distribuição de rotas, sendo necessárias políticas para controlar o fluxo de tráfego.

Os Sistemas Autónomos (ASes, *Autonomous Systems*) desempenham um papel fundamental no que respeita ao encaminhamento (*routing*) intra-domínio e inter-domínio, uma vez que a sua configuração influencia a forma como o tráfego é tratado no caminho ascendente ou descendente, i.e. do cliente para o operador ou do operador para o cliente. Isto também se reflecte nas políticas adoptadas com outros ASes.

Na Internet um Sistema Autónomo (AS) é um conjunto de redes IP sobre o controlo de um ou mais operadores de rede que apresentam uma política de encaminhamento única e claramente definida [RFC1930]. Esta definição deve-se ao facto de que múltiplas organizações poderem "executar" o protocolo de encaminhamento BGP (*Border Gateway Protocol*) usando números privados de AS para um ISP (*Internet Service Provider*) que depois interliga todas essas organizações à Internet. Temos como exemplo em Portugal a GigaPIX da FCCN (Fundação para a Computação Científica Nacional).

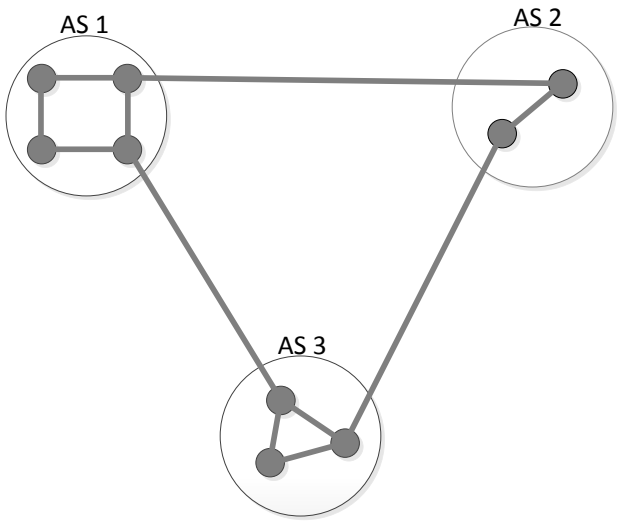

Figura 2. 3 – Exemplo simples de uma rede com três Sistemas Autónomos

A figura 2.3 mostra um exemplo simples de uma rede com três ASes. Os círculos maiores caracterizam o AS em si, enquanto os círculos mais pequenos (preenchidos a cinzento) representam os encaminhadores internos ao AS. Os encaminhadores que interligam os ASes uns aos outros chamam-se de encaminhadores de fronteira *(border-routers*). É nestes encaminhadores que a maioria das políticas entre Sistemas Autónomos é configurada.

O trabalho de Gao [Gao01] identificou que na maioria dos casos essas políticas são comuns e classificou-as como pertencendo a um de três tipos:

1. Provider – Customer (Fornecedor – Cliente, representado por p2c e c2c)
2. Peer to Peer (Par – Par, p2p)
3. Sibling to Sibling (Irmão – Irmão)

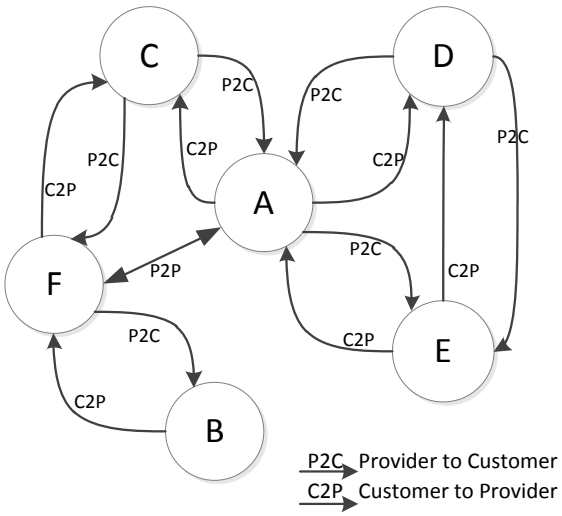

Figura 2. 4– Exemplos de políticas entre ASes

Tipicamente, um AS local não dispõe de rotas para todas as redes da Internet. Para tal, paga a outro AS para que este as forneça, estamos perante uma política do tipo fornecedor – cliente (*provider-customer*). As políticas do tipo fornecedor cliente dividem-se em duas políticas: direcção fornecedor para cliente (p2c, *Provider to Customer*) em que um AS com esta política configurada, sabe que todas as rotas que receber desse AS não as exporta para outro AS em que a sua relação com ele seja do tipo cliente fornecedor (c2p, *Customer to Provider*). Desta forma, um fornecedor transporta tráfego entre os seus clientes, mas um cliente não transporta tráfego de dois fornecedores. Por exemplo, na figura 2.4 o fornecedor D pode transportar o tráfego dos ASes A e E em caso de falha na ligação directa, mas o AS A não transporta tráfego entre os ASes C e D.

As políticas do tipo p2p definem um acordo mútuo entre dois ASes para que possam passar o tráfego exclusivamente entre os seus clientes. Por exemplo, na figura 2.4 os ASes A e

F têm uma política deste tipo entre eles. Desta forma, estas duas entidades podem exportar as rotas dos seus clientes B e E, baixando assim o custo monetário dado que tipicamente as ligações do tipo c2p são de preço mais elevado.

Finalmente as políticas s2s (p2p com conectividade total) definem um acordo mútuo também entre dois ASes para permitir conectividade total entre eles. No entanto este tipo de relação pode também apenas ser usado como uma protecção (p2p total de *backup*) um do outro. Neste caso, assegura o transporte total de pacotes no caso de um deles, perder a conectividade com o resto da Internet.

Resumidamente, o modelo de negócios da Internet assenta essencialmente sobre estes tipos de políticas entre os Sistemas Autónomos. Como é perceptível, os ASes necessitam uns dos outros para que possam chegar a todas as redes da Internet. Dado que estamos a falar de ligações pagas e muitas vezes a custo muito elevado, estas políticas estão bem definidas em ambos os ASes, regulando a utilização de recursos na Internet.

2.2.2. Caracterização Topológica da Internet.

Uma característica importante para os protocolos de encaminhamento, não só para os actuais como também para os que possam porvir, é o conhecimento da natureza topológica da Internet. Este conhecimento é relevante para podermos aproximar mais da realidade as simulações e testes na área dos protocolos inter-domínio.

Desenhar uma imagem da Internet não é uma tarefa fácil porque as relações entre Sistemas Autónomos não estão explicitamente descritas. Portanto torna-se necessário garantir que existem fontes de dados adequadas para deduzir essas relações. A secção 2.2.2.1 fala sobre diferentes tipos de fontes e ferramentas para obter informações sobre o estado das redes; por sua vez na secção 2.2.2.2 mostra-se uma visão geral sobre os estudos feitos por vários autores que deduziram a estrutura topológica da Internet.

2.2.2.1. Fontes de Informação e Níveis de Abstracção

Várias fontes de informação podem ser usadas para inferir as características a partir da Internet, embora cada fonte possa ter um nível de abstracção diferente. Nesta secção são considerados três níveis: o nível IP, encaminhador e AS.

Começando com a abstracção ao nível do encaminhador, uma vez que este é constituído por múltiplas interfaces de rede, cada uma é captada como sendo um nó distinto ao nível IP.

São utilizadas técnicas de sondagem para diferenciar os encaminhadores, que permitem associar um conjunto de endereços IP virtuais a cada uma interface do encaminhador.

O *Mercator* [GT00] é um exemplo de uma ferramenta que usa pacotes ICMP [Bra89] para pesquisar as interfaces de um encaminhador de modo a tentar inferir a sua topologia de rede. De forma complementar temos o *Ally* [NSW02] que procura semelhanças entre os encaminhadores, usando os seus nomes e a informação proveniente do DNS (*Domain Name System*). Estas abordagens aumentam a carga na rede para detectarem as interfaces de rede, e também podem falhar, pois alguns encaminhadores podem estar configurados para não responder a pacotes ICMP por questões de segurança.

Passando para a abstracção ao nível IP, podemos usar a ferramenta *trace-route* para sondar vários nós na Internet, identificando o caminho de um pacote UDP ou ICMP na rede, para um determinado destino IP. A CAIDA, por exemplo, desenvolveu um programa chamado *Skitter* [BH02] que recolhe informações de vários pontos de observação distribuídos pelo mundo inteiro, explorando assim o mesmo destino com um conjunto de endereços IPV4.

Embora o *trace-route* possa parecer simples, ele tem limitações conhecidas, nomeadamente, os caminhos descobertos numa direcção podem diferir na direcção oposta [DF07].

Por fim, foca-se a abstracção ao nível do AS. A este nível é possível obter a informação de duas fontes: Informação das *routing registries* (registos de encaminhamento) e a informação do protocolo de encaminhamento BGP.

Regional Internet Registries [rir09] é um exemplo de routing registries, onde a informação de inter-domíno é exposta através do protocolo WHOIS [Dai04]. É também possível obter dados normalizados na linguagem RPSL (*Routing Policy Specification Language*) [CA99] utilizando um *Internet Routing Registry* (IRR) [Dat09]. Apesar de a acessibilidade aos *Routing Registries* ser contínua, o seu conteúdo não revela as falhas temporárias da rede [NCC09].

Como alternativa existe a informação das tabelas de encaminhamento BGP. Esta informação pode ser encontrada nos servidores de rota (*route servers*) dos ASes. Os projectos *Route Views* [UO09] ou *RIPE NCC* são exemplos de informações de encaminhamento retiradas de encaminhadores BGP de todo o Mundo. A informação do encaminhamento BGP tem uma vantagem em relação as *Routing Registries*: como a informação é retirada dos encaminhadores BGP, cada um retrata a sua visão do estado actual da rede, embora seja difícil tirar qualquer conclusão sobre a relação entre ASes.

2.2.2.2. Deduzindo a Internet e seus Geradores

Nesta secção apresentam-se de uma formal geral e cronológica alguns estudos para definir a estrutura topológica da internet.

Em 1999, Faloutsos [FFF09] mostrou que a estrutura da Internet segue uma Lei de Potência, (*Power Law*). Uma Power Law *c* pode ser descrita pela equação:

$$c \alpha A^{t},$$

onde *A* representa a métrica que segue uma Power Law, de acordo com um valor característico *t*. O gerador de topologias *BRITE* [bri09] é assente neste propósito, embora o seu desenho seja apenas adequado para redes de grande escala.

Mais tarde em 2001 *Gao* [Gao01] inferiu as relações entre ASes baseada no pressuposto que a Internet seria hierárquica, através dos dados do *Route Views* assumindo que as informações de encaminhamento transmitidas seguem caminhos livres de vales (i.e. uma informação de rota vinda de um AS vizinho (*peer*) ou fornecedor (*provider*) não é retransmitida para outro vizinho ou provedor). No entanto na sua pesquisa, ela encontrou dados inconsistentes.

O trabalho de *Gao* foi continuado em 2002 por *Subramanian* [SARK02], que inferiu as relações entre ASes a partir de múltiplos pontos de vista. O seu estudo permitiu ainda apresentar um mecanismo de classificação dos ASes em diferentes níveis numa hierarquia.

No mesmo ano, *Vazquez* [VPSV02] estudou as propriedades hierárquicas da Internet e a sua correlação com a conectividade entre nós, apresentando uma distribuição sem escala que segue uma Power Law com característica t variando entre 2 e 3. Temos como exemplos de geradores de topologias hierárquicas: o *GT-ITM* [gti09] e o *IGen* [ige09].

Estudos mais recentes [MKF+06] mostram que as métricas na Internet que seguem uma Power Law não estão correlacionadas com os níveis hierárquicos. No seu trabalho os autores concluíram também que os geradores de topologias hierárquicas não são adequados para criar topologias artificiais. O estudo mostra que nem todas as métricas da Internet seguem uma distribuição de *Power Law* pura, mas sim uma distribuição sem escala (*scale-free*).

Um ano mais tarde, os mesmos autores questionaram, se as políticas comuns existentes seriam capazes de inferir ou modelar uma topologia [DKF+07]. Para responder a esta questão, os membros da CAIDA contactaram pequenos administradores de ASes e verificaram que as relações estabelecidas com outros domínios podem ser do tipo híbrido, isto é, uma relação p2p com outro AS pode ser usada como uma relação de *sibling* para efeitos de redundância ou protecção (*backup*).

2.3. Protocolos de Encaminhamento

Uma vez compreendido o modelo de negócios da Internet e como esta está estruturada, é a altura de focar os protocolos de encaminhamento. Eles são responsáveis por encaminhar e fazer chegar a todo lado os dados que viajam pelas redes em forma de pacotes de dados como vimos no início deste capítulo.

Os protocolos de encaminhamento podem-se dividir em dois níveis, Intra-Domínio e Inter-Domínio. O primeiro diz respeito aos protocolos que são aplicados dentro de um AS, o nível inter-domínio define o tipo de protocolo que é usado para inter-ligar os vários ASes. Na secção 2.3.1 fala-se sobre os protocolos mais comuns nos dois níveis, ao passo que na secção 2.3.2 apresentam-se novas soluções propostas pelo mundo académico para fazer face aos problemas conhecidos dos protocolos tradicionais.

2.3.1. Protocolos de Encaminhamento Tradicionais

Nesta secção são apresentados os protocolos de encaminhamento mais conhecidos ao nível do intra-domíno, referindo as suas vantagens e desvantagens. Essencialmente estes protocolos pertencem a duas classes de algoritmos: classe de vector de distâncias (DV, *Distance Vector*) e de estado de linha (LS, *Link State).*

2.3.1.1. Vector de Distâncias

O primeiro protocolo usado na ARPANET [MW77] foi baseado no algoritmo de Bellman-Ford [cn02], chamando-se *Routing Information Protocol* (RIP) [Mal98].

O RIP é um protocolo de vector de distâncias dado que cada encaminhador mantém uma tabela com três entradas para cada destino D_i com custo C_i na forma (D_i, C_i, N_i) sendo N_i o próximo vizinho (*next hop*) pelo qual consegue chegar ao destino D_i. Para garantir a acessibilidade de todos os nós na rede, cada um envia a sua tabela para os seus vizinhos periodicamente contendo os parâmetros (D_i, C_i).

Quando um nó recebe uma tabela de um vizinho D_k, compara os seus valores (D_i, C_i) e faz a soma $(D_i, C_{ki} + C_k)$ para cada destino D_i sendo C_{ki} o custo passando pelo vizinho D_k para o destino D_i. Se o custo C_i for superior a $(C_{ki} + C_k)$ a tabela será alterada como o custo menor. Por exemplo, ao se analisar a figura 2.5, inicialmente sem quebra de link entre A e B, verifica-se as tabelas de encaminhamento de cada nó são: A=[(C,8,C);(B,1,B)]; B=[(A,1,A);(C,1,C)]; C=[(A,8,A);(B,1,B)]. No entanto passado algum tempo o nó B envia a

C a sua tabela [(A,1); (B,1)] e C recebe também de A a sua tabela [(C,8);(B,1)]. Desta forma, quando C calcular a sua nova tabela fica a saber que tem um caminho mais curto para A passando por B. Assim sendo, C instala a suas rotas finais para A e B com a seguinte tabela [(A,2,B);(B,1;B)].

Embora o conceito seja bastante simples e escalável, padece de um problema conhecido denominado de *contagem para o infinito* [CRK89]. Para melhor se perceber este problema, voltemos a observar a figura 2.5 considerando agora a quebra de ligações entre os nós A e B. Uma vez que C não tem conhecimento desta falha, ele envia para B a informação que consegue chegar a A com um custo de 2. Como B sabe pela sua tabela que está à distância de C com custo 1, ele actualiza a rota para A com um custo de 3; quando C recebe a tabela de B ele actualiza a rota para A passando por B para um custo de 4 e por assim em diante, até que o custo para chegar ao nó A na tabela de encaminhamento de C acaba por chegar ao valor que define a perda de rota (infinito).

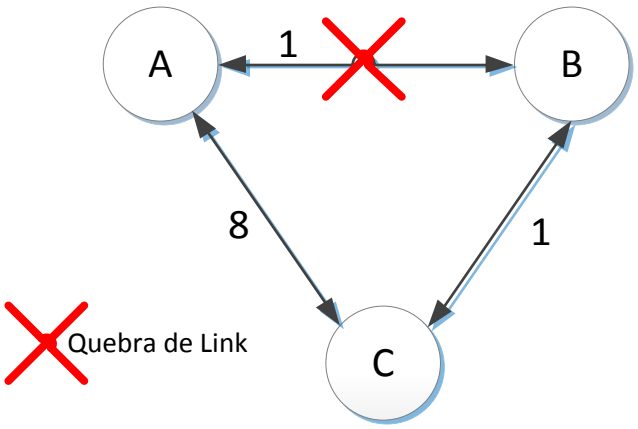

Figura 2. 5– Problema da contagem até ao infinito num loop

Numa tentativa de reduzir o problema de contagem para o infinito, o RIP foi dotado de duas técnicas principais, *split horizon with reverse poison* (separação de horizontes com envenenamento inverso) e *hold down* (manutenção de rota em baixo) [ceig09]. Se um nó aprender uma rota por um vizinho, a técnica diz que não volta a enviar a mesma rota de volta para o vizinho que a enviou. Embora estas técnicas tenham melhorado o protocolo, este continua a ter problemas de convergência lenta.

2.3.1.2. Estado de Linha

Após vários anos de utilização de um protocolo de vectores de distância na ARPANET, apareceu uma nova classe com o nome Estado de Linha (LS, *Link State*). Esta classe deu

origem a dois novos protocolos de intra-domínio: *Open Shortest Path Firts* (OSPF) [Moy98] e o *Intermediate System – to Itermediate System* (IS-IS) [RFC1142]. Uma das principais preocupações destes protocolos prende-se com a manutenção da visão geral da rede com um tempo de convergência baixo, em contraste com a classe de vector de distâncias que padece de reacção lenta a más notícias, como por exemplo a quebra de uma ligação.

O tempo de convergência baixo dos protocolos desta classe, deve-se ao facto de cada nó periodicamente inundar de uma forma fiável toda a rede com uma lista contendo o estado de linha com os seus vizinhos. Quando um nó recebe do seu vizinho uma lista com o estado das ligações (LSA – *Link State Advertisement*), ele passa essa lista para a sua tabela interna e reenvia a mesma mensagem para todos os seus vizinhos excepto para o mesmo por onde recebeu essa mesma lista. Para prevenir mensagens LSA repetidas, cada nó verifica se a mensagem recebida têm um número de sequência mais recente, caso contrário a mensagem é descartada. Desta forma, a actualização das rotas que passavam por uma ligação que falhe é feita numa única operação, sem necessitar de múltiplas trocas de vectores.

Uma vez que cada nó possui a mesma visão global da rede, dado pelo mecanismo de inundação de rede, o cálculo da tabela de encaminhamento faz-se recorrendo ao cálculo da árvore de caminho mais curto para cada destino, tipicamente recorrendo ao algoritmo Dijkstra [Dij59]. Como a principal característica deste tipo de protocolos é manter a visão geral da rede, cada vez que existe uma alteração na topologia, por exemplo uma quebra de ligação, desencadeia-se uma nova inundação da rede e uma nova árvore de caminho mais curto tem de ser calculada em cada nó, para prevenir mensagens repetidas cada nó.

Embora a convergência deste protocolo seja muito mais rápida do que o DV, ele falha na escalabilidade, dado que cada nó tem que armazenar a topologia da rede e calcular a árvore de caminho mais curto cada vez que uma mensagem for recebida, se pretender ter sempre a tabela de encaminhamento actualizada. Desta forma consome mais recursos do que o DV [JI92]. Este problema dá origem a outro problema ainda maior se uma ligação entre dois nós mudar de estado constantemente toda a rede fica a saber desta falha [OBOM03] e todos os nós têm de constantemente estar a "correr" o algoritmo Dijkstra de forma a calcular as suas árvores de caminhos mais curtos, ocupando assim recursos de cada nó.

2.3.1.3. Border Gateway Protocol (BGP)

O encaminhamento ao nível do inter-domínio é muito semelhante ao intra-domínio com a diferença de que o elemento básico agora é o AS em vez do encaminhador. O *Border Gateway Protocol* (BGP) [RLH06], actualmente na sua versão quatro, é o protocolo padrão para o encaminhamento inter-domínio.

O protocolo comporta-se como um algoritmo de vector de caminhos, as mensagens de encaminhamento que exporta para os seus vizinhos são caminhos para destinos alcançáveis. Quando um encaminhador recebe um caminho válido, isto é, livre de ciclos, ele adiciona a sua identificação ao caminho e reenvia esse caminho para os seus vizinhos válidos. Desta forma é possível evitar ciclos no encaminhamento de pacotes.

Para que a troca de rotas seja possível, é necessário que dois encaminhadores que "correm" o protocolo BGP estabeleçam uma ligação TCP entre eles, designada de sessão. Existem dois tipos de sessões: internas (iBGP, *Internal Border Gateway Protocol*) e Externas (eBGP, *External Border Gateway Protocol*). A primeira é usada para distribuir rotas BGP dentro do AS, enquanto a última é usada para troca de caminhos entre ASes. As rotas que estes encaminhadores trocam entre si estão estruturadas na forma: [Destino de prefixo IP, Caminho para chegar ao destino, Atributos do caminho]. O caminho descreve uma lista ordenada por ordem de passagem de cada AS usado para chegar ao destino, enquanto os atributos são usados nas decisões de encaminhamento.

Quando uma mensagem de encaminhamento é recebida, o encaminhador compara-a com um grupo de rotas instaladas para o mesmo prefixo; o caminho melhor é depois escolhido e instalado baseado num conjunto de regras. Estas regras mostradas na tabela 2.1, são processadas de forma ordenada para a decisão de escolha de rota. As regras 1 e 3 usam os valores dos atributos LOCAL_PREF e MED respectivamente, que fazem parte da mensagem de encaminhamento enviada ou recebida.

Nº	Regra	Quem define o valor?
1	Maior atributo LOCAL_PREF	Router Local
2	Menor comprimento de caminho	Router Vizinho
3	Menor atributo Multi Exit Discriminator (MED)	Router Vizinho
4	eBGP sobre iBGP	Ninguém
5	Menor custo rota intra-AS	Router Local
6	Menor ID de encaminhador	Ninguém

Tabela 2.1 – Resumo das regras do BGP para selecção de rota

O valor do atributo LOCAL_PREF é configurado localmente no encaminhador BGP pelos administradores dos ASes. Este atributo define a preferência do caminho e pode ser usado, por exemplo, para controlar o tráfego de saída. Por sua vez, o valor do atributo MED é configurado pelos encaminhadores vizinhos. Este atributo revela o quanto uma rota exportada deve ser discriminada, e é usado para configurar o tráfego de entrada. No entanto, a capacidade do atributo MED pode ser anulada pelo atributo de LOCAL_PREF, dado que é analisado na primeira regra.

Uma vez finda a selecção de rota, se o caminho recebido for seleccionado como o melhor, é então acrescentado à mensagem com as rotas exportadas para outros encaminhadores. Durante a exportação de rotas, os encaminhadores podem manipular os atributos do caminho para diversos fins. As 3 técnicas de manipulação mais usadas são: AS-*path prepend, route aggregation* e o atributo de *community*.

A técnica de AS-*path prepend* (preceder caminho de ASes) consiste na adição do identificador do AS, uma ou mais vezes no caminho da rota de forma a fazer com que esta rota seja menos preferível, dado que o BGP durante o seu processo de selecção de rotas prefere sempre caminhos mais curtos para um dado destino. No entanto, o uso desta técnica pode ser anulada pelo atributo local de preferência (LOCAL_PREF).

A técnica de agregação de rotas (*route aggregation*), como o próprio nome sugere, é usada para aglomerar rotas e exportá-las num único prefixo, ajudando assim a melhorar a escalabilidade do protocolo. Desta forma os encaminhadores que recebem as mensagens instalam apenas uma rota em vez de várias provenientes de diferentes prefixos. Por exemplo, imaginemos um fornecedor com um prefixo /16 atribui um prefixo /24 a um cliente. Desta forma pode apenas anunciar o seu prefixo em vez dos dois, dado que o prefixo /24 pode ser incluído dentro do prefixo /16. No entanto, esta técnica acarreta consequências para os clientes que pertencem a dois provedores em simultâneo (*multi-homed customers*). Estes clientes irão receber tráfego de rotas não agregadas em vez das agregadas pois ao nível das decisões de encaminhamento, o BGP prefere reencaminhar pacotes para um prefixo mais específico.

Por último a técnica do atributo de comunidade (*community*), pode ser usada para marcar rotas exportadas com um identificador conhecido e acordado entre dois ASes. Na recepção de uma rota marcada com o atributo de comunidade, os encaminhadores aplicam um conjunto de acções para a comunidade definida. Por exemplo, as relações de negócio entre sistemas autónomos podem ser obtidas recorrendo a esta técnica. Para melhor se perceber esta prática a figura 2.6 ilustra uma pequena rede com quatro ASes: A,B.C e D, com os seus respectivos

encaminhadores de fronteira R1, R2, R3 e R4. Vai-se assumir que o AS D é cliente do AS A, e que entre D e C foi estabelecido um acordo de só trocar rotas entre os seus clientes (i.e. é uma ligação p2p).

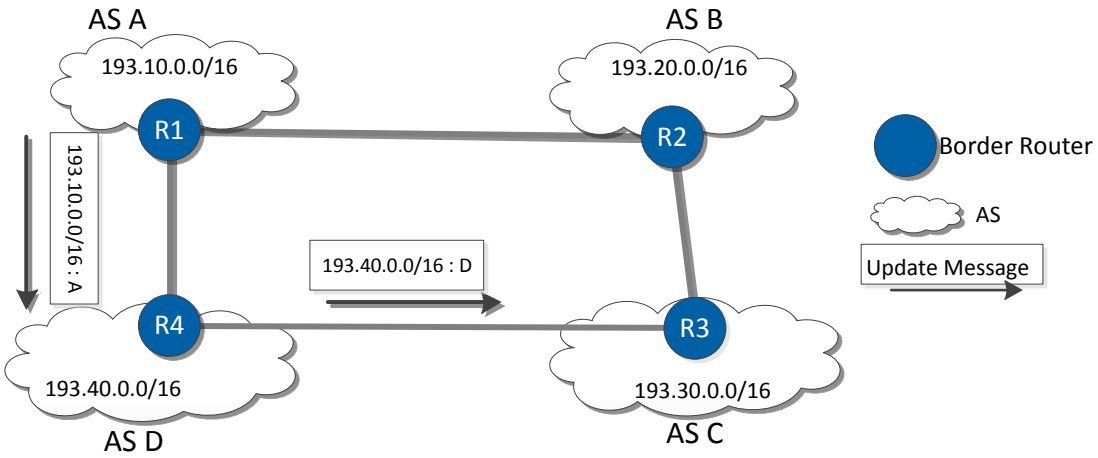

Figura 2. 6– Aplicação relações de negócio com *communities*

Quando o AS A envia uma mensagem de actualização de rota para D, o encaminhador R4 marca esta rota com um comunidade conhecida, que tem como acção associada não exportar esta rota para o exterior. Desta forma a rota será apenas conhecida internamente no AS. Do mesmo modo, quando D envia uma mensagem para C, o encaminhador R3 marca esta rota com uma comunidade do mesmo tipo. Como as ligações são pagas entre os ASes, o AS C não vai estar interessado em suportar o trânsito de tráfego vindo de D para B porque provavelmente D não pagou para usar a ligação entre C e B. Deste modo, os administradores dos sistemas autónomos recorrem ao uso de comunidades para evitar violações de políticas previamente acordadas.

No entanto, o uso de comunidades entre ASes pode levar a decisões não expectáveis se mal configuradas. Mais grave, o uso deste atributo leva também a um aumento do número de entradas nas tabelas de encaminhamento [DB08].

O BGP faculta alguns mecanismos de prevenção contra falhas de ligações: O *Minimum Route Advertisement Interval* (MRAI, intervalo mínimo entre anúncio de rotas) e o *route-flap dampening* (atenuação de batimento de rota). O MRAI [RLH06] funciona como uma espécie de cão de guarda para cada actualização de um prefixo - só permite o anúncio / retracção de actualizações de rotas depois de um intervalo mínimo de tempo configurável após a alteração anterior (o valor recomendado é de 30 segundos). Se um certo prefixo começa a oscilar, isto é, constantemente a ser actualizado, o MRAI vai suster o anúncio de todas a rotas que usam essa

ligação, melhorando deste modo a estabilidade. No entanto o MRAI poderá atrasar o anúncio de actualizações importantes [GP01].

O mecanismo de *route-flap dampening* ignora rotas que se alteram com muita frequência. Deste modo, evita a comunicação de rotas provenientes de encaminhadores com ligações que mudam de estado frequentemente [VCG98]. Da mesma forma como o MRAI, este mecanismo pode atrasar a convergência da rede [MZ02].

O BGP foi pensado como um protocolo de acessibilidade, e não está preparado para todos os requisitos das redes modernas. No entanto o seu uso tem sido estendido para várias finalidades além do simples encaminhamento, tais como a engenharia de tráfego (*traffic engineering*).

Como o protocolo está apenas preocupado com a acessibilidade, é complicado lidar com requisitos de qualidade de serviço (QoS) a nível do encaminhamento. Por exemplo, os mecanismos de estabilidade de rotas podem introduzir várias perdas a nível de pacotes e latência (*jitter*), uma vez que estes mecanismos atrasam a própria convergência da rede [SKM09]. Além da convergência de rotas, a estabilidade de rotas é também necessária.

O BGP tem também limitações para ASes com mais de um fornecedor (*multihomed*). Como estes ASes podem receber diferentes rotas para o mesmo prefixo, as entradas nas tabelas de encaminhamento aumentam consideravelmente [YMBB05]. Além disso, um AS *multihomed* não consegue tirar proveito do encaminhamento multi-caminho, uma vez que o BGP apenas selecciona e exporta o melhor caminho para fins de encaminhamento.

2.3.2. Soluções Académicas para novos Protocolos

Na secção anterior apresentaram-se alguns dos problemas de hoje no encaminhamento inter-domínio. Nesta secção apresentam-se algumas propostas do mundo académico que pretendem resolver os problemas ao nível do inter-domínio e intra-domínio. Começa-se por analisar duas soluções que se focam em melhorar os aspectos dos protocolos intra-domínio, no entanto a sua lógica pode ser também aplicada ao routing ao nível do inter-domínio. Ambos XL [LVPS08] e FCP [LCR+07] propõem extensões para reduzir ou suprimir as actualizações que diminuem o tempo de convergência de uma rede.

2.3.2.1. *Approximate Link-State* (XL)

O XL é protocolo da classe estado de linha, cuja lógica pode ser aplicada a qualquer protocolo padrão deste tipo de classe como por exemplo, o OSPF. A principal preocupação

deste protocolo é difundir as actualizações de estado para os seus vizinhos de forma selectiva, desde que cumpra por ordem qualquer uma destas regras:

1- A actualização é usada para anunciar um aumento de custos (falha de ligação);
2- O vizinho é usado na árvore de caminho mais curto do protocolo;
3- O custo para alcançar um dado destino melhorou por um factor de $1 + \varepsilon$.

Se uma actualização não cumprir nenhuma destas regras, é anulada. Os autores mostraram que assim diminuem o número de actualizações, reduzindo o tempo de convergência da rede. A supressão de rotas tem a desvantagem de se poder estar a usar rotas de menor qualidade (i.e. não óptimas). Além disso, não suporta encaminhamento multi-caminho, tornando-o assim inútil se vários caminhos estão aptos a serem usados.

2.3.2.2. *Failure Carrying Protocol* (FCP)

Com uma lógica diferente, o FCP pode ser usado, como um mecanismo de encaminhamento de último recurso, para qualquer protocolo da classe estado de linha ou mesmo o BGP, suprimindo todas as actualizações. O protocolo assume que cada nó tem um mapa da rede confiável inundado e distribuído por um ou vários coordenadores [CCF+05].

O protocolo assenta sobre a marcação dos pacotes de dados com as ligações que falharam em cada encaminhador atravessado. Desta forma, é possível evitar ciclos com um tempo de convergência nulo, enquanto todos os nós compartilham a mesma perspectiva da rede. Sem uma visão consistente da rede, os seus autores recomendam o uso do encaminhamento a partir da origem. Um novo caminho a partir da origem é recalculado quando o caminho de origem contém ligações inválidas.

O BGP pode ser estendido com o FCP, assinalando no pacote as ligações inválidas resultantes de violações de política de rota. No entanto, esta técnica tem um desempenho limitado por não armazenar nenhuma falha de ligação temporária, diminuindo assim a sua robustez para consequentes decisões de encaminhamento que ignoram as falhas anteriores.

Embora a opção de suprimir o tempo de convergência seja atractiva, a marcação das rotas com informações sobre falha de ligações aumenta a dimensão dos pacotes. Como solução, os seus autores sugerem o uso de rótulos conhecidos em vez das ligações. No entanto, para usar estas etiquetas é requerida alguma coordenação e normalização dos rótulos ao nível do inter-domínio. Além disso, o protocolo tem o mesmo problema de usar rotas não óptimas da mesma forma que o XL.

2.3.2.3. Novos Protocolos na Área do Inter-Domínio

O foco principal desta dissertação é na área do encaminhamento inter-domínio. A este nível existem duas grandes contribuições: a *New Inter-domain Routing Architecture* (NIRA) [YCB07] e o *Hybrid Link-State Protocol* (HLP) [SCE+05]. Ambos visam resolver os problemas de escalabilidade e de convergência baseados na hipótese de que a Internet segue uma estrutura hierárquica.

O HLP assume uma estrutura hierárquica de relações do tipo fornecedor-cliente com raiz num fornecedor de nível-1 (Tier-1), Um provedor deste tipo é sempre um AS do núcleo da Internet, ou seja, é um AS que não tem nenhum fornecedor. Cada fornecedor forma a sua própria hierarquia, portanto um AS que pertence a várias hierarquias é um AS *muti-homed*.

O protocolo anuncia as rotas com base na identificação do AS. Desta forma reduz a dimensão da tabela de encaminhamento, melhorando a estabilidade da tabela de encaminhamento. Também difere do BGP dado que as políticas estão explicitamente publicadas e formam uma hierarquia. Alterações topológicas são anunciadas dentro da hierarquia recorrendo a mensagens de estado de linha (*Link-State*), ao passo que entre hierarquias é usado uma espécie de BGP chamado *Fragmented Path-Vector* (FPV, Vector de caminhos fragmentado).

Este tipo de mensagens, difere do BGP na medida que não inclui o caminho total na mensagem: desde o AS de origem até ao AS de destino, suprime os AS dentro da hierarquia. Uma mensagem FPV é constituída por dois campos (P_i, C_i), em que P_i apresenta um caminho ordenado dos AS de fronteira atravessados até ao destino i, e C_i representa o custo associado para esse destino. Quando uma mensagem de FPV chega a um AS de fronteira, a sua informação é actualizada e disseminada para dentro da hierarquia recorrendo a uma nova mensagem de estado de linha.

Este protocolo dispõe ainda de um mecanismo de supressão de anúncios de caminhos baseado no custo: se um AS dentro da hierarquia perder a conectividade devido a uma falha de ligação, o encaminhador adjacente procura uma nova rota para esse destino; caso a encontre e o seu novo custo não ultrapassar um valor configurável de Δ, o encaminhador decide não anunciar essa falha.

Mas detalhes sobre este protocolo serão explicados e analisados no próximo capítulo, onde é realizada uma proposta original que estende este protocolo. Esta dissertação tem como objectivo principal, estudar a estabilidade e escalabilidade deste protocolo, e adaptá-lo as regras comerciais da Internet.

O NIRA, tal como o HLP, adopta uma estrutura hierárquica. Cada hierarquia tem um encaminhador de *Tier-1* como fornecedor que pertence à região do núcleo. Cada um destes fornecedores atribui recursivamente prefixos de IPV6 aos seus clientes. Esta hierarquia é classificada como a rede de acesso dos clientes, ou simplesmente um grafo ascendente. Em oposição ao núcleo, relações do tipo p2p (*peer to peer*) podem ter endereços não visíveis ao núcleo e atribui-os recursivamente aos seus clientes.

Para propagar informação de encaminhamento, o NIRA usa um protocolo chamado *Topology Information Propagation Protocol* (TIPP). É composto por dois componentes: um vector de caminhos (*path-vector*) que anuncia as rotas ao nível do fornecedor; e um componente de estado de linha, que é usado para controlar mudanças topológicas dentro da hierarquia.

Para estabilidade e convergência um domínio pode configurar o TIPP para proibir o anúncio de rotas entre domínios. Desta forma o protocolo apenas reenvia as mensagens de estado de linha entre domínios que servem para circulação para tráfego.

Até agora o conceito do NIRA é muito semelhante ao HLP. No entanto ele dá liberdade ao utilizador para escolher as rotas para os seus pacotes, limitando-o o conjunto de fornecedores usados. Dessa forma, se um utilizador enviar dados para um dado destino, estes serão encaminhados com base nos endereços do utilizador e destino num sentido hierárquico: em primeiro no sentindo ascendente de acordo com o grafo do cliente e em seguida, no sentido descendente em direcção ao destino com base no grafo ascendente do destino.

O NIRA suporta encaminhamento multi-caminhos (*multipath*). Quando um utilizador deseja utilizar rotas alternativas, ele pode consultar o servidor de *Name-to-Route Lookup Service* (NRLS) que funciona de modo muito semelhante ao servidor de nomes (DNS), e assim obter as rotas disponíveis para o destino desejado. Apesar de dar liberdade ao utilizador para escolher as suas rotas, o NIRA falha no modelo actual de negócios da Internet, dado que os estudos recentes (como se viu na secção 2.2.2), mostram que a Internet tende seguir uma estrutura sem escala, em oposição à ideia de uma hierarquia.

No entanto se tivermos em conta a realidade das relações multi-provedor (*multi-homed*) da Internet, o NIRA é um modelo mais capaz de lidar com hierarquias do que o HLP, dado que deixa o utilizador escolher o seu próprio conjunto de hierarquias.

2.4. Resumo

Neste capítulo começou-se por introduzir um pouco da história da Internet, como é que tudo evoluiu, passando pelo crescimento da própria rede até ao factor mais importante que faz mover a Internet e inovar a tecnologia inerente: a comercialização da tecnologia.

De seguida entrou-se noutra área mais importante para esta dissertação, os protocolos de encaminhamento. Caracterizou-se a estrutura da Internet, fazendo uma breve menção ao modelo de negócios a que a Internet está sujeita. Também se falou sobre os vários estudos efectuados para se perceber a organização da Internet, bem como os programas desenvolvidos nesta área para gerar topologias que ajudam ao desenvolvimento e melhoramento dos protocolos de encaminhamento.

Por fim, referiram-se os protocolos mais comuns e utilizados na rede, quer a nível do intra como inter-domínio com as suas vantagens e desvantagens. Mencionaram-se, também duas grandes contribuições completamente inovadoras para um protocolo de encaminhamento ao nível do inter-domínio. É sobre este nível e sobre um destes protocolos que assenta o tema principal desta dissertação, o estudo da estabilidade e escalabilidade sobre o HLP. No próximo capítulo é explicado detalhadamente um novo protocolo de encaminhamento proposto nesta dissertação, que estende o HLP para considerar o modelo de negócios da Internet.

Capítulo 3.

ARQUITECTURA DO PROTOCOLO

3.1. Introdução

Este capítulo apresenta a arquitectura proposta para estender o HLP (Hybrid Link-State Path-Vector Protocol) de forma a torná-lo compatível com o modelo de negócios da Internet. Como se viu no capítulo anterior, o protocolo BGP (Border Gateway Protocol) é muito complexo e apesar da sua flexibilidade, a manipulação de certos atributos pode-se tornar num paradigma entre flexibilidade e complexidade. Os administradores dos Sistemas Autónomos tendem a manipular os atributos do BGP com o intuito de alterar os mecanismos de encaminhamento e não só; por exemplo, o BGP não distribui informação sobre as políticas de encaminhamento, estas são implementadas localmente por filtros cujo conteúdo é mantido secreto. Como resultado o BGP sofre de problemas algorítmicos, incluindo fraca escalabilidade, pouco isolamento a falhas na rede, e baixa convergência resultante da exploração de caminhos (*Path Exploration*) não uniformizada.

Desenhar um protocolo ao nível do Inter-Domínio que satisfaça ambos os requisitos de políticas e algorítmica, representa um grande desafio. Existe sempre o inerente conflito entre a necessidade económica de manter as políticas de encaminhamento privadas e a carência estrutural para desenhar algoritmos robustos.

No trabalho de *Subramanian* [SCE+05] é proposta uma nova arquitectura para o encaminhamento a nível do Inter-Domínio, HLP (Hybrid Link-State Path-Vector Protocol), baseado na assumpção que a Internet segue um modelo hierárquico, os encaminhadores usam dois protocolos para anunciar e aprender rotas, dentro da hierarquia o anúncio de rotas é feito com recurso ao protocolo estado de linha (Link-State Protocol) e entre hierarquias usa uma versão modificada do protocolo vector de caminhos (Path Vector). Esta dissertação visa estender o trabalho de *Subramanian* por forma a tornar o protocolo HLP compatível com o

modelo de negócios actual da Internet. Primeiro é apresentada uma visão geral sobre os fundamentos do HLP e de seguida uma extensão que o torna compatível com o modelo de negócios da Internet.

3.2. HLP – Fundamentos Básicos

O protocolo HLP introduz uma nova abordagem ao encaminhamento Inter-Domínio cujas principais preocupações, segundo os seus autores, vão de encontro às áreas onde o BGP necessita de mais modificações: Estrutura de encaminhamento, políticas, granularidade e estilo de encaminhamento.

3.2.1. Estrutura de Encaminhamento

Para suportar políticas baseadas em caminhos, o BGP revela informações sobre o caminho completo desde a origem até ao destino, isto é, todos os Sistemas Autónomos (ASes - Autonomus Systems) atravessados são incluídos na rota. Como resultado não só prejudica a escalabilidade com também torna difícil isolar eventos de encaminhamento dado que falhas locais na rede podem ser visíveis globalmente. Além disso, a interdependência resultante entre ASes torna a Internet vulnerável a problemas de configuração um simples erro de configuração ou um encaminhador comprometido podem afectar o resto da rede.

Para evitar estes problemas, o HLP "esconde" alguma informação no caminho revelado ao seu par entre hierarquias. Fá-lo assumindo a natureza hierárquica da estrutura de encaminhamento resultante das relações típicas entre ASes (pares, clientes e provedores) omitindo assim pequenas ocorrências resultantes de eventos dinâmicos no interior da hierarquia, entre os nós vizinhos que definem a raiz.

3.2.2. Políticas

No que diz respeito as políticas, completamente oposto ao revelar todo caminho ao seu vizinho, o BGP mantém em privado a configuração das políticas. No entanto como se viu no capítulo anterior, grande parte das relações entre ASes pode ser inferida e categorizada como pares, clientes ou provedores. Em aproximadamente 99% dos Ases, as definições de políticas para regras de exportação e preferência de rotas seguem duas orientações simples baseadas nas inter-relações entre os mesmos:

- Regra de exportação de rota – Não reencaminhar rotas anunciadas por um par ou um fornecedor para outro par ou fornecedor;
- Regra de preferência de rota – Preferir rotas por clientes em vez das rotas anunciadas pelos seus pares ou fornecedores.

Enquanto estas políticas predominam no uso das configurações dos ASes, o BGP recusa-se a revelá-las explicitamente. Isto significa que o BGP é incapaz de distinguir entre uma política mal configurada e uma genuína tornando-o complicado de gerir e de diagnosticar erros, estando assim mais vulnerável a erros de configuração e ataques. Adicionalmente, na ausência de orientações estritas na definição de políticas, pode levar a conflitos de políticas que resultam em baixa convergência e instabilidade de encaminhamento.

O HLP, em contraste, publica explicitamente as relações fornecedor-cliente e restringe o conjunto normal de caminhos disponíveis para um dado destino aos que obedecem às hierarquias definidas por essas relações. Como resultado temos um protocolo de encaminhamento que, no caso comum, é capaz de reconhecer erros de configuração e limitar a propagação de rotas.

3.2.3. Granularidade de Encaminhamento

O BGP usa encaminhamento baseado em prefixos. Enquanto o modelo inicial promoveu a agregação de prefixos para melhorar a escalabilidade, o seu uso hoje em dia é dominado pelo fenómeno oposto de desagregação para engenharia de tráfego, múltiplos caminhos e políticas de encaminhamento. Nos últimos anos mais de 11.000 redes desagregaram o seu prefixo. Em combinação com o advento de muitas redes /24, isto resultou num aumento alarmante no número de prefixos distintos numa tabela de encaminhamento, já que um único evento desencadeia uma actualização separada para cada prefixo.

Embora o encaminhamento BGP por prefixos resulte em grande instabilidade nas rotas e em tabelas de caminhos de grande dimensão, ainda assim geralmente não suporta múltiplos caminhos. O HLP foi desenhado para encaminhar baseado na identificação do AS em vez de prefixos, originando tabelas de encaminhamento substancialmente mais pequenas. Além de reduzir a instabilidade de rotas face a falhas, o encaminhamento ao nível do AS tem outros benefícios auxiliares. O mapeamento entre prefixos de endereços e localizações (identificado pelo AS) é mais estático do que a topologia da rede; deste modo podem ser usados mecanismos de transporte de segurança mais apropriados, não só para a informação topológica da rede bem como também para o mapeamento entre AS e os prefixos que ele

detém. Por sua vez, permite detectar facilmente erros de configuração quando erroneamente um AS reivindica a posse de um prefixo que pertence a outro AS.

3.2.4. Estilo de Encaminhamento

O BGP usa como protocolo de encaminhamento vector de caminhos (PV – *Path Vector*). Este protocolo admite políticas complexas (uma vez que permite ao AS aplicar as suas próprias políticas em todo o caminho) e facilmente suprime ciclos. No entanto, o pior caso de convergência no protocolo de vector de caminhos cresce exponencialmente com o comprimento do caminho [CAF00]. Este protocolo também introduz interdependência desnecessária que impede a escalabilidade e as propriedades de isolamento do protocolo, dado que um simples evento de mudança de estado de uma ligação, desencadeia actualizações de rotas a todos os caminhos para ASes que atravessam essa ligação, expondo desta forma uma enorme fracção de eventos globalmente visíveis.

Em alternativa a este protocolo temos os bem conhecidos protocolos: Vector de Distâncias (DV- *Distance Vector*) e o Estado de Linha (LS – *Link-State*), mas nenhum é adequado para suportar encaminhamento baseado em políticas. O Vector de Distâncias (DV) não revela qualquer informação do caminho até ao destino, o que impede o suporte de políticas. Por outro lado o Estado de Linha (LS) pode violar as políticas de privacidade dado que revela toda e qualquer actividade a todos os ASes de destino.

Aparte das políticas, ambos protocolos têm as suas vantagens e limitações. O LS tem uma convergência rápida e incorre numa baixa instabilidade, dado que as actualizações são para eventos de ligações e não nas alterações nas rotas (nos protocolos como o DV e PV um evento de uma ligação pode causar múltiplas alterações nas rotas). Além disso, o diagnóstico de falhas é fácil de analisar com o LS uma vez que fornece uma visão global do estado da rede. No entanto, esta visibilidade global limita a escalabilidade e isolamento de falhas. Em contraste o PV pode ser adaptado para fornecer bom isolamento de falhas, como se pode ver na próxima secção. No entanto o diagnóstico de falhas torna-se difícil.

Nenhuma das abordagens é uma solução ideal, mas cada uma beneficia dos seus méritos. Por isso o HLP é uma solução híbrida que usa os protocolos LS e PV para encaminhamento inter-domínio.

3.3. HLP – Modelo de Encaminhamento

3.3.1. Estrutura

O HLP foi desenhado assumindo a existência de uma estrutura hierárquica maioritariamente formada por relações do tipo fornecedor-cliente entre ASes, ilustrada na figura 3.1.

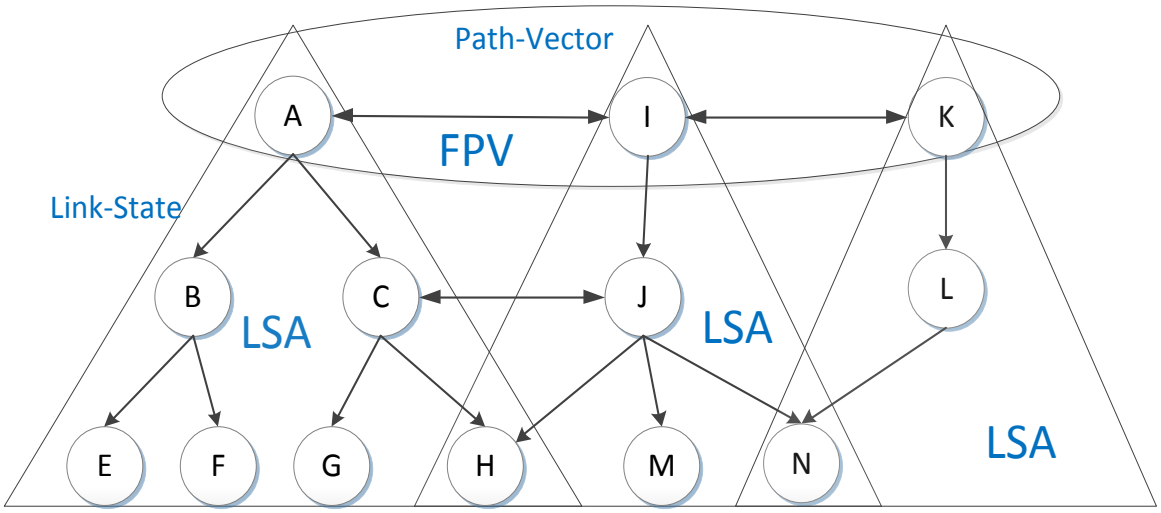

Figura 3. 1– Exemplo de hierarquia entre ASes

Classifica-se como nível 1(Tier-1) o AS raiz da hierarquia fornecedor-cliente. Note-se que esta classificação difere da terminologia convencional de ISP (*Internet Service Provider*) tier 1, em que um ISP de nível inferior poderia ser classificado como sendo de nível 1. Pela definição do HLP, um AS só é de nível 1 se não for um cliente para outro fornecedor. Por exemplo, os ASes A,I e K da figura 3.1 são os ASes raízes, e por sua vez de nível 1, das suas respectivas hierarquias dado que não possuem qualquer fornecedor, isto é, não são clientes de nenhum outro AS

Um AS com múltiplos fornecedores (*multihomed* AS) pode fazer parte de uma ou mais hierarquias fornecedor-cliente. Como se pode ver na figura 3.1, o sistema autónomo H pertence em simultâneo às hierarquias de A e I. Por outro lado, os AS das diferentes hierarquias fornecedor-cliente podem-se interligar entre eles usando ligações do tipo par-a-par (peer-to-peer) e estas ligações poderão ocorrer a diferentes níveis da hierarquia, como é o caso dos ASes A e C da figura 3.1. Assume-se que não existem ciclos na hierarquia.

3.3.2. Modelo de Propagação de Rota

Baseado numa estrutura de encaminhamento hierárquica, o HLP usa uma combinação de encaminhamento por estado de linha LS (*link-state*) dentro da hierarquia com encaminhamento por vector de caminhos (*path-vector*) entre as hierarquias.

Dentro de uma hierarquia quando um evento de encaminhamento inter-AS ocorre, os outros ASes são notificados através de um anúncio de estado de linha. Essa comunicação é feita a nível da granularidade do AS e não dos encaminhadores (routers). Cada AS mantém informações de estado de linha sobre as ligações fornecedor-cliente dentro da sua hierarquia, que são actualizadas através de mensagens de actualização.

Entre hierarquias o HLP usa encaminhamento por vector de caminhos, similar ao BGP, onde um AS propaga a informação de acessibilidade para um dado destino através de um vector de caminho marcado com todos os AS envolvidos no percurso. A distinção principal é que o HLP usa um vector de caminhos fragmentado, FPV (*Fragmented Path Vector*), que contém apenas uma parte do caminho para um dado destino. O FPV omite a parte do caminho dentro da hierarquia. Como o comprimento do caminho não tem significado de encaminhamento, cada anúncio FPV inclui também a métrica de distância.

Ilustra-se agora, através de um exemplo, o modelo básico de propagação de rotas dentro e entre hierarquias. Cada AS mantém uma base de dados sobre a topologia LS e uma tabela de encaminhamento para o vector de caminhos. Dois tipos de mensagens são trocados pelos nós; anúncios de estado de linha (LSA – *Link State Advertisements*) e vectores de caminhos fragmentados FPV.

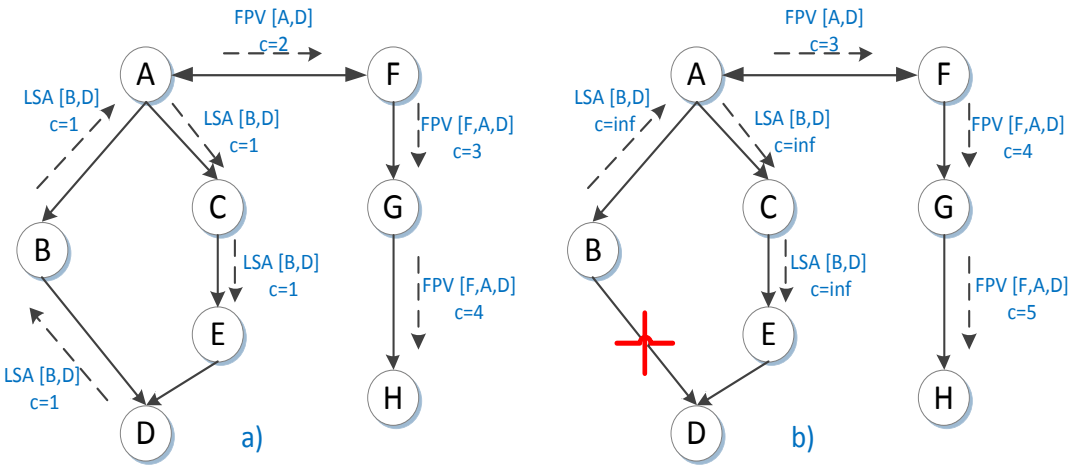

Figura 3. 2– Exemplo simples de propagação de rota em caso de falha de link

Considere-se o exemplo topológico da figura 3.2 (a) composta por duas hierarquias com raiz em A e F e a ligação entre B e D. A rota para D através desta ligação torna-se conhecida inicialmente através de uma mensagem de LSA que notifica todos os nós na hierarquia de A da existência do custo da ligação entre B e D (considera-se para efeitos deste exemplo que todas as ligações têm custo 1). O AS A, ao receber este LSA propaga um vector de caminho para F com um FVP (B,D) com um custo de métrica 2. Por sua vez, quando este chega a F é distribuído dentro da hierarquia sem qualquer modificação do caminho; note-se que nenhum dos caminhos dentro da hierarquia de A ou F aparecem no caminho do FPV, apenas o custo é actualizado.

No cenário modificado e ilustrado na figura 3.2 (b), quando a ligação entre B e D falha, todos os ASes contidos na hierarquia de A recebem a notificação por LSA da falha da ligação. No entanto, como A tem um caminho alternativo dentro da sua própria hierarquia para D, é enviado um FPV para F com actualização do custo, correspondendo à operação de remoção de rota no BGP (*route withdrawal*). Deste modo, F propaga o FPV com o custo actualizado dentro da sua própria hierarquia. No entanto, se A não tivesse uma rota alternativa, iria eliminar a rota para D, propagando a informação para F. Complementarmente, o HLP possui um mecanismo de supressão de actualizações de rota, descrito na próxima secção, que visa reduzir a sinalização na rede. Com este mecanismo, a alteração de custo da rota para D poderia ser omissa a F uma vez que A tem uma rota alternativa dentro da sua hierarquia.

Os anúncios de mensagens FPV podem ser propagados através de mais que uma ligação para um par (*peering link*); o reenvio de mensagens deste tipo permite ao HLP exportar rotas entre hierarquias, como é o caso do AS I na figura 3.1 que anuncia rotas de A para K. Em tais situações estamos perante o cenário de pares indirectos. Neste cenário o caminho na mensagem de FPV inclui todos os ASes envolvidos no percurso, evitando-se ciclos de encaminhamento e o problema de contagem para o infinito.

Teorema 1: *Na ausência de ciclos na topologia fornecedor-cliente, se cada AS seguir as regras de propagação do HLP e cada AS escolher uma rota de cliente, se existir, então o protocolo de encaminhamento é desprovido de ciclos não transitórios de encaminhamento e do problema de contagem para infinito.*

A prova deste teorema [LCM+04] faz uso de simples rótulos nas ligações. Associa-se um rótulo 3 a qualquer ligação do tipo cliente-fornecedor que apareça ao longo do caminho, rótulo 2 a uma ligação par e rótulo 1 às ligações fornecedor-cliente. As regras de propagação do HLP asseguram que qualquer caminho válido é sempre não crescente.

3.3.3. Supressão de Actualizações de Caminhos

O modelo descrito na secção anterior de propagação de rotas, é insuficiente para obter uma boa escalabilidade e isolamento a falhas. Para aperfeiçoar estas duas métricas é necessário ocultar as actualizações de rotas aos ASes vizinhos, propagando-as apenas quando for necessário. Esse objectivo é alcançado recorrendo ao conceito de custo "escondido". Assume-se que o custo num caminho é uma métrica aditiva. Seja *R* o caminho primário para um dado destino, quando um AS observa um incremento no custo do caminho ou uma falha em *R*, analisa se possui um caminho alternativo *S* com um custo comparável com *R*. Em caso positivo, o AS comuta para a nova rota *S*, suprimindo assim as actualizações para os ASes, que poderiam ser desencadeadas pela alteração no caminho *R*.

Dois caminhos dizem-se comparáveis se a diferença de custos entre os dois é menor que um valor limite Δ, definido pelo próprio AS.

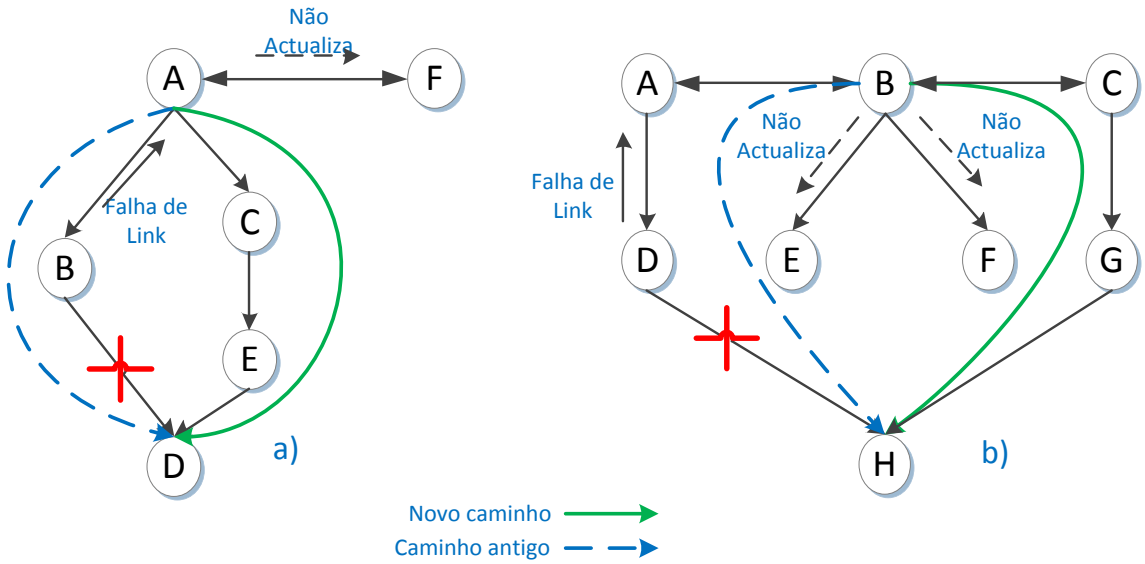

Figura 3. 3– Formas de suprimir actualização do custo de um caminho

O conceito de custo comparável relaxa a noção do encaminhamento pelo caminho mais curto e ajuda a alcançar um melhor isolamento a falhas e uma melhor escalabilidade. No entanto, é necessário ter cuidado pois sempre que um AS anula uma actualização na mudança de custo num caminho para o seu vizinho (rotas através deste AS) o estado da tabela de encaminhamento mantida pelo AS contíguo torna-se obsoleto. Se a supressão não for efectuada correctamente, pode-se introduzir ciclos de encaminhamento não transitórios. O HLP usa explicitamente a hierarquia do AS e as regras definidas na secção 3.2.2 de forma a evitar ciclos de encaminhamento não transitórios.

O HLP define apenas três regras para suprimir a actualização da alteração de custo num caminho:

1) Não propagar alterações de custo de rotas para clientes, inferiores a valor máximo Δ . A figura 3.3 (a) ilustra como o processo é feito: o AS A não propaga para F a alteração do custo da rota para D, uma vez que possui um rota alternativa via C com custo 3. Como, não informa o vizinho AS F sobre o novo custo para D, para todos os efeitos, do ponto vista do AS D é como se não tivesse existido qualquer falha.
2) Não propagar alterações de custo através de um vizinho par (se o AS anterior no caminho também é um vizinho par) para os seus clientes, inferiores a um valor máximo Δ. Pela figura 3.3 (b) pode-se ver que o AS B não actualiza os seus clientes E e F sobre a falha entre D e H; simplesmente comuta para o novo caminho para D via o AS C.
3) Suprimir notificações na falha de uma das múltiplas ligações pares entre dois ASes. Ao contrário de 1 e 2 que envolvem a supressão de custos pelos ASes mais acima na hierarquia dos que originaram a falha. Neste caso a falha é local aos dois ASes, é inteiramente opção do AS se propaga ou não a alteração do custo.

3.3.4. Manipulação de Relações Complexas

Na prática, nem todas as relações inter-domínio são puramente do tipo fornecedor- cliente ou par-a-par. Dois exemplos de relações complexas podem ser: a) relação de irmão entre dois ASes (*sibling relationship*) que são propriedade da mesma administração; b) dois ASes pretendem ter relações diferentes para destinos distintos ou confinados a uma localização geográfica, por exemplo, ter relação par-a-par para ligações dentro da Europa e fornecedor-cliente para ligações fora da Europa.

No HLP este tipo de relações complexas são modeladas como se fossem ligações par-a-par dentro da hierarquia. Desta forma, ao tratar essas ligações como sendo par-a-par, o HLP emula o comportamento do BGP mantendo assim a compatibilidade e estado actual. Por outro lado um AS envolvido neste tipo de ligação não necessita de revelar a natureza da complexidade.

3.3.5. Síntese

Nas secções anteriores, descreveu-se o HLP no seu modelo padrão. Apresentou-se a estrutura, o modelo de propagação de rotas e o factor principal deste protocolo, a supressão de actualizações. No entanto, este protocolo não é compatível com o actual modelo de negócios da Internet descrito no capítulo 2. Para o tornar conciliável apresenta-se na próxima secção uma extensão ao protocolo.

3.4. HLP ++ Compatibilização com o Modelo de Negócios da Internet.

3.4.1. Enquadramento.

Como se viu no Capítulo 2 secção 2.2.1, o modelo de negócios da Internet assenta em relações do tipo fornecedor-cliente, cliente-fornecedor, par-a-par e irmão-a-irmão. Estas relações são necessárias na medida em que nenhum AS por si próprio consegue servir aos seus clientes acessibilidade a todos pontos da Internet sem recorrer a outros ASes Este tipo de relações envolve contratos monetários entre os ASes para transporte de tráfego. Como é perceptível, estas entidades não estão interessadas em que os contratos sejam violados, ou dito de outra forma, se um AS não pagou por uma ligação então não a poderá usar.

Os protocolos de encaminhamento têm de ser compatíveis com as políticas que definem o modelo de negócios da Internet. Para tal é necessário restringir o anúncio de rotas efectuado pelo protocolo, de maneira a não se anunciar rotas inválidas. No presente capítulo, nas secções anteriores apresentou-se o protocolo HLP na sua versão padrão. Na área do encaminhamento inter-domínio, este protocolo não satisfaz alguns dos requisitos dada a natureza dos dois protocolos que usa de uma forma híbrida. Na próxima secção apresentam-se as incompatibilidades.

3.4.2. Incompatibilidades

Como se viu na secção 3.3.2, o HLP emprega dois protocolos de encaminhamento de uma forma híbrida. Cada nó usa em simultâneo os protocolos estado de linha (LS) e vector de caminhos (PV).

Dentro de uma hierarquia, o anúncio de rotas é realizado com recurso ao LS. No HLP as mensagens LSA (*Link State Advertisements*) diferem das usadas no LS típico, na medida em que transportam informação adicional necessária ao protocolo. A estrutura de mensagens é apresentada na próxima secção.

Entre hierarquias as rotas são anunciadas usando o PV numa versão modificada e identificada como FPV (*Fragmented Path Vector*). A diferença em relação ao PV original é que o caminho nas mensagens não contém todos os ASes envolvidos no percurso.

3.4.2.1. Ao Nível do LSA

O LSA é um protocolo do tipo LS onde cada AS envia uma mensagem com a lista dos seus vizinhos para todos os nós, utilizando um algoritmo de inundação. O HLP torna-se incompatível com o actual modelo de negócios ao nível do LS, dado que este protocolo foi desenhado para o cenário em que todos os ASes, têm a mesma visão da rede. Com as políticas do modelo de negócios esta propriedade pode torna-se falsa.

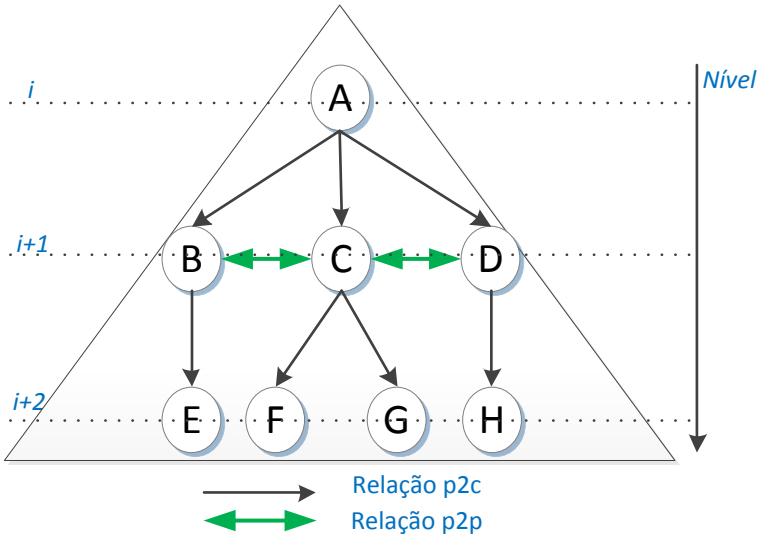

Figura 3. 4– Exemplo de incompatibilidade com o LSA

Observe-se a figura 3.4 que apresenta uma hierarquia com diferentes relações entre ASes. Sem o modelo de negócios, todos os ASes teriam a mesma visão da rede: o AS E para chegar ao AS H saberia que dispõe de dois caminhos através de B; por outro lado B saberia que para chegar a H teria uma rota via C e outra via A. No entanto, as políticas associadas ao modelo de negócios tornam a visão de B diferente, uma vez que B e C possuem uma relação do tipo p2p, significando que ambos podem anunciar rotas dos seus clientes, mas não dos seus pares ou fornecedores.

O AS C não está interessado em transportar tráfego de B para D ou de B para A uma vez que a ligação de p2p negociada entre os dois destina-se exclusivamente ao tráfego entre os seus clientes. Pode-se ver, desta forma, que a visão que o AS B tem da rede é diferente da visão de C. Segundo as regras da relação p2p, B consegue chegar a H apenas por A; por outro lado, C consegue chegar a H por A e D. No entanto, C não pode anunciar a B estes caminhos, uma vez que não transporta o tráfego de B destinado a H ou A pelos motivos anteriormente explicados.

Pelos motivos expostos, o LSA tem de ser modificado para que as políticas do modelo de negócios actual da Internet sejam respeitadas.

3.4.2.2. Ao Nível do FPV

O FPV é um protocolo do tipo PV. Ao contrário dos protocolos LS, os protocolos PV trocam tabelas de caminhos entre os nós. No HLP este protocolo é usado para anúncio de rotas entre hierarquias, ou seja, os ASes entre hierarquias trocam entre si as suas tabelas de encaminhamento. Numa topologia com uma hierarquia exclusivamente com relações p2c, os ASes que são as raízes da hierarquia conhecem os caminhos para todos os ASes dentro da sua hierarquia. No entanto, as ligações FPV entre hierarquias podem ocorrer entre ASes num nível mais baixo da hierarquia. Dito de outra forma, as ligações p2p entre hierarquias onde se corre FPV não são exclusivas dos ASes raízes. Este facto é demonstrado no capítulo 4, onde é feito um estudo sobre a natureza topológica da Internet actual.

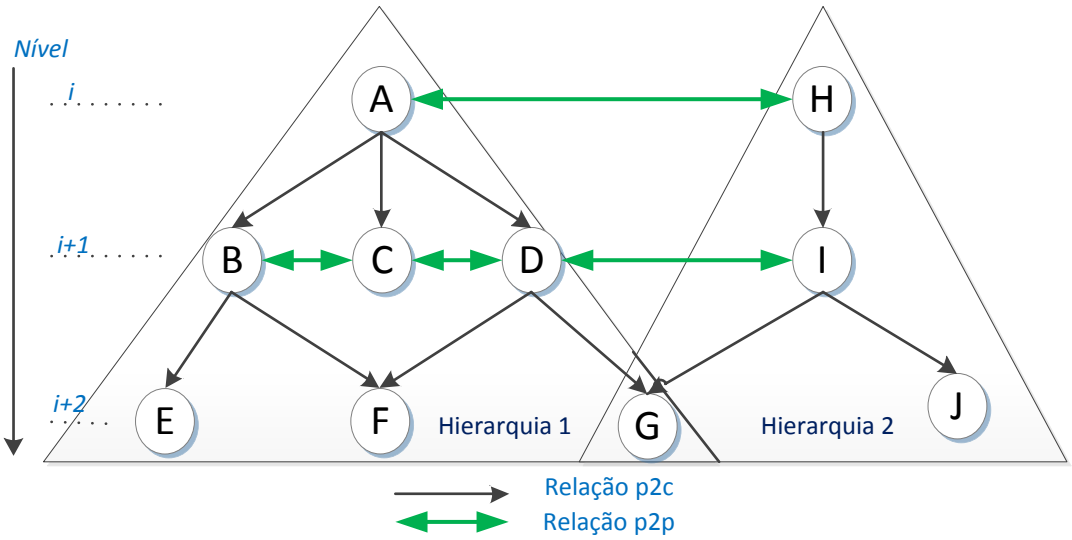

Figura 3. 5– Exemplo de incompatibilidade com o FPV

A figura 3.5 ilustra um exemplo onde uma relação p2p entre hierarquias é estabelecida a um nível inferior na hierarquia, através da ligação entre o AS D e I. No cenário sem modelo

de negócios, o AS D envia sua tabela de encaminhamento para o AS I, incluído os caminhos para os ASes A e C do mesmo modo o AS I manda a sua tabela para D abrangendo a rota para o AS H.

Nesta situação, as tabelas de encaminhamento trocadas entre os AS D e I não é compatível com o actual modelo de negócios da Internet, uma vez que a ligação entre eles é do tipo p2p. Pelo mesmo motivo da incompatibilidade no LSA, tanto o AS D como o I não estão interessados em transportar tráfego com origem no cliente do seu vizinho para o seu fornecedor, por exemplo tráfego com origem em J e destino em A não pode usar a ligação entre D e I. Esta ligação deve ser usada exclusivamente para tráfego entre os seus clientes; F,G e J.

Pelas razões expostas, o protocolo FPV deve ser alterado de modo a que o modelo de negócios seja honrado.

3.4.3. Protocolo HLP ++.

Na secção anterior apresentaram-se as limitações do protocolo HLP em relação ao modelo de negócios actual da Internet. Nesta secção são apresentadas as modificações realizadas ao HLP, designadas genericamente de HLP++

3.4.3.1. Regras de Exportação de Caminhos.

O HLP++ modifica as regras de exportação de caminhos. A nível do LSA o HLP usa duas regras para o reenvio de caminhos recebidos pelos ASes vizinhos:

1- Se a mensagem de actualização de rota vem de um cliente, é reenviada para todos os ASes vizinhos excepto pelo cliente que a enviou.
2- Se a mensagem de actualização vem de um fornecedor, reenviada exclusivamente para os clientes.

O HLP++ acrescenta uma nova regra para lidar com mensagens recebidas pelos nas relações p2p, onde o protocolo HLP é omisso. Como foi visto na secção 3.4.2.1, os ligações do tipo p2p não podem ser utilizadas para transportar outro tráfego que não seja apenas entre os clientes dos dois ASes que estabeleceram a relação. Assim, é acrescentada uma nova regra ao nível do LSA:

3- Se a actualização vem de um par, reenvia-se exclusivamente para os clientes.

A nível do FPV, o protocolo usa duas regras para o reenvio de rotas, que são suficientes em termos de compatibilidade com o modelo de negócios:

1- Se a mensagem chega através de um fornecedor, reenvia-se apenas para os clientes, nunca para outros fornecedores ou pares.
2- Se a mensagem provém de um par, reenvia-se para os clientes e outros pares, mas nunca para outros fornecedores.

3.4.3.2. Mensagens

Dado que o protocolo HLP++ utiliza o LSA e FPV de uma forma híbrida, são necessários diversos tipos de mensagens para troca de informações entre os ASes de forma a transportar informação sobre as relações entre ASes. A tabela 3.1 sumariza os tipos de mensagens enviadas ou recebidas pelos ASes

Mensagem	Descrição	Protocolo
CUSTOMER_LSA	Mensagem enviada por um cliente dentro da hierarquia	LSA
PROVIDER_LSA	Mensagem enviada por um fornecedor dentro da hierarquia	LSA
PEER_FPV	Mensagem enviada por um par de outra hierarquia	FPV
PROVIDER_FPV	Mensagem enviada por um fornecedor dentro de uma hierarquia com origem em outra hierarquia	FPV
PEER_LSA	Mensagem enviada por um par dentro da hierarquia	LSA

Tabela 3.1 – Mensagens enviadas pelos ASes

O protocolo HLP usava para troca de rotas entre os ASes apenas os primeiros quatro tipos de mensagens incluídos na tabela 3.1, onde se pode ver o tipo de mensagem, descrição e a que classe de protocolo pertence. Atendendo à alteração efectuada ao LSA, referida na secção anterior, foi necessário adicionar um tipo de mensagem novo, PEER_LSA, para suportar as trocas de rotas entre ASes com relações do tipo p2p. Este tipo de mensagem vem garantir a aplicação da regra 3 adicionada ao protocolo a nível do LSA, activada quando um AS recebe uma mensagem deste tipo.

3.4.3.3. Alteração ao LSA

Como se viu na figura 3.4 da secção 3.4.2.1, o LSA tem de ser alterado de forma a suportar as relações p2p e torná-las compatíveis com o actual modelo de negócios da Internet. Na secção 3.4.3.1 apresentou-se uma das modificações a nível das regras de exportação de caminhos. No entanto, esta alteração não é suficiente. A figura 3.6 ilustra um exemplo onde, devido à natureza do protocolo LSA e do seu algoritmo de cálculo de caminho mais curto Dijkstra [Dij59], o protocolo se torna incompatível com o modelo de negócios.

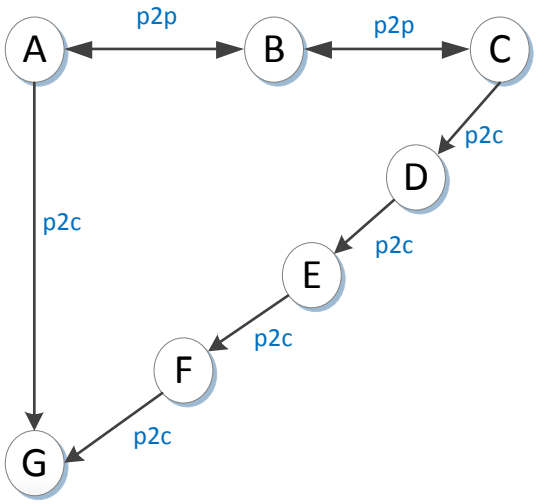

Figura 3. 6– Exemplo de incompatibilidade com LSA, Dijkstra

No cenário apresentado na figura 3.6, sem alteração ao LSA, o tráfego com origem no AS G e destino C, tem como caminho [G,A,B,C]. No entanto, pelo modelo de negócios da Internet, o caminho de G a C deve ser [G,F,E,D,C]. Esta incompatibilidade acontece devido ao algoritmo de cálculo seleccionar o caminho mais curto, e o caminho pelas ligações do tipo p2c ser mais longo do que o caminho pelas ligações p2p. Mas neste caminho o AS B transporta tráfego entre os ASes A e C.

Para contornar este problema foi necessário alterar o LSA de forma que quando o AS B envia a lista dos seus vizinhos ao AS A omite o AS C. A modificação do LSA consiste em adicionar uma verificação intermédia no processo de inundação, em que a lista de vizinhos a enviar depende do tipo de relação que tem com o vizinho a quem vai enviar a lista, e.g. se a relação com o vizinho for do tipo p2p, exclui da lista todos os vizinhos do tipo par e fornecedor.

A figura 3.7 ilustra o algoritmo LSA utilizado no HLP++. Este algoritmo é activado a partir do processo de encaminhamento principal, designado HLP++, após a recepção ou o envio de mensagens. O processo de envio da lista de vizinhos da figura 3.7 é uma versão modificada do método de inundação de rede do protocolo LS, como já foi referido anteriormente. Durante o processo de construção da tabela de vizinhos a enviar, o AS analisa o tipo de relação que possui com o vizinho a quem vai enviar a sua tabela: para cada vizinho a incluir na tabela verifica a sua relação com ele e aplica as regras descriminadas no esquema da figura 3.7: se o AS de destino for um AS do tipo par, durante o processo de construção da tabela de vizinhos a enviar, todos dos ASes do tipo par ou fornecedor são excluídos da tabela,

se o AS destino for do tipo fornecedor exclui da lista todos os vizinhos do tipo par, caso contrário (se for do tipo cliente) envia todos os vizinhos.

A mensagem enviada pelo AS com a tabela é marcada com um tipo de mensagem que pode ser: PROVIDER_LSA, PEER_LSA e CUSTOMER_LSA. Estes rótulos servem para identificar o tipo de relação que o AS tem com o vizinho que vai receber a mensagem, facilitando o processo de reenvio após recepção da mesma pelo AS vizinho.

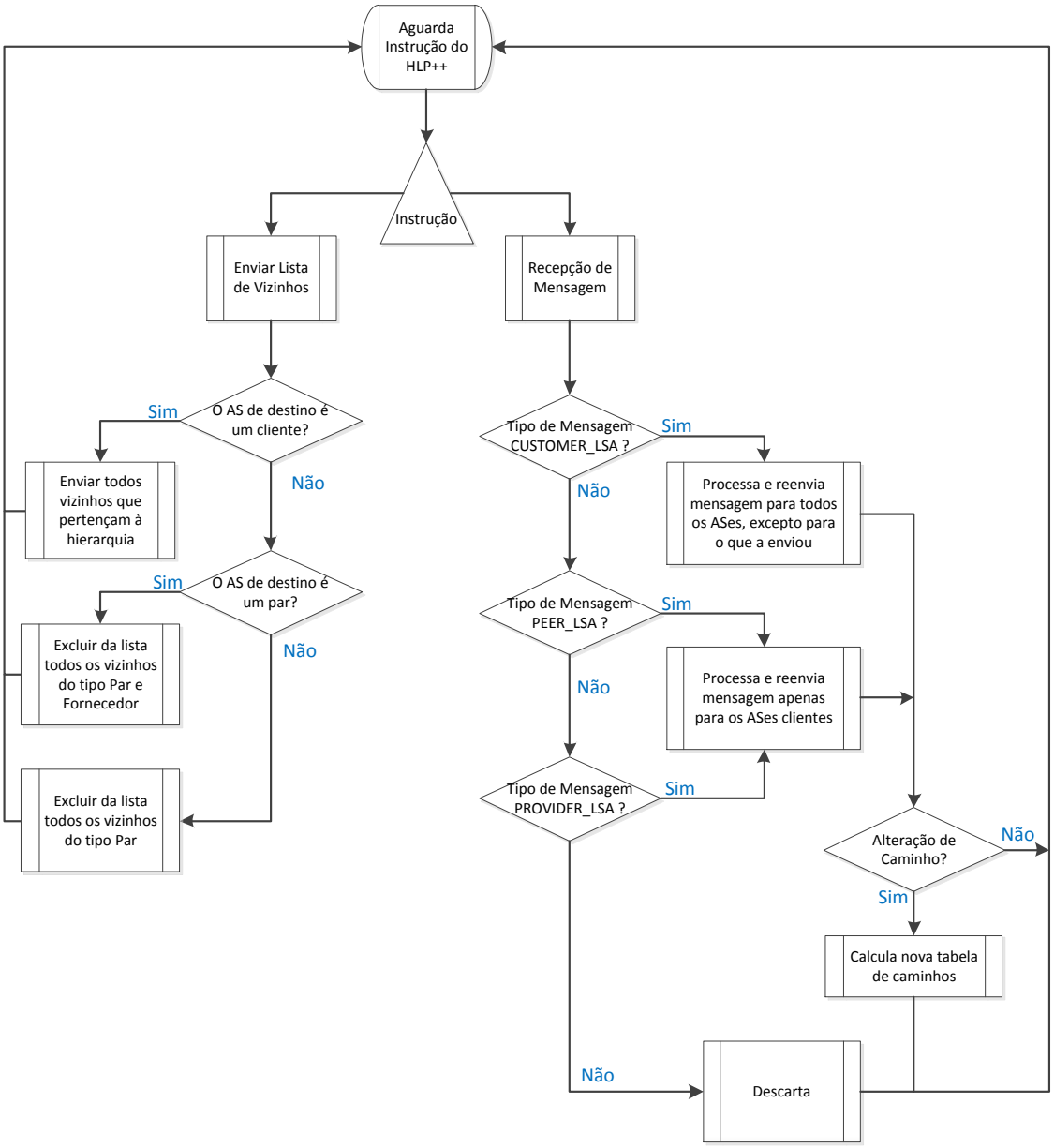

Figura 3. 7– Fluxograma do algoritmo LSA utilizado

O método de recepção de mensagem também foi modificado. No protocolo padrão LS quando um nó recebe uma mensagem, esta é reenviada para todos os seus vizinhos. No caso do LSA para HLP++ o procedimento é alterado e, o reenvio de mensagens é feito de acordo

com o tipo de mensagem recebida. Como se pode ver pela figura 3.7, se a mensagem for do tipo PEER_LSA quer dizer que o AS recebeu uma tabela de vizinhos proveniente de um vizinho do tipo par; neste caso a mensagem é reenviada apenas para vizinhos do tipo cliente.

3.4.3.4. Alteração ao FPV

Na secção 3.4.2.2 referiram-se as incompatibilidades ao nível do protocolo FPV com o modelo de negócios da Internet, nesta secção explicam-se as modificações efectuadas ao protocolo FPV no HLP++.

No HLP padrão o nó que possui uma interligação com outro vizinho de outra hierarquia envia a sua tabela de caminhos. Devido ao modelo de negócios da Internet o AS não pode simplesmente enviar todas as rotas. É necessário filtrar as rotas para os seus fornecedores e pares. A figura 3.8 ilustra o algoritmo FPV utilizado no HLP++.

No procedimento de envio da tabela de caminhos é feita a verificação se o AS tem algum fornecedor (i.e. se ele não é um nó raiz da hierarquia). Se ele não for raiz, as rotas para os seus fornecedores e pares são retiradas da tabela antes de esta ser enviada para o vizinho da outra hierarquia; caso contrário toda a tabela é enviada.

O processamento de recepção de mensagens não foi alterado significativamente em relação ao HLP. Neste caso, as regras definidas no protocolo HLP e explicadas na secção 3.4.3.1 são suficientes para assegurar a compatibilidade com o modelo de negócios da Internet. Após a recepção de uma mensagem do tipo PEER_FPV, é invocada a operação "executa envio de PROVIDER_FPV". Nesta operação, o algoritmo copia a mensagem recebida, adiciona o seu identificador de AS ao caminho e envia exclusivamente para os seus clientes. Desta forma, é responsável por distribuir as rotas recebidas através do protocolo FPV.

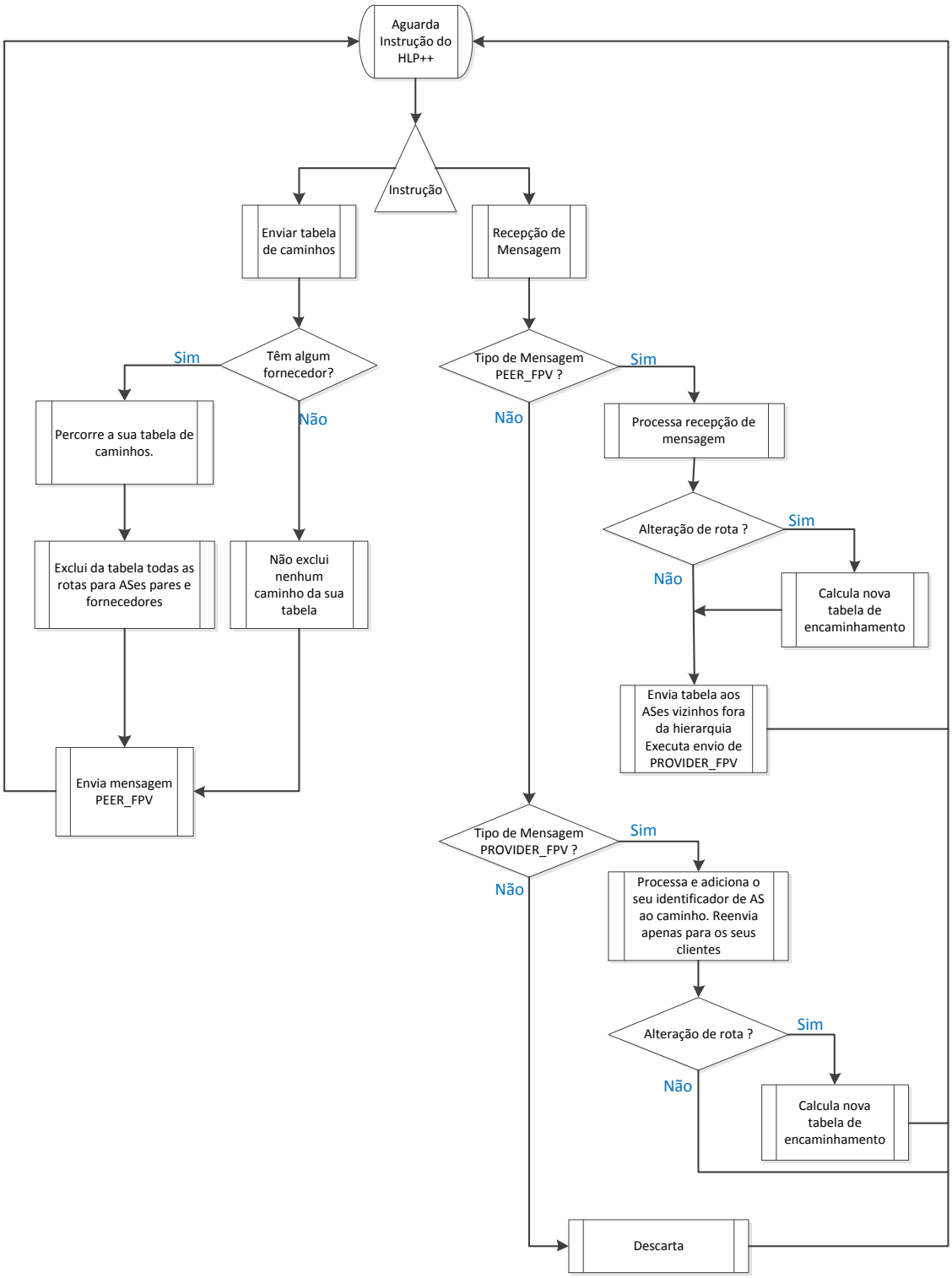

Figura 3. 8– Fluxograma do algoritmo FPV utilizado

3.4.3.5. Algoritmo de controlo do HLP++

Familiarizados com as alterações aos protocolos LSA e FPV realizadas no HLP++, vai-se agora descrever o algoritmo de controlo do protocolo HLP++, que realiza o controlo total do encaminhamento do AS.

A figura 3.9 descreve em detalhe o algoritmo utilizado. O primeiro passo consiste na leitura das configurações do AS. Como em qualquer protocolo existem, variáveis e dados que têm de ser fornecidos a partir das configurações no AS. No caso do HLP++ assume-se como parte de configuração os seguintes dados da tabela 3.2.

Propriedade	Descrição
AS_ID	Identificador do AS
Vizinho	Identificador do AS vizinho
Hierarquia	Identificador de Hierarquia
Relação com o vizinho	Fornecedor, cliente ou Par
Custo X	Custo limite para envio de actualização de rota

Tabela 3.2 – Propriedades de configuração do AS

Após a leitura das configurações do protocolo, dá-se início ao processo de encaminhamento do protocolo HLP++, com o arranque dos protocolos LSA e FPV. A tabela de caminhos inicial que o FPV envia aos seus vizinhos pares das hierarquias contíguas inclui apenas as informações iniciais dos seus vizinhos.

De seguida são realizadas as inicializações dos processos LSA e FPV, antes de entrar no estado de operação. No estado de operação, o processo HLP++ encontra-se apto a receber eventos, que na sua maioria são de *link down/link up* e recepção de mensagens. O processo também gere os eventos associados a temporizações dos protocolos LSA e FPV, nomeadamente tempos de reenvio de mensagens, tempos de espera e tempos de *keep alives* (geração de tráfego mínimo para manutenção de ligações).

Quando um evento ocorre, o seu tipo é analisado pelo processo de controlo. No caso ser uma recepção de mensagem, o processo chama o método responsável pelo processamento da mesma, que poderá pertencer a método do processo LSA ou FPV. Em ambos os casos é dada a instrução de recepção de mensagem ao respectivo protocolo, no entanto se for uma mensagem do tipo LSA são necessárias verificações adicionais.

Após recepção de mensagem por parte do protocolo LSA, o processo de controlo necessita de verificar se houve alteração à tabela de encaminhamento do AS e se este tem

vizinhos de outra hierarquia. Esta verificação é fundamental na medida em que se o AS estiver conectado a outra hierarquia, é necessário averiguar se também é necessário notificar a alteração à hierarquia vizinha através do envio de mensagens FPV.

Caso se verifique a alteração de rota, o AS analisa a sua tabela de encaminhamento conferindo se possui uma rota alternativa para a mesma actualização. Caso possua calcula o custo da rota alternativa, e se for superior a um custo X (que faz parte das configurações iniciais do protocolo), é enviada uma actualização FPV para o AS vizinho; caso contrário a alteração é omitida.

A análise do custo para a nova rota é feita com recurso a duas tabelas de encaminhamento mantidas pelo AS: tabela local (LSA) e a tabela de caminhos enviados por FPV. O algoritmo compara o custo para um dado destino entre as duas tabelas e caso o custo seja superior ao valor configurado, ou não se encontre um determinado caminho que está na tabela de FPV na tabela local LSA, é gerada uma mensagem de actualização para o vizinho fora da hierarquia.

No que respeita aos eventos de *link up e link down* , são invocados métodos nos processos FPV e LSA para o envio de caminhos, dependendo do vizinho estar ou não dentro da hierarquia.

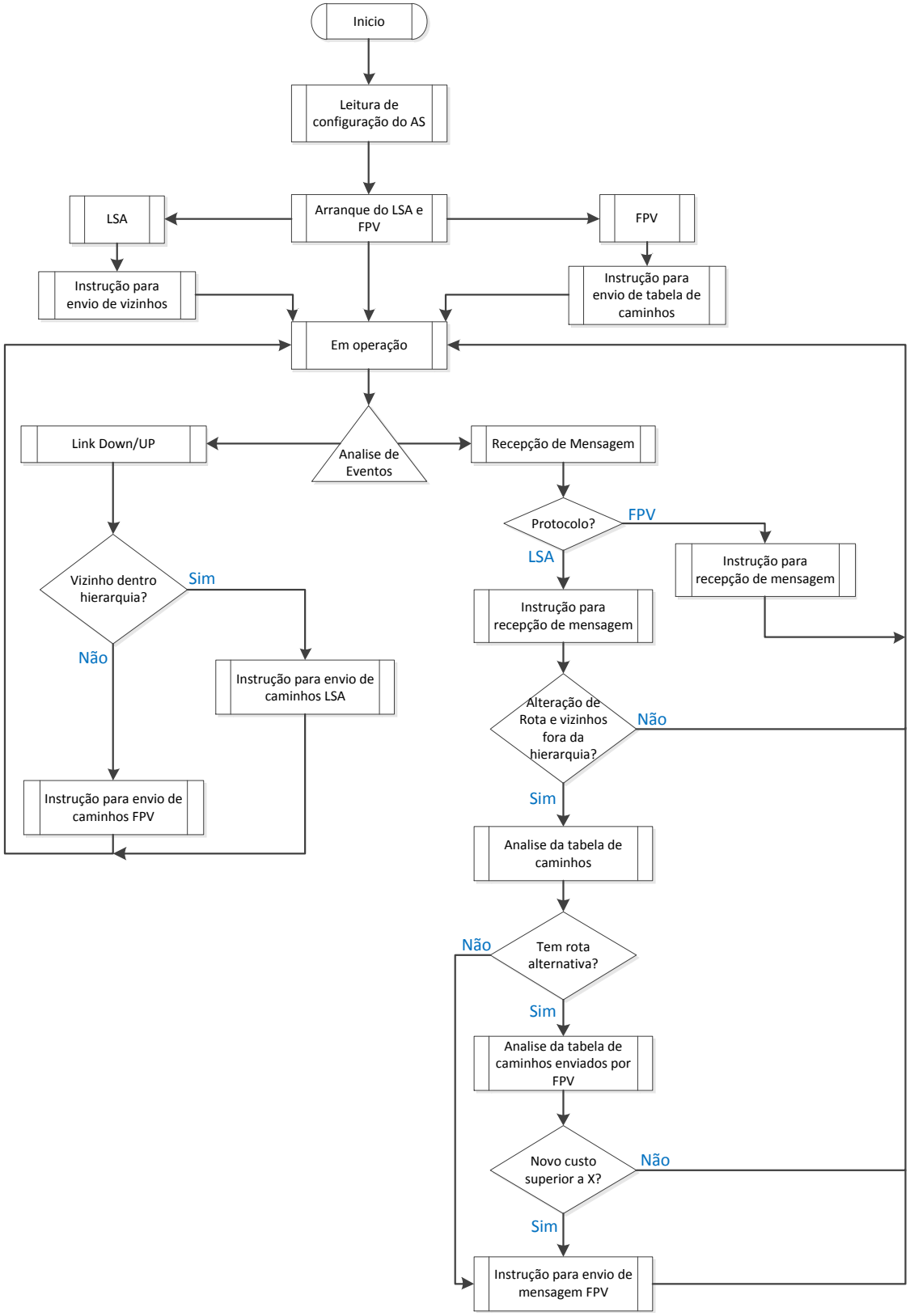

Figura 3. 9– Fluxograma do algoritmo HLP++ utilizado

3.4.4. Resumo

Neste capítulo introduziu-se a necessidade de um novo protocolo na área do encaminhamento inter-domínio. Apresentaram-se os fundamentos básicos do protocolo HLP, assim como as suas limitações face ao modelo actual de negócios da Internet. De seguida foram apresentadas as modificações propostas ao protocolo HLP para o tornar compatível com o modelo de negócios da Internet; a esta extensão deu-se o nome de HLP++. Foram apresentados e descritos os algoritmos utilizados e o modo de funcionamento.

No próximo capítulo procede-se à análise de desempenho do protocolo HLP++.

Capítulo 4.

ANÁLISE DO DESEMPENHO

4.1. Introdução

Este capítulo apresenta a análise do desempenho do protocolo implementado e descrito no capítulo anterior. Para uma melhor equidade na análise, comparam-se os resultados com o protocolo BGP (*Border Gateway Protocol*). Usou-se o BGP++ [bg09], uma implementação do protocolo BGP para o simulador ns-2 [ns09], construída a partir da implementação GNU/Zebra [gnu09].

Uma vez que o protocolo desta dissertação foi implementado recorrendo ao simulador ns-2, a secção 4.2 apresenta uma breve descrição sobre o simulador, bem como as alterações introduzidas de forma a suportar o HLP++.

A secção 4.3 apresenta as características da topologia usada para testar o HLP++. Esta topologia foi definida com base em dados recolhidos da base de dados da CAIDA (*Cooperative Association for Internet Data Analysis*). Por fim a secção 4.4 apresenta os resultados das experiências efectuadas com o HLP++.

4.2. Realização do HLP++ no simulador de redes ns-2

Esta secção apresenta o simulador ns-2 bem como as alterações inseridas para suportar o HLP++.

4.2.1. Introdução ao ns-2

O ns-2 é uma ferramenta poderosa para testes de protocolos de redes com ou sem fios. Este simulador oferece uma base adequada para criar ou modificar mecanismos em cada

camada do modelo OSI [osi09]. O Código fonte é aberto, o que dá flexibilidade suficiente aos utilizadores para modificar ou corrigir qualquer protocolo existente. No entanto tais modificações implicam um grande conhecimento do software e dos seus mecanismos. Nesta secção assume-se que o leitor tem um conhecimento básico[2] sobre os mecanismos do ns-2.

O ns-2 é realizado recorrendo a duas linguagens de programação: uma de script Tcl [tcl09] e outra de programação C++. O Tcl é usado não só para criar os scripts de simulação, mas também como interface para executar os comandos Tcl em objectos C++ associados. Durante uma simulação ns-2 há comandos Tcl que chamam rotinas programadas em código C++, que por sua vez chamam novamente código Tcl. A linguagem C++ é usada essencialmente para definir os mecanismos do protocolo. Por exemplo, um script Tcl pode conter uma instrução que acciona um evento para desactivar fisicamente uma ligação. Por sua vez este evento pode chamar uma rotina em C++ para avisar o protocolo de encaminhamento que uma ligação está em baixo; como consequência o protocolo de encaminhamento pode recalcular as novas rotas para todos os destinos.

Para testar um cenário de rede, o ns-2 recebe um script de simulação em Tcl e executa uma simulação em estado discreto. Para cada protocolo, independentemente da sua camada, é criado um objecto Tcl associado a cada nó. Esse objecto designa-se normalmente como agente e serve como ponto de recepção de pacotes. A interacção entre camadas é suportada por um módulo especial chamado de classificador (*classifier*). Um classificador também é utilizado para outros fins, tais como enviar pacotes para outros nós.

O simulador ns-2 dispõem de origem de vários protocolos de encaminhamento. No entanto, nenhum destes se adequa para um cenário de encaminhamento a nível do inter-domínio.

4.2.2. Realização do HLP++

Nesta dissertação definiu-se um novo protocolo, HLP++, reutilizando parcialmente o protocolo de estado de linha (*LS - Link State Protocol*) existente, efectuando as alterações descritas no capítulo 3.

Os comandos de Tcl são utilizados em cada agente de encaminhamento para se definir o tipo de ligação, o identificador do vizinho e a hierarquia a que o nó pertence.

[2] O manual do software pode ser consultado em [nsM09], um tutorial está também disponível na página de Marc Greis [nsT09].

Nº	Comando Tcl
1	set rtobj [$node rtObject ?]
2	set rtproto [$rtobj rtProto? HLP++]
3	$rtproto cmd setHierarchy ID
4	$rtproto cmd setProvider ID
5	$rtproto cmd setPeer ID

Tabela 4.1 – Excerto de código Tcl para configuração de um AS

A tabela 4.1 apresenta um excerto de script Tcl para configurar um AS no ns-2. O comando número 1 é utilizado para obter o objecto de encaminhamento do nó; um objecto de encaminhamento contém todos os agentes de protocolos de encaminhamento em utilização pelo nó. O segundo comando serve para obter o agente de encaminhamento do protocolo HLP++. Os restantes comandos configuram o agente de encaminhamento informando o agente espelhado em C++ sobre propriedades do AS tais como: a hierarquia que este pertence (terceiro comando), quais os seus fornecedores e pares.

A tabela anterior refere-se a comandos de script Tcl associados a um nó. Para efeitos de simulação e análise de resultados são necessários os três comandos adicionais, representados na tabela 4.2, relacionados com o controlo da topologia da rede.

Nº	Comando Tcl
1	$rtproto cmd sendUpdates
2	$rtproto cmd startRouting
3	$rtproto cmd showIpRoute

Tabela 4.2 – Comandos Tcl para simulação e análise

Todos os comandos são executados a nível do agente de encaminhamentos reflectindo as ordens no agente homólogo em C++; o primeiro (sendUpdates) ordena ao agente para começar o processo de inundação do LS (Link State) entre os nós que fazem parte da simulação, o segundo (startRouting) instrui o agente para começar o processo de encaminhamento do protocolo implementado HLP++. O processo de encaminhamento não começa sem que seja realizada uma inundação inicial. No início os nós não sabem a que hierarquia pertencem os seus vizinhos, sendo necessária esta inundação para que todos os nós tenham a visão da rede (numa situação real isso não acontece dado que essas propriedades são trocadas entre os AS durante o processo de negociação entre eles). Por último, o comando

showIpRoute serve para que o agente de encaminhamento mostre a tabela de encaminhamento do nó em questão.

Cada função existente no código C++ do LS foi subsequentemente alterada de forma a reflectir as modificações explicadas no capítulo anterior secção 3.4.3.3.

4.3. Características da topologia

Nesta secção apresenta-se a topologia utilizada durante as simulações do protocolo. Esta topologia foi obtida a partir das relações entre ASes calculadas no projecto CAIDA [cai09]. Foi desenvolvido software em Java para analisar os dados obtidos do projecto CAIDA cruzados com dados obtidos do RIPE (*Réseaux IP Européens*) num ficheiro de texto do tipo; "AS_ID *policy* AS_ID" em que AS_ID é o identificador de AS e *policy* reflecte o tipo de relação entre os dois ASes. Por exemplo, a linha AS6939 p2p AS3549 define uma relação par-a-par (p2p – *peer-to-peer*) entre o AS6939 e o A3549. Estas relações também podem ser do tipo p2c (fornecedor-cliente) ou c2p (cliente-fornecedor) como foi visto no capítulo 2.

Para escolher a topologia, utilizou-se o comando *trace route* a partir de uma máquina na rede da Portugal Telecom para descobrir o seu AS fornecedor; posteriormente analisou-se o ficheiro da CAIDA através do software Java desenvolvido para se obter os fornecedores do AS anteriormente seleccionado até se atingir um AS raiz, isto é, um AS que não tem qualquer outro fornecedor.

De seguida o mesmo software foi utilizado aplicando outras regras para descobrir os pares do AS raiz que seriam potencialmente outros ASes raiz, dado que a natureza da Internet é existir uma quase *full mesh* entre ASes de nível 1. A condição de paragem foi atingir um número significativo de ASes raiz. A topologia escolhida contém 54 ASes e inclui 10 ASes raiz.

Poder-se-ia ter escolhido um número mais elevado de ASes. No entanto, a realização do protocolo BGP em ns-2 consome demasiados recursos, nomeadamente memória, para a realização de simulações com número superior de nós em tempo útil. A figura 4.1 ilustra a topologia já inserida no simulador ns-2, visualizada através do visualizador gráfico nam integrado no simulador. Para esta topologia foram ainda definidas 2 hierarquias (H0 e H1) representadas na figura 4.1, em que os nós a amarelo correspondem à hierarquia H0 enquanto os restantes pertencem à hierarquia H1. A divisão hierárquica foi efectuada com base nos 10 ASes de nível 1. Seleccionaram-se 6 para a hierarquia H0 e os restantes 4 para a hierarquia

H1. Depois utilizaram-se as relações p2c destes para chegar até aos ASes de nível mais baixo. A correspondência entre o número de nó em ns-2 e o número de AS real é feito na tabela 4.3.

Embora se trate apenas de uma reduzida amostra da Internet real, a figura 4.1 mostra uma rede massivamente ligada. No sentido descendente a figura mostra os ASes ligados por ligações do tipo p2c representadas com cor cinzenta. As ligações do tipo p2p estão representadas com cor azul. No topo da figura encontram-se os 10 ASes de nível 1. É notória uma grande quantidade de ligações do tipo p2p entre ASes de nível 1 com ASes de níveis inferiores, mas também é evidente uma substancial predominância de ligações p2p face às p2c.

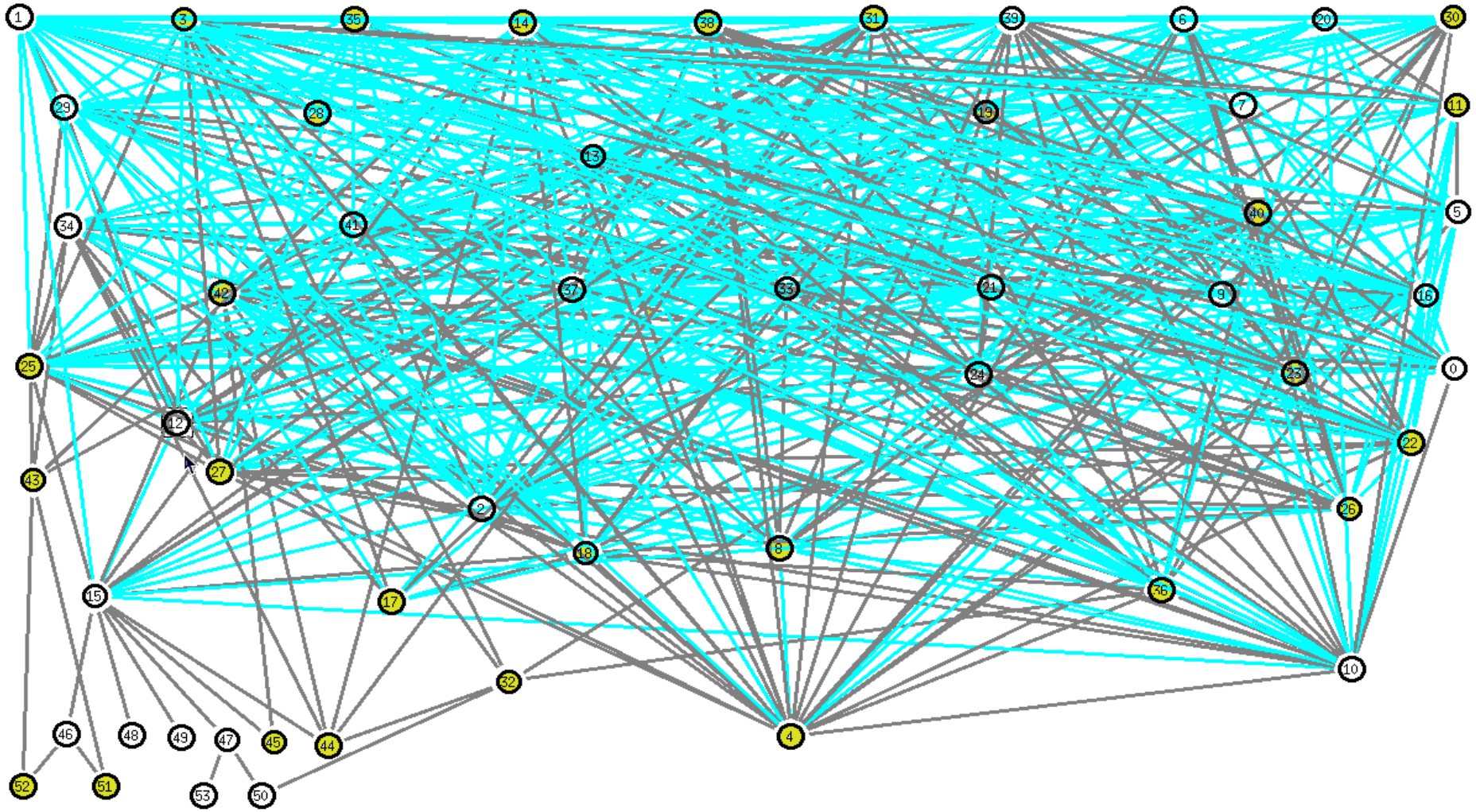

Figura 4. 1– Topologia utilizada vista em ns/nam

Nó em ns-2	Número de AS	Nome da Instituição	Hierarquia Atribuida
0	8708	RDSNET	H1
1	6939	Hurricane Electric	H1
2	2497	Internet Initiative Japan Inc.	H1
3	3549	Global Crossing	H0
4	12956	Telefonica	H0
5	6830	UPC Broadband	H1
6	4323	TW Telecom Holdings	H1
7	9002	RETN Limited	H1
8	5400	BT Global Services	H0
9	4766	Korea Telecom	H1
10	6762	Telecom Italia Sparkle	H1
11	22773	Cox Communications	H0
12	5413	GX Networks	H1
13	1299	TeliaSonera AB Networks	H1
14	174	Cogent Communications	H0
15	8657	Portugal Telecom	H1
16	3303	SWISSCOM	H1
17	3216	Golden Telecom	H0
18	1273	Cable and Wireless IP GSOC Europe	H0
19	19151	WV FIBER LLC	H0
20	2828	XO Communications	H1
21	13237	LambdaNet Communications	H1
22	2516	KDDI Corp.	H0
23	3786	LG DACOM Corporation	H0
24	8928	Interoute Communications	H0
25	286	KPN Internet Solutions	H0
26	6539	Bell Canada	H0
27	3491	Beyond The Network America	H0
28	20932	IP-MAN.Net Engineering	H1
29	6461	MFN - Metromedia Fiber Network	H1
30	7018	AT&T WorldNet Services	H0
31	701	MCI Communications Services	H0
32	2860	Novis	H0
33	3561	Savvis	H1
34	702	MCI Communications Services	H1
35	209	Qwest Communications Company	H0
36	5511	France Telecom - Orange	H0
37	3257	Tinet SpA	H1
38	1239	Sprint	H0
39	3356	Level 3 Communications	H1
40	3320	Deutsche Telekom AG	H0
41	2914	NTT America, Inc.	H1
42	6453	TELEGLOBE IP ENGINEERING	H0
43	9186	ONI TELECOM	H0
44	13156	CABOVISAO	H0
45	12542	TVCABO	H0
46	3243	TELEPAC	H1
47	15525	PT PRIME	H1
48	15457	Cabo Tv Madeirense	H1
49	42863	TMN	H1
50	35038	INESC	H1
51	34873	IGIF- Ministério da Saúde	H0
52	25253	Caixa Geral de Depósitos	H0
53	43643	Tap Air Portugal	H1

Tabela 4.3 – Identificação dos nós da figura 4.1

Antes de se continuar a caracterização da figura 4.1, vai-se introduzir a definição de grau de um nó. O grau de um nó é caracterizado pelo número de ligações que o nó tem para os seus vizinhos [FFF99].

O gráfico da figura 4.2 ilustra em percentagem cumulativa o grau dos ASes da figura 4.1., com um grau médio de 19,1 ligações por AS.

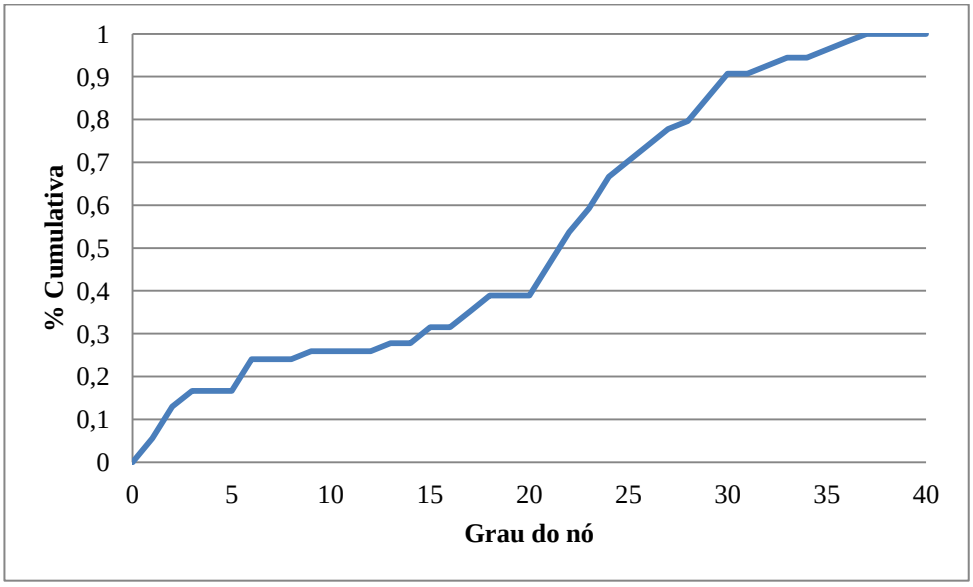

Figura 4. 2– Distribuição do grau do nó em percentagem cumulativa

A topologia da figura 4.1 também foi usada nas simulações com duas hierarquias. Os gráficos das figuras 4.3 e 4.4, comparam em percentagem cumulativa o grau dos ASes das hierarquias H0 e H1 em termos de ligações dentro e para fora da hierarquia. Mediu-se um grau médio de ligações dentro da hierarquia de 9 e 9,8 respectivamente para H0 e H1. Para as ligações para ASes da outra hierarquia mediu-se um grau médio de 10,1 e 9,4 respectivamente para H0 e H1.

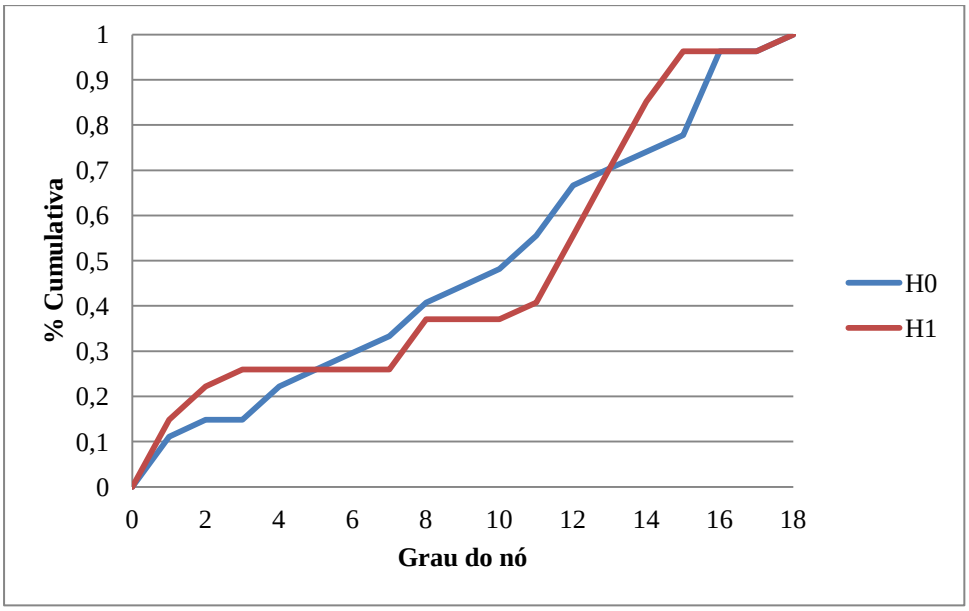

Figura 4. 3– Comparação grau do nó para ligações dentro das hierarquias H0 e H1

Comparando o grau dos nós para o caso das ligações dentro da hierarquia ilustrado na figura 4.3, é possível observar que as duas hierarquias não estão igualmente distribuídas. Esta desigualdade deve-se a casos como por exemplo, a pequena quantidade de ASes *stub* (terminais) presentes no canto inferior esquerdo da figura 4.1. Este grupo de ASes incluídos na hierarquia H1 são responsáveis por uma pequena quantidade de ligações entre hierarquias dos ASes sob domínio da rede Portuguesa.

No que respeita a hierarquia H0 temos o caso completamente oposto: O AS número 4 que representa o AS associado à rede da *Telefonica* (AS12956), possui 22 ligações do tipo c2p (cliente-fornecedor) e apenas 8 ligações do tipo p2p (par-a-par). Estas disparidades ajudam a perceber as diferenças entre as duas hierarquias representadas na figura4.3. Pode-se também concluir com estes dois exemplos que o grau de ligações multi-caminho (*multihomed*) de um AS na rede representada não está correlacionado com o nível do AS [AGA+09]. Este facto está associado à diferente segmentação em ASes existentes nas redes da Portugal Telecom e na rede da Telefónica

Analisando o gráfico da figura 4.4 é perceptível que alguns dos ASes da hierarquia H1 não possuem ligações para fora da sua hierarquia. Estes domínios são ASes do tipo *stub*. Desta forma, tal como está representado na figura 4.1, eles foram associados à mesma hierarquia dos seus fornecedores; caso contrário estes ASes não teriam qualquer caminho para destinos na sua hierarquia.

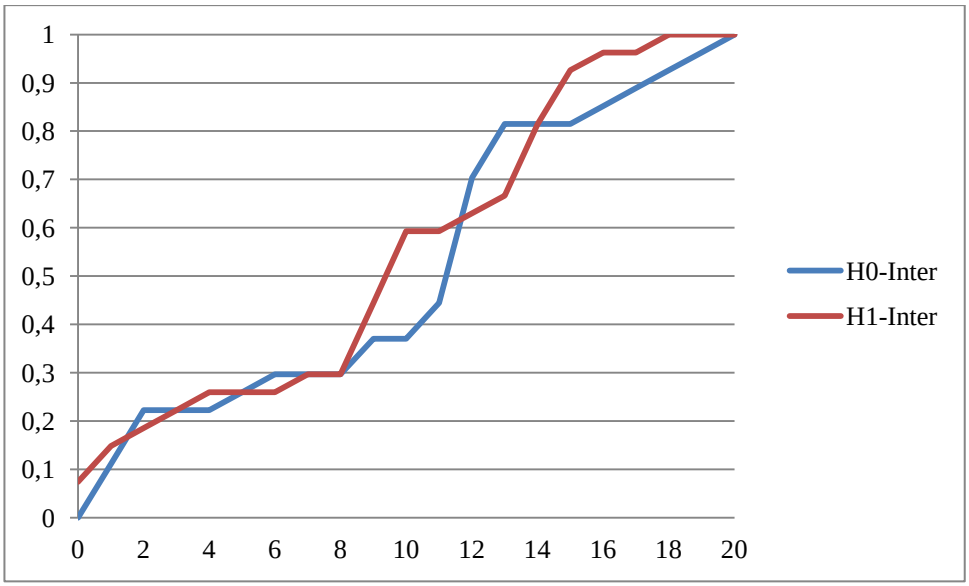

Figura 4. 4– Comparação grau do nó para ligações entre as hierarquias H0 e H1

A tabela4.4 apresenta o número de ligações entre hierarquias distinguindo o tipo de ligação.

Tipo de Ligação	Hierarquia H0	Hierarquia H1	Total de Ligações
c2p	56	38	94
p2p	86	73	159

Tabela 4.4 – Número de ligações entre hierarquias

Como se pode observar pela tabela 4.4, o número total de ligações entre hierarquias é de 253. Quando se considera a mesma rede como uma única hierarquia, o número total de ligações é 517. Portanto o número de ligações entre hierarquias é quase metade do número total de ligações da topologia utilizada. Repare-se que esta topologia está muito longe da topologia hierárquica considerada no desenho original do protocolo HLP, onde o número de ligações entre hierarquias é muito reduzido. Desta forma, é previsível uma enorme interdependência entre hierarquias, que leva à propagação de eventos FPV de actualização topológica entre as duas hierarquias.

É necessário distinguir o tipo de ligações entre hierarquias dado que são estas que garantem acesso global à rede e não apenas a um conjunto restrito de destinos dentro da hierarquia. Pela tabela 4.4 é perceptível que a hierarquia H0 possui mais ligações do que H1. Como resultado H0 tem menos probabilidade de perder conectividade para H1do que H1 perder conectividade para H0.

4.4. Experiências com o HLP++

Esta secção apresenta o resultado das experiências efectuadas em ns-2 com o HLP++. As experiências centraram-se na análise das mensagens de encaminhamento transportadas em pacotes de sinalização. Os resultados são comparados com o desempenho do protocolo BGP, medido usando o BGP++ [bg09]. A figura 4.5 apresenta os resultados da experiência do HLP++ com a topologia ilustrada na figura 4.1 quando é usada apenas uma única hierarquia (HLP++) e quando são usadas duas hierarquias (HLP++ 2H). Para o BGP e para o HLP++ com uma hierarquia foram seleccionadas aleatoriamente 70 falhas de ligação entre ASes. Para cada falha isolada foi registado o número de ASes afectados, isto é, o número de ASes que receberam uma mensagem de actualização devido à falha de ligação. Para o HLP++ com duas hierarquias foram escolhidas aleatoriamente 177 falhas de ligação, das quais 93 dizem respeito a falhas de ligação entre hierarquias e as restantes 84 repartidas igualmente dentro de cada uma das hierarquias. No que respeita as configurações do BGP, as relações comerciais foram implementadas recorrendo ao atributo de comunidade, como exemplificado no capítulo 2 secção 2.3.1.3. Na tabela 4.5 são apresentados o número médio de ASes afectados para o conjunto de simulações realizadas para cada configuração do protocolo testada.

BGP	HLP++	HLP++ 2H
26,3	41,6	45,3

Tabela 4.5 – Número médio de ASes afectados

A figura 4.5 apresenta a percentagem cumulativa do número de ASes afectados por falhas de ligação. Verifica-se que, o desempenho do HLP++ é inferior ao do BGP para as duas configurações, como seria expectável. No BGP há omissão de actualização quando a falha não afecta nenhuma rota seleccionada. O HLP++ usa como protocolo de encaminhamento dentro da hierarquia o LS (*Link State*). Assim todos os ASes são notificados após falha de ligação, excepto quando há nós acessíveis apenas através de rotas que violam as restrições comerciais. Com duas hierarquias poderia haver algum isolamento de falhas dentro da hierarquia, mas a enorme densidade de interligações entre as duas hierarquias na rede real

considerada leva a que o efeito obtido seja precisamente o contrário – o número médio de nós afectados aumenta com a introdução das duas hierarquias.

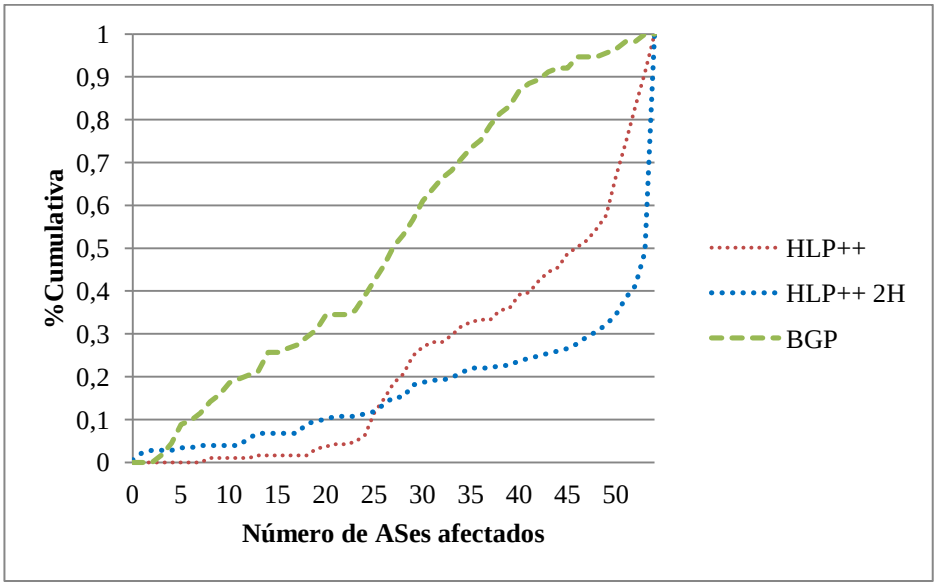

Figura 4. 5– Percentagem cumulativa dos ASes afectados

Estes resultados justificam-se pelo facto de o HLP++ ter sido desenhado para operar numa rede puramente hierárquica, o que não é o caso da Internet como mostra a figura 4.1. Embora o HLP++ disponha de um mecanismo de supressão de actualização de caminhos, este não é suficiente para uma topologia com as características da Internet. O elevado número de ligações *multihome,* como se pode ver por exemplo para o nó número 4 da figura 4.1, fazem com que as mensagens de LS se propaguem às outras hierarquias.

No entanto, o HLP++ é um protocolo que funciona bem em topologias com uma hierarquia pura, sem ligações do tipo p2p entre níveis de hierarquia e sem ASes *multihomed.*

Na figura 4.6 é apresentada uma experiência efectuada apenas com os ASes *stub* representados na figura 4.1. Nesta experiência seleccionaram-se aleatoriamente 14 falhas de ligação entre os ASes de nível mais baixo na hierarquia, tendo-se obtido um número médio de 25,6 ASes afectados pelas falhas. Este número é comparável com o número de ASes médio em cada hierarquia, igual a 26,5. Desta forma, é possível confirmar que neste cenário aumenta o número de actualizações de estado devido a falhas contidas na sua hierarquia que não se propagam às hierarquias vizinhas.

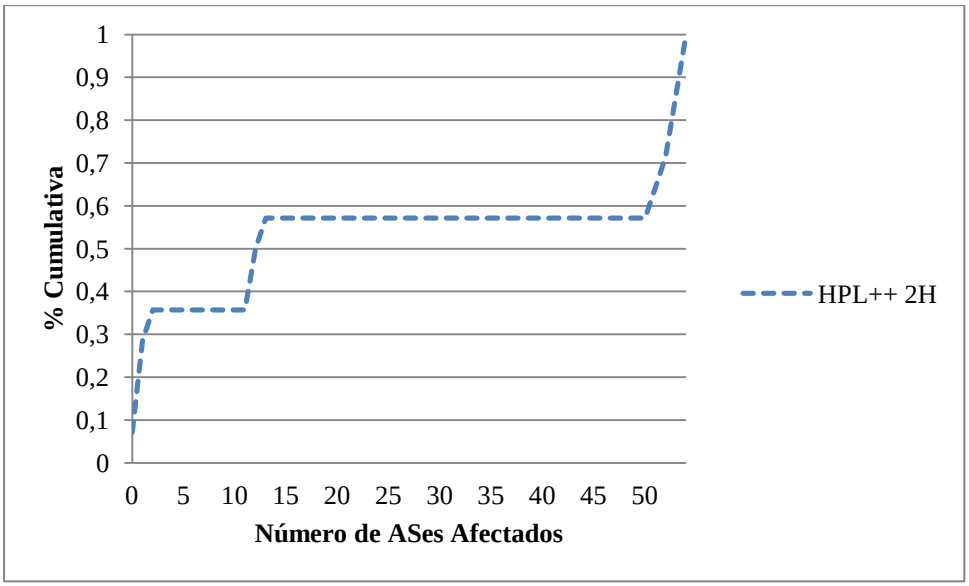

Figura 4. 6– Percentagem cumulativa dos ASes afectados com falhas em níveis inferiores na hierarquia

Capítulo 5.

CONCLUSÕES

Neste último capítulo é realizada uma síntese geral sobre a hipótese formulada anteriormente e respectivas ideias face ao trabalho desenvolvido ao longo da dissertação. De seguida são resumidos os principais contributos da dissertação, bem como direcções para trabalho futuro.

No início da dissertação, na secção 1.2 foi apresentada uma hipótese de inovação sobre protocolos de encaminhamento inter-domíno. Assuntos como a natureza topológica da Internet e sua evolução, bem como as relações entre ASes (*Autonomous Systems*) são apenas breves noções sobre o trabalho investigado e desenvolvido nesta dissertação.

A dissertação foca o caso particular da adaptação de um dos novos protocolos propostos pelo mundo académico, de forma a que honre as relações comerciais existentes no actual modelo de negócios da Internet. O desenvolvimento da modificação ao protocolo foi apresentando ao longo do capítulo 3, conjugando um conjunto de ideias aproveitadas de protocolos existentes, apresentados no capítulo 2, bem como o estudo da natureza topológica da Internet, conjugados com alguns factores inovadores. O desempenho exibido no capítulo 4 foi obtido com o protocolo proposto utiliza a topologia definida também nesta dissertação de 54 ASes (um subconjunto da rede Internet). O protocolo foi integralmente implementado no simulador ns-2.

5.1. Conclusões

Nesta dissertação foram abordados temas sobre uma área considerada desafiante nos dias actuais, o encaminhamento inter-domínio. Em particular foram discutidos assuntos sobre arquitecturas, técnicas e mecanismos sobre protocolos, tratando-se de uma área de investigação com alguma sensibilidade comercial. Existem dois factores que tornaram o desenvolvimento desta dissertação tão aliciante. O maior deles foi ser investigação numa área do mundo das telecomunicações, presente nos dias de hoje na vida quotidiana das pessoas: a Internet. Contudo, também foi motivante a implementação de um novo protocolo com uma abordagem e arquitectura totalmente diferentes do protocolo actualmente em operação na Internet (BGP), e sua extensão de forma a respeitar as relações comerciais entre os vários operadores.

Ao se observarem os protocolos propostos pela comunidade de investigação (NIRA e HLP) pode-se concluir que ambos apresentam virtudes num ambiente específico. No entanto, quando se enfrenta um cenário próximo da realidade, os resultados finais são diferentes. Foi deste prisma que partiu uma das ideias desta dissertação - o estudo da natureza topológica da Internet. No capítulo 2 foram discutidas diversas perspectivas sobre a estrutura da Internet. O estudo realizado nesta dissertação com recurso aos dados de um projecto que estudou as relações entre ASes da CAIDA [cai09], revela que a Internet não segue uma estrutura hierárquica pura. A liberdade e flexibilidade do protocolo BGP permitem aos administradores dos ASes negociarem entre si diversos tipos de relações que servem melhor os seus propósitos, independentemente do nível hierárquico em que se encontram. Levam assim a que a topologia da Internet se assemelhe mais a uma rede sem escala.

As experiências realizadas no capítulo 4 com o protocolo proposto nesta dissertação e sua comparação com o BGP mostram que, em ambos os casos o BGP revela um melhor desempenho. Estes resultados explicam-se pelo facto de o HLP ter sido desenhado para uma rede hierárquica pura. Na realidade a topologia da rede Internet observada contém múltiplas ligações a múltiplos fornecedores (ligações *multihomed*), sendo substancialmente diferente de uma rede hierárquica. No entanto, para situações onde se verifica e aproxima a uma hierarquia o HLP apresenta um bom desempenho, como é verificado no capítulo 4 para o caso da experiência realizada com uma parte da topologia apresentada dos ASes associados à sub-rede da Portugal Telecom.

É de realçar que parte dos resultados apresentados nesta dissertação, foram publicados na conferência internacional *IEEE Globecom'09* [AGA+09].

5.2. Trabalho Futuro

Muitos protocolos propostos pela comunidade de investigação têm por base a hipótese que a Internet é uma rede hierárquica pura. Como se viu, esta hipótese não é válida sendo necessário identificar novas abordagens alternativas mais adaptadas à Internet.

O estudo sobre a natureza topológica da Internet revela uma grande presença de ligações *multihomed*. Nenhuma das soluções propostas tira partido dessas ligações, isto é, embora se possua uma rota para o mesmo destino por pontos diferentes, nenhum protocolo consegue usar essas ligações em simultâneo – apenas uma das ligações é seleccionada.

Algumas arquitecturas também propõem encaminhamento ao nível do identificador do AS e não dos seus prefixos. Embora esta técnica reduza substancialmente o tamanho das tabelas de encaminhamento, é necessário um mecanismo eficiente para efectuar a associação entre o identificador do AS e seus prefixos, que deve ser alvo de trabalho futuro.

Evoluções nas três áreas identificadas acima poderiam contribuir para virar uma nova página na área do encaminhamento inter-domínio.

BIBLIOGRAFIA

[AGA+09] P. Amaral, F. Ganhão, C. Assunção, L. Bernardo, e P. Pinto. Scalable multi-region routing at inter-domain level. IEEE Globecomm, Novembro 2009.

[bg09] Bgp++ ver http://www.ece.gatech.edu/research/labs/MANIACS/BGP++/, Setembro 2009.

[BML09] B. Leiner, V. Cerf, D. Clark, R. Kahn, L. Kleinrock, D. Lynch, J. Postel, L. Roberts, and S. Wolff, A brief history of the internet. SIGCOMM Comput. Commun. Rev. 39, 5 (Oct. 2009), 22-31.

[Bra89] R. Braden. Requirements for internet hosts. communication layers. Technical report, IETF, RFC 1122, Outubro 1989.

[bri09] Brite: Boston university representative topology generator. ver http://www.cs.bu.edu/brite/, Agosto 2009.

[BH02] B. Huaker. Topology discovery by active probing. Em Proc. Symp. Applications and the Internet (SAINT), Janeiro 2002.

[CA99] C. Alaettinogluoglu. Routing policy specification language (rpsl). Technical Report, IETF, RFC 2622, Junho 1999.

[cai09] Caida as relationships data research project. Ver http://www.caida.org/data/active/as-relationships/, Setembro 2009.

[CAF00] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian. Delayed Internet Routing Convergence. Em Proc. ACM SIGCOMM (2000).

[CCF+05] M. Caesar, D. Caldwell, N. Feamster, J. Rexford, A. Shaikh, and J. Van der Merwe. Design and implementation of a routing control platform. Em NSDI, 2005.

[ceig09] Cisco - eigrp. ver http://www.cisco.com/en/US/tech/tk365/ technologies_white_paper09186a0080094cb7.shtml, Setembro 2009.

[CRK89] C. Cheng, R. Riley, and S. Kumar. A loop-free extended bellman-ford routing protocol without bouncing effect. ACM SIGCOMM Computer Communication Review, Volume 19, No.4, Setembro 1989.

[Dai04] L. Daigle. Whois protocol specification. Technical report, IETF, RFC 3912, Setembro 2004.

[Dat09] Merit Network Routing Assets Database. Internet routing database. Ver ftp://ftp.ra.net/, Junho 2009.

[DB08] B. Donnet and O. Bonaventure. On bgp communities. ACM SIGCOMM Computer Communication Review,Volume 38, No, 2, Abril 2008.

[DF07] B. Donnet and T. Friedman. Internet topology discovery: A survey. *IEEE Communications Surveys, Volume 9, No. 4*, 4th Quarter, 2007.

[Dij59] E. Dijkstra. A note on two problems in connexion with graphs.
Numerische Mathematik, 1, pp. 269-271, 1959.

[DKF+07] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker,Y. Hyun, K. klaffy, and G. Riley. As relationships: Inference and validation. ACM SIGCOMM Computer Communication Review, Volume 37, No. 1, Janeiro 2007.

[FFF99] M. Faloutsos, P. Faloutsos, and C. Faloutsos.
On power-law relationships of the internet topology.ACM SIGCOMM, 1999.

[Gao01] L. Gao. On inferring autonomous system relationships in the internet.
IEEE/ACM Transactions on Networking, Volume 9, No. 6, Dezembro 2001.

[gnu09] Gnu/zebra. Ver http://www.zebra.org/, September 2009.

[GP01] T. Griffin and B. Premore. An experimental analysis of bgp convergence time.
IEEE International Conference on Network Protocols, 2001.

[GT00] R. Govindan and H. Tangmunarunkit. Heuristics for internet map discovery.
Em *Proc. IEEE INFOCOM*, Março 2000.

[gti09] Gt-itm: Georgia tech internetwork topology models.
Ver http://www.cc.gatech.edu/projects/gtitm/, Setembro 2009.

[HC02] H. Chang et al. Towards capturing representative as-level internet topologies.
Em Proc. ACM SIGMETRICS, Junho 2002.

[ige09] Igen: Topology generation through network design heuristics. Ver
http://www.info.ucl.ac.be/~bqu/igen/, Setembro 2009.

[JI92] J. Internem. Dynamics of link-state and loop-free distance-vector routing algorithms.
Journal of Internetworking: Research and Experience, volume. 3, pp. 161-188, 1992.

[KKK07] N. Kushman, S. Kandula, and D. Katabi. Can you hear me now?!
it must be bgp. ACM SIGCOMM Computer Communication Review, Abril 2007.

[LCM+04] L.Subramanian, M. Caesar, M. Handley, M. Mao, S. Shenker, and I. Stoica.
HLP: A Next Generation Inter-domain Routing Protocol, Novembro 2004. UC Berkeley Technical Report No. CSD-04-1357

[LCR+07] K. Laksminarayanan, M. Caesar, M. Rangan, T.Anderson, S. Shenker, and I. Stoica.
Achieving convergence-free routing using failure-carrying packets. ACM SIGCOMM Computer Communication Review, Agosto 2007.

[LK08] L. Kleinrock. History of the Internet and its flexible future. IEEE, *Wireless Communications*, volume .15, no.1, pp.8-18, Fevereiro 2008.

[LVPS08] K. Levchenko, G.Voelker, R.Paturi, and S. Savage. Xl: An efficient network routing algorithm. ACM SIGCOMM Computer Communication Review,Agosto 2008.

[Mal98] G. Malkin. Rip version 2. Technical report, IETF - Network Working Group,
Novembro 1998.

[met] Predicting the internet's catastrophic collapse and ghost sites galore in 1996.
Ver Infoworld, Dezembro 4, 1995.

[MKF+06] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker,X. Dimitropoulos, K klaffy, and A. Vahdat. The internet as-level topology:Three data sources and one definitive metric. ACM SIGCOMM Computer Communication Review, Volume 36, Issue 1, 2006.

[Moy98] J. Moy. Ospf version 2. Technical report, IETF - Network Working Group,1998.

[MW77] J. McQuillan, and D. Walden. "the arpanet design decisions". Computer Networks, vol. 1, Agosto 1977.

[MZ02] Z. Mao. Route application damping exacerbates internet routing convergence. Em ACM SIGCOMM, 2002.

[NCC09] RIPE NCC. Routing registry consistency check reports. Ver http://www. ripe.net/projects/rrcc/, Julho 2009.

[rir09] R. i. registries. Ver http://www.isoc.org/briefings/021/, Julho 2009.

[ns09] ns-2 - network simulator 2.Ver http://www.isi.edu/nsnam/ns/, Setembro 2009.

[nsM09] ns-2 manual. Ver http://www.isi.edu/nsnam/ns/ns-documentation.html, Setembro 2009.

[NSW02] R. Mahajan, N. Spring, and D. Wetheral. Measuring isp topologies with rocketfuel. Em Proc. ACM SIGCOMM, Agosto 2002.

[nsT09] Marc Greis tutorial for the network simulator "ns". Ver http://www.isi.edu/nsnam/ns/tutorial/, Setembro 2009.

[OBOM03] Y. Ohara, M. Bhatia, N. Osamu, and J. Murai. Route flapping effects on ospf. Em Applications and the Internet Workshops, 2003.

[osi09] Osi: Open system interconnection. Ver http://standards.iso.org/ittf/PubliclyAvailableStandards/s020269_ISO_IEC_7498-1_1994(E).zip, Setembro 2009

[QC02] Q. Chen. The origin of power laws in internet topologies revisited. Em Proc. IEEE INFOCOM'02, New York, Junho 2002.

[RLH06] Y. Rekhter, T. Li, and S. Hares. Rfc 4271 - a border gateway protocol 4 (bgp-4). Technical report, IETF - Network Working Group, 2006.

[RFC1142] D. Oran. "osi is-is intra-domain routing protocol". Technical report, IETF, Rfc 1142

[SARK02] L. Subramanian, S. Agarwal, J. Rexford, and R. Katz. Characterizing the internet hierarchy from multiple vantage points.Em IEEE INFOCOM 2002 , New York, Junho 2002.

[SCE+05] L. Subramanian, M. Caesar, C. Ee, M. Handley, M. Mao, S. Shenker, and I. Stoica. Hlp: A next generation inter-domain routing protocol. ACM SIGCOMM Computer Communication. Revisão, Agosto 2005.

[SKM09] A. Sahoo, K. Kant, and P. Mohapatra. Bgp convergence delay after multiple simultaneous router failures: Characterization and solutions.Em Computer Communications, Volume 32, Maio 2009.

[SR95] S. Ruthfield. The Internet's history and development: from wartime tool to fish-cam. Crossroads 2, 1 (Sep. 1995), 2-4.

[tcl09] Tcl developer site. Ver http://www.tcl.tk/, Setembro 2009.

[UO09] University of Oregon. Route views, university of oregon route views project. Ver http://www.routeviews.org/, Julho 2009.

[VCG98] C. Villamizar, R. Chandra, and R. Govindan. Rfc 2439 - bgp route application damping. Technical report, IETF - Network Working Group, 1998.

[VPSV02] A.Vazquez, R. Satorras, and A Vespignani. Large-scale topological and dynamical properties of the internet. Physical Review, Volume 65, No 6, 066130, Junho 2002.

[YCB07] X. Yang, D. Clark, and A. Berger. Nira: A new inter-domain routing architecture. IEEE/ACM Transactions on Networking, Volume 15, No 4, Junho 2007.

[YMBB05] M. Yannuzi, X. Bruin, and O.Bonaventure. Open issues in interdomain routing: A survey. IEEE Network, Novembro/Dezembro 2005.