

UNIVERSIDADE FEDERAL DO PARÁ
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

TESE DE DOUTORADO

**ESTRATÉGIA DE PLANEJAMENTO DE REDES DE ACESSO DE
ALTA CAPILARIDADE BASEADAS EM TECNOLOGIA DSL: UMA
ABORDAGEM HÍBRIDA EMPÍRICA E PROBABILÍSTICA**

DIEGO LISBOA CARDOSO

BELÉM/PA

2010

UNIVERSIDADE FEDERAL DO PARÁ
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

DIEGO LISBOA CARDOSO

**ESTRATÉGIA DE PLANEJAMENTO DE REDES DE ACESSO DE
ALTA CAPILARIDADE BASEADAS EM TECNOLOGIA DSL: UMA
ABORDAGEM HÍBRIDA EMPÍRICA E PROBABILÍSTICA**

Tese submetida à Banca Examinadora do Programa de Pós-graduação em Engenharia Elétrica da UFPA para a obtenção do Grau de Doutor em Engenharia Elétrica, elaborada sob a orientação do Prof. Dr. Carlos Renato Lisboa Francês.

UFPA / ITEC / PPGEE

BELÉM/PA

2010

**ESTRATÉGIA DE PLANEJAMENTO DE REDES DE ACESSO DE
ALTA CAPILARIDADE BASEADAS EM TECNOLOGIA DSL: UMA
ABORDAGEM HÍBRIDA EMPÍRICA E PROBABILÍSTICA**

DIEGO LISBOA CARDOSO

Esta tese foi julgada adequada para o **Exame de Defesa de Doutorado em Engenharia Elétrica**, na área de Computação Aplicada, e aprovado na sua forma final pela banca examinadora designada pelo Programa de Pós-Graduação em Engenharia Elétrica do Instituto de Tecnologia da Universidade Federal do Pará em 26 de março de 2010.

Prof. Dr. Carlos Renato Lisboa Francês (PPGEE - UFPA) - ORIENTADOR

Prof. Dr. Solon Venâncio de Carvalho (LAC - INPE) - MEMBRO EXTERNO

Prof. Dr. Marcelo Eduardo Vieira Segatto (UFES) - MEMBRO EXTERNO

Prof. Dr. Gervásio Protásio dos Santos Cavalcante (PPGEE - UFPA) - MEMBRO

Prof. Dr. Antonio Jorge Gomes Abelém (PPGCC - UFPA) - MEMBRO

VISTO:

Prof. Dr. Marcus Vinicius Alves Nunes - Coordenador do PPGEE/CT/UFPA

UFPA / ITEC / PPGEE

BELÉM/PA

2010

S232p Cardoso, Diego Lisboa
de planejamento de redes de alta capilaridade baseadas em
tecnologias DSL: Uma abordagem híbrida empírica e
probabilística Projeto e Implementação de um Sistema de Suporte à
Decisão para o Observatório de Saúde da Amazônia. /

Diego Lisboa Cardoso. Belém, 20105.

63p.: il.

Dissertação (Mestrado) – Universidade Federal do Pará.
Centro Tecnológico. Programa de Pós-Graduação em Engenharia

Elétrica. 200510.

“Nenhum trabalho de qualidade pode ser feito sem concentração, auto-sacrifício, esforço e dúvida.”

Max Beerbohm

A Deus, meus pais e amigos, por transformarem todo o processo deste trabalho em momentos de superação e a Ravenna, minha namorada, pela compreensão e suporte.

AGRADECIMENTOS

Li que um dos melhores momentos no processo de escrever uma tese é aquele em que o autor tem a oportunidade de agradecer àqueles que o ajudaram, porque raramente um pesquisador faz seu trabalho sozinho. Esta tese é um exemplo desta afirmativa.

Agradeço a Deus, por ter iluminado meu caminho em todos os dias de minha vida, sempre me protegendo e a toda minha família.

Aos meus pais e irmão, que sempre apoiaram meu estudo e o caminho que decidi trilhar, mostrando o que é ser um homem digno, entendendo meus momentos de infelicidade e tristeza e proporcionando momentos de pura felicidade.

À Ravenna, pelo amor demonstrado e por estar sempre junto a mim nos momentos de alegria e tristeza.

Ao Prof. Carlos Renato Lisboa Francês, por sua orientação e pela confiança em mim depositada, mostrando não ser apenas um orientador, mas também um amigo que está disposto a ajudar.

Ao Prof. João Crisóstomo Weyl Albuquerque da Costa, por disponibilizar tudo que estava ao seu alcance e que juntamente com o Prof. Renato Francês e todos os demais profissionais do programa de Pós-graduação em Engenharia Elétrica da Universidade Federal do Pará, mostram exemplo em educação e profissionalismo.

Aos amigos Ádamo Santana, Edvar da Luz Oliveira, Jorge Rocha, Marcos Seruffo e Marcelino Silva, companheiros trabalho, de diversão, conversas, conselhos, estudo e brincadeiras, sempre presentes e incentivando meu trabalho de várias formas. Em especial ao meu amigo Ádamo, que nunca desistiu de mim.

Ao CNPQ, pelo apoio financeiro dado a este trabalho e que foi de grande importância para sua conclusão.

ÍNDICE

LISTA DE FIGURAS	v
LISTA DE TABELAS	vi
LISTA DE ABREVIATURAS	viii
RESUMO	xi
ABSTRACT	xii
1. INTRODUÇÃO.....	1
2. TECNOLOGIAS DE ACESSO.....	7
2.1 - <i>CONSIDERAÇÕES INICIAIS</i>	7
2.2 – <i>TECNOLOGIAS NÃO CABEADAS (WIRELESS)</i>	8
2.3 – <i>TECNOLOGIAS CABEADAS (WIRED)</i>	9
2.4 - <i>CONSIDERAÇÕES FINAIS</i>	21
3. TÉCNICAS DE SUPORTE À DECISÃO	22
3.1 - <i>CONSIDERAÇÕES INICIAIS</i>	22
3.2 – <i>PROCESSO DE KDD (KNOWLEDGE DISCOVERY DATABASE)</i>	23
3.3 – <i>ALGORITMOS DE INTELIGÊNCIA COMPUTACIONAL</i>	29
3.3.1 – <i>Redes Bayesianas</i>	31
3.3.1.1 - <i>Construção de Redes Bayesianas</i>	33
3.3.1.3 - <i>Aprendizado de Redes Bayesianas</i>	34
3.3.1.5 - <i>Aprendizado da Estrutura de Redes Bayesianas</i>	35
3.3.2 – <i>Algoritmos Genéticos</i>	39
3.4 - <i>CONSIDERAÇÕES FINAIS</i>	49
4. TRABALHOS CORRELATOS	50
4.1 - <i>CONSIDERAÇÕES INICIAIS</i>	50
4.2 – <i>TRABALHOS CORRELATOS</i>	50
4.3 – <i>CONSIDERAÇÕES FINAIS</i>	57
5. ESTRATÉGIA DE PLANEJAMENTO DE REDES DE ACESSO DE ALTA CAPILARIDADE BASEADAS EM TECNOLOGIA DSL: UMA ABORDAGEM HÍBRIDA EMPÍRICA E PROBABILÍSTICA.....	58
5.1 - <i>CONSIDERAÇÕES INICIAIS</i>	58

5.2 - ABORDAGEM	59
5.2.1 – Definição de arquitetura e equipamentos envolvidos	62
5.2.2 – Definição das variáveis a serem analisadas	63
5.2.3 – Criação do testbed	66
5.2.4 – Testes Empíricos	68
5.2.5 – Análise dos Resultados obtidos	69
5.2.6 – Estudo de correlação Bayesiano	74
6. CONCLUSÃO	83
REFERÊNCIAS BIBLIOGRÁFICAS	88
APÊNDICE A	94
PUBLICAÇÃO CIENTÍFICA	94

LISTA DE FIGURAS

<u>Figura 2.1 – Visão dos sistema PLC (HARSNICA et al., 2004)</u>	10
<u>Figura 2.2 - Utilização do acoplador (HARSNICA et al., 2004)</u>	11
<u>Figura 2.3 - Estrutura básica da Linha Digital do Assinante</u>	12
<u>Figura 2.4 - Bandas de frequências do POTS e ADSL (STARR et al., 1999)</u>	13
<u>Figura 2.5 – Comparativo das tecnologias DSL (CLAVENNA, 2004)</u>	14
<u>Figura 2.6 - Aplicação típica da tecnologia ADSL para acesso Banda Larga</u>	15
<u>Figura 2.7 – Ruídos de cabeamento DSL (KERPEZ; KINNEY, 2008)</u>	17
<u>Figura 2.8 – Exemplo de Next e Fext</u>	18
<u>Figura 3.1 - Principais tarefas de mineração de dados (REZENDE, 2003)</u>	23
<u>Figura 3.2 - Principais tarefas de mineração de dados (REZENDE, 2003)</u>	25
<u>Figura 3.3 - Os três principais passos para a manipulação de incerteza com técnicas de inteligência computacional</u>	27
<u>Figura 3.4 - Exemplo de uma rede Bayesiana (HECKERMAN, 1997)</u>	31
<u>Figura 3.5 - Algumas estruturas possíveis de redes bayesianas para as variáveis A, B e C</u>	36
<u>Figura 3.6 - Distribuição das probabilidades da aptidão dos indivíduos</u>	43
<u>Figura 3.7 - Exemplo da aplicação do operador de crossover</u>	45
<u>Figura 3.8 - Exemplo de aplicação do crossover de dois pontos</u>	46
<u>Figura 3.9 - Exemplo de aplicação do Crossover Uniforme</u>	46
<u>Figura 3.10 - Exemplo da aplicação do operador de mutação</u>	47
<u>Figura 5.1 - Etapas do processo de mineração de dados (REZENDE, 2003)</u>	59
<u>Figura 5.2 – Exemplo de um <i>testbed</i> padrão</u>	61
<u>Figura 5.3 –<i>testbed</i> implementado</u>	62
<u>Figura 5.3 – QoE e QoS (ZAPATER; BRESSAN, 2007)</u>	64
<u>Figura 5.4 – Comportamento do MOS VoIP</u>	70
<u>Figura 5.5 – Perda de pacotes por aplicação</u>	71
<u>Figura 5.6 – Comportamento do VoIP para 3500m</u>	71
<u>Figura 5.7 – Comportamento da vazão da aplicação de vídeo e dados</u>	73
<u>Figura 5.8 – Rede Bayesiana para o fluxo <i>Triple Play</i> sobre ADSL2+</u>	75
<u>Figura 5.11 – Comportamento da vazão da aplicação de dados</u>	79
<u>Figura 2.1 – Visão dos sistema PLC (HARSNICA et al., 2004)</u>	11

<u>Figura 2.2 - Utilização do acoplador (HARSNICA et al., 2004)</u>	12
<u>Figura 2.3 - Estrutura básica da Linha Digital do Assinante</u>	13
<u>Figura 2.4 - Bandas de frequências do POTS e ADSL (STARR et al., 1999)</u>	14
<u>Figura 2.5 – Comparativo das tecnologias DSL (CLAVENNA, 2004)</u>	15
<u>Figura 2.6 - Aplicação típica da tecnologia ADSL para acesso Banda Larga</u>	16
<u>Figura 2.7 – Ruídos de cabeamento DSL (KERPEZ; KINNEY, 2008)</u>	18
<u>Figura 2.8 – Exemplo de Next e Fext</u>	19
<u>Figura 3.1 - Principais tarefas de mineração de dados (REZENDE, 2003)</u>	24
<u>Figura 3.2 - Principais tarefas de mineração de dados (REZENDE, 2003)</u>	26
<u>Figura 3.3 - Os três principais passos para a manipulação de incerteza com técnicas de inteligência computacional</u>	28
<u>Figura 3.4 - Exemplo de uma rede Bayesiana (HECKERMAN, 1997)</u>	32
<u>Figura 3.5 - Algumas estruturas possíveis de redes Bayesianas para as variáveis A, B e C</u>	37
<u>Figura 3.6 - Distribuição das probabilidades da aptidão dos indivíduos</u>	44
<u>Figura 3.7 - Exemplo da aplicação do operador de crossover</u>	46
<u>Figura 3.8 - Exemplo de aplicação do crossover de dois pontos</u>	47
<u>Figura 3.9 - Exemplo de aplicação do Crossover Uniforme</u>	47
<u>Figura 3.10 - Exemplo da aplicação do operador de mutação</u>	48
<u>Figura 5.1 - Etapas do processo de mineração de dados (REZENDE, 2003)</u>	60
<u>Figura 5.2 – Exemplo de um <i>testbed</i> padrão</u>	62
<u>Figura 5.3 – QoS e QoE (ZAPATER; BRESSAN, 2007)</u>	65
<u>Figura 5.4 – Comportamento do MOS VoIP</u>	70
<u>Figura 5.5 – Perda de pacotes por aplicação</u>	71
<u>Figura 5.6 – Comportamento do VoIP para 3500m</u>	72
<u>Figura 5.7 – Comportamento da vazão da aplicação de vídeo e dados</u>	73
<u>Figura 5.8 – Rede Bayesiana para o fluxo <i>Triple Play</i> sobre ADSL2+</u>	75
<u>Figura 5.9 – Probabilidade de ocorrência para uma determinada configuração VoIP</u>	77
<u>Figura 5.10 – Comportamento da vazão do vídeo</u>	78
<u>Figura 5.11 – Comportamento da vazão da aplicação de dados</u>	79

LISTA DE TABELAS

<u>Tabela 2.1 – Resumos das tecnologias xDSL</u>	15
<u>Tabela 3.1 – Ordenação da aptidão interpolada para o intervalo [0.00; 2.00]</u>	44
<u>Tabela 5.1 – Classificação do MOS (TR-126, 2006)</u>	65
<u>Tabela 5.2 – Medidas para Vídeo e Dados em 3500m</u>	72
<u>Tabela 5.3 – Medidas para Vídeo e Dados em 4500m</u>	73
<u>Tabela 5.4 – Valor obtido para os atributos</u>	81

LISTA DE ABREVIATURAS

3GPP2	<i>3rd Generation Partnership Project 2</i>
ADSL	<i>Asymmetric Digital Subscriber Line</i>
AG	Algoritmos Genéticos
AP	<i>Access Point</i>
ARQ	<i>Automatic Repeat Request</i>
CATV	<i>Cable Television</i>
COM	<i>Copper Plant Manager</i>
DPC	Distribuição de Probabilidade Conjunta
DSL	<i>Digital Subscriber Line</i>
DSLAM	<i>Digital Subscriber Line Access Multiplexer</i>
DSM	<i>Dynamic Spectrum Management</i>
DWDM	<i>Dense wavelength division multiplexing</i>
EDA	<i>Ethernet DSL Access</i>
FEC	<i>Forward Error Correction</i>
FEXT	<i>Far-end Crosstalk</i>
FTP	<i>File Transfer Protocol</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
IA	Inteligência Artificial
IP	<i>Internet Protocol</i>
IPTV	<i>Internet Protocol Television</i>
ITU-T	<i>International Telecommunication Union - Telecommunication Standardization Sector</i>
KDD	<i>Knowledge Discovery Database</i>
LAN	<i>Local Area Network</i>

LABIT	<i>Laboratório de Inovação Tecnológica em Telecomunicações</i>
LAPS	<i>Laboratório de Processamento de Sinais</i>
LEA	<i>Laboratório de Eletromagnetismo Aplicado</i>
LPRAD	<i>Laboratório de Planejamento de Redes de Alto Desempenho</i>
MAN	<i>Metropolitan Area Networks</i>
MD	<i>Mineração de Dados</i>
MDL	<i>Descrição de Comprimento Mínimo</i>
MOS	<i>Mean Opinion Score</i>
MPEG	<i>Moving Picture Experts Group</i>
NEXT	<i>Near-end Crosstalk</i>
NGN	<i>Next Generation Network</i>
PDA	<i>Personal Digital Assistant</i>
PLC	<i>Power Line Communication</i>
POTS	<i>Plain Old Telephony System</i>
PQ	<i>Priority Queue</i>
PSNR	<i>Peak Signal Noise Ratio</i>
QoE	<i>Quality of Experience</i>
QoS	<i>Quality of Service</i>
RB	Redes Bayesianas
RNA	Redes Neurais Artificiais
RTP	<i>Real-time Transport Protocol</i>
SDH	<i>Synchronous Digital Hierarchy</i>
SHDSL	<i>Symmetric High-speed Digital Subscriber Line</i>
SNR	<i>Signal to Noise Ratio</i>
SONET	<i>Synchronous optical networking</i>
TPC/IP	<i>Transmission Control Protocol / Internet Protocol</i>
UWB	<i>Ultra-Wide Band</i>
VCEG	<i>Video Coding Experts Group</i>
VDSL	<i>Very-high-bit-rate Digital Subscriber Line</i>

VLAN	<i>Virtual Local Area Network</i>
VLC	<i>Video Lan Client</i>
VOIP	Voz sobre IP
WAN	<i>Wide Area Network</i>
WDM	<i>Wavelength-division Multiplexing</i>
WFQ	<i>Weighted Fair Queuing</i>
WFQ-LL	<i>Weighted Fair Queuing Low Latency</i>
WIMAX	<i>Worldwide Interoperability for Microwave Access</i>
WLAN	<i>Wireless Local Area Network</i>
WMAN	<i>Wireless Metropolitan Area Network</i>
WPAN	<i>Wireless Personal Area Network</i>

RESUMO

A tecnologia DSL- *Digital Subscriber Line* permitem um acesso universal de banda larga, com redução de custo e tempo de implantação necessários, quando comparadas a outras redes de acesso. DSL pode ser considerada uma rede de banda larga de grande capilaridade, uma vez que utiliza uma combinação de infra-estrutura de telefonia existente e a tecnologia de transmissão. Não obstante, a rede DSL precisa ser amplamente compreendida, de maneira que suas principais vantagens não sejam suplantadas pela ineficiência geral do sistema. Esta tese apresenta uma abordagem baseada em estratégias que possibilitem o planejamento de redes de comunicação que *a priori* não foram implementadas para suportar fluxos *Triple Play* (pré-requisito obrigatório diante do perfil atual de usuários da Internet). Será mostrado que, a partir do uso de medidas reais e análises probabilísticas, é possível elaborar o planejamento de redes de comunicação, considerando os parâmetros físicos e lógicos, tais como: o caminho de transmissão, a influência do ruído na comunicação, os protocolos utilizados e tráfego de *Triple Play* (voz, vídeo e dados).

Palavras-chave: Redes; NGN; DSL; Triple Play; Qualidade de Serviço.

ABSTRACT

Recently, the Next Generation Network (NGN) have emerged as a viable alternative to merge the data, voice and video transmission over a single network, while maintaining reasonable standards for quality of service. However, despite the improvements brought by the NGN philosophy, many developing countries lack the infrastructure necessary to implement such technology. In this context, technologies such as DSL (Digital Subscriber Line) allows universal access to broadband, with reduced cost and time, when compared to others access networks. In many third world countries, DSL can be considered a broadband network of high capillarity, since it uses a combination of the existing telephony infrastructure and transmission technologies. However, the DSL network needs to be widely understood, so that its main advantages are not outweighed by the inefficiency of the system. Therefore, the implementation of a planning policy to aid in this process, taking into account the current needs of applications (voice, video and data) is of major importance. This thesis presents an approach based on strategies to facilitate the planning of communication networks which, a priori, were not implemented to support triple play streams (a prerequisite, nowadays, for the current profile of Internet users). It will be shown that, from the use of actual measurements and probabilistic analysis, it is possible to enable the planning of communication networks, considering physical and software parameters, such as the transmission path, the influence of noise in the communication, the protocols used, and traffic of Triple Play (voice, video and data) applications.

Keywords: Networking; NGN; DSL; Triple Play; Quality of Service.

1. INTRODUÇÃO

A disseminação da informação por meios digitais traz à tona novas realidades no cenário mundial e, com isso, novas tecnologias vêm se fortalecendo no sentido de dinamizar o processo de difusão da informação e melhoria da qualidade de vida da população. As redes de nova geração (The Next Generation Network - NGN) têm surgido como alternativa viável para mesclar a transmissão de dados, voz e vídeo em uma única rede, mantendo-se padrões razoáveis de qualidade de serviço. Todavia em muitos países, especialmente os de terceiro mundo, infra-estruturas de telecomunicações necessárias ao suporte de NGN são precárias ou, no pior caso, inexistem. Para estes cenários, há uma preocupação adicional à necessidade de *backbone* ou a camada de transporte; a rede de última ou de acesso.

O termo “última milha” se tornou uma palavra-chave para indicar a tecnologia que conecta o usuário final para ao backbone da rede.

Na maior parte da América do Norte e Europa Ocidental, a penetração da Internet é alta e a maior parte da população tem acesso à mesma. No entanto, esse fato não é verdade nos países em desenvolvimento, onde apenas uma pequena porcentagem da população tem acesso à grande rede, ainda que com velocidades muito baixas e custos elevados (AMBROSI et al., 2005). Segundo (IWS, 2009), mais de 70% da população dos países em desenvolvimento não têm acesso à internet devido à falta de infra-estrutura, além disso, em países como China, Índia e Brasil, com dimensões continentais, a construção de uma rede de telecomunicações nova, utilizando, por exemplo, fibra óptica, torna-se um processo oneroso e quase impraticável. Neste contexto, tecnologias alternativas devem ser buscadas, considerando-se que as mesmas ofereçam desempenho e custo mais atrativos.

Existem várias abordagens para prover serviço de acesso ao usuário final (XIAO et al., 2007). Dentre as disponíveis, uma alternativa com menor tempo de implantação e menor custo denota o uso de uma combinação de infra-estruturas existentes, tais como redes elétricas e redes de telefonia, as quais estão amplamente disponíveis para os usuários finais, na maioria dos países em desenvolvimento (PAPAGIANNI et al., 2009). Para as áreas onde tais redes possuem baixa penetração, redes sem fio despontam como a melhor solução melhor; entretanto, pressupõe-se a necessidade de instalação de uma infra-estrutura básica (torre, estações de base, antenas, etc.).

Redes de telefonia foram originalmente construídas para a comunicação de voz analógica, transportando sinais de voz em até 4 KHz na banda de frequência, e não para comunicação digital de dados. Apesar de sua origem, podem ser consideradas redes de banda larga com grande capilaridade, pois podem utilizar uma combinação de infra-estrutura de enlaces de cobre existentes e tecnologias de transmissão digital (*Digital Subscriber Line*), permitindo assim um acesso universal em banda larga.

A tecnologia DSL (*Digital Subscriber Line*) permanece como a tecnologia de acesso dominante com 65% dos assinantes mundiais de banda larga, contra 33% de conexão por fibra ótica.

A tecnologia DSL também está na vanguarda do acesso banda larga na Europa, América Latina e em países em desenvolvimento como a Índia (OLSEN et al., 2006)(ARENAS et al., 2006) (FAUDON et al, 2006), representando 83% das conexões de banda larga e na China, que possui um crescimento constante, alcançando 93.549,000 assinantes (POINT TOPIC, 2009). Na América Latina a tecnologia DSL é responsável por cerca de 77% de todo o acesso banda larga realizado. No final de 2005 haviam quase 5.300.000 assinantes de tecnologia ADSL (*Assimetric Digital Subscriber Line*) na América Latina (ARENAS et al., 2006).

Particularmente, no Brasil houve um grande crescimento de acesso via banda larga. Em 2005, dos usuários que acessam em domicílio, 52.1% tinham somente a conexão discada; 41.2% unicamente a banda larga e 6.7% ambas as formas de acesso. Em 2008 essa estatística mudou consideravelmente, sendo que o acesso banda larga subiu para 58% contra 31% do acesso discado, onde o acesso via DSL corresponde a 23% do acesso via banda larga (CETIC, 2008).

É fato que o acesso à banda larga muda as necessidades dos usuários, que no princípio apenas se dispunham da tecnologia para o acesso a *web sites*. Nos dias de hoje, estes usuários evoluíram em seus padrões de acesso, estando agora interessados no acesso a uma vasta gama de serviços (e.g. vídeo, voz e dados) simultaneamente; e estes, não mais separadamente, através de provedores diferentes, e sim através de um único provedor de serviços. Dessa forma, as companhias telefônicas (*Telcos*) buscam ofertar o chamado tráfego *Triple Play* (transmissão de voz, vídeo e dados) através de televisão por IP (ou *Internet Protocol Television - IPTV*), a fim de competir em par com as empresas de televisão a cabo, as quais

entraram no mercado de voz e acesso à internet. Portanto, é imperativo estudar as aplicações em tais infra-estruturas, as quais não foram projetadas para este objetivo.

A manutenção de rede de última milha é outro fator importante, realizada por meio de uma série de sistemas; testes manuais realizados por técnicos e testes automatizados, que são desenvolvidos para o serviço de telefonia antigo (POTS - *Plain Old Telephony System*), os quais ignoram as frequências DSL acima de 4 KHz. A manutenção é baseada em estimativas aproximadas do comprimento do enlace telefônico e não levam em consideração as características individuais de cada enlace. Tais problemas tendem a crescer com a necessidade de manutenção do serviço de DSL, que passam a transmitir serviços *Triple Play*: Dados, Protocolos de voz sobre o IP (VoIPs) e IPTV.

À luz deste indicativo é importante utilizar uma estratégia de testes que possibilitem identificar com precisão e informar a origem de um problema na rede; se esse problema vem do provedor de serviços Internet (ISP), a central telefônica (CO), dos enlaces, do modem ou do computador do usuário. Testes DSL não estão limitados somente a medição de parâmetros elétricos do enlace de cobre, mas também incluem uma análise comparativa dos dados extraídos com limites previamente conhecidos (KERPEZ, 2008).

A convergência entre tecnologias existentes e emergentes de banda larga tem sido considerada como um grande desafio, principalmente no que concerne a entrega de conteúdo multimídia nestas tecnologias. Isso se deve ao fato das novas aplicações, tais como o IPTV, demandarem grandes larguras de banda; as quais normalmente não estão disponíveis devido a, por exemplo, altas distâncias dos enlaces DSL, ruídos na linha, congestionamento de dados ou deficiência dos protocolos implementados.

Quanto à transmissão na rede, o protocolo TCP/IP (*Transmission Control Protocol / Internet Protocol*) é considerado, por censo geral, como o protocolo padrão para intercomunicação dos diferentes tipos de rede, entretanto, dentre os diversos problemas que uma rede IP pode possuir, um dos principais é o fornecimento de QoS fim-a-fim. Para isso, alguns serviços utilizam o protocolo de transporte TCP (*Transmission Control Protocol*) que, desde seu advento, tem sido considerado o principal protocolo de comunicação. Todavia em redes com alto grau de perda de pacotes este protocolo torna-se falho, pois ele habilita o controle de congestionamento do protocolo desnecessariamente, uma vez que certas aplicações podem tolerar um certo grau de perdas na comunicação.

Observa-se então que a atual comunicação de dados que utiliza principalmente o protocolo TCP/IP para aplicações como o tráfego “puramente” HTTP (*Hyper Text Transfer Protocol*) é apropriado, uma vez que mantém a compatibilidade entre as redes existentes (roteadores, gateways, etc), além de realizar o controle de fluxo e de congestionamento. Entretanto, para as aplicações atuais, conhecidas como *Triple Play*, que possuem fluxos sensíveis à QoS, o protocolo TCP/IP mostra-se deficiente.

O fato do protocolo TCP/IP ser dividido em camadas facilita a implementação de novas aplicações, além de estar de acordo com os princípios relacionados à engenharia de software; contudo, atualmente, esta característica tornou-se negativa, uma vez que as novas aplicações necessitam de níveis mínimos para seu funcionamento. Por sua vez. Estes níveis mínimos para aplicações específicas em detrimento de outras, poderiam ser mantidos através de uma maior interação entre os protocolos.

Alternativamente, o modelo *crosslayer* (entre camadas) visa implementar uma interação entre as camadas do protocolo, fornecendo cada vez mais qualidade de serviço ao usuário, que a cada dia torna-se menos avesso ao mundo informatizado e, conseqüentemente, cada vez mais exigente quanto à tecnologia. Através desta abordagem, por exemplo, torna-se possível adaptar parâmetros de desempenho em tempo real da camada de aplicação, tais como vazão e *jitter*, desde que a camada de transporte informe uma perda de pacotes excessiva, justificando uma adaptação da aplicação que esta sendo utilizada, baseado nas características do meio no qual esta sendo realizada a transmissão.

Por conseguinte, é fundamental investigar as tecnologias (tanto de última milha, como de protocolos utilizados para transmissão) com vistas a realizar um planejamento da melhor estratégia de implantação dentre os diversos cenários possíveis. Essas inferências de possíveis cenários, via de regra, são realizadas utilizando-se uma combinação entre prototipação e modelagem para avaliação de desempenho.

Este trabalho tem por objetivo precípua contribuir com as áreas envolvidas, propondo o desenvolvimento de uma metodologia para planejamento e avaliação de desempenho de redes com ADSL2+ com tráfego *Triple Play* (vídeo, voz e dados). Tal estratégia envolve uma atividades de testes, a fim de gerar um arcabouço de dados consistente, empregando-se uma análise probabilística e abordagem *crosslayer*, permita planejar a implementação de redes de última milha baseadas em tecnologia DSL, de forma a maximizar o desempenho das

aplicações e minimizar os custos de implantação; além de auxiliar nos processos decisórios decorrentes de serviços e na manutenção de redes existentes. Os testes implementados foram idealizados sobre o *backbone* das Telecomunicações Brasileiras S.A. (Telebrás) e com base em normas específicas de comunicação DSL.

Dentre os objetivos específicos desta tese, é possível destacar:

- Implementação de protótipos para realização de testes, seguindo normas internacionais;
- Elaboração de testes exaustivos com base nos cenários propostos, de modo a criar uma base de dados consistente, a qual será utilizada para estudos de correlação e análise de cenários antecipatórios;
- Aplicação do processo de MD em um domínio real, mais especificamente no de tecnologias de última milha. Para tanto, são aplicados modelos de inteligência computacional (redes bayesianas - RB e algoritmos genéticos - AG) para estabelecer um sistema de suporte à decisão que permita a tomada de ações de maneira antecipada, baseada na exploração dos cenários que representem os aspectos essenciais da abordagem *crosslayer*.
- Aplicação em cenários reais, de modo a simular tomada de decisões em fases tais como: projeto de implantação de uma infra-estrutura de telecomunicações, a fim de qualificar os investimento e maximizar os retornos; diagnóstico de redes existentes, possibilitando a análise da implantação de novos serviços aos usuários, desde que sejam mantidos padrões mínimos de qualidade;

De maneira a efetivar as proposições desta Tese, este trabalho está estruturado da seguinte forma: no segundo capítulo serão apresentadas algumas tecnologias utilizadas na implementação de infra-estruturas de rede, entre as quais se encontra a tecnologia foco desta pesquisa, assim como alguns conceitos e definições referentes ao tema.

No terceiro capítulo, são apresentadas as técnicas que serão utilizadas para apoiar esta proposta de tese, nas quais são abordados as técnicas de suporte a decisão e otimização.

O quarto capítulo é destinado aos trabalhos e estudos relacionados às áreas englobadas neste trabalho, que sintetizam parte do atual “estado-da-arte” das áreas envolvidas nesta tese.

No quinto capítulo, é apresentada a proposta deste trabalho, denotando a metodologia utilizada para implementação de qualidade de serviço para aplicações *Triple Play*. Apresentam-se, também, exemplos e estudos de aplicação juntamente com a análise de desempenho e resultados obtidos.

Por fim, o capítulo seis apresenta as conclusões de toda a pesquisa realizada, enfocando nas contribuições da tese para as diversas áreas envolvidas, assim como as dificuldades encontradas ao longo do desenvolvimento da pesquisa e trabalhos futuros como desdobramento da pesquisa desenvolvida.

2. TECNOLOGIAS DE ACESSO

2.1 - CONSIDERAÇÕES INICIAIS

A crescente demanda por serviços de telecomunicações, o aumento na disseminação da informação através da Internet e o interesse em investimentos na inclusão digital geraram uma nova realidade onde a disponibilidade de tais serviços é indispensável e na qual as redes de banda larga possuem um papel fundamental.

Redes de banda larga são aquelas que provêem acesso à Internet com altas velocidades, possibilitando aos usuários finais um acesso com qualidade e que atenda às suas necessidades. Através destas, são implementadas as tecnologias de acesso, também conhecidas como tecnologias de última milha, nome dado às redes que interligam o usuário final às redes centrais.

Estas infra-estruturas têm a função de oferecer o serviço IP fim-afim com alto desempenho à camada de aplicação do protocolo TCP/IP, nas quais os principais requisitos relacionados às redes são: largura de banda adequada para os serviços, inteligência para prover QoS (Qualidade de Serviço) assegurando requisitos mínimos e arquitetura robusta que suporte *multicast* (pertinente para aplicações de vídeo).

As tecnologias de última milha podem ser divididas em duas grandes classes, as não cabeadas (*wireless*) e as cabeadas (*wired*). Nas tecnologias de acesso sem fio (*wireless*) destacam-se a comunicação via satélites, sistemas celulares, redes locais sem fio (WLAN's) e as redes sem fio de longas distâncias (Wimax). Já nas tecnologias cabeadas predominam a fibra ótica (para acesso e não para *backbone*), PLC (*Power Line Communications*), CATV (*cable TV*), xDSL e linha discada.

Neste capítulo, serão apresentados os principais conceitos que permeiam algumas destas tecnologias e seus funcionamentos, bem como uma breve apresentação dos principais componentes utilizados.

2.2 – TECNOLOGIAS NÃO CABEADAS (*WIRELESS*)

As redes sem fio classificam-se por distância em: Redes Pessoais (PANs – *Personal Area Networks*), Redes Locais (LANs – *Local Area Networks*), Redes Metropolitanas (MANs – *Metropolitan Area Networks*) e Redes de Longa Distância (WANs – *Wide Area Networks*). Cada tipo de rede possui aplicações específicas e diferem entre si quanto à largura de banda, potência dos dispositivos de hardware, serviços oferecidos, etc.

As PANs sem fio (*Wireless PANs – WPANs*) possuem alcance máximo de 10 metros, onde destacam-se padrões como o *Bluetooth* e o UWB (*Ultra-Wide Band*). O padrão *Bluetooth* (IEEE – *Institute of Electrical and Eletronics Engineers* 802.15.1) suporta taxa de transmissão de até 1 Mbps e é voltado para computação pessoal sem fio (KUROSE; ROSS, 2006), integrações entre PDAs e computadores e comunicação entre periféricos. Para serviços multimídia, como a distribuição de áudio ou vídeo digital no interior de uma casa, pode ser utilizado o padrão UWB (IEEE 802.15.3), que suporta taxas em torno de 400 Mbps.

As LANs sem fio (*Wireless LANs - WLANs*) suportam um número maior de aplicações e usuários, são baseadas no padrão IEEE 802.11 e cobrem distâncias de até 100 metros. As redes 802.11 podem ser constituídas através de duas topologias: infra-estrutura e ad-hoc. No modelo baseado em infra-estrutura, cada estação da rede se comunica com um elemento central chamado ponto de acesso (AP – *Access Point*), desta forma, todo tráfego de rede entre as estações passa pelo AP que, por sua vez, pode estar integrado a uma rede local fixa. A topologia ad-hoc segue outra linha de aplicação, nela as estações comunicam-se diretamente entre si, não existindo um elemento centralizador (ponto de acesso ou AP). Este tipo de rede é muito utilizada em ambientes corporativos, domésticos, aeroportos, etc. Os padrões associados às WLANS IEEE 802.11 são: 802.11a (velocidades de até 54 Mbps), 802.11b (velocidades de até 11 Mbps) e 802.11g (velocidades de até 54 Mbps).

Em função do aumento na área de abrangência e nas taxas de transmissão as redes 802.11 estão presentes nos segmentos das LANs e MANs. O uso de antenas direcionais (ponto-a-ponto) ou da topologia *mesh* tem permitido ao 802.11 alcançar distâncias de até 10Km com taxas de 54 Mbps.

As redes metropolitanas sem fio (WMANs, IEEE 802.16) foram inicialmente propostas para atender à chamada última milha com qualidade comparável às tecnologias

DSL e aos circuitos dedicados, todavia devido possuir características tais como um maior alcance, mecanismos para estabelecimento de qualidade de serviço, níveis de segurança, entre outros, este padrão tem sido utilizado na implementações de *backbones*. As redes 802.16 permitem atingir a taxa de 75 Mbps com alcance de 50 Km, utilizando topologias ponto-a-ponto e ponto-multiponto (EKLUND et al., 2006). Nesta última, ocorre a comunicação entre as estações de usuário e uma estação base, que coordena o acesso ao meio, de forma similar aos APs das redes 802.11.

A outra topologia suportada pelo 802.16 é a rede *mesh*. O WiMAX Fórum (WIMAX FÓRUM, 2009) reúne provedores de serviços sem fio e fabricantes de equipamentos os quais atestam a interoperabilidade dos produtos baseados no padrão IEEE 802.16 e da mesma forma como aconteceu com as redes WI-FI e a tecnologia 802.11, os termos WiMAX e 802.16 tem sido empregados como sinônimos. Além de prover aos sistemas fixos (e.g. um prédio) a possibilidade acesso sem fio por banda larga.

2.3 – TECNOLOGIAS CABEADAS (*WIRED*)

Existem diversas tecnologias cabeadas disponíveis no mercado as quais podem ser aplicadas como tecnologia de última milha, cada uma com sua especificidade e aplicabilidade. Alguns exemplos de tecnologias são fibra ótica, PLC, DSL, entre outras.

2.3.1 - Fibra ótica

Desde que foram desenvolvidas, as fibras óticas representaram uma revolução na forma de transmitir informações, sendo utilizada para transmitir voz, televisão e sinais de dados via ondas de luz, por meios de fios finos e flexíveis, constituídos de vidro ou plástico que, comparados com fios metálicos, apresentam inúmeras vantagens.

O funcionamento da fibra ótica se dá através da propagação da luz no interior de uma fibra ótica fundamentada na reflexão total da luz. Quando um raio de luz se propaga em um meio cujo índice de refração é n_1 (núcleo) e atinge a superfície de outro meio com índice de refração n_2 (casca), onde $n_1 > n_2$, e desde que o ângulo de incidência (em relação à normal) seja maior ou igual ao ângulo crítico, ocorrerá o que é denominado de reflexão total, do que resulta o retorno de raio de luz ao meio com índice de refração n_1 . A luz é injetada na fibra

por uma de suas extremidades sob um cone de aceitação, em que este determina o ângulo por que o feixe de luz deverá ser injetado, para que ele possa de propagar ao longo da fibra óptica.

A composição básica de fibras ópticas é de materiais dielétricos com uma estrutura cilíndrica, composta de uma região central, que denominamos núcleo, que é por onde a luz trafega, e uma região periférica, denominada casca, que envolve completamente o núcleo. As dimensões vão variar conforme o tipo da fibra, podendo ser de 8 micrometros até 200 micrometros, e a casca de 125 micrometros até 240 micrometros.

A composição básica de uma fibra ótica é a que segue: núcleo (fibra de vidro), casca que envolve o núcleo (fibra de vidro); película que recobre a casca, chamado de acrilato; um tubo onde as fibras são comportadas, chamadas de tubete; os fios de aramida, para atuar como proteção e tração; bastão de Kevlar, utilizado para dar resistência mecânica ao cabo; e a capa, constituída por um polímero.

Através de tecnologias como SDH (*Synchronous Digital Hierarchy*) e SONET (*Synchronous Optical Network*) é realizado o transporte de dados em longas distâncias. Além disso, a utilização de DWDM (*Dense Wavelength Division Multiplexing*) na camada ótica possibilita maximizar a utilização das fibras, permitindo a multiplexação de diversos comprimentos de onda em uma única fibra (MORROW; VIJAYANADA, 2003). Devido seu alto custo ela é pouco utilizada como última milha.

2.3.2 - PLC (*Power Line Communication*)

Power Line Communication é uma tecnologia de acesso que utiliza a infra-estrutura da rede elétrica para transmissão de dados. Para o funcionamento do PLC, os cabos elétricos são utilizados como meio de transmissão e a instalação elétrica *indoor* comporta-se como uma rede de dados onde cada tomada elétrica é um ponto de conexão da rede. O PLC utiliza redes de distribuição secundária, onde estão conectados os consumidores, com abrangência de alguns quarteirões por circuito. Esta tecnologia requer baixo investimento, pois as tomadas de energia elétrica já serão os pontos de entrada e saída de dados.

A largura de banda disponível é compartilhada entre dezenas de usuários simultaneamente (usuários que estiverem ligados ao mesmo transformador), logo o desempenho de uma conexão varia de acordo com o número de pessoas que estiverem utilizando, assim como em outras tecnologias de acesso. Em função desse compartilhamento, é necessário proteger a privacidade do tráfego individual; para tal se deve empregar técnicas

como a VLAN (*Virtual Local Area Network*), baseada na IEEE 802.1Q, que assegura divisão de domínios de *broadcast*. No modelo atual, cada *head end* (HE) é constituído de 254 canais individuais, ou seja, são 254 usuários que compartilham a mesma rede física, porém em redes lógicas diferentes.

Uma rede PLC é composta por elementos básicos tais como: o modem PLC, responsável pela conexão entre os equipamentos do usuário final e a rede elétrica; a estação base PLC que é o equipamento do provedor de serviço de comunicação responsável por conectar a rede de acesso PLC com o *backbone* do provedor (Figura 2.1). A conexão com o *backbone* pode ser feita através de interfaces padrões com altas taxas de transmissão como, por exemplo, *Fast Ethernet* e WDM (*Wavelength-Division Multiplexing*), e a conexão com o canal de transmissão PLC é realizado através de um acoplamento eletromagnético semelhante ao utilizado no modem PLC.

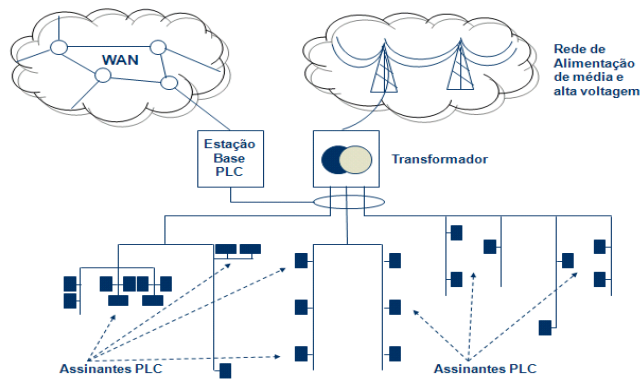


Figura 2.1 – Visão dos sistema PLC (HARSNICA et al., 2004)

O acoplamento é realizado através de um equipamento chamado de *Coupler* (Figura 2.2) e deve garantir uma separação galvânica, atuando como um filtro passa alta, distinguindo o sinal de comunicação (acima de 9 kHz) do sinal elétrico de potência (50 ou 60 Hz), para que assim possa enviar e receber o sinal de informação através da rede de distribuição de energia elétrica.

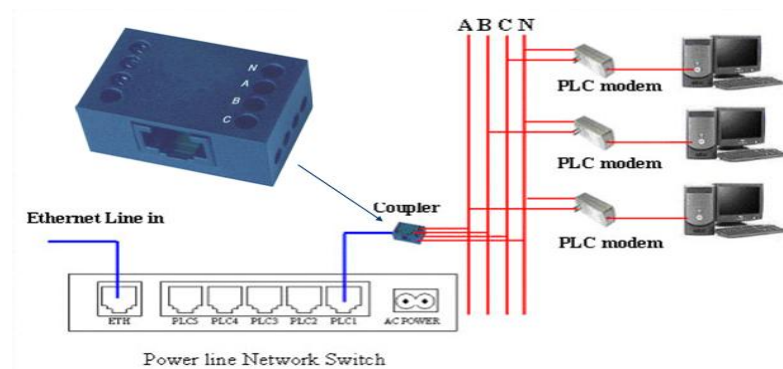


Figura 2.2 - Utilização do acoplador (HARSNICA et al., 2004)

Os repetidores são utilizados em casos que a distância entre a estação base e o modem é muito longa, impossibilitando a comunicação entre estes dois dispositivos, devido à grande atenuação do sinal.

O *gateway* difere de um repetidor, pois além de amplificar e injetar o sinal na rede em outra frequência é capaz de realizar a gerência da sub-rede formada por ele. De maneira geral, um gateway pode ser inserido na rede para segmentá-la em diversas sub-redes, na qual cada sub-rede é gerenciada por um *gateway* e todos são gerenciados pela estação base.

2.3.3 - DSL (*Digital Subscriber Line*)

Digital Subscriber Line (simplesmente DSL ou ainda xDSL) é uma família de tecnologias que fornecem altas taxas de transmissão de dados, utilizando a própria infraestrutura de telefonia que chega na maioria das residências possibilitando a comunicação de voz e dados ao mesmo tempo.

As tecnologias DSL utilizam os enlaces locais de cobre (linha telefônica) para transportar os sinais DSL do modem do assinante até a central telefônica – CO (*Central Office*) - as quais se conectam, em paralelo, em um equipamento multiplexador denominado de DSLAM (*Digital Subscriber Line Access Multiplexer*). Do DSLAM, os dados movem-se, em alta velocidade, diretamente para as redes digitais de *backbones* que, por sua vez, se conectam aos servidores de aplicação, como os de serviços encontrados na internet. Esta arquitetura pode ser observada na Figura a seguir.

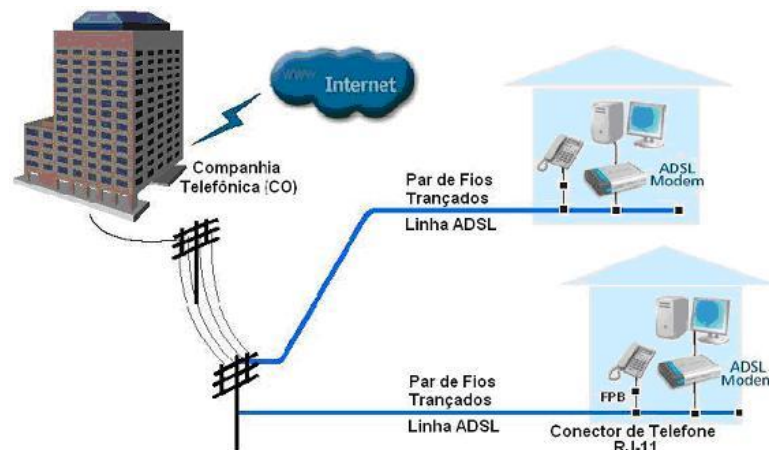


Figura 2.3 - Estrutura básica da Linha Digital do Assinante

A sigla DSL descreve a tecnologia enquanto que xDSL representa a família da tecnologia DSL. A identificação “x” assume diferentes significados dependendo da velocidade e aplicação da família. Cada família desta tecnologia possui características específicas, as quais devem ser levadas em consideração no momento de sua implantação.

O ADSL é uma tecnologia de transmissão de dados que permite taxas de até 7 Mbps para *downstream* (sentido central-assinante) e de até 800 Kbps para *upstream* (sentido assinante-central) (STARR et al., 1999). O ADSL começou a ser desenvolvido no início da década de 90 e a sua primeira versão comercial ocorreu em 1995. Em 1998, o ITU-T (*International Telecommunication Union – Telecommunication*) aprovou um conjunto de recomendações para ADSL, chamado recomendação G.992.1. Este conjunto de recomendações é praticamente idêntico à recomendação ANSI T1.413 versão 2.

A tecnologia ADSL é dita assimétrica porque o canal de *downstream* ocupa muito mais a banda de frequência do que o canal de *upstream*. Uma das razões para essa assimetria se deve ao fato de que as aplicações residenciais estão mais voltadas para o *download* de dados na rede.

Na ADSL usual, isto é, onde o usuário tem um serviço de comunicação por voz e um link ADSL na mesma linha, a banda de *upstream* opera na região de frequência entre 26 kHz e 138 kHz e a banda de *downstream* entre 142 kHz e 1,1 MHz. O serviço de telefonia – POTS usa as frequências próximas do nível DC (*Direct Current*) até aproximadamente 4 kHz (Figura 2.4); de modo que existe um intervalo de frequências entre a banda do POTS e a banda de upstream da ADSL, esse intervalo é conhecido como banda de guarda.

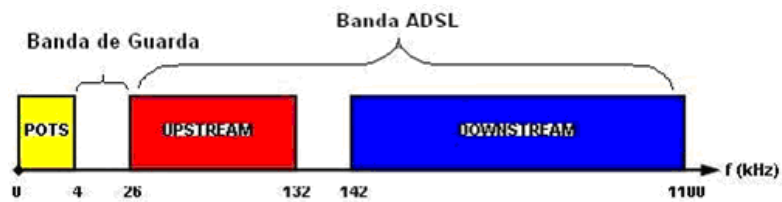


Figura 2.4 - Bandas de frequências do POTS e ADSL (STARR et al., 1999)

Objetivando alcançar maiores distâncias e funções extras como, por exemplo, diagnóstico do enlace, em julho 2002, o ITU-T finalizou as recomendações G.992.3 *Asymmetric Digital Subscriber Line Transceivers 2 - ADSL2 (G.dmt.bis)* e G.992.4 *Splitterless Asymmetric Digital Subscriber Line Transceivers 2 - Splitterless ADSL2 (G.lite.bis)*. Os dois novos padrões para a tecnologia ADSL, e que foram denominados ADSL2. Entretanto, apesar de estar definido nessa recomendação que 8 Mbps é aproximadamente a máxima taxa de bits, taxas de até 11 Mbps podem ser alcançadas quando se usa o equipamento ADSL2 existente.

No ano seguinte, a ITU-T liberou a recomendação G992.5 que descreveu o ADSL2+, uma tecnologia que usa uma largura de banda duas vezes maior que a do ADSL2 e que pode alcançar taxas de bits na banda de *downstream* de até 25 Mbps. A maior extensão da largura de banda está associada ao canal de *downstream*, entretanto existem modos de operação que também estendem o canal de *upstream* e tornam possível taxa de bits de até 2,5 Mbps para *upstream*. As soluções que usam o ADSL2+ geralmente são do tipo multimodo, ou seja, permitem operação conjunta, na mesma infra-estrutura, do ADSL, ADSL2, e também do ADSL2+.

A Figura 2.5 e a Tabela 2.1 sintetizam as peculiaridades das variantes xDSL. Os termos *downstream* e *upstream* significam a comunicação entre a central telefônica e o assinante e entre o assinante e a central telefônica, respectivamente.

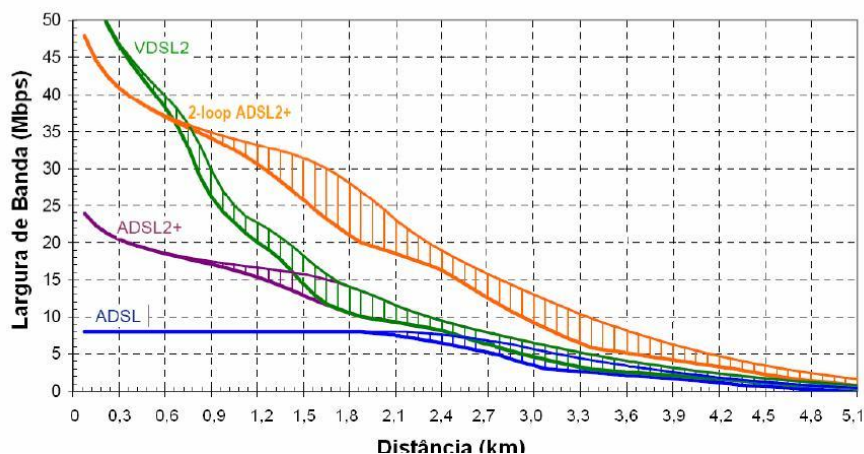


Figura 2.5 – Comparativo das tecnologias DSL (CLAVENNA, 2004)

Tabela 2.1 – Resumos das tecnologias xDSL.

Tecnologia	Significado	Ano de padronização	Velocidades Máximas
ADSL	<i>Asymmetric Digital Subscriber Line</i>	1999	7 Mbps (<i>downstream</i>) 800 kbps (<i>upstream</i>)
ADSL2	ADSL2	2002	8 Mbps (<i>downstream</i>) 1 Mbps (<i>upstream</i>)
ADSL2+	ADSL2+	2003	24 Mbps (<i>downstream</i>) 1 Mbps (<i>upstream</i>)
ADSL2-RE	<i>ADSL Reach Extended</i>	2003	8 Mbps (<i>downstream</i>) 1 Mbps (<i>upstream</i>)
SHDSL	<i>Symmetric High-Bit Rate Digital Subscriber Line</i>	2003	5,6 Mbps (<i>downstream/upstream</i>)
VDSL	<i>Very-high-data-rate DSL</i>	2004	55 Mbps (<i>downstream</i>) 15 Mbps (<i>upstream</i>)
VDSL2-12 MHz	VDSL 2 Alcance Longo	2005	55 Mbps (<i>downstream</i>) 30 Mbps (<i>upstream</i>)
VDSL2-30 MHz	VDSL 2 Alcance Curto	2005	100 Mbps (<i>downstream/upstream</i>)

O modelo de referência para um sistema ADSL é definido pela norma ANSI T1.413 e pelo relatório técnico TR-001 do DSL Fórum. Tal modelo é apresentado na Figura 2.6 (STARR et al., 1999).

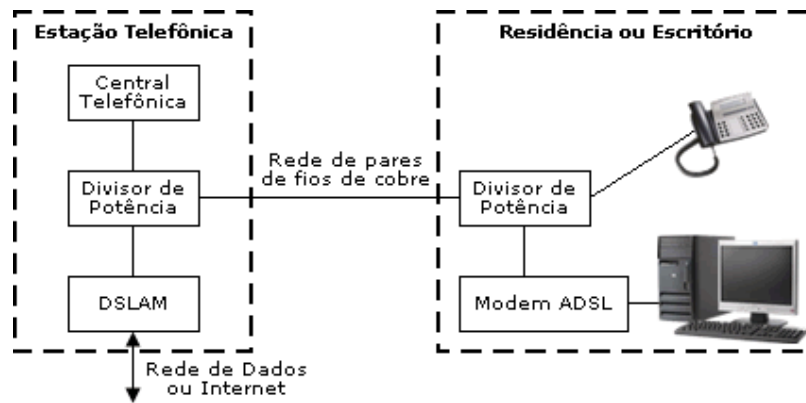


Figura 2.6 - Aplicação típica da tecnologia ADSL para acesso Banda Larga

A rede ADSL (e suas variantes) apresenta basicamente os seguintes componentes:

- **Modem ADSL:** Equipamento instalado no cliente para conexão com um PC ou com uma rede local;
- **Divisores de potência:** São divisores de potência e filtros colocados na residência do usuário e na Estação Telefônica, e que permitem a separação do sinal de voz da chamada telefônica do tráfego de dados ADSL;
- **DSLAM:** multiplexador de acesso DSL cuja função é concentrar o tráfego de dados das várias linhas telefônicas com modem DSL e conectá-lo com a rede de dados;
- **Rede de dados:** A rede de dados a que se conecta o DSLAM poderá ser a rede do provedor de conexão a internet ou qualquer outro tipo de rede de dados. Para acompanhar a tendência de migração de tecnologia, os DSLAMs evoluíram da tecnologia ATM para *Ethernet/IP* (THOMPSON, 2005).

A utilização da tecnologia DSL para transmissão de dados em alta velocidade permite um rápido fornecimento do serviço, principalmente devido ao fato de existir uma rede externa e um cabeamento com par trançado de cobre, cuja cobertura é ampla em praticamente todas as áreas e nichos de mercado. A idade desta rede não é homogênea, podendo coexistir redes novas e redes de até 50 anos. Dessa forma, a rede existente que, por um lado é uma vantagem competitiva extraordinária, por outro lado pode se configurar em um problema.

O crescimento rápido do acesso via DSL, gera uma alta demanda no setor de telecomunicações por equipamentos que podem prever, com precisão, o desempenho do

acesso DSL através de uma linha telefônica (também referida como uma linha local do assinante).

A linha do assinante é um *link* de par trançado metálico entre o cliente e o Escritório Central telefônico – *Central Office* (CO). Existentes equipamentos de análise de enlaces DSL que são capazes de avaliar o desempenho das mesmas, contudo exigem uma operação nos dois pontos (envio de sinais de teste de uma extremidade do circuito e medição dos sinais no outro extremo), envolvendo o envio de um veículo de serviço. Isto leva a processos de teste caros, sendo, portanto, uma solução indesejável para os fornecedores de acesso DSL.

Além disso, existem itens que podem impedir a utilização do serviço, tais como perdas na linha (*loop loss*); existência de extensões na linha (*bridge taps*); ruídos específicos de enlaces de par trançado; altas distâncias do assinante à central telefônica (comprimento do *loop*); entre outros, que podem diminuir o desempenho do serviço. Normalmente as especificações são dadas para cabeamento 24 ou 26 AWG com distâncias médias de 2090m (SANDRI et al., 2007); todavia, é difícil identificar com exatidão tais trechos, que tipo de cabo foi utilizado e suas distâncias, o que torna necessária a medição. Tais itens são apresentados mais especificamente na seção a seguir.

Uma vez que o DSL utiliza frequências de espectro relativamente altas, o sinal é suscetível a fontes de ruído externo. Assim, a compreensão sobre o comportamento de diferentes tipos de ruído e seus efeitos sobre o desempenho da rede são extremamente úteis na concepção e implantação de sistemas DSL (ADSL, ADSL2+), bem como os da próxima geração (VDSL, VDSL2). Durante os últimos anos, o *crosstalk* tem sido considerado o maior prejudicador de serviços DSL. No entanto, outros tipos de ruído também ganharam grande atenção, como a interferência de rádio frequência (RFI), o ruído impulsivo (IN), o ruído de impulso repetitivo elétrico (REIN), entre outros (TELEBRAS, 2007)(WALLACE et al., 2005).

A Figura 2.7 ilustra as deficiências dos enlaces DSL (STOLLE, 2002)(KERPEZ et al., 2003), que são principalmente *loop* e *bridge taps*, *crosstalk*, interferência eletromagnética (EMI), o ruído impulsivo, distorção harmônica e ruído de fundo.

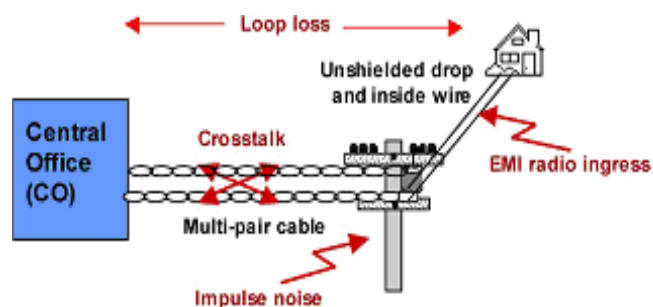


Figura 2.7 – Ruídos de cabeamento DSL (KERPEZ; KINNEY, 2008)

Ruídos em linhas telefônicas normalmente ocorrem devido a um desequilíbrio do cabo de par trançado. Existem muitos tipos de ruídos que causam este desequilíbrio, os mais comuns são ruído *crosstalk*, o ruído de rádio e ruído impulsivo.

Ruído de rádio são sinais de transmissão sem fio remanescente em linhas de telefone, especialmente emissores de rádio AM e transmissões amadoras (HAM).

Ruído impulsivo é um *crosstalk* não-estacionário de eventos eletromagnéticos (tais como o toque de telefones em linhas compartilhando o mesmo *binder*, e descargas elétricas atmosféricas) que podem ser de banda estreita ou larga e que ocorrem aleatoriamente. Ruídos impulsivos podem ser dezenas de milivolts em amplitude e podem durar centenas de microssegundos (STARR et al., 2003)(CIOFFI, 1999).

As fontes de ruído impulsivo são as mais diversas possíveis e inclui ignição de motores elétricos, controle de voltagem de elevadores, dispositivos elétricos domésticos e industriais, etc. (STARR et al., 1999). A sua natureza aleatória e não estacionária faz com que o ruído impulsivo seja de difícil modelagem e previsão. Técnicas como correção de erros no receptor (FEC – *forward error correction*) e *Interleave* tentam minimizar o impacto do ruído impulsivo na transmissão de sistemas DSL.

Crosstalk é causada pela radiação eletromagnética de outras linhas telefônicas nas proximidades, via de regra, dentro do mesmo cabo. O *crosstalk* é o principal fator de limitação da taxa de bits e da distância alcançada em tais sistemas (STARR et al., 1999). O entrelaçamento dos cabos metálicos entre si tem por objetivo reduzir tal com acoplamento e com isso, melhorar o desempenho do sistema (STARR et al., 1999). Existem dois tipos de *crosstalk*: o FEXT e NEXT.

O *Next* (*Near-end Crosstalk*) é o maior impedimento para sistemas que compartilham a mesma banda de frequência entre o *upstream* e o *downstream* (STARR et al., 1999). O *Next* é o ruído visto pelo receptor localizado na mesma extremidade do cabo onde um dado transmissor é a fonte de ruído (Figura 2.8). O *Next* pode ser minimizado utilizando diferentes bandas de frequência para *upstream* e *downstream*.

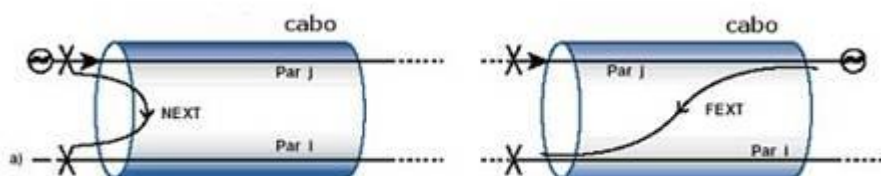


Figura 2.8 – Exemplo de Next e Fext

O *Fext* (*Far-end Crosstalk*) é o ruído detectado pelo receptor localizado na extremidade mais distante do transmissor que é a fonte de ruído (Figura 2.8). O *Fext* é menos prejudicial que o *Next* devido o mesmo ser atenuado em razão da distância que tal ruído precisa percorrer para chegar ao receptor.

Técnicas como o DSM (*Dynamic Spectrum Management* – Gerenciamento Dinâmico de Espectro) buscam uma forma controlada de injeção de espectro em sistemas DSL, de tal maneira que o *crosstalk* resultante assumam valores aceitáveis do ponto de vista de desempenho (STARR et al., 2003).

O ruído branco é o sinal cujo espectro cobre toda a faixa de frequências, ou seja, vai de menos infinito ao mais infinito. Pode-se dizer que o ruído branco possui componentes em todas as frequências. Denominado pela cor branca por ser equivalente à junção de todas as demais.

O ruído branco aparece somado ao sinal, na recepção, devido principalmente ao movimento aleatório de elétrons nos pares telefônicos e também nos componentes eletrônicos, que também é conhecido como ruído térmico. É desejável que a contaminação de ruído branco no sinal recebido seja a menor possível, além de que a relação entre as potências de sinal e ruído seja a maior possível; visto que, quanto maior for à relação sinal/ruído, menos erros ocorrerão na detecção do sinal. As relações entre potências normalmente são expressas em dB.

O conceito de relação Sinal/Ruído (*Signal to Noise Ratio* – SNR) se aplica a qualquer meio físico de transmissão, seja ele por par de cobre ou por rádio frequência (RF), por exemplo. Ele também pode ser designado como qualidade do sinal.

$$S/R(dBm) = 10 * \log(Ps/Pr) \quad (3.1)$$

Onde:

S/R = Relação Sinal/Ruído em dB;

PS = Potência do Sinal;

Pr = Potência do ruído.

Muitas pesquisas têm sido feitas sobre o impacto do ruído não-estacionário em sistemas DSL; entretanto, poucos estudos têm sido realizados abordando o seu impacto em termos de análise experimental. Isso pode ser creditado à falta de acesso a uma infra-estrutura adequada para lidar com experiências práticas.

2.4 - CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentadas algumas tecnologias de acesso disponíveis atualmente, com ênfase especial para a tecnologia DSL, a qual será utilizada posteriormente para estudos que validam a metodologia proposta nesta Tese. No que concerne a tecnologia DSL, foram apontados fatores que influenciam na qualidade de serviço do meio, como ruído e *crosstalk*. Essas características serão alvo dos estudos realizados no capítulo 5.

No próximo capítulo, serão apresentadas técnicas de suporte a decisão que permitem a realização de inferências a cerca dos parâmetros de qualidade de serviço em redes DSL.

3. TÉCNICAS DE SUPORTE À DECISÃO

3.1 - CONSIDERAÇÕES INICIAIS

Estudos para o desenvolvimento de meios e métodos que permitam a eficiência da gestão do conhecimento tem tornado-se imprescindíveis, em especial devido à escassez dos recursos financeiros disponíveis, tornando-se necessário que esses poucos recursos existentes sejam aplicados em ações que efetivamente possam gerar impacto na vida da população. Para tanto, é necessário poder quantificar e qualificar os investimentos.

O suporte à decisão está associado frequentemente a aplicações envolvendo análise e exploração dos dados e informações históricas de uma organização, de modo a prover um mecanismo de alto nível para auxiliar os processos de tomada de decisão. É imperativo que esses mecanismos sejam implementados em sistemas computacionais face à grande quantidade de dados que são manipulados e à própria condição humana, relacionada com a limitação para a análise de grande volume de informações, estresse, erro, desatenção, etc. As soluções computacionais utilizadas são compostas, em sua maioria, por um conjunto de métodos, técnicas e ferramentas de áreas como bancos de dados, estatística, e inteligência computacional. Um exemplo típico e proeminente dessas soluções se enquadra na área denominada Mineração de Dados (MD).

As técnicas de MD podem ser utilizadas para diversos propósitos, tais como classificação, regressão, modelagem de dependência, entre outras. Esta seção discutirá sobre o processo de descoberta de conhecimento e aprofundar-se-á na etapa de mineração de dados, na qual destacamos, dentre os métodos existentes, devido sua aplicabilidade e eficiência, as redes bayesianas e os algoritmos genéticos.

A primeira se destaca como uma das melhores técnicas de inteligência computacional, tendo se tornado um dos principais métodos para se trabalhar com incerteza no campo da inteligência artificial (HUANG et al., 2004). Particularmente, esse fato é devido às suas propriedades analíticas excepcionais de representar domínios, correlacionar e estudar as dependências entre suas variáveis, o que permitem mais facilmente visualizar e compreender as relações entre as variáveis do ambiente; consistindo em um fator decisivo e de grande valia

para a representação e análise do domínio pelos usuários. A segunda é uma ferramenta que utiliza o princípio da seleção natural e sobrevivência do mais apto, baseando-se no trabalho de Charles Darwin. Os AGs buscam a melhor solução para um dado problema, percorrendo o espaço de possíveis soluções, utilizando-se das informações do próprio domínio para encontrar soluções cada vez melhores.

A motivação para incorporar alguma inteligência e automação de tarefas na infraestrutura de gerenciamento da rede é direcionada pela necessidade de aumentar o gerenciamento e o tamanho da rede sem aumentar a equipe de operadores; pela necessidade de fornecer uma alta qualidade de serviço para os usuários garantindo que os objetivos do nível de serviço para rede serão mantidos; pela necessidade de correlacionar eventos das camadas inferiores de gerenciamento para determinar estratégias nas superiores (*crosslayer*); pela necessidade de aumentar a sofisticação requerida para diagnóstico, devido à maior complexidade do sistema; e pela necessidade de detectar falhas na rede rapidamente e se possível, a recuperação automática.

3.2 – PROCESSO DE KDD (*KNOWLEDGE DISCOVERY DATABASE*)

O processo de KDD consiste em produzir conhecimento a partir de uma base de dados inserida pelo usuário, a fim de encontrar informações coerentes ao que é desejado. Tem por principal objetivo encontrar padrões válidos e potencialmente úteis, de modo que possam ser utilizados pelos usuários nos níveis decisórios.

O KDD desponta como uma tecnologia capaz de cooperar amplamente na busca do conhecimento embutido nos dados, de modo a apresentarem padrões que destoam, muitas vezes, do restante da massa de dados. O KDD consiste da combinação de métodos e ferramentas de estatística, Inteligência Artificial, visualização e banco de dados para encontrar padrões e regularidades nos dados (MANNILA, 1997).

O uso e aplicação dos termos de KDD e *data mining* é muito semelhante na literatura, sendo muitas vezes utilizados para representar a mesma idéia (SYMEONIDIS; MITKAS, 2006). Basearemos-nos aqui no conceito de *data mining* como etapa componente do processo de KDD.

O KDD, na verdade, é um conjunto de atividades contínuas que geram um conhecimento a respeito da base de dados, sendo dividido em etapas: a de limpeza, a de enriquecimento, a de codificação, a de mineração de dados (*Data Mining*) e a de interpretação do conhecimento descoberto. Um esquema apresentado por Fayyad (1996) das fases que compõem o processo KDD pode ser visto na Figura 3.1.

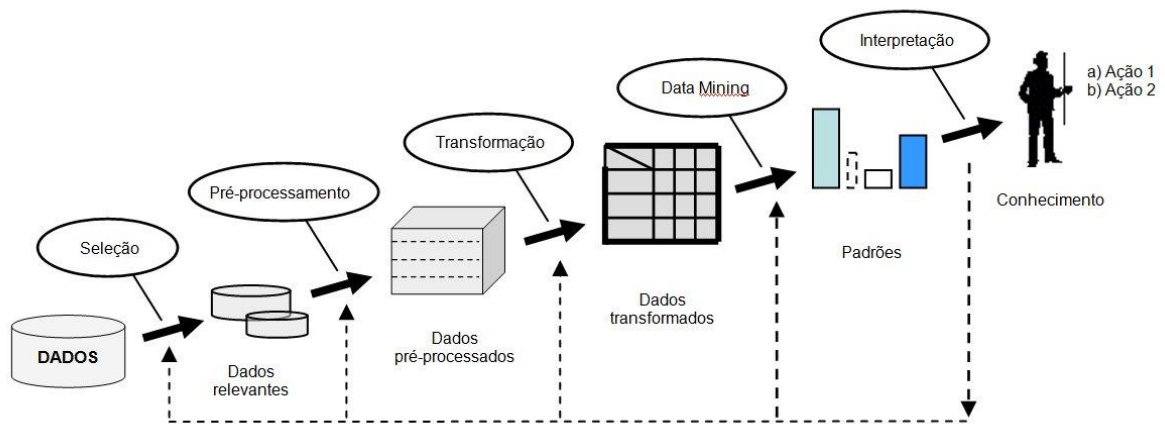


Figura 3.1 - Principais tarefas de mineração de dados (REZENDE, 2003)

A primeira fase, conhecida também como seleção, visa agrupar os dados que são pertinentes ao domínio que está sendo estudado e reduzir o domínio dos dados que serão analisados a uma amostra que represente fielmente a base de dados nos aspectos relevantes à análise desejada, exigindo logo de início uma boa compreensão do domínio por parte do responsável pela aplicação do processo KDD.

O pré-processamento é uma parte crucial no processo de KDD, pois a qualidade dos dados vai determinar a eficiência dos algoritmos de mineração. Nesta etapa deverão ser realizadas tarefas que eliminem dados redundantes e inconsistentes, recuperem dados incompletos e avaliem possíveis dados discrepantes ao conjunto (*outliers*). O auxílio de um especialista do domínio é fundamental, pois na maioria dos casos apenas alguém que realmente entende do assunto é capaz de dizer se um dado é um *outlier* ou um erro de digitação.

Na fase de transformação, a base de dados já pré-processada é formatada de tal forma que a ferramenta/algoritmo de *data mining* possa analisá-lo, extraíndo o conhecimento desejado. Nesta etapa, o conjunto de dados deve ser organizado segundo as requisições e

particularidades da técnica de extração de conhecimento que será utilizada na etapa de *data mining*.

A fase de *data mining* é considerada o núcleo do processo de KDD, em razão de ser nela que são efetivamente aplicados os métodos e algoritmos que possibilitam a extração de conhecimentos a partir da base de dados.

Esta fase é a responsável pela transformação de dados em informações. A mineração de dados está relacionada com a descoberta de novos fatos, regularidades, restrições, padrões e relacionamentos e não apenas consultas complexas e elaboradas com a finalidade de confirmar uma hipótese em função dos relacionamentos existentes.

A mineração de dados surgiu da necessidade do emprego de técnicas e ferramentas que permitissem transformar, de maneira inteligente e automática, os dados disponíveis em informações úteis.

As tarefas básicas, realizadas via processo de MD, são comumente classificadas em duas categorias:

- Descritivas: concentram-se em encontrar padrões que descrevam os dados, caracterizando as propriedades gerais desses dados, de forma interpretável pelos seres humanos;
- Preditivas: realizam inferência nos dados correntes para construir modelos, que serão utilizados para reedições do comportamento de novos dados.

Na Figura 3.2, são apresentadas as principais tarefas de MD e suas classificações quanto às categorias supracitadas.

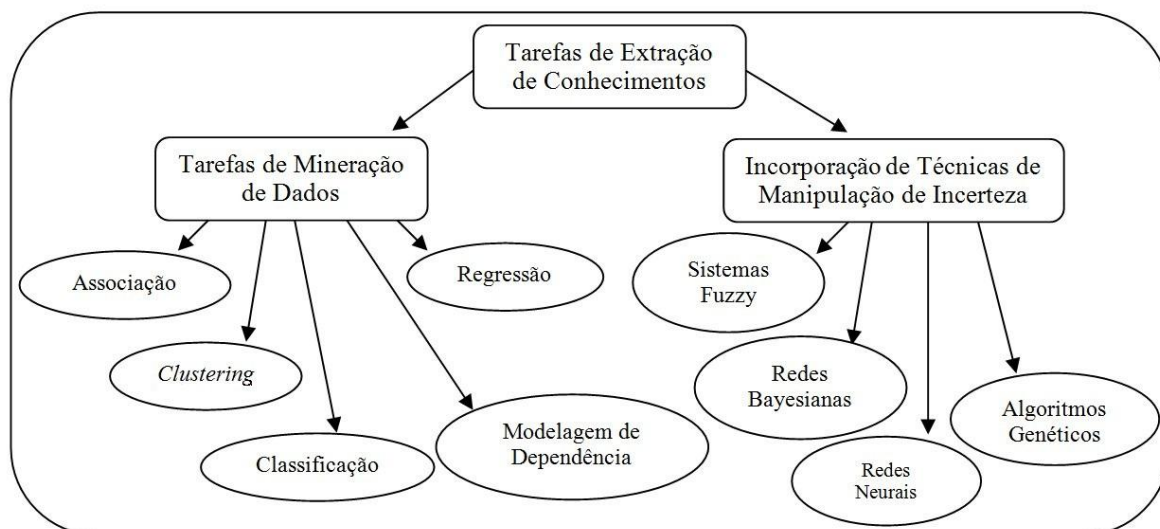


Figura 3.2 - Principais tarefas de mineração de dados (REZENDE, 2003)

A classificação visa o mapeamento de um determinado caso (registro do conjunto de dados) dentro de uma das várias classes pré-definidas (e.g. regras de classificação a respeito de doenças podem ser extraídas de um conjunto de casos conhecidos e usadas para fazer diagnóstico em novos pacientes, baseados em seus sintomas). De modo simplificado, a tarefa de regressão é conceitualmente a mesma da classificação, sendo que na atividade de classificação, a classe é categórica, enquanto na regressão, a classe representa um valor contínuo (e.g. determinar o consumo de energia elétrica de uma determinada unidade consumidora em um mês futuro).

As regras de associação determinam as relações entre os atributos de uma base de dados (e.g. regras de associação podem descrever que itens são comumente comprados juntamente com outros em um supermercado).

Outra tarefa descritiva a ser destacada é a de *clustering*. Também conhecida como análise de agrupamentos, essa tarefa permite criar, a partir de grandes conjuntos de dados, agrupamentos de dados menores e com características semelhantes. Nas tarefas de *clustering*, não há necessidade da determinação prévia de classes (e.g. investigar quais os diferentes grupos de consumidores de drogas, com base em suas características pessoais).

Por fim, a tarefa de modelagem de dependência, na qual estão incluídas as RBs, permite codificar o relacionamento probabilístico entre os atributos de um conjunto de dados.

A partir da definição da tarefa de MD a ser realizada, é possível determinar qual tipo de algoritmo será utilizado para realizar o processo de extração propriamente dito. Os modelos induzidos dos dados seguem geralmente os padrões estatísticos, neurais, simbólicos, de dependências probabilísticas ou baseadas na teoria de algoritmos genéticos.

Um modelo estatístico típico é gerado pelo método de regressão (e.g. regressão linear) e pode ser representado por um sistema de equações. Um modelo neural é representado como uma arquitetura (e.g. rede *feedforward*) de nós e conexões (com pesos) entre eles, além de uma função de aprendizado. Já os modelos simbólicos são geralmente representados por regras do tipo SE-ENTÃO ou árvores de decisão. Enquanto os modelos de dependências típicas, cujos maiores expoentes são redes bayesianas, estabelecem uma estrutura de relacionamento causal entre as variáveis aleatórias de um domínio de aplicação. Por fim, os modelos baseados em algoritmos genéticos têm sido aplicados nas operações de classificação e sumarização.

Além disso, são amplamente utilizadas soluções híbridas, as quais combinam os modelos supracitados com vistas à otimização do processo de extração de conhecimento em termos, por exemplo, do aprendizado, da estimativa de parâmetros e da melhoria da interpretabilidade dos padrões descobertos (GOLDSCHMIDT; PASSOS, 2005).

A escolha do melhor algoritmo para MD é freqüentemente crítica, pois é sabido que nenhum deles tem desempenho ótimo em todos os domínios de aplicação (SALZBERG, 1997). A seleção desses algoritmos é realizada pelo analista e deve ser definida nas restrições do domínio e/ou nas preferências do usuário final (e/ou especialista no domínio). Considerando essas restrições, o analista pode selecionar o algoritmo baseado em alguns parâmetros como, por exemplo, o tipo de aprendizado, linguagens de representação e desempenho.

É importante ressaltar que, além da observação desses parâmetros, as avaliações experimentais desempenham um papel fundamental na seleção de um algoritmo, uma vez que não existem métodos formais para decidir qual o melhor algoritmo para um determinado domínio de aplicação (SALZBERG, 1997).

O processo de KDD não termina quando os padrões nos dados de entrada são extraídos. É preciso realizar ainda uma etapa, comumente chamada pós-processamento ou interpretação, a qual deve permitir que o usuário entenda e possa julgar a utilidade do

conhecimento extraído, contrastando-o com o conhecimento do especialista do domínio. Essa interação pode facilitar a busca das causas de possíveis erros ocorridos ao longo de todo esse processo.

Os conhecimentos extraídos são apresentados ao usuário para interpretá-los, avaliando-os quanto à sua qualidade e/ou utilidade, para que possam ser utilizados na tomada de decisões.

Segundo (TURBAN; ARONSON, 2001), as técnicas de inteligência computacional tratam incerteza como um processo contendo três passos básicos, conforme apresentado na Figura 3.3.

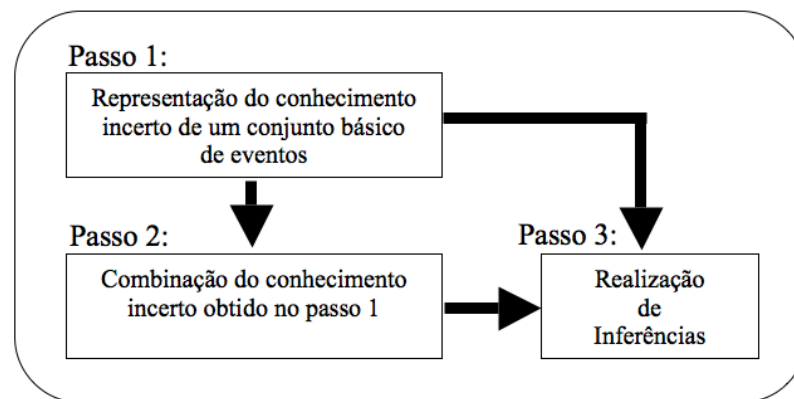


Figura 3.3 - Os três principais passos para a manipulação de incerteza com técnicas de inteligência computacional

No passo 1, o conhecimento inexato é obtido, por meio de interações com o especialista de um domínio específico e/ou induzido a partir de dados, e representado geralmente de modo numérico (e.g. valores de probabilidade), gráfico (e.g. diagrama de influência) ou simbólico (e.g. regras *fuzzy*). O conhecimento básico modelado no passo 1 pode ser diretamente usado para a realização de inferências (passo 3). Entretanto, na maioria dos casos reais, é necessário combinar os padrões obtidos no passo 1 a partir do uso de robustas técnicas de raciocínio incerto, tais como sistemas *fuzzy* e redes bayesianas. Por fim, o passo 3 funciona como uma máquina de inferência para induzir o conhecimento obtido a partir dos passos 1 e 2.

Desse modo, para que o fluxo de execução do processo mostrado na Figura 3.3 possa ser realizado satisfatoriamente, é necessário prover modelos cada vez mais confiáveis e de fácil interpretação do conhecimento extraído, de forma que seja possível, identificar o quanto

esse conhecimento é interessante e relevante para o suporte de seus processos decisórios. Alguns destes modelos serão apresentados de forma mais geral e posteriormente, quando necessário, serão aprofundados.

3.3 – ALGORITMOS DE INTELIGÊNCIA COMPUTACIONAL

O uso de algoritmos de inteligência computacional busca, tipicamente, a imitação em computador de diferentes aspectos da inteligência humana, a qual é obtida por intermédio de diferentes estratégias e metodologias. É a área do conhecimento que lida com o desenho de algoritmos inspirados em fenômenos biológicos e que permitem a resolução de problemas complexos.

O cenário mundial na área de computação aponta, cada vez mais, para o desenvolvimento e a aplicação de sistemas computacionais inteligentes. Esses sistemas utilizam e integram uma série de técnicas da Inteligência Artificial (IA) para reproduzir aspectos do comportamento humano como percepção, raciocínio, adaptação e aprendizado, em um novo paradigma para a IA: a Inteligência Computacional. Essas técnicas abrangem:

- Redes Neurais;
- Algoritmos Genéticos;
- Lógicas Clássica, Não Monotônicas e Nebulosa;
- Redes Bayesianas;

As Redes Neurais Artificiais (RNAs) são dispositivos computacionais não-lineares, inspirados na funcionalidade dos neurônios biológicos. Elas têm atraído uma enorme atenção por possuírem uma variedade crescente de aplicações, como reconhecimento de padrões, otimização, interpretação de imagens, previsões em sistemas complexos, etc. As características de plasticidade e generalização das Redes Neurais capacitam-nas a trabalhar também satisfatoriamente em tarefas como percepção, aprendizado, aquisição de conhecimento e processamento de linguagem natural.

Os algoritmos genéticos são procedimentos de busca baseados na mecânica da genética e da seleção natural, com aplicabilidade em problemas de otimização em geral e, mais particularmente, no aprendizado automático (HOLLAND, 1975). A evolução das

espécies se dá pela seleção dos indivíduos mais aptos ao ambiente em que vivem. Estes têm mais chance de sobreviver e de gerar filhos que herdaram suas características. Esta forma de otimização das espécies através das gerações que os AGs imitam, onde o ambiente corresponde ao problema a ser otimizado e os indivíduos às possíveis soluções do problema.

Os algoritmos genéticos têm sido usados para resolver problemas em várias áreas de estudo, principalmente na área de otimização combinatória, como por exemplo, o problema clássico do caixeiro viajante, alocação de recursos, geração automática de programas e, mais recentemente, treinamento e definição de topologias de redes neurais.

A lógica matemática é responsável por, através de sua linguagem formal, justificar e parametrizar o processo de aprendizado e de raciocínio feitos sobre o conhecimento, pelas técnicas aqui apresentadas. A lógica nebulosa tem a capacidade de modelar o raciocínio incompleto e impreciso, sendo aplicada na solução de problemas de senso comum. Mais especificamente, a lógica *fuzzy* tem encontrado uma gama muito grande de aplicações, como em interpolação espacial, sistemas de apoio a decisão, raciocínio aproximado e aprendizado.

Uma RB é um grafo acíclico dirigido, composto por uma estrutura qualitativa, representando as dependências entre suas variáveis; e quantitativa, avaliando, em termos probabilísticos, essas dependências (CHEN, 2001). Juntos, esses componentes propiciam uma representação eficiente da distribuição de probabilidade conjunta (DPC) de um conjunto de variáveis X_i de um determinado domínio (PEARL, 1988).

No contexto da análise e mineração de dados, abordaremos mais profundamente as RBs e os AG. Inicialmente, as RBs foram elegidas por despontarem como um dos principais métodos para a modelagem da incerteza, permitindo tanto a previsão quanto o diagnóstico de eventos; além de possibilitarem a codificação dos relacionamentos probabilísticos entre as variáveis que definem um determinado domínio (RUSSEL; NORVIG, 2003) e utilizadas para representar processos probabilísticos e causais. E, os AGs, por serem empregados em problemas complexos de otimização fundamentalmente em virtude de fatores como: funcionam tanto com parâmetros contínuos como discretos, ou ainda com uma combinação dos mesmos; realizam buscas simultâneas em várias regiões do espaço de busca, pois trabalham com uma população e não com um único ponto; utilizam informações de custo ou recompensa e não derivadas ou outro conhecimento auxiliar; otimizam parâmetros de funções objetivos com superfícies complexas, reduzindo a incidência de mínimos locais.

Importante ressaltar que outro fator determinante foi à expertise da equipe dos laboratórios associados (LPRAD – Laboratório de Planejamento de Redes de Alto Desempenho, LEA – Laboratório de Eletromagnetismo Aplicado e LAPS – Laboratório de Processamento de Sinais) nas RBs e AGs os quais foram amplamente aplicados em projetos de pesquisa, trabalhos de conclusão de curso, dissertações de mestrado, teses de doutorado e artigos publicados internacionalmente.

3.3.1 – Redes Bayesianas

Como apontado previamente, pode-se entender uma RB como um modelo gráfico que representa as relações de causalidade das variáveis de um determinado problema, ela é composta por:

- Um conjunto de nós que representam as variáveis aleatórias $X = \{X_1, \dots, X_n\}$ do domínio (atributos da base de dados);
- Conjunto de arcos ou setas conectando os nós da rede, definindo assim a relação de dependência condicional e de causalidade;
- Uma tabela de probabilidades condicional (TPC) para cada nó, que quantifica os efeitos que os nós-pais (P_a) exercem sobre esse nó. Os nós-pais de um nó n são todos aqueles que possuem setas apontando para n . Um exemplo de RB é descrito na Figura 3.4.

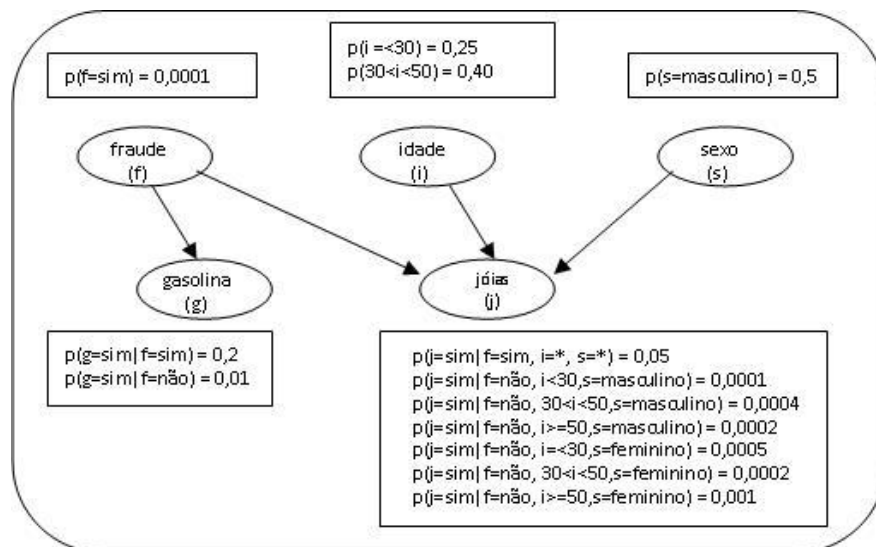


Figura 3.4 - Exemplo de uma rede Bayesiana (HECKERMAN, 1997)

Neste exemplo, os nós fraude, idade, gasolina, jóias e sexo correspondem às variáveis do domínio, dessa forma $X = \{fraude, idade, sexo, gasolina, jóias\}$.

As variáveis fraude, idade e sexo, por serem raiz, possuem apenas as respectivas probabilidades a priori $P(X_i = verdadeiro)$ e $P(X_i = falso)$. Os nós gasolina e jóias por sua vez apresentam distribuições de probabilidade condicional $P(X_i | Pais(X_i))$, que são relações de dependência quantitativa entre os nós da rede.

É possível inferir sobre as dependências condicionais que se estabelecem entre as variáveis através do gráfico da rede, bastando evidenciar a ocorrência de um determinado estado em uma ou mais variáveis da rede, propagando, dessa forma, o efeito das observações pela rede.

Através das inferências, qualquer análise, pergunta ou hipótese que se deseje fazer ao modelo torna-se possível. Os questionamentos levantados inicialmente sobre o domínio podem agora ser estudados, evidenciando os devidos eventos nos atributos da rede e, dessa forma, computando e atualizando as probabilidades das demais variáveis da RB.

Para calcular a inferência, são utilizados os cálculos de probabilidade condicional e probabilidade condicional conjunta (representados a seguir) do teorema proposto por Bayes. A regra de Bayes fundamenta atualmente a base matemática de todos os sistemas que trabalham com inferência probabilística.

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad (3.1)$$

$$P(A | B \cap E) = \frac{P(B | A \cap E)P(A | E)}{P(B | E)} \quad (3.2)$$

Generalizando, tem-se:

$$P(A = v_i | B) = \frac{P(B | A = v_i)P(A = v_i)}{\sum_{k=1}^{n_A} P(B | A = v_k)P(A = v_k)} \quad (3.3)$$

Uma das grandes vantagens das RBs é a semântica deste modelo, a qual facilita, dada a inerente representação causal dessas redes, o entendimento e o processo de tomada de decisão, por parte dos usuários desses modelos (CHEN, 2001). Isto se deve, basicamente, ao

fato de as relações entre as variáveis X_i do domínio poderem ser visualizadas graficamente, além de proverem um mecanismo de inferência que permite quantificar, em termos probabilísticos, a magnitude dos efeitos dessas relações.

Por meio da representação gráfica da RB é possível visualizar e compreender mais facilmente as relações entre as variáveis do ambiente, representando a distribuição de probabilidade conjunta para todas as variáveis. Além disso, a RB também permite que se reduza a complexidade da inferência.

3.3.1.1 - Construção de Redes Bayesianas

Para ilustrar o processo de construção de uma RB, considere o problema bastante simplificado da detecção de fraude em compras com cartões de crédito. Na Figura 3.4, esse exemplo é mostrado, na qual os arcos são desenhados da causa para o efeito, os quadros mostram a distribuição de probabilidade local associada a cada nó da RB, e os asteriscos representam a atribuição de quaisquer valores, dentre os possíveis, às variáveis. Desse modo, a estrutura da RB pode ser obtida seguindo os passos:

1. Selecionar o conjunto de variáveis que descrevem o domínio. Uma possível escolha para o problema de detecção de fraude seria: fraude(f), gasolina(g), jóias(j), idade(i) e sexo(s), representando se a compra é ou não fraudulenta, se houve ou não compra de gasolina nas últimas 24 horas, se houve ou não compra de jóias nas últimas 24 horas, o sexo e a idade do usuário do cartão de crédito, respectivamente.
2. Eleger uma ordem para essas variáveis. Uma possível ordem das variáveis para o problema de detecção de fraude seria: f, i, s, g e j.
3. Enquanto houver variáveis a serem consideradas:
 - a. Adicionar um nó à rede para cada variável.
 - b. Estabelecer um conjunto de nós-pais de X_i sobre o conjunto mínimo de nós já incluídos na rede, que satisfazem a propriedade de independência condicional. No exemplo, utilizando a ordem f, i, s, g e j, sugerida no passo 2, obtém-se:

$$P(i|f)=P(i)$$

$$P(s|f,i)=P(s)$$

$$P(f|i,s)=P(f)$$

$$P(f|i,s)=P(f)$$

- c. Definir a TPC para X_i .

Vale ressaltar que, apesar de serem mostrados como uma seqüência simples, na prática, esses passos são geralmente interligados e complexos. Por exemplo, julgamentos de independência condicional e/ou causa-efeito podem influenciar na construção da RB.

Além disso, esse processo de construção considera apenas o conhecimento prévio (ou de fundo) do domínio. Em outras palavras, a rede deve ser concebida por intermédio da interação com o especialista do domínio. Essa interação tem a finalidade de identificar os relacionamentos entre as variáveis de interesse, para em seguida codificá-los na rede. Por exemplo, o especialista no problema da detecção de fraude pode identificar a influência direta que a idade exerce sobre a aquisição de jóias e até mesmo quantificar essa dependência (através dos valores de probabilidades condicionais). Na próxima seção, é mostrado o aprendizado da estrutura e dos parâmetros (probabilidades) de uma RB a partir de dados.

3.3.1.3 - Aprendizado de Redes Bayesianas

A aprendizagem de RBs consiste em induzir, a partir de um conjunto de dados, as distribuições de probabilidades condicionais e identificar as relações de independência existentes nesse conjunto. Esse processo de aprendizagem considera dois aspectos: aprendizagem da estrutura, quando não se tem a priori definido pelo especialista do domínio, tal estrutura; e a aprendizagem dos parâmetros, após a obtenção da estrutura, por intermédio das interações com o especialista ou induzida a partir dos dados.

Para um melhor entendimento desse processo, considere o exemplo de detecção de fraude em compras com cartão de crédito mostrado na Figura 3.4. Primeiramente, a RB que representa as relações entre as variáveis desse problema é especificada, por exemplo, pelo usuário. Em seguida, é necessário que seja especificado como a distribuição de probabilidade de cada nó será representada. No caso do problema da detecção de fraude, as variáveis foram discretizadas em um número de estados (valores) para que cada distribuição de probabilidade

possa ser representada em uma tabela (TPC), por exemplo, idade foi discretizada dentro dos valores (≤ 30 , $30 - 50$, ≥ 50).

Finalmente, o algoritmo tenta estimar as probabilidades (parâmetros) da TPC baseado no conjunto de dados de treinamento. Por exemplo, a célula $P(i \leq 30)$ da TPC da variável i pode ser simplesmente calculada a partir do número de clientes do conjunto de dados de treinamento que tenham idade igual ou inferior a 30 anos. O parâmetro $P(j=sim/f=sim, i \leq 30, s=feminino)$ pode ser computado por intermédio da fração dos exemplos de treinamento onde a compra seja uma fraude, a idade do cliente seja igual ou inferior a 30 anos e o sexo seja feminino e que tenha comprado jóias nas últimas 24h.

Um ponto que merece ser destacado no processo de aprendizado de RBs é se todos os valores das variáveis são observados (não há valores de atributos ausentes) no conjunto de dados de treinamento ou se algumas não são consideradas. Assim, os métodos de aprendizagem devem considerar as seguintes situações:

- Estrutura da rede conhecida e conjunto de dados completos;
- Estrutura conhecida e conjunto de dados incompletos;
- Estrutura desconhecida e conjunto de dados completos;
- Estrutura desconhecida e conjunto de dados incompletos.

3.3.1.5 - Aprendizado da Estrutura de Redes Bayesianas

O aprendizado da estrutura de RBs a partir de dados completos pode ser realizado, via de regra, considerando dois paradigmas: busca e pontuação e baseado em independência condicional (CHENG et al., 1997).

No paradigma de busca e pontuação, a aprendizagem se realiza buscando uma estrutura que seja aderente aos dados. Em geral se inicia com um grafo sem arcos, então, usa-se algum método de busca gulosa que adicione um arco ao grafo. O passo seguinte consiste em usar uma função de pontuação para determinar se a nova estrutura é melhor que a anterior. Caso seja melhor, o novo arco é mantido. Esse processo continua até que nenhuma nova estrutura seja melhor que as anteriores.

Diferentes critérios de pontuação estão disponíveis na literatura para avaliar uma estrutura, tais como os descritos em (COOPER; HERSKOVITS, 1992), (LAM; BACCHUS, 1994) (HECKERMAN et al., 1995). Já o processo de busca por novas estruturas é realizado via métodos heurísticos. Para reduzir o espaço de busca, a maioria dos algoritmos requer ordenamento a priori dos nós.

Os algoritmos que utilizam o paradigma de análise de independência condicional procuram descobrir as dependências a partir dos dados, e então usam essas dependências para inferir a estrutura. As relações de dependência são avaliadas pelo uso de alguma classe de teste de independência condicional. Detalhes mais aprofundados destes tipos de algoritmos podem ser encontrados nos algoritmos descritos em (SPIRITES et al., 2001) e (CHENG et al., 1997).

Como forma de compreender os processos de aprendizado da estrutura de uma RB e em razão de ser mais comumente utilizado, serão destacados a seguir os principais básicos do paradigma de busca e pontuação.

Para isso, é necessário entender que o aprendizado da estrutura de RBs pode ser visto como processo de busca por uma estrutura que codifique a TPC para um conjunto de variáveis aleatórias X , dado um conjunto de dados D . Isto é, encontrar uma estrutura, em um possível espaço de hipóteses de estruturas S^h , avaliando as probabilidades a priori das hipóteses S^h - $P(S^h)$. Então, dado um conjunto de dados D , calcular as probabilidades condicionais que maximizam $P(S^h/D)$ (estrutura) e $P(\theta_{ijk}/D, S^h)$ (probabilidade condicionais das TPCs dos nós da estrutura da RB S^h).

O cálculo de $P(\theta_{ijk}/D, S^h)$ pode ser realizado conforme descrito na subseção anterior. O cálculo de $P(S^h/D)$ pode ser feito, via Regra de Bayes, da seguinte maneira:

$$P(S^h/D) = \frac{P(S^h)P(D/S^h)}{P(D)} \quad (3.4)$$

Sendo que $P(D)$ é independente da estrutura S^h . Dessa forma, para determinar a distribuição de probabilidade condicional das possíveis estruturas, é necessário calcular a verossimilhança marginal dos dados ($P(D/S^h)$) para cada estrutura (considerando a probabilidade a priori $P(S^h)$ de cada uma das estruturas) e então determinar a estrutura que codifica a TPC para X baseado nos valores máximos de verossimilhança. Este método, conhecido como abordagem Bayesiana completa (por considerar todas as possíveis estruturas de RB do espaço de hipóteses), é freqüentemente inviável (HECKERMAN, 1997). Por

exemplo, para se ter uma idéia, o espaço de hipóteses das estruturas das RBs para um conjunto formado por apenas três variáveis A , B e C é formado por 25 diferentes estruturas. Na Figura 3.5 são mostradas algumas dessas estruturas.

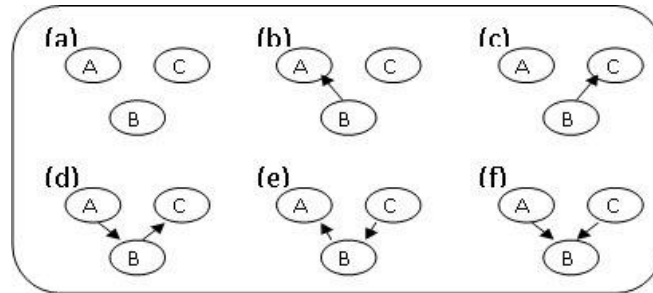


Figura 3.5 - Algumas estruturas possíveis de redes bayesianas para as variáveis A , B e C .

Em problemas com um grande número de variáveis a serem consideradas é necessário o uso de outras abordagens. Entre as mais referenciadas na literatura e mais utilizadas pelos métodos de aprendizado de RBs, destaca-se a abordagem de Seleção do Modelo.

A tarefa básica da Seleção do Modelo é selecionar um modelo “bom” (i.e. uma “boa” hipótese de estrutura) a partir dos modelos possíveis e considerá-lo como se fosse o modelo “correto”. Esta abordagem pode suscitar várias questões:

- É possível obter resultados precisos na seleção do modelo com essa abordagem?
- Como decidir se o modelo é “bom” ou não?
- Como encontrar modelos “bons”?

A questão da precisão é difícil de ser respondida. No entanto, várias pesquisas mostram experimentalmente que a seleção de um único modelo “bom” frequentemente produz resultados bastante satisfatórios (HECKERMAN et al., 1995).

Quanto à segunda questão, é necessário estabelecer algum critério para medir o grau em que uma determinada estrutura de RB combina adequadamente o conhecimento de fundo (estrutura da rede fornecida pelo especialista) com o conhecimento embutido nos dados. Um modelo bastante utilizado é o da probabilidade condicional relativa, calculada do seguinte modo:

$$L(S^h) = \log P(S^h) + \log P(D/S^h) \quad (3.5)$$

Como pode ser observado na equação 3.5, esse critério possui dois componentes: o *log a priori* ($\log P(S^h)$) e o *log de verossimilhança marginal* ($\log P(D/S^h)$).

Vários métodos podem ser encontrados na literatura para o cálculo de $\log P(S^h)$. O caso mais simples considera qualquer hipótese de S^h igualmente comum (não considerando, portanto, o valor de $\log P(S^h)$ para o cálculo de $\log P(D, S^h)$). Um outro método, proposto por (HECKERMAN et al., 1995), usa uma estrutura de rede dada a priori (conhecimento de fundo). A idéia básica deste método é penalizar a probabilidade ($P(S^h)$) de cada estrutura de acordo com alguma medida de desvio entre uma determinada estrutura e a rede considerada a priori. A métrica MDL (Descrição de Comprimento Mínimo), proposta por Lam e Bacchus (1994), faz balanceamento entre o ajuste aos dados e a complexidade do modelo. A adição de uma variável *pai* causa aumento do *log de verossimilhança*, bem como da penalização. Haverá adição de um arco caso o incremento na verossimilhança seja relevante.

Segundo Dawid (1984), o cálculo do $\log P(D/S^h)$ poderia ser feito utilizando a equação 3.6:

$$\log P(D/S^h) = \sum_{e=1}^n \log P(x_e/x_1, \dots, x_{e-1}, S^h) \quad (3.6)$$

O termo $P(x_e/x_1, \dots, x_{e-1}, S^h)$ é a predição para x_e feita para o modelo S^h . O modelo com maior *log de verossimilhança marginal* ($\log P(D/S^h)$), considerando as probabilidades *a priori* das estruturas iguais, representa um “bom” modelo de estrutura para o conjunto de dados D .

Finalmente, com relação à terceira pergunta, várias pesquisas têm sido direcionadas para o uso de algoritmos de busca heurística, a fim de encontrar uma boa estrutura a partir do espaço de hipótese de todas as possíveis estruturas de RBs. Um exemplo desses algoritmos é o de *busca gulosa*, que inicia com a escolha de uma estrutura qualquer de RB. Então, avalia o *log de verossimilhança marginal* (através, por exemplo, da equação 3.6) para todas as possíveis mudanças m (e.g. adicionar ou retirar um arco da estrutura). Em seguida, realiza a mudança m cujo valor do *log de verossimilhança marginal* seja máximo. A busca é concluída quando não existe m que proporcione um valor maior para o *log de verossimilhança marginal*. Este algoritmo é utilizado por vários métodos de aprendizado de RBs como, por exemplo, o K2 (COOPER; HERSKOVITS, 1992) e *Bayesian Knowledge Discovery* (RAMONI; SEBASTIANI, 1997).

Para ilustrar o aprendizado da estrutura de redes Bayesianas, considere o exemplo do K2. Este método aprende a estrutura de uma RB a partir de um conjunto de dados completo, cuja ordem das variáveis deve ser fornecida pelo usuário desse método. O K2 adota um método para calcular (selecionar) a estrutura com máxima verossimilhança e um algoritmo de

aprendizado para encontrá-la (aproximadamente). Essa busca é iterativa e inicia com uma estrutura de rede bem simples, onde todas as variáveis são independentes uma das outras. Em seguida são avaliadas as verossimilhanças marginais de cada rede resultante de possíveis mudanças (e.g. adicionar uma nova ligação entre dois nós dessa rede), e então é aplicada a melhor das alterações antes de uma nova iteração. Esse processo continua até que o algoritmo não consiga encontrar nenhuma alteração simples que melhore (aumente) o valor de máxima verossimilhança.

Esse algoritmo, classificado como de busca e pontuação e um dos mais utilizados, permite encontrar a mais provável estrutura de rede de crença S a partir de um determinado conjunto de dados D . O algoritmo K2 aplica a pontuação Bayesiana segundo a equação 3.7.

$$P(S | D) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(r_i + N_{ij} - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (3.7)$$

Na qual:

n é o número de nós;

q_i é o número de configurações dos pais da variável X_i ;

r_i é o número de possíveis valores do nó X_i ;

N_{ijk} é o número de casos em D onde o atributo X_i é instanciado com o seu valor k , e a configuração dos pais de X_i é instanciada com o valor j ;

N_{ij} denota o número de observações em que a configuração dos pais de X_i é instanciada com o valor j , sendo $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$.

Após a construção da rede (a partir do conhecimento fundo do domínio, dos dados ou da combinação de ambos), são necessários mecanismos de inferência para computar as eventuais probabilidades de interesse.

3.3.2 – Algoritmos Genéticos

Um Algoritmo Genético é um método de otimização e busca inspirado em mecanismos de populações de seres vivos foi introduzido por John Holland, em 1975, e popularizado por um de seus alunos, David Goldberg, em 1989 (GOLDBERB, 1989). Esses

algoritmos destacam-se por possuírem amplos espaços de busca e por serem do nível de complexidade de problemas NP-Completo.

Os AGs seguem o princípio da seleção natural e sobrevivência do mais apto e se baseiam no trabalho de Charles Darwin. Os AGs buscam a melhor solução para um dado problema, tentando várias soluções e utilizando-se das informações do próprio domínio para encontrar soluções cada vez melhores.

Os AGs têm sido empregados em problemas complexos de otimização fundamentalmente em virtude dos seguintes fatores:

- Funcionam tanto com parâmetros contínuos como discretos, ou ainda com uma combinação dos mesmos;
- Realizam buscas simultâneas em várias regiões do espaço de busca, pois trabalham com uma população e não com um único ponto;
- Utilizam informações de custo ou recompensa e não derivadas ou outro conhecimento auxiliar;
- Otimizam parâmetros de funções objetivos com superfícies complexas, reduzindo a incidência de mínimos locais;
- Trabalham com uma codificação do conjunto de parâmetros e não com os próprios parâmetros;
- Fornecem uma lista de parâmetros ótimos e não uma simples solução;
- São fáceis de serem implementados em soluções computacionais;
- São modulares e portáteis, no sentido que o mecanismo de evolução é separado da representação particular do problema considerado. Assim, eles podem ser transferidos de um problema para outro;
- São também facilmente combinados com outras técnicas heurísticas.

As técnicas de AG resolvem problemas de otimização e de busca, onde os métodos tradicionais falham. As técnicas tradicionais, frequentemente, iniciam o processamento com um único candidato (indivíduo) manipulado, utilizando alguma heurística, na maioria das

vezes estática, diretamente ligadas ao problema a ser solucionado. No caso de AG, são realizadas operações em paralelo sobre uma população de candidatos (vários indivíduos) e a busca é feita, em diferentes áreas do espaço de solução, selecionando um número apropriado de membros para a busca, em várias regiões.

A idéia básica do funcionamento de um AG pode ser resumida nos seguintes passos:

1. Gerar a população inicial de cromossomos.
2. Avaliar cada cromossomo e lhe atribuir uma nota (aptidão).
3. Selecionar os melhores indivíduos (cromossomos mais aptos).
4. Aplicar operações de *crossover* (combinação) e mutação, gerando uma nova população (descendentes).
5. Repetir dos itens 2 a 4 até encontrar uma solução satisfatória.

Conforme apresentado, a base do funcionamento de um AG são os operadores genéticos, utilizados para gerar novas soluções (gerações da população) em analogia à evolução natural. Essas operações bem como a terminologia empregada nesse método de otimização, serão descritas nas subseções a seguir.

3.3.2.1 - Terminologia

Um Algoritmo Genético representa uma metáfora da teoria da evolução, permitindo o uso de termos próprios da biologia. Desta forma, para uma melhor compreensão das relações destes termos com a teoria dos AGs, os mesmos são descritos a seguir:

- Cromossomo e Genoma: genoma é o conjunto completo de genes de um organismo. Um genoma pode ter vários cromossomos. Ambos representam a estrutura de dados para codificar uma solução do problema proposto. Um cromossomo ou genoma representa um ponto no espaço de busca;
- Gene: unidade de hereditariedade transmitida pelo cromossomo, o qual controla as características do organismo. Nos AGs, é um parâmetro codificado no cromossomo, um elemento do vetor cromossomo;

- **Indivíduo:** um simples membro da população. É formado pelo cromossomo e sua aptidão;
- **Genótipo:** é a composição contida no genoma. Na Computação Evolucionária, representa a informação contida no cromossomo ou genoma;
- **Fenótipo:** representa o objeto, estrutura ou organismo construído pelas informações contidas no genótipo (cromossomo decodificado);
- **Alelo:** é uma das formas alternativas de um gene. Em AG, representa os valores do parâmetro codificado em um gene.

3.3.1.2 – Representação

A representação de um cromossomo pode conter várias formas, sendo a mais usual a representação binária, na qual as informações são codificadas em grandes cadeias de bits, como forma de utilizar os operadores de *crossover* e mutação tradicionais. Entretanto, é possível empregar outras formas de representação, dependendo do tipo de aplicação desenvolvida.

Vários pesquisadores têm discutido qual a melhor representação e muitos deles têm mostrado experimentos favoráveis à representação real (MICHALEWICZ, 1994), principalmente por ser mais facilmente compreendida pelo ser humano. Todavia, a representação binária é mais simples de ser implementada, além de ser amplamente utilizada para a representação dos cromossomos (possíveis soluções de um problema). A escolha do tipo de representação é pautada, via de regra, pelo tipo de problema que está sendo modelado.

3.3.1.3 - População Inicial

Os parâmetros de entrada de um AG incluem uma população inicial de cromossomos. Estes representam as estruturas de dados usados para representar uma possível solução do problema proposto. Cada cromossomo se encontra no espaço de busca de soluções, o qual representa o conjunto de todas as configurações que um cromossomo pode assumir. A cada cromossomo é atribuída uma aptidão, ou seja, uma nota ou valor para medir a qualidade da solução codificada.

Quando não existe um conhecimento a priori do espaço de busca, usa-se uma população inicial aleatória, porém, uma população inicial pequena pode não representar todas as regiões do espaço de busca. As possíveis soluções para este problema incluem (GOLDBERG, 1989):

- Gerar uma população inicial uniforme (com pontos espaçados igualmente entre si), distribuída por todo o espaço de busca;
- Gerar metade da população aleatoriamente e a outra metade invertendo os bits da primeira;
- Gerar uma população inicial maior que as subseqüentes, melhorando a representação do espaço de busca;
- Utilizar a técnica denominada *seeding*, que consiste em colocar na população inicial, soluções encontradas por outros métodos de otimização, garantindo soluções melhores para o AG que as geradas por esses outros métodos.

3.3.1.4 - Avaliação e Seleção

Após a criação da população inicial, o AG seleciona os cromossomos com melhor aptidão. Os cromossomos mais aptos são os escolhidos para gerar os cromossomos filhos e são denominados de população intermediária. Indivíduos mais aptos têm mais oportunidades de serem reproduzidos (produzindo descendentes cada vez mais aptos).

A determinação da aptidão varia de acordo com o problema em questão. Por exemplo, caso se deseje determinar o máximo global de uma função complexa com vários pontos de máximo e mínimo, pode-se utilizar o valor da função objetivo, interpolando estes valores para um intervalo determinado.

Os cromossomos podem ser escolhidos com probabilidade proporcional a sua aptidão, implicando na possibilidade de existência de indivíduos (cromossomos) duplicados na população intermediária. Esse procedimento, conhecido como “Método da Roleta”, (ou "Roda da Roleta"), funciona do seguinte modo: com os cromossomos ordenados por aptidão, calculam-se as aptidões acumuladas, gerando um número aleatório r entre 0 (zero) e a soma total das aptidões. O cromossomo escolhido será o primeiro que possuir aptidão acumulada

maior que r, como pode ser visto no exemplo ilustrado na Tabela 3.1. Repete-se o processo até ser atingido o número de cromossomos desejados na população intermediária.

Tabela 3.1 – Ordenação da aptidão interpolada para o intervalo [0,00; 2,00].

Posição	Aptidão	Aptidão Acumulada
1	2,00	2,00
2	1,50	3,50
4	1,00	4,50
5	0,50	5,00
6	0,00	5,00

A utilização deste método de seleção não trabalha com números negativos, sendo necessário o escalonamento dos valores das aptidões em um intervalo positivo, para tal utiliza-se a equação 3.8.

$$f_i = Min + (Max - Min) * \frac{(N - i)}{(n - 1)} \quad (3.8)$$

Ao serem plotados em um gráfico, os dados da equação 3.8 se apresentam como mostrados na Figura 3.6. Por este método, a probabilidade de escolha de cada cromossomo pode ser calculada pelo intervalo da aptidão acumulada. Enquanto o cromossomo 1 será escolhido para qualquer número r entre [0; 2[, o cromossomo 2 só será selecionado para valores de r entre [2; 3,50[, ou seja, dividindo-se o intervalo de 2 do cromossomo 1 pelo total de 5 obtemos 0,4, enquanto o cromossomo 2, obtém probabilidade de $1,5/5 = 0,3$, e assim sucessivamente até o último cromossomo.

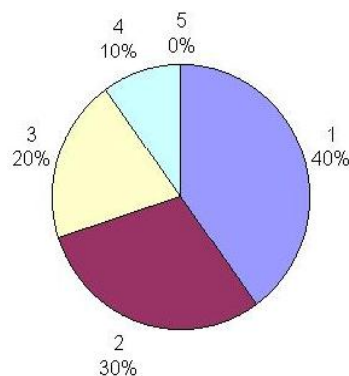


Figura 3.6 - Distribuição das probabilidades da aptidão dos indivíduos.

Na Figura 3.6, pode-se observar o fator de probabilidade agindo sobre a escolha dos indivíduos da população, enquanto o cromossomo 1 tem 40% de chance de ser escolhido, por

possuir um intervalo maior, o cromossomo de posição 5 nunca será escolhido, pois se $4,50 < r \leq 5,00$ será escolhido o cromossomo 4 e não o cromossomo 5.

Deve-se notar que o procedimento tende a causar convergência prematura do AG, pois a população intermediária pode ficar saturada de elementos com alta aptidão, mas não com a melhor aptidão, resultando em soluções não tão boas quanto o esperado.

O problema de convergência prematura pode ser combatido utilizando, principalmente, as seguintes estratégias:

- Aumentar da taxa de mutação. A diversidade é proporcional à taxa de mutação, porém, a mutação é conhecida por destruir informação genética dos cromossomos, por isso é necessário cautela quando do ajuste da taxa de mutação;
- Evitar cromossomos duplicados na população;
- Controlar o número de filhos do super-indivíduo (indivíduo com alta aptidão, mas não com aptidão ótima) usando, por exemplo, técnicas como seleção por torneio.

A Seleção por Torneio evita o problema de convergência prematura. No torneio, escolhe-se aleatoriamente e com probabilidades iguais, n cromossomos e o de maior aptidão entre eles é selecionado para a população intermediária. O processo é repetido até se preencher a população intermediária. Na seleção por torneio, não existe a necessidade de escalonamento da aptidão ou seu ordenamento, o que acarreta uma economia de recursos e tempo de processamento.

3.3.1.5 - Operador de *Crossover*

O operador genético *crossover* é responsável pela geração de novas populações. Esse operador é aplicado a um par de cromossomos da população intermediária para gerar descendentes, o que permite a exploração do espaço de busca das soluções do problema tratado.

A idéia do *crossover* é mesclar em um novo cromossomo as melhores qualidades de dois indivíduos. Levando em consideração que as melhores características dos indivíduos não são conhecidas, o melhor a fazer é combinar as características aleatoriamente. O AG trata estas características como blocos de construção e tenta combiná-los em melhores indivíduos

via *crossover*. Algumas vezes, o *crossover* combina as piores qualidades dos indivíduos e, neste caso, seus descendentes não sobreviverão muito tempo na população. O melhor caso, no entanto, consiste em combinar as melhores qualidades dos pais, gerando um filho com aptidão ainda melhor que a de seus progenitores, desde que essas características sejam compatíveis entre si.

O operador de *crossover* divide as cadeias de bits dos cromossomos pais em uma posição aleatória, produzindo um par de “cabeças” e outro de “caudas”, a permuta de “caudas” dessas cadeias de bits entre os cromossomos pais dá origem aos cromossomos filhos, como pode ser visto na Figura 3.7.

pai1	101100011101011	111000110110110
pai2	001001101100101	001110010110011
filho1	101100011101011	001110010110011
filho2	001001101100101	111000110110110

Figura 3.7 - Exemplo da aplicação do operador de *crossover*.

O operador de *crossover* não ocorre em todos os pares selecionados. A probabilidade de aplicação do operador de *crossover* pode variar de acordo com uma taxa, conhecida como taxa de *crossover*, geralmente entre 60% e 99%. A não ocorrência de *crossover* implica na cópia dos cromossomos pais para a próxima geração, preservando algumas soluções da geração anterior. Uma implementação possível para a taxa de *crossover* poderia usar números pseudo-aleatórios no intervalo $[0, 1]$, aplicando o *crossover* caso o número gerado seja menor que a taxa de *crossover*.

Os tipos do operador de *crossover* são determinados pelo número de pontos de corte utilizados, sendo os mais conhecidos: *crossover* de um-ponto, *crossover* multiponto e *crossover* uniforme. O *crossover* de um ponto divide as cadeias de bits dos cromossomos pais em um ponto escolhido aleatoriamente. O *crossover* de n pontos escolhe n pontos aleatórios nas cadeias de bits dos pais, onde n é um número par, e as duas seções entre estes dois pontos são permutadas. O *crossover* multiponto mais utilizado tem sido o de dois pontos, pois estes tendem a preservar esquemas de maior comprimento. Um exemplo do *crossover* de n pontos pode ser visto na Figura 3.8.

pai1	1011000111	0101111100	0110110110
pai2	0010011011	0010100111	0010110011
filho1	1011000111	0010100111	0110110110
filho2	0010011011	0101111100	0010110011

Figura 3.8 - Exemplo de aplicação do crossover de dois pontos.

O *crossover* uniforme utiliza uma máscara de bits aleatórios gerada para cada par de cromossomos escolhidos para gerar novos filhos. Este processo pode ser implementado da seguinte forma: se o bit da máscara possui valor 1, o bit correspondente do pai1 é copiado para o filho1, caso contrário, o bit do pai2 é copiado para o bit do filho1. Este processo se repete por toda a cadeia de bits da máscara. Para a geração do filho2, este processo é invertido. Na Figura 3.9, é ilustrado o uso do *crossover* uniforme.

Máscara de bits	0110110110	1011000111	0010110011
pai1	1011000111	0101111100	0110110110
filho1	0010001111	0001100100	0010110011
pai2	0010011011	0010100111	0010110011

Figura 3.9 - Exemplo de aplicação do Crossover Uniforme.

Enquanto o *crossover* de n pontos gera um filho com metade dos bits de cada pai, o *crossover* uniforme gerará um filho, no qual seus genes terão um número aleatório de bits do pai1 e do pai2, devido à utilização da máscara aleatória de bits.

Vários estudos têm investigado a diferença de desempenho para os diversos tipos de *crossover* (ESHELMAN et al., 1989); (BEASLEY et al., 1993) e (GREFENSTETTE, 1986). O que se pode deduzir é que não há grandes diferenças, pois a robustez do AG não apresenta alteração significativa em seu desempenho em uma faixa relativamente larga de variação de parâmetros (taxa de crossover e mutação, tamanho da população, etc).

3.3.1.6 - Operador de Mutação

O operador de mutação é utilizado após o *crossover* e é considerado o melhor mecanismo para produzir variações nas características dos indivíduos da população. A mutação é aplicada, com certa probabilidade, a cada bit dos cromossomos filhos, invertendo o valor do bit sobre o qual é aplicado, criando novas tentativas de solução com pequenas mudanças aleatórias na representação anterior. Estas perturbações, nas cadeias dos cromossomos, darão origem a uma nova cadeia, evitando que a busca fique estagnada em sub-regiões do espaço de busca, possibilitando, também, que qualquer ponto do espaço seja atingido (REZENDE, 2003).

Este operador contribui para aumentar a diversidade de cromossomos na população, entretanto, destrói informação contida no cromossomo. Dessa forma, deve-se utilizar uma taxa de mutação pequena, mas suficiente para assegurar a diversidade, normalmente entre 0,1% e 5%. Na Figura 3.10, é apresentado um exemplo do uso deste operador.

Filho1	101100 <u>0</u> 1110101100 <u>1</u> 110010110011
Filho2	001001101100101111000110 <u>1</u> 10110
Filho1	10110 <u>1</u> 01110101100 <u>0</u> 110010110011
Filho2	001001101100101111000110 <u>0</u> 10110

Figura 3.10 - Exemplo da aplicação do operador de mutação.

Após a aplicação dos operadores de *crossover* e mutação, os cromossomos filhos são então avaliados e o processo se repete em ciclos. Estes ciclos são chamados de gerações. O número de repetições (ou gerações) pode ser pré-determinado; indicado pela ocorrência do caso ideal (critério de parada), se este for conhecido; ou na convergência do algoritmo, ocorrendo quando as aptidões dos indivíduos de uma população se tornam muito parecidas.

A substituição dos cromossomos da população atual pode ocorrer de forma geracional ou *steady-state*. A primeira substitui toda a população, criando N filhos para substituir N pais, enquanto que a última cria um conjunto com os melhores pais e filhos para a próxima geração; neste tipo de substituição são criados um ou dois filhos em cada geração para

substituir os piores cromossomos da população. Alternativamente podem-se substituir os pais ou os indivíduos mais velhos, pois estes já transmitiram seus genes à população.

3.4 - CONSIDERAÇÕES FINAIS

Neste capítulo, foram descritos os fundamentos básicos do processo de KDD que dão suporte aos processos de decisão, com destaque para os requisitos e as ações comumente realizadas nas etapas que compõe esse processo.

Para realização da extração de padrões em MD, é muito comum o uso de técnicas para raciocínio incerto com o objetivo de criar modelos do conhecimento embutido nos dados, em domínios reais com alguma inexatidão em suas premissas e, assim, dar suporte ao processo de decisão relacionado às ações futuras sobre esse domínio. Apesar das técnicas de manipulação de incerteza não estarem exclusivamente associadas à resolução de problemas de mineração de dados, são mecanismos eficientes e amplamente utilizados para análise de dados.

Desta forma, duas de suas principais técnicas foram apresentadas, focando nos principais conceitos e definições referentes ao modelo de inteligência computacional conhecido como redes bayesianas e os algoritmos genéticos. Foram expostas suas principais características e definições, que servirão como base para as análises apresentadas nas seções seguintes.

4. TRABALHOS CORRELATOS

4.1 - CONSIDERAÇÕES INICIAIS

Neste capítulo serão apresentados alguns trabalhos levantados na literatura especializada que serviram de base no estudo da proposta deste trabalho. As referências são apresentadas de modo a situar o estado-da-arte das áreas aqui envolvidas. São apontadas também algumas limitações destes estudos e os aspectos que são desenvolvidos neste trabalho, definindo de forma clara onde se encontra a contribuição desta tese para a área.

Os trabalhos apontados neste capítulo estão divididos de acordo com os fundamentos das abordagens apresentadas: implementação de metodologias de testes, aferições e avaliação de tecnologias de acesso, principalmente ADSL2+, voltados a implementação de qualidade de serviço. E por fim, técnicas de *crosslayer* visando a otimizar o desempenho de aplicações *Triple Play* sob tecnologias de acesso.

4.2 – TRABALHOS CORRELATOS

A implementação de QoS em redes cabeadas ou não cabeadas tem sido um dos grandes desafios dos últimos anos, principalmente com a popularização das redes de banda larga, que possibilitaram maiores larguras de banda e com isso, maiores demandas dos cliente por serviços.

A nova demanda de serviços em tempo-real com qualidade tem influenciado no desenvolvimento de pesquisas que implementem tais necessidades. Para isso, técnicas de qualidade de serviço têm sido utilizadas, e não somente elas, diversas outras técnicas são aplicadas para chegar a este objetivo.

Dentre as abordagens existentes, a implementação de metodologias de testes possibilita um maior conhecimento das funcionalidades e da estabilidade do sistema, permitindo assim uma otimização dos recursos, tanto em operação normal, como em condições sobrecarregadas, de forma a garantir um bom desempenho mesmo em situações extremas. Os testes de desempenho não podem, nem devem ser realizados a esmo, correndo o risco de gerar resultados que não representam totalmente o sistema ou, no pior dos casos,

inválidos e incoerentes. Estes devem ser realizados conforme especificações técnicas e para isso, estratégias devem ser implementadas envolvendo os principais elementos (e.g. hardware, software, protocolo, etc), de modo a avaliar o sistema de forma eficiente.

Em (SANDRI et al., 2007), foi proposta uma metodologia para avaliação de serviços de banda larga ADSL. Esta estratégia considera diferentes variações de testes, possibilitando avaliar diversas funcionalidades dos equipamentos e serviços DSL, como funcionalidades, desempenho, conformidade, etc. Foi implementado um *testbed* típico e elaborados testes de carga dos equipamentos, serviços oferecidos e de capacidade da rede, bem como definido que variáveis e parâmetros deveriam ser obtidos e/ou analisados.

Uma abordagem equivalente foi aplicada em (KERPEZ; KINNEY, 2008), no qual são descritos testes automatizados, características a serem mensuradas pertencentes a elementos da rede, análises e operações ADSL2+, descrevendo métodos integrados de medição de todos os elementos envolvidos em uma comunicação deste tipo, bem como os serviços oferecidos. Objetiva-se que problemas físicos e de camadas superiores possam ser identificados e que esta combinação de medições, coleta de dados, rotinas de análise e administração possam reduzir o tempo para detecção de falhas e reparo das mesmas.

Outro trabalho nesta linha é apresentado em (PINILLA et al., 2008) onde é apresentado o conceito de COM (*Copper Plant Manager*), uma combinação de qualificação de loops existentes e ferramentas de monitoramento, que permite uma melhor gestão dos recursos através de enlaces DSL. Um conjunto de estratégias é apresentada, as quais possibilitam que dados e serviços possam ser oferecidos com maior confiabilidade e precisão, além de possibilitar o provisionamento de recursos através de uma gerência de fim-a-fim da última milha.

Tais trabalhos apresentam estratégias bem completas, inclusive abrangendo testes de diversas camadas do protocolo TCP/IP (e.g. camada física, enlace e rede), as quais serviram de base para esta tese, contudo estão limitados apenas à idealização, levantamento e definição dos testes. A execução e aferição, tarefas que induzem a diversos outros problemas, não foram realizadas. Deixando em aberto problemas como métodos de armazenamento, seleção de dados, dimensionamento das amostras, avaliação estatística, correlação das variáveis obtidas, entre outras.

No que concerne à avaliação das aplicações envolvidas na comunicação utilizando tecnologias DSL, diversos trabalhos abordam aplicações específicas, no qual o canal está dedicado para esta aplicação e/ou um conjunto destas aplicações, contudo poucos utilizam uma abordagem dos fluxos *Triple Play* (voz, vídeo e dados) em um mesmo canal de comunicação. Em (TAN et al., 2009) foi apresentada uma avaliação do comportamento de aplicações de voz (VoIP) no sentido de *upstream*, considerando que a alocação de banda para cada assinante tem-se mostrado problemática no desempenho aplicações sensíveis ao atraso.

Em (PAPAGIANNI et al., 2009) são apontados problemas como a mal exploração da largura de banda disponível, proporcionando assim uma baixa qualidade de serviço para assinantes de tecnologias xDSL que utilizam serviços tais como vídeo, voz, e dados. O desempenho destas aplicações foi analisado considerando diversas técnicas de escalonamento de pacotes (PQ – Priority Queue, WFQ – Weighted Fair Queuing, WFQ-LL - Weighted Fair Queuing with low Latency), além da proposição de um novo tipo de escalonamento. Foram consideradas também as diferentes formas de empacotamento da aplicação de vídeo (codificação). O desempenho destes algoritmos de escalonamento é avaliado através de simulações, a fim de propor a solução mais adequada para aplicações *Triple Play*.

Um estudo equivalente é realizado em (LIU et al., 2007), onde é apontada a dificuldade de alocar largura de banda em redes NGN para serviços *Triple Play*. Nele é proposta a divisão dos fluxos em classes, objetivando maximizar a utilização do canal e o atendimento de requisitos mínimos de QoS das aplicações através da aplicação de métodos de Lagrange com condições de KKT. Resultados foram apresentados de forma a definir larguras de bandas ideais (mínimas e máximas) para cada aplicação.

Em (GONÇALVES et al., 2009) é apresentada a experiência de construção de um *testbed* e implementação de testes para transmissão de vídeo sob a tecnologia ADSL2+, onde foram utilizados situações adversas tais como a presença de tráfego e erros de transmissão impactando na qualidade do vídeo. Tais testes foram implementados com o objetivo de mostrar que para utilizar tal tecnologia devem ser implementadas políticas de QoS. Foi detectado que a codificação é de suma importância para obter uma qualidade na comunicação, onde se destaca o codificador H.264, codificador considerado como padrão para este tipo de transmissão, e que será mais bem apresentado em seções posteriores.

Dentre as técnicas encontradas na literatura que buscam aumentar a qualidade na comunicação, no que concerne aos protocolos envolvidos, destaca-se a técnica de *crosslayer*. Esta abordagem busca melhorar o funcionamento da pilha de protocolos agregando certa “inteligência” a ela, possibilitando a interação entre as camadas do protocolo com a finalidade de otimizar recursos e melhorar o desempenho da rede.

Esta abordagem foi idealizada inicialmente para redes sem fio, devido possuírem inconvenientes tais como largura de banda limitada, grande incidência de ruídos na comunicação, problemas de propagação do sinal, grande perda de pacotes, necessidade de gerenciamento para mobilidade dos clientes, entre outros. Tais problemáticas poderiam ser diminuídas através de uma maior interação entre as camadas do protocolo TCP/IP. Em (OTHMAN et al., 2006) são propostas técnicas para prover QoS através de *crosslayer*, onde implementou-se uma comunicação entre as camadas físicas (PHY) e a camada de acesso ao meio - MAC, possibilitando o controle dos diferentes tipos de dados através da política WFQ, na qual são criadas filas por classe de serviços. O comportamento do algoritmo era determinado com base nos dados da camada física.

Um estudo similar foi realizado em (KLIAZOVICH; GRANELLI, 2004), onde foi implementado uma otimização no desempenho da camada de transporte do protocolo TCP em um ambiente sem fio. É introduzido um agente ARQ (*Automatic Repeat Request*) dentro da pilha de protocolos de um hospedeiro, criando a possibilidade de exploração de mensagens ARQ na camada de enlace dentro do esquema de reconhecimento do TCP, com o intuito de maximizar a vazão da transmissão.

(BOUAM et al., 2007) apresentou um novo protocolo que busca prover QoS para aplicações tais como voz, vídeo e dados em redes 802.11. Através de dados obtidos na camada física do protocolo TCP/IP (condições de propagação dos dados, por exemplo), características de outras camadas do protocolo são alteradas em tempo de execução (tamanho do pacote IP). O trabalho apresenta resultados favoráveis obtidos através de simulações realizadas, no qual a vazão dos fluxos é maximizada, todavia sua aplicação se restringe as redes 802.11.

Ainda sobre redes 802.11 destaca-se (ARAÚJO et al., 2007) os quais propuseram estratégias para determinar quanto uma rede sem fio influencia em outra na mesma área de cobertura, além disso, bem como esta influência interfere nos parâmetros de QoS. Obtidos os

dados foi utilizada uma RB para obter informações de como os atributos se relacionam, são eles atraso, vazão, *jitter*, MOS, potência do sinal e distância. Entretanto tais trabalhos não apresentaram técnicas que possibilitassem a garantia de QoS, apenas como os atributos reagem após inferências realizadas nas RB; além de que, novamente, a tecnologia estudada foi a sem fio.

Estas aplicações não se restringiram apenas para redes locais. O segundo projeto para redes de terceira geração (3GPP2) foi utilizado como estudo de caso em (BOUROUHA et al., 2004) no qual foi garantido um alto valor de largura de banda e suporte de QoS para aplicações integradas, utilizando, através de um controle de admissão que possibilite a admissão de usuários de acordo com seus perfis; e através de um controle de alocação de recursos foi alcançado o objetivo de determinar um conjunto de parâmetros que possibilitassem uma vazão, atraso, *jitter* entre outras características dentro de padrões pertinentes, entretanto apenas para aplicações sem fio.

Desta forma, nota-se que diversos trabalhos utilizaram tais técnicas para garantir padrões mínimos de QoS para as aplicações através de *crosslayer*, entretanto limitando-se a redes sensíveis como as sem fio, devido as redes cabeadas serem menos suscetíveis a interferência por ruídos. Entretanto, não são consideradas tecnologias de última milha tais como PLC e ADSL2+, que possuem altas taxas de interferências (e.g. ruídos e desgastes de sinal), as quais trabalham em condições adversas, o que influencia diretamente nas aplicações mais sensíveis como voz e vídeo; principalmente em redes de alta capilaridade (grandes áreas de abrangência), as quais, embora não tendo sido desenvolvidas com o objetivo de transmissão digital de dados, são atualmente utilizadas para este propósito.

Por intermédio dos trabalhos encontrados na literatura, firma-se que é necessário um estudo mais aprofundado, que possibilite estabelecer informações e relações sobre o comportamento das aplicações em infra-estruturas de rede já estabelecidas e/ou que serão criadas. No caso de redes já em funcionamento, o comportamento e desempenho de novas aplicações devem ser analisados, de forma a determinar como será o seu comportamento ao serem inseridas. Além disso, é necessária a definição de uma estratégia de testes que seja abrangente e aborde as principais variáveis e parâmetros que influenciam em uma comunicação de banda larga típica, de maneira a criar um conjunto de dados válido e representativo do domínio.

A grande motivação para incorporar alguma inteligência no planejamento e gerenciamento da rede é dada pela necessidade de correlacionar eventos das camadas inferiores de gerenciamento para determinar estratégias nas superiores (*crosslayer*) e pela necessidade de aumentar a sofisticação requerida para diagnóstico, devido à maior complexidade do sistema e pela necessidade de detectar falhas na rede rapidamente e se possível a recuperação automática.

As RBs oferecerem um mecanismo natural de modelagem de diagnósticos. A grande aplicação dessa técnica está associada à resolução dessa classe de problemas nos mais diversos campos de atuação. Por exemplo, no caso do domínio dos sistemas elétricos, há um emprego maciço em diagnósticos de falhas de equipamentos, além da utilização desse método para prover análises qualitativas e quantitativas, para a avaliação da confiabilidade dos sistemas elétricos, com vistas às melhorias dos processos de planejamento e operação, conforme pode ser visto em (YONGLI et al., 2008) e (WILSONA; HUZURBAZAR, 2007).

Sua aplicabilidade também pode ser observada no projeto de estradas e rodagens (BAYRAKTAR; HASTAK, 2009), as quais são um desafio para as agências rodoviárias estaduais e contratadas, uma vez que estas estradas estão localizadas, frequentemente, em áreas urbanas e gera um alto impacto no tráfego local, na comunidade empresarial e nas regiões periféricas, levando a um envolvimento multipartidário. O trabalho apresenta um sistema de suporte à decisão para auxiliar na seleção de estratégias adequadas para contratação de projetos rodoviários, considerando, em nível macro, as inter-relações entre as partes interessadas, bem como os fatores críticos que impactam no projeto.

Percebe-se então a aplicabilidade das técnicas de MD, mais especificamente das RBs na implementação de estratégias e diagnósticos dos mais variados problemas, contudo vale destacar algumas peculiaridades que diferem este trabalho dos apresentados anteriormente:

- Aplicação em um inédito escopo de trabalho, as redes de acesso, mais especificamente redes ADSL2+, através da implementação de uma estratégia de planejamento, diagnóstico e predição de possíveis comportamentos destas redes através de uma abordagem *crosslayer*. Não sendo encontrados, no desenvolvimento desta tese, trabalhos relacionados a este tipo de abordagem;
- Uma abordagem a qual esta fundamentada em testes empíricos e interações com especialistas no domínio;

- As referências encontradas na literatura se caracterizam por oferecer um caráter exploratório para as análises, isto é, as decisões são tomadas a partir do diagnóstico da situação (ou problema); sendo o método aqui implementado capaz de realizar análises antecipatórias. Assim, a partir de uma meta pré-estabelecida, podem-se identificar quais os cenários que propiciam obter tal meta, ou seja, quais os procedimentos devem ser adotados para alcançar os objetivos traçados antecipadamente.

4.3 – CONSIDERAÇÕES FINAIS

Procurou-se abordar neste capítulo alguns trabalhos e referências com relação às técnicas e estudos abordados nesta tese, de modo a ilustrar o estado-da-arte da área. Com isso, pretende-se validar a utilização das mesmas. Os trabalhos apresentados discorrem sobre a criação de estratégias de testes, manutenção e monitoramento de redes de acesso DSL, além de demonstrarem a aplicabilidade de técnicas de *crosslayer* para implementação de QoS em redes de última milha.

Além disso, alguns trabalhos que discorrem sobre a aplicação de algoritmos de inteligência computacional foram apresentados, mostrando suas capacidades de classificação e diagnóstico. Todavia, aponta-se aqui a sua limitação na descoberta de quais fatores podem corroborar com a realização de determinadas ações, dada a força da correlação existente entre essas ações e os demais aspectos que a influenciam.

Desse modo, no próximo capítulo, é apresentado uma metodologia de planejamento de redes de alta capilaridade desenvolvidas nesta tese, que possibilita, utilizando-se uma base de dados obtida por aferição, diagnosticar comportamentos de aplicações e, de enlaces de comunicação e, através de um método híbrido que combina técnicas que possam estabelecer relações de causa e efeito (RBs) e de otimização (algoritmos genéticos), encontrar os estados de determinadas variáveis que possam estabelecer uma condição ótima almejada para uma variável meta.

5. ESTRATÉGIA DE PLANEJAMENTO DE REDES DE ACESSO DE ALTA CAPILARIDADE BASEADAS EM TECNOLOGIA DSL: UMA ABORDAGEM HÍBRIDA EMPÍRICA E PROBABILÍSTICA

5.1 - CONSIDERAÇÕES INICIAIS

No Brasil, o setor de telecomunicações encontra-se com um duplo desafio: a) a pressão da demanda pela expansão dos serviços já existentes; b) a demanda por novos tipos de serviços, decorrente da rápida inovação técnica que caracteriza o setor, e da alta tecnologia aplicada em um ambiente competitivo.

O mercado e os demais setores da economia apresentam-se cada vez mais competitivos, as organizações têm de recorrer às suas bases de dados e às informações externas que a influenciam, para extrair padrões que melhor expressem suas ações futuras, como forma de aperfeiçoar os seus processos decisórios e, em decorrência, o andamento de seu negócio

A preocupação cada vez maior das empresas em investirem em tecnologias de armazenamento e processamento de informações, bem como de transformá-las em conhecimento, tem atraído interesse pela investigação de mecanismos que possam realizar essa transformação, proporcionando um auxílio efetivamente inteligente ao processo de tomada de decisão.

No que concerne as empresas de telecomunicações, esta situação ocorre devido à necessidade de aumentar o gerenciamento e o tamanho da rede sem acarretar em maiores investimentos, pela necessidade de fornecer uma maior qualidade no serviço que está sendo oferecido aos usuários, e pela necessidade de correlacionar eventos das camadas inferiores de gerenciamento para determinar estratégias nas superiores (*crosslayer*).

Essas investigações podem ser enquadradas em uma área denominada como Mineração de Dados, que representa uma fonte de maduras tecnologias, amplamente inseridas nos processos organizacionais dessas corporações. Como visto na seção 3, a MD é parte componente do KDD, e pode ser compreendido como um processo interativo e iterativo, não

trivial, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados (FAYYAD et al., 1996).

O sucesso da aplicação do processo de MD está intimamente ligado a diversos aspectos, desde o entendimento do domínio da aplicação, passando pela qualidade dos dados a partir dos quais a extração de conhecimento vai ser realizada e das escolhas das técnicas e ferramentas utilizadas, até a facilidade de compreensão dos padrões descobertos. Portanto, é pouco provável que os padrões descobertos por meio de dados inconsistentes ou com valores ausentes, sem que haja um devido tratamento desses problemas, garantam aos gestores decisões confiáveis. Assim, a confiabilidade e compreensibilidade dos padrões descobertos decorrem de escolhas bem feitas das tarefas e técnicas de mineração utilizadas.

Este trabalho visa implementar, através de técnicas de *crosslayer*, estratégias para planejamento e avaliação de desempenhos de redes ADSL2+, as quais implementem níveis mínimos de QoS para aplicações *Triple Play*. Esta abordagem será alcançada através de um conjunto de técnicas como: aferição de dados, modelagem, otimização, simulação, etc, criando assim, um arcabouço de informações que nortearão a implantação de aplicações *Triple Play* e/ou infra-estrutura de redes banda larga.

5.2 - ABORDAGEM

Como ditado anteriormente, a descoberta de conhecimento a partir de dados é entendida como um processo contendo, pelo menos, as seguintes grandes etapas: (1) compreensão do domínio da aplicação; (2) pré-processamento dos dados; (3) extração de padrões (4) avaliação do conhecimento extraído ou pós-processamento e (5) consolidação e utilização do conhecimento extraído. Um esquema representativo contendo todas essas etapas é ilustrado na Figura 5.1.



Figura 5.1 - Etapas do processo de mineração de dados (REZENDE, 2003)

Desta forma, uma estratégia foi montada buscando seguir este esquema, contudo diferencia-se do método tradicional devido utilizar duas abordagens específicas.

A primeira é a adição um passo anterior a todo o processo, a obtenção dos dados através da aferição, de forma a gerar uma base de dados real e concisa, que envolva os principais elementos da comunicação, bem como as principais situações que podem ocorrer no processo de transmissão de fluxos *Triple Play*.

A segunda é a utilização, na fase de MD, de variáveis que mensuram o desempenho das aplicações no mais diversos níveis do protocolo TCP/IP, gerando assim uma abordagem *crosslayer*. Isso ocorre devido ao protocolo TCP/IP, que é considerado, por censo geral, como o protocolo para intercomunicação dos diferentes tipos de rede, possuir uma grande deficiência: o fornecimento de QoS fim-a-fim.

Para prover serviços com maior confiabilidade, alguns utilizam o protocolo de transporte TCP, que desde seu advento tem sido considerado o principal protocolo de comunicação, todavia em redes com perda de pacotes este protocolo torna-se falho, pois ele habilita o controle de congestionamento do protocolo desnecessariamente, uma vez que certas aplicações podem tolerar perdas na comunicação (KIM; CHUNG, 2007).

Observa-se então que a atual comunicação de dados que utiliza principalmente o protocolo TCP/IP para aplicações costumeiras como o tráfego HTTP é apropriada, pois mantêm a compatibilidade entre as redes existentes (roteadores, *gateways*, etc.). Entretanto para as aplicações atuais, conhecidas como *Triple Play*, o protocolo TCP/IP mostra-se deficiente. O fato do protocolo TCP/IP ser dividido em camadas facilita a implementação de

novas aplicações, além de estar de acordo com os princípios relacionados à engenharia de software. Contudo devido às aplicações atuais necessitarem de condições mínimas para que a transmissão seja realizada satisfatoriamente, o protocolo oferece algumas restrições, mostrando-se ineficiente.

O modelo *crosslayer* visa implementar uma interação entre as camadas do protocolo, fornecendo cada vez mais qualidade de serviço ao usuário, que a cada dia torna-se menos avesso ao mundo informatizado e, conseqüentemente, cada vez mais exigente quanto à tecnologia. Através desta técnica, por exemplo, torna-se possível adaptar parâmetros de desempenho em tempo real da camada de aplicação, tais como vazão e *jitter*, desde que a camada de transporte informe uma grande quantidade de pacotes perdidos, possibilitando assim, uma adaptação da aplicação que está sendo utilizada baseada nas características do meio no qual esta transmitido os dados.

É importante destacar que esta variação da abordagem do KDD foi montada seguindo recomendações de normas específicas para *testbeds*, tamanho dos enlaces, ruídos e padrões mínimos que servirão de patamares de qualidade para as aplicações.

A estratégia e a metodologia a serem utilizadas nos testes estão divididas nos seguintes tópicos:

1. Definição de arquitetura e equipamentos envolvidos;
2. Definição das variáveis a serem analisadas;
3. Criação do *testbed*;
4. Configuração dos equipamentos e preparação para execução dos testes;
5. Testes empíricos;
6. Análise dos resultados obtidos;
7. Estudo de correlação Bayesiano;

Os tópicos de um a cinco envolvem a aferição e coleta de dados, os quais passarão pelas fases de conhecimento do domínio e pré-processamento. Os tópicos seis e sete estarão

envolvidos nas fases de extração de padrões, pós-processamento e utilização do conhecimento.

Cada passo será detalhado posteriormente, apresentando, quando for pertinente, os principais elementos envolvidos assim como suas características.

5.2.1 – Definição de arquitetura e equipamentos envolvidos

De forma a avaliar a comunicação *Triple Play* em uma rede DSL, deve-se utilizar um modelo padrão para testes como o sugerido por (XIAO, et al., 2007)(KERPEZ; KINNEY, 2008) e (SANDRI et al., 2007) (Figura 5.2), os quais definem a seqüência de elementos interligados.

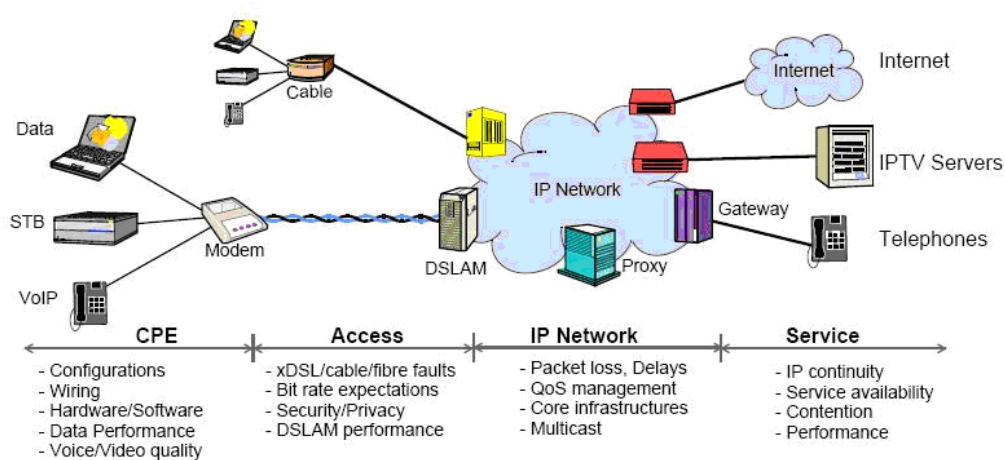


Figura 5.2 – Exemplo de um *testbed* padrão

A arquitetura a ser utilizada envolverá os seguintes itens:

- xDSL Modems, incluindo ADSL, ADSL2, ADSL2+;
- DSLAM;
- Simulador de cabos padrão Europeu;
- Analisador de protocolos.

O cenário de testes utilizado é composto de modems, DSLAM, cabos de telecomunicações, gerador de ruído e computadores e o mesmo foi montado no LABIT (Laboratório de Inovação Tecnológica em Telecomunicações) da UFPA.

A geração dos ruídos é feita pelo DSL 5500, um gerador de ruído da *Spirent Communications*, sendo que o ruído é gerado na faixa de operação do ADSL2+ (4,3125 kHz a 2,208 MHz).

Um analisador de protocolo da marca RADCOM® (RADCOM, 2009) também foi utilizado com a função de filtrar os pacotes que irão trafegar na rede, isolando fluxos específicos para geração de medidas de desempenho.

O conjunto DSLAM / EDA (*Ethernet DSL Access*) compõe o equipamento existente no lado da central telefônica, permitindo assim a comunicação de dados via um enlace DSL. O computador conectado ao DSLAM é o responsável por gerar fluxos de vídeo que serão distribuídos através de *multicast* entre os clientes.

Foi utilizado um simulador de enlaces ADSL2 + ETSI DLS 410E3 da *Spirent Communications*. O mesmo é responsável por reproduzir as características do AC e DC do cabo de par trançado telefônico utilizando circuitos passivos (R, L & C).

A metodologia a ser aplicada é a convencionalmente utilizada para benchmarking de altas camadas de protocolo em quaisquer tipos de dados a serem transmitidos, onde os dados a serem alterados são dados específicos de tecnologia DSL, os quais foram anteriormente apresentados e são eles: tamanho de enlace e aplicações as quais são utilizadas.

5.2.2 – Definição das variáveis a serem analisadas

O tráfego a ser estudado nesta proposta, como dito anteriormente, é o *Triple Play* (voz, vídeo e dados); existem diversos parâmetros disponíveis para mensurar a qualidade de uma transmissão, seja ela de dados ou multimídia, entretanto alguns parâmetros podem gerar maior ou menor influência na comunicação e, portanto, é necessário definir quais métricas devem ser analisados para cada aplicação. Para isso, é importante apresentar o conceito de QoS e QoE (*Quality of Experience*) que permeiam desde o serviço entregue ao assinante até a definição de toda a infra-estrutura de comunicação.

Um dos grandes desafios das operadoras para fornecer o serviço de IPTV é oferecer aos seus clientes o nível de qualidade pelo menos equivalente aos sistemas de TV a cabo ou via satélite. Esta tarefa se torna complexa a partir do momento que se deseja transportar fluxos de vídeo em larga escala através de uma infra-estrutura de redes IP onde os múltiplos serviços compartilham recursos.

Aspectos voltados à rede e aos sistemas são conhecidos como parâmetros de QoS caracterizando uma medida de desempenho no nível de pacotes, ou seja, é um conjunto de requisitos e mecanismos utilizados para implementar uma infra-estrutura de rede capaz de satisfazer os requisitos das aplicações, como por exemplo, oferecer o gerenciamento dos efeitos do congestionamento no tráfego de pacotes. Os parâmetros mais comuns relacionados à rede são: atraso (*delay*), *jitter* (variação do atraso), perda de pacotes e largura de banda (TR-126, 2006). Entretanto tais parâmetros não são suficientes para mensurar a experiência do usuário, uma vez que o usuário de um serviço IPTV não se importa com as restrições de desempenho da rede, e sim com características mais elevadas, como por exemplo no sincronismo das imagens, tempo de atraso entre mudança de canais, entre outros. Tal abordagem, que se refere à percepção do usuário, é conhecida como QoE.

Assim, com o modelo de QoS por si só não abrange todos os aspectos necessários para avaliar o serviço, é necessário utilizar modelos de QoE e QoS combinados em um modelo de camadas no qual o domínio de parâmetros de QoE estão em um nível superior aos de QoS (Figura 5.3). Tal modelo pode ser definido em 3 (três) camadas (ZAPATER; BRESSAN, 2007):

- **Serviço:** uma pseudo camada acima da camada de aplicação, a qual é exposta ao usuário e na qual os parâmetros de QoE são medidos. Um exemplo de medida é o MOS (*Mean Opinion Score*), que avalia o grau de percepção do usuário em relação a um serviço e sua degradação;
- **Aplicação:** parâmetros relacionados a aplicações como as de vídeo (e.g. resolução, compressão, codecs, etc.).
- **Transporte:** parâmetros relacionados à camada de transporte (e.g. *jitter*, atraso, perda de pacotes, etc.).

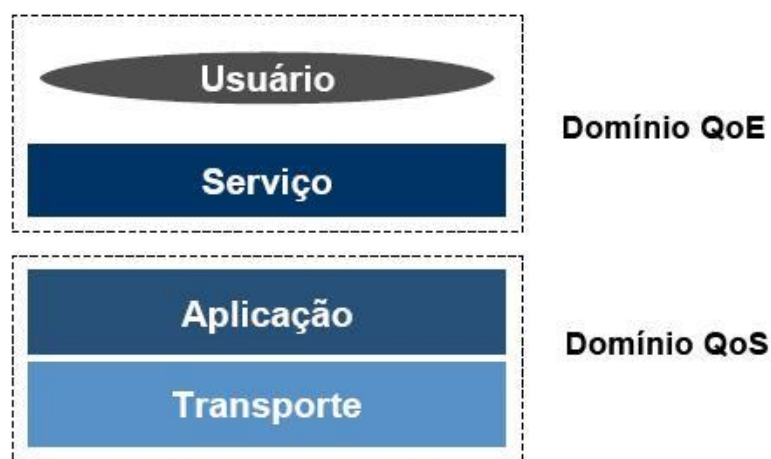


Figura 5.3 – QoE e QoS (ZAPATER; BRESSAN, 2007)

Em (TR-126, 2006) foi apresentado um relatório técnico específico para testes utilizando este tipo de tráfego (*Triple Play*) sobre tecnologia DSL, onde foram definidas três abordagens que podem ser consideradas para análise das aplicações que utilizam estes fluxos, são elas:

Subjetiva, na qual são utilizados experimentos controlados e utilização de participantes nos experimentos, os quais irão definir parâmetros de satisfação em relação à aplicação. Toma-se aqui como exemplo o MOS que é uma métrica definida a partir de um conjunto de testes subjetivos, no qual vários ouvintes avaliaram a qualidade do áudio com base na clareza e cadência dos sons. A avaliação é feita conforme a Tabela 5.1.

Tabela 5.1 – Classificação do MOS (TR-126, 2006)

MOS	Qualidade do MOS	Grau de interferência
5	Excelente (excellent)	Imperceptível
4	Boa (good)	Perceptível mas não incômodo
3	Apropriado (fair)	Pouco incômodo
2	Pobre (poor)	Incômodo
1	Ruim (bad)	Muito incômodo

Objetiva, utilizando equipamentos eletrônicos para mensurar experiência no processo de apresentação do serviço (e.g., carregamento total de um site).

Indireta, caracterizando a análise dos parâmetros de desempenho da rede (vazão, atraso, etc.).

Desta forma, destacam-se para o tráfego de dados parâmetros como taxa de perda de pacotes, atraso e vazão; uma vez que atrasos ou perda de pacotes muito grandes geram retransmissões desnecessárias que, por sua vez, podem acarretar congestionamentos na rede.

Para aplicações multimídia como videoconferência e VoIP, consideradas aplicações sensíveis e de tempo real, parâmetros como *jitter* (variação estatística do atraso), atraso, MOS e perda devem ser analisados, pois podem comprometer a recepção do fluxo.

Isto posto, as medidas de desempenho obtidas para este estudo de caso estão divididas por aplicação, como segue:

1. Fluxo de voz: *Jitter* (*jitter_VoIP*), perda de pacotes IP (*Perda_VoIP*), MoS (*Mean Opinion Score*) (*MOS_VoIP*), número de tentativas com sucesso (*Tentativas_VoIP*).
2. Fluxo de vídeo: *Jitter* (*jitter_Video*), vazão do vídeo (*Vazao_Video*) e perda de pacotes IP (*Perda_Video*).
3. Fluxo de dados (FTP – Fila Transfer Protocol): Atraso (*Atraso_FTP*), *jitter* (*Jitter_FTP*), perda de pacotes IP (*Perda_FTP*) e vazão (*Vazão_FTP*).

Onde:

Def.1: O termo “*throughput*” ou vazão é definido como a taxa de bits máxima que é permitida em uma transmissão fim-a-fim sem ocorrer nenhuma perda de pacotes durante os testes.

Def.2: O atraso de pacotes IP em um sentido é o tempo que um pacote IP (de um certo tamanho) precisa para viajar da origem ao destino. Em alguns casos, também conhecido como latência.

Def.3: A perda de pacotes IP é a relação entre o número de pacotes perdidos e pacotes transmitidos entre origem e destino durante um longo período de tempo.

5.2.3 – Criação do *testbed*

A presença de enlaces de grandes distâncias e diferentes bitolas gera obstáculos na transmissão do sinal elétrico, degradando a qualidade da recepção e aumentando a probabilidade de erros na comunicação. O nível de sinal pode variar significativamente devido

à atenuação do sinal em cabos metálicos gerado por grandes distâncias conseqüentemente retransmissões são produzidas e o desempenho das aplicações é influenciado.

Outro fator que deve ser levado em consideração são os ruídos que podem ocorrer durante uma comunicação, os quais podem influenciar na relação sinal-ruído (qualidade da comunicação). Devem ser analisados os ruídos mais freqüentes, com diferentes potências, objetivando um estudo aprofundado do impacto dos mesmos sobre as aplicações.

Para realização dos testes, foi criado um *testbed* que possibilitou a obtenção de dados estatísticos, os quais foram utilizados em análises e serviram como entrada de dados para fases posteriores.

Para os testes, utilizou-se um cenário típico de transmissão IPTV apresentado na Figura 5.2. Nesse sistema, há máquinas comunicantes trafegando o serviço disponibilizado pelo IPTV - que é a convergência entre voz, vídeo e dados (AL-HEZMI et al., 2006).

O conjunto DSLAM / EDA compõe o equipamento existente no lado da central telefônica, permitindo assim a comunicação de dados via um enlace DSL. O computador conectado ao DSLAM possui a função de gerar fluxos de vídeo que serão distribuídos através de *multicast* entre os clientes.

Como meio de acesso, escolheu-se utilizar um simulador que possibilita a determinação do comprimento dos cabos. Desta forma, pode-se alterar a distância entre a central e os clientes.

O comprimento padrão de enlaces telefônicos é definido pela norma brasileira padrão Telebrás 225-540-788 de abril de 1997, a qual propõe a distância de 2090 metros para um enlace de testes, todavia foram utilizados diferentes tamanhos de enlace (2500 m, 3000 m, 3500 m, 4000 m e 4500 m) com bitola de $\varnothing=0.4$ mm PE, objetivando alcançar todas as variantes de tamanhos de enlaces existentes na infra-estrutura de telefonia brasileira. Para cada tamanho de *loop* foram realizados testes sem nenhum ruído (denominado Caso0), nível de ruído branco de $W= -140$ dB e 24 DSL (ISDN) (chamado de Caso1), nível de ruído branco de $W= -130$ dB e 24 DSL (ISDN) (denominado Caso2) e nível de ruído branco de $W= -120$ dB e 24 DSL (ISDN) (chamado Caso3). Conjuntos de ruídos específicos de comunicação DSL recomendados pela (TR-48, 2002) e (ITU-T RECOMMENDATION G.992.5, 2005), são utilizados.

Todos os testes, para cada tamanho de enlace, foram repetidos 10 vezes com uma duração de 120 segundos. O analisador de protocolo RADCOM® também foi utilizado com a função de filtrar os pacotes que irão trafegar na rede, isolando fluxos específicos para geração de medidas de desempenho.

Foram desenvolvidas aplicações para tratamento dos arquivos *trace* gerados pelo analisador de protocolos, evitando que novas licenças de software tivessem que ser obtidas. Tais ferramentas foram desenvolvidas em linguagem C e Matlab para cálculo de atraso, vazão, *jitter*, perda de pacotes e estão disponíveis no site do LPRAD (www.lprad.ufpa.br/~diego).

É importante ressaltar que se pretende definir o comportamento das aplicações através de padrões e curvas típicas, importantes para definição e delimitação do modelo a ser estudado, assim como a definição da distribuição de probabilidade para cada medida de desempenho estudada (e.g. atraso, probabilidades de bloqueio e vazão). Esta preocupação se dá, principalmente, em situações onde é necessário modelar o estudo de caso.

5.2.4 – Testes Empíricos

Para a realização da análise desta tecnologia de última milha, utilizar-se-á um cenário típico de transmissão IPTV, onde serão disponibilizados serviços tais como voz, vídeo e dados (PAPAGIANNI et al., 2009).

Um dos grandes desafios que a aplicação VoIP oferece às operadoras no início de seu planejamento de rede é escolher a codificação padrão de voz mais adequada a fim de proporcionar boa qualidade de voz e eficiência da rede. Existem diversos codificadores (*codecs*) que podem ser utilizados, tais como o G.711 para uma compressão de 64 Kbps, 16kbps com G.726, G.729 com 8 Kbps e o altamente comprimido G.723.1 com 5.3 Kbps; desta forma, os prestadores de serviços VoIP pode escolher o nível de compressão de voz que será aplicado aos seus clientes. Neste estudo em particular o codec G.711 foi empregado. G.711 é o padrão internacional para codificação de voz para um canal de 64 Kbps. Para o transporte de voz, foi utilizado o Callgen, uma ferramenta VoIP desenvolvida pelo projeto OpenH232 (OPENH323, 2009), largamente utilizada para testes e com suporte ao codec G.711 (PAPAGIANNI et al., 2009).

O codec de vídeo padrão utilizado em redes DSL é o MPEG-4 H.264, o qual está sendo adotado por grandes operadoras de serviços de vídeo, e é largamente utilizado na

avaliação do desempenho dos serviços *Triple Play* em redes de última milha xDSL (ITUTREC.H.264, 2007). Foi desenvolvido em conjunto pela UIT *Video Coding Experts Group* (VCEG) e ISO *Moving Picture Experts Group* (MPEG). O H.264 é usado em redes fixas e móveis.

O H.264 provou ser mais resistente às redes propensas a erros através do uso da ordenação de macroblocos flexíveis e particionamento de dados. Além de fornecer uma alta capacidade de compressão, capacidade de fornecer boa qualidade de vídeo, possibilita taxas de transmissão substancialmente menores do que outros *codecs*. Comparado ao codec de vídeo MPEG-2, corta pela metade a taxa de transmissão, enquanto o ganho de codificação em comparação ao H.263 e H. 263+ está na faixa de 24% até 47% (KAMACI; ALTUNBASAK, 2003).

Para a geração de tráfego de vídeo, utilizou-se o VLC (*VideoLAN Client*) (VLC, 2009), um reprodutor multimídia que suporta vários formatos de vídeo, além de vários protocolos de *streaming*. Foi utilizado o protocolo RTP (*Real Time Protocol*) para transmissão do vídeo com codec H.264 com taxa de 1.200 Kbps para vídeo e codec AAC com taxa de 192 Kbps para o áudio.

Para o tráfego de dados, foi utilizada uma ferramenta denominada IPERF (IPERF, 2009), através da qual foi simulado um tráfego FTP contínuo de modo a ocupar o restante da banda disponível. Esta ferramenta é dedicada para análise de desempenho de redes e amplamente utilizada em testes (RAO et al., 2009) e (PRIMET et al., 2002).

As avaliações a serem realizadas serão classificadas em testes de capacidade da rede (*Basic Connection*) e comportamento das aplicações (*Network Services Capacity*) onde são estudados os comportamentos dos protocolos envolvidos.

5.2.5 – Análise dos Resultados obtidos

Alguns resultados podem ser retirados e observados diretamente da base de dados, como os apresentados a seguir. Tais resultados, mesmo que simplistas mostram o comportamento das aplicações.

Na figura 5.4 é apresentado o comportamento da qualidade da comunicação VoIP. O ruído de -120db e 24D (Caso3) notoriamente influencia negativamente na comunicação, contudo percebe-se que a influência de ruídos na comunicação de voz afeta principalmente a

distâncias de 4000 m em diante, onde ocorrem quedas de qualidade de até 69% em relação à distância de 2500m.

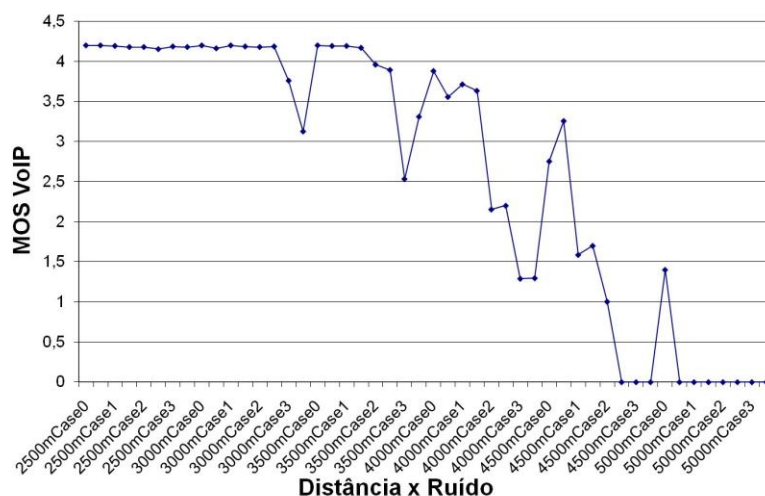


Figura 5.4 – Comportamento do MOS VoIP.

A perda de pacotes é uma das principais características que afetam a qualidade dos fluxos *Triple Play*, principalmente para aplicações que não utilizam protocolos confiáveis de comunicação (principalmente de voz e de vídeo). Aplicações como as de dados, que utilizam protocolos confiáveis e que possibilitam retransmissões, garantem a chegada das informações com integridade, mesmo com baixas taxas de transmissão. A Figura 5.5 demonstra o comportamento das aplicações quanto à perda de pacotes. Os resultados obtidos ilustram a relação direta entre distância, ruído e degradação dos fluxos, principalmente em distâncias de 3500 a 4500 m, as quais estão mais suscetíveis a ruídos. Tais distâncias são bastante utilizadas em países com grandes dimensões geográficas, como Brasil, Índia e China, e que já possuem uma infra-estrutura de telefonia já estabelecida.

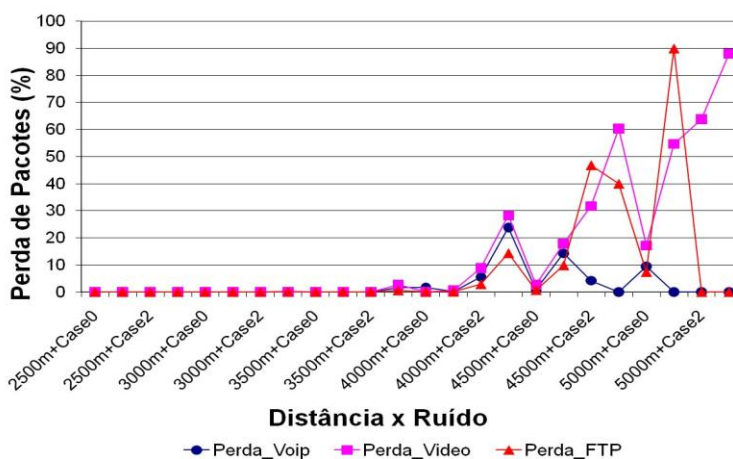


Figura 5.5 – Perda de pacotes por aplicação

Este comportamento é percebido com maior efetividade nas Tabelas 5.2 e 5.3, que apresentam o comportamento (em níveis percentuais) das aplicações em relação à comunicação sem ruído. A aplicação de voz não sofreu perda de pacotes em todas as variações de ruídos, contudo, para o conjunto ruído branco com -130 db e 24D (Caso2) obteve um crescimento de 44% no *jitter*, o que impactou diretamente em seu MOS, ocasionando uma degradação de 6% na qualidade da comunicação, aspectos observados na Figura 5.6. Esse impacto foi maior na combinação de ruído branco de -120 db e 24D (Caso3), que gerou uma degradação de aproximadamente 40% na qualidade de comunicação, alcançando um MOS que inicialmente era de 4.2 para, em média, 2.9.

As aplicações de vídeo e dados quase não sofreram impactos, todas as suas medidas ficaram próximas, com variações de no máximo 3%, conforme dados da Tabela 5.2, com exceção da vazão do fluxo de dados, que obteve uma queda de 2 Mbps (sem ruídos na comunicação) para aproximadamente 317 Kbps (ruído branco de -120 db e 24 D – Caso3), ou seja, uma queda de 600%.

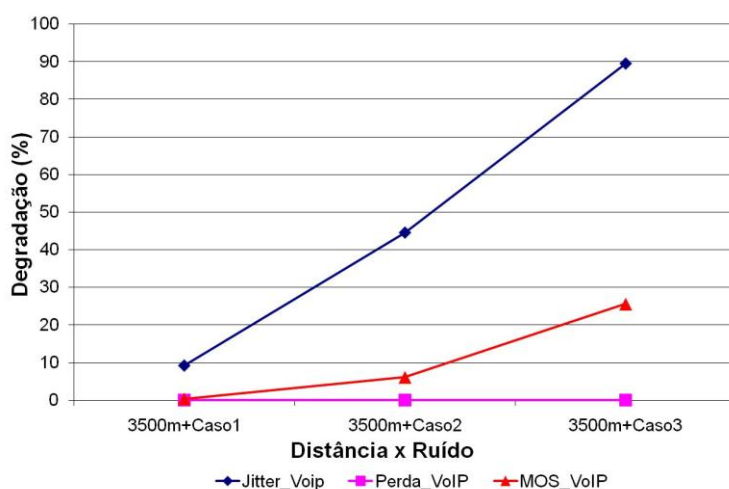


Figura 5.6 – Comportamento do VoIP para 3500m.

Tabela 5.2 – Medidas para Vídeo e Dados em 3500m.

Medida / Distância & Ruído	Jitter Vídeo (ms)	Perda Vídeo (%)	Vazão Vídeo (Mbps)	Jitter FTP (ms)	Perda FTP (%)	Atraso FTP (ms)	Vazão FTP (Mbps)
3500m rb140db&24D	7.9	0.02	1.41	5.07	0.11	0.073	2.06
3500m rb130db&24D	7.39	0.03	1.42	5.37	0.18	0.079	1.98
3500m rb120db&24D	4.55	2.7	1.39	12.01	0.56	0.28	0.317

Para a distância de 4000 m, devido ser uma distância bem aquém do que o desejável, observa-se, conforme Tabela 5.3, que as aplicações de VoIP e Vídeo sofreram impactos consideráveis com a inserção de ruídos (mais especificamente os de -130db e -120db) em relação a comunicação sem ruído (em níveis percentuais). Contudo percebe-se que ocorreu uma melhora em certas variáveis de desempenho de algumas aplicações (e.g vídeo); este fato se dá pela incapacidade de manter a transmissão do fluxo FTP com tais ruídos, o que possibilitou que as demais aplicações utilizem toda a banda disponível. Como a aplicação de vídeo possui uma maior necessidade de largura de banda, é esperado que ela possua uma maior variação que as demais aplicações, como observado na tabela abaixo.

Importante ressaltar que os valores negativos caracterizam o quanto a comunicação com ruído foi pior do que a sem ruído (e.g vazão FTP sem ruído foi de 1.34 Mbps, após inserção do ruído houve uma queda na vazão para 661.8 Kbps, ou seja, uma redução de 50.67 %).

Tabela 5.3 – Medidas para VoIP, Vídeo e Dados em 4000m (em percentuais).

Medida / Distância & Ruído	Jitter Voip	Perda VoIP	MOS VoIP	Jitter Vídeo	Perda Vídeo	Vazão Vídeo	Jitter FTP	Perda FTP	Atraso FTP	Vazão FTP
4000m rb140db&24 D	127.9 6	91.50	1.21	-6.85	2166.67	0.56	63.43	3.27	105.59	-50.67
4000m rb130db&24 D	408.3 5	373.18	40.74	-22.1	29283.33	-7.93	252.8 9	2196	373.72	-90.41
4000m rb120db&24 D	357.4 8	329.60	40.58	8.00	94066.67	-29.48	520.5 3	11347	357.73	-99.46

Quanto ao impacto do conjunto distância do enlace e ruídos nas aplicações, pode-se analisar este comportamento através das figuras a seguir. A Figura 5.7 apresenta o comportamento da vazão das aplicações de dados e vídeo. Enquanto a aplicação de vídeo se mantém constante, até a distância de 4500m com ruído de -140db e 24D (Caso1), onde, percebe-se uma queda vertiginosa. A aplicação de dados possui um comportamento instável devido utilizar o protocolo TCP que possui característica de ser adaptativo e utiliza a banda disponível para transmissão.

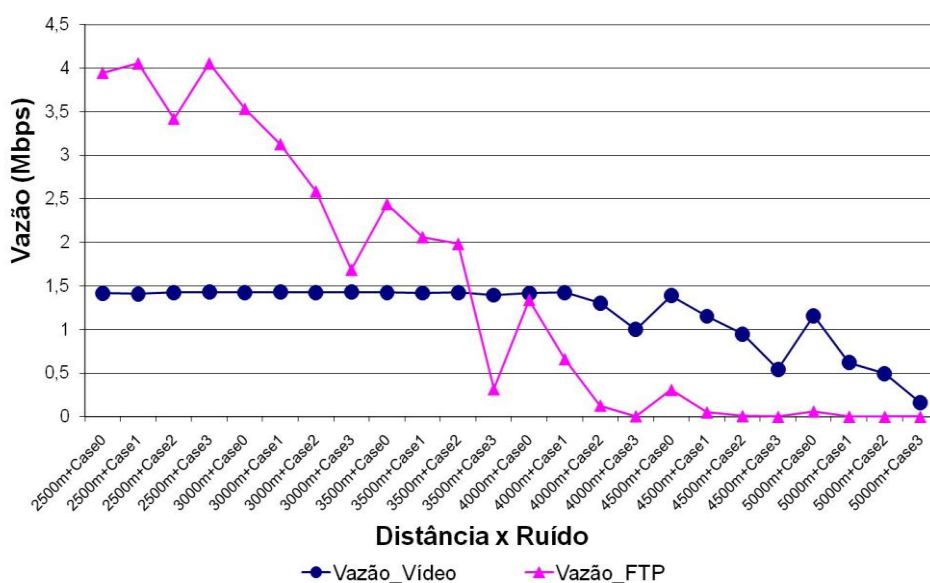


Figura 5.7 – Comportamento da vazão da aplicação de vídeo e dados.

Por intermédio das medidas de desempenho e com auxílio de um especialista, podem-se realizar conjecturas sobre o comportamento e funcionalidades das aplicações, contudo esse estudo acaba por não ser completo uma vez que fatores tais como a influência de um atributo em outro, não são observados; essa característica é observada através de correlações. A

correlação de eventos é importante para reduzir custo com manutenção da rede, melhorar a disponibilidade e o desempenho dos serviços da rede. Dados brutos são interpretados e analisados, levando em consideração um conjunto de critérios pré-estabelecidos, ou definidos dinamicamente em função do processo de gerência.

5.2.6 – Estudo de correlação Bayesiano

Dentre as diversas técnicas de inteligência computacional para correlação de variáveis, as redes bayesianas se destacam como modelos gráficos probabilísticos para representação de conhecimento e raciocínio em domínios onde existem incertezas. O seu caráter unificado torna possível comparar diferentes hipóteses sobre os dados e a natureza intuitiva do formalismo gráfico de RBs a torna um dos melhores métodos analíticos disponível para a tomada de decisão (ROCHA et al., 2007).

Uma vez montada a rede, podemos verificar o seu comportamento futuro, dadas inúmeras variações que os seus atributos possam apresentar, dependendo da abordagem que o usuário queira utilizar. Assim, o usuário pode verificar como irá se comportar a rede dada determinada modificação em algum de seus atributos, isso porque uma vez modificado um certo dado, os seus valores (probabilidades relacionadas com o nó em questão) são propagados para os demais nós de acordo com a sua dependência, ajustando a rede de acordo com a modificação feita, e permitindo-se verificar o comportamento e as variações ocorridas na rede.

Dessa forma, através das RBS, podemos prever como o fluxo *Triple Play* irá se comportar nas redes de última milha ou quais seriam as características físicas que a rede deveria possuir para que conseguisse atender a essa nova necessidade, possibilitando assim que empresas de telecomunicações possam realizar investimentos com mais segurança, atendendo a necessidade do usuário de qualidade e eficiência na prestação do serviço.

A Figura 5.8 representa a RB montada contendo todos os atributos obtidos através dos testes empíricos (vide seção 5.2). Cada nó possui uma tabela de probabilidades associada a ele (e.g. Atraso_FTP), além dos nós estão também representadas as suas dependências através da direção de suas ligações (e.g. o fato de haver ruído na comunicação influi na probabilidade de haver alguma variação no *jitter* da aplicação VoIP, e esta por sua vez no MOS do VoIP).

Ao se realizar alguma modificação na rede (e.g. é aplicada uma evidência de haver um ruído branco de -140db e 24D na comunicação como iminente ou 100%), tal mudança seria

repassada como uma reação em cadeia para toda a rede, onde os valores dos nós que se relacionassem, seja diretamente ou indiretamente, com o nó alterado seriam modificadas para se adequar à alteração feita, prevendo o comportamento da rede caso tal evento se concretizasse.

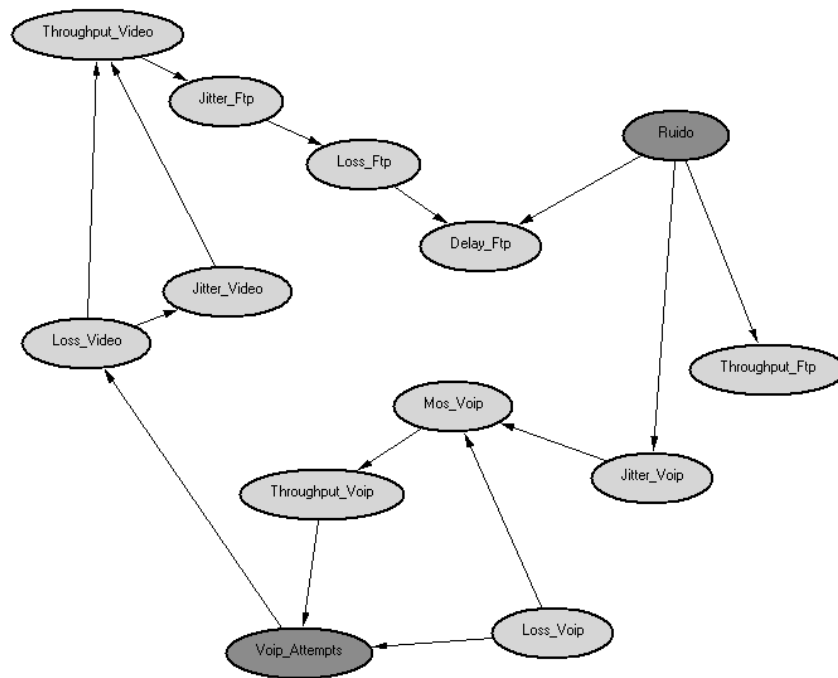


Figura 5.8 – Rede Bayesiana para o fluxo *Triple Play* sobre ADSL2+.

Para Análise dos cenários obtidos, foram implementados alguns estudos de casos que demonstram a utilização desta abordagem para planejamento de redes.

No primeiro estudo de caso foram utilizados parâmetros de QoS mínimos, obtidos através de normas internacionais, que definem a qualidade destas aplicações e quais são as medidas de desempenho desejadas. Objetiva-se encontrar a distância máxima e o conjunto de ruídos que nos possibilite realizar a transmissão dos fluxos com qualidade. Desta forma, poder-se, por parte das Telcos, avaliar se seus enlaces teriam condições de dar suporte a estas aplicações. Os resultados são comparados com amostras adquiridas das redes de telefonia brasileira (TELEBRAS, 1997).

No segundo estudo de caso, realiza-se o processo inverso. É dado como entrada (inferenciado) uma determinada distância de enlace e observa-se o impacto desta definição

nas medidas de desempenho das aplicações, comparando seus resultados com as respectivas normas.

Objetiva-se, com estes estudos, mostrar que alterações graves no nível físico afetam as camadas superiores. Através das informações sobre a interrelação entre as variáveis é possível oferecer ao protocolo a oportunidade de adaptá-lo e assim, diminuir o impacto desta mudança na qualidade do serviço prestado ao usuário. Caso ocorra um ruído específico na transmissão (nível físico do protocolo TCP/IP) poder-se-ia diminuir a taxa de transmissão (atributo da camada de rede) e desta forma alterar a qualidade do vídeo apresentado ao usuário (camada de aplicação do protocolo TCP/IP).

Seguindo este objetivo, uma vez definida a RB e de posse dos dados disponibilizados por ela (relacionamento entre os atributos, probabilidades de estar em determinados estados, etc.), definem-se os parâmetros mínimos a serem estabelecidos, ou seja, as faixas de valores para determinados atributos.

5.2.6.1 – Estudo de Caso 1: Aplicação VoIP

Segundo (PAPAGIANNI et al., 2007) uma comunicação VoIP deve possuir parâmetros mínimos de qualidade que devem ser analisados sem os quais a comunicação VoIP ficaria inviabilizada. Utilizando o protocolo G.711, uma comunicação VoIP é viável se possui características como: *jitter* de 60 ms, 1% de perda de pacotes e, no máximo, 150 ms de atraso. Este foi utilizado nos testes por ser o protocolo padrão para comunicação de 64kbps.

A comunicação VoIP necessita de uma baixa largura de banda, contudo é muito suscetível a problemas de comunicação, como gargalos, atrasos, perdas e ruídos na comunicação, desta forma, é a aplicação de maior preocupação em uma comunicação *Triple Play*. Isto posto, através de inferências realizadas na RB, observou-se qual seria o comportamento das outras aplicações (vídeo e dados) para que se mantivesse a qualidade necessária para a comunicação de voz.

Inicialmente, no que diz respeito à camada física, conforme observado na Figura 5.9, as distâncias que possibilitaram que tais parâmetros fossem mantidos foram 3000m com ruído branco de -130db e 24D (Caso2), e 3500m sem nenhum tipo de ruído ou com ruído branco de -140db e 24D (Caso1).

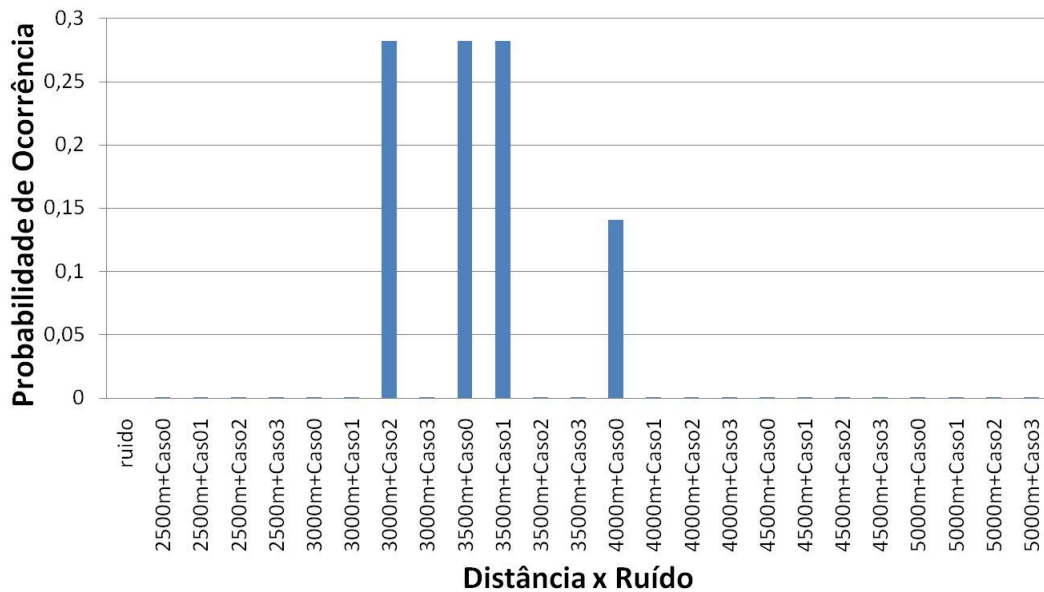


Figura 5.9 – Probabilidade de ocorrência para uma determinada configuração VoIP.

Uma vez que a distância média dos enlaces telefônicos é de 2090 m, de acordo com a norma (TELEBRAS, 1997), observa-se que existe a possibilidade de estender tais distâncias em 66.9%, ou seja, em 1404 m, possibilitando assim menores investimentos pelas Telcos em repetidores (*bridge taps*) os quais geram novas interferências na comunicação ou na definição exata das distâncias necessárias para atender a estas necessidades.

A aplicação de vídeo possui 88.7% de probabilidade de possuir largura de banda entre 1.2 Mbps a 1.4 Mbps, *jitter* com 99% de estar na faixa menor de 20 ms e 87.9% de estar ter perda de até 10%. A aplicação de FTP, que utiliza o protocolo adaptativo TCP, ficou com 91.2 % de chance de ficar com atraso médio de até 100 ms, 90.7% de perda menor de 10% e vazão entre 1.6 Mbps a 2.4 Mbps. Todos os níveis aceitáveis segundo as normas internacionais (TR-126, 2006).

5.2.6.2 – Estudo de Caso 2: Tamanho do Enlace

Neste caso será utilizada a análise inversa, na qual é apresentada uma distância específica e se analisa o comportamento do fluxo *Triple Play* para tal situação. Utilizou-se a distância de 4500m, com ruído branco de -120db e 24D (Caso3), ou seja, influência de ruído em uma distância que representa 11% (4 a 4.5 km) dos enlaces existentes na rede de telefonia fixa (TELEBRAS, 1997) sob influência de um ruído de alta intensidade.

Para este conjunto de distância e ruídos percebeu-se através das correlações obtidas que a comunicação VoIP ficou impossibilitada, devido a disputa pelo canal ser alta e não existirem canais reservados para cada aplicação. A disputa pelo canal é realizada de forma justa na qual todos possuem a mesma chance de obter o canal; a aplicação de vídeo acaba por ocupar quase que totalmente a banda disponível, mesmo que de forma precária, como podemos observar nos dados abaixo.

A aplicação de vídeo poderá obter larguras de banda variáveis entre 400 a 550 Kbps, contudo com uma perda de pacotes entre 50% e 60% do total, ou seja, acima do valor máximo estipulado pela normal TR-126 (TR-126, 2006) que é de 10%. A aplicação de TCP possuirá um desempenho instável, com uma vazão de até 400 Kbps e perda de pacotes de até 10%. Nas figuras a seguir se observa o comportamento da aplicação de vídeo e FTP, mostrando a precisão dos resultados obtidos através desta inferência.

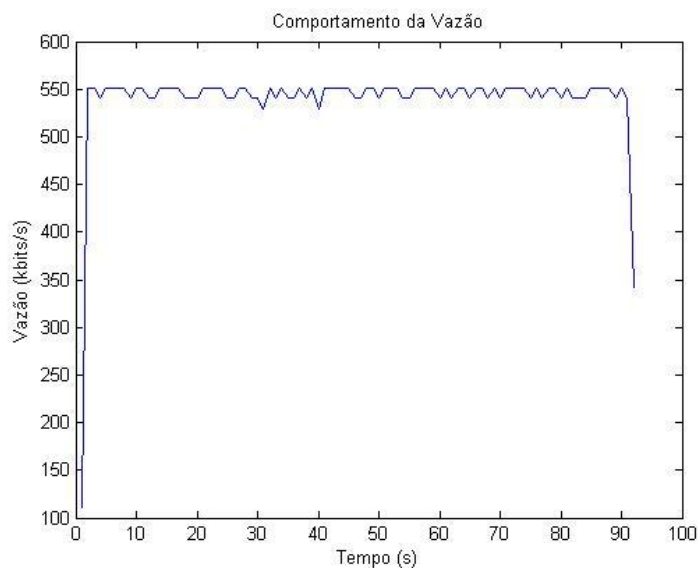


Figura 5.10 – Comportamento da vazão do vídeo.

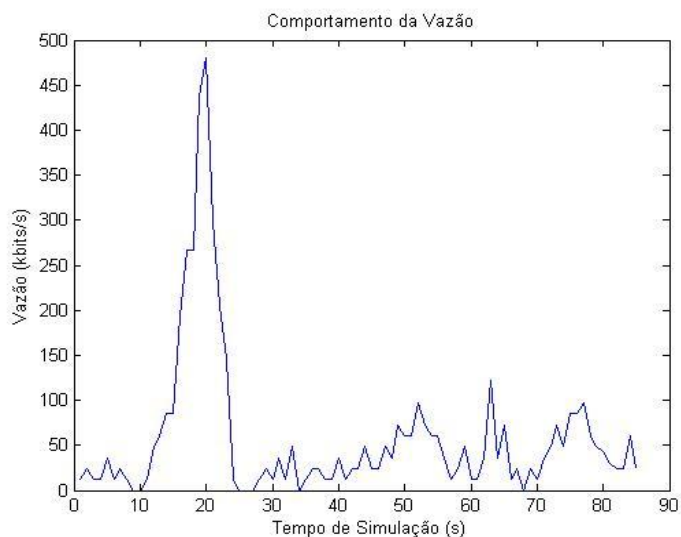


Figura 5.11 – Comportamento da vazão da aplicação de dados.

Uma solução para tal situação é a implementação de QoS na última milha, na qual é definida prioridade de roteamento entre os pacotes; desta forma o fluxo VoIP poder ser enviado, mesmo que as demais aplicações sofressem perdas de desempenho e/ou qualidade. A aplicação de vídeo teria de ser adaptada para uma nova qualidade de imagem e som, alcançando assim a largura de banda disponível. Através de *codecs* como o H.264 (ITU-TREC.H.264, 2007) é possível compactar vídeos com alta qualidade e transmiti-los com larguras de banda variando entre 256 Kbps a 10Mbps.

5.2.6.3 – Procura pela configuração de estado Otimizada

Como dito anteriormente, dentre as muitas técnicas de mineração de dados encontrados na literatura, as RB são consideradas as mais proeminentes, ao se considerar a facilidade na interpretação de conhecimentos obtidos a partir de um domínio com incerteza. A razão disto é devido elas fornecerem um mecanismo para representar o modelo causal de um determinado conjunto de dados (ROCHA et al., 2007).

No entanto, as RB apresentam uma restrição para determinar a análise de cenários ótimos, i.e. definir a melhor combinação de estados de determinadas variáveis (discreta ou contínua) que permitisse alcançar certa exigência (estado de uma ou mais variáveis do domínio). Em muitas aplicações reais, a busca de situações que conduzam à realização de determinados objetivos é extremamente importante. Para tratar este tipo de problema foi utilizada um sistema híbrido combinando o sistema de busca implementado pelos AGs com as características (parâmetros e estrutura de correlação) das RBs.

O objetivo deste modelo é identificar a melhor configuração, dentre os possíveis valores de variáveis existentes no domínio, o que maximiza um determinado atributo, identificando inicialmente as outras variáveis que apresentam uma dependência dele.

Com as variáveis obtidas a partir dos testes empíricos (ver seção 5.2) e da RB criada utilizando o algoritmo de busca e pontuação K2 (ver seção 3.3.1.5), onde todos os atributos foram discretizados em vinte estados, de acordo com a frequência de seus valores, permitindo assim verificar a probabilidade associada a cada um deles, bem como as probabilidades condicionais existentes entre as variáveis; o passo seguinte é, fazendo uso dos dados fornecidos pela RB, buscar o cenário que maximize o atributo desejado (ou uma combinação de atributos). Nesta etapa, utilizamos um AG modificado, como consta no modelo híbrido apontado.

Aqui, em vez de uma função de custo para validar os indivíduos da população, um algoritmo de inferência Bayesiana foi implementado (Equação 5.1); isto é, a RB é utilizada como uma função de custo. Dessa forma, cada um dos indivíduos do algoritmo genético representa uma configuração de inferência na RB, gerados aleatoriamente (e.g., evidenciando o ruído, variáveis com estado 18, *Jitter_VoIP* com o estado 1, Perda de voz (*Loss_VoIP*) no estado 7 e vazão no estado 4 gera o indivíduo 2-1-7-4). Cada indivíduo é, então, para a sua classificação, submetido ao módulo de inferência Bayesiana, a fim de verificar a probabilidade de que o atributo escolhido(s) será maximizada; obtendo, ao final das iterações, a melhor configuração possível de inferências sobre a RB para a maximização do atributo escolhido.

$$P(x_i | c_1, c_2, \dots, c_n) = P(x_i) \prod_{k=1}^n P(c_k | x_i) \quad (5.1)$$

onde c_1, c_2, \dots, c_n são possíveis eventos evidenciados; e x_i é o evento que queremos observar.

Ao final desta etapa (após a análise do algoritmo genético) são obtidos os respectivos estados para esta maximização (para cada atributo).

Estudo de Caso: Tamanho do Enlace

Para demonstrar a aplicabilidade do modelo, foi implementado um estudo de caso que utiliza uma combinação de cenários das seções 5.2.6.1 e 5.2.6.2, simulando uma situação em

que o comprimento do cabo telefônico que conecta o usuário final a CO tem 4500 m, com nível de ruído branco $W = -140$ dB e 24 DSL (ISDN) (Caso1) e é necessária uma comunicação VoIP com uma qualidade aceitável (valores de MOS entre 3 e 4). Assim, com base nestas necessidades, os atributos de ruído e MOS_VoIP foram definidas com estados 18 (4500m + case1) e 7 (MOS entre 3 a 4) respectivamente.

Usando a técnica apresentada foram obtidos os resultados abaixo:

Tabela 5.4 – Valor obtido para os atributos.

Atributos	Estados
<i>Jitter_VoIP</i>	378 para 425 ms
Perda_VoIP	2.5 para 5 %
Vazão_VoIP	153 para 170 Kbps
Tentativas_VoIP	2
Perda_Video	0 para 8.7 %
<i>Jitter_Video</i>	28.5 para 30.78 ms
Vazão_Video	0.547 para 0.675 Mbps
<i>Jitter_FTP</i>	0 para 0.1 ms
Perda_FTP	54.6 para 63.7 %
Atraso_FTP	0 para 0.1 ms
Vazão_FTP	1.6 para 2.02 Mbps

Os valores obtidos mostram que a inferência foi possível, contudo com algumas restrições. Para a comunicação VoIP, apenas 2 das 4 chamadas foram efetuadas com sucesso. A aplicação de vídeo terá uma largura de banda disponível entre 547 e 675 Kbps e o FTP com uma largura de banda entre 1.6 e 2.02 Mbps, porém com uma perda de pacotes de 54 a 63%, considerado um valor muito alto de acordo com (TR-126, 2006).

Os resultados mostraram a difícil manutenção destas aplicações ou, no pior cenário, a impossibilidade de utilização, a menos que seja implementado algum tipo de QoS (via *hardware* ou *software*) ou de ajuste no enlace de comunicação. Com isso, o diagnóstico necessário para sistemas complexos e a rápida detecção de falhas na rede e suas origens podem ser aprimoradas, possibilitando uma recuperação automática e/ou mais efetiva.

5.3 - CONSIDERAÇÕES FINAIS

Neste capítulo, foi descrita uma estratégia metodológica utilizada para planejamento de redes de acesso ADSL2+ sobre influência de ruídos e outros fenômenos da camada física nas camadas superiores, admitindo-se a abordagem *crosslayer*; além de explicar sobre a necessidade, funcionamento e os elementos envolvidos na estratégia. Estudos de caso foram apresentados de modo a apresentar e validar a metodologia.

Adicionalmente a isso, foi apresentada uma proposta de otimização através de uma abordagem híbrida entre redes bayesianas e algoritmos genéticos. Além de possibilitar a quantificação do relacionamento causal entre as variáveis, por meio da descoberta dos valores que compõem uma combinação ótima, de modo a obter uma determinada meta, por exemplo, de maior distância aceitável para o perfeito funcionamento de uma determinada aplicação. As variações do método proposto foram criadas, principalmente, a partir das necessidades encontradas na literatura especializada.

Ambas as contribuições possuem caráter de ineditismo e efetivamente são propostas que vão no sentido de implementar uma estratégia de planejamento de redes de acesso, bem como ampliar a capacidade de aplicação em problemas do “mundo real”, utilizando modelos de inteligência computacional como métodos de otimização híbridos.

6. CONCLUSÃO

6.1 - CONSIDERAÇÕES DA TESE

A evolução tecnológica da informática e das telecomunicações, culminando com a integração entre computadores e telecomunicações, encurtou distâncias e permitiu a reprodução e o repasse de conteúdo com surpreendente rapidez e alcançando grandes distâncias. A introdução de novas tecnologias está exigindo das organizações a mudança em sua forma de relacionamento no mercado, com seus clientes e serviços prestados por elas.

O *Triple Play* é um serviço fornecido pelas operadoras de telefonia fixa ou provedoras de serviços de TV a cabo que mistura o fornecimento de voz, vídeo e dados através da infraestrutura de banda larga. Esta é a verdadeira convergência digital já anunciada e esperada pelos consumidores. Neste contexto, o *Triple Play* tem sido foco de diversas pesquisas, isso está relacionado com o desenvolvimento das redes de computadores, que evoluem no sentido de aumentar as bandas de comunicação e estabelecer garantias de transmissão através da inserção do conceito de Qualidade de Serviço e Qualidade de Experiência. Este cenário engloba uma grande disputa entre: empresas de telecomunicações e empresas de CATV, onde as telcos vêem o *Triple Play*, mais especificamente o IPTV, como uma oportunidade de manter o seu mercado.

Uma vez que os sistemas DSL se firmam como tecnologia base para sistemas banda larga, um adequado estudo dos efeitos nocivos (e.g., ruídos estacionários e não estacionários, interferências eletromagnéticas, etc.) ao desempenho de tal sistema faz-se necessário.

Esta tese tem por objetivo apresentar uma estratégia para planejamento e avaliação de desempenhos de redes ADSL2+, que através da técnica de *crosslayer*, possibilita a manutenibilidade de níveis mínimos de QoS para aplicações *Triple Play*. Por intermédio de técnicas como aferição de dados, modelagem, otimização, entre outras, busca-se criar um base de dados que sirva de referência para a implantação destas aplicações e/ou infra-estrutura de redes banda larga.

Este objetivo é alcançado através da correlação de variáveis obtidas nas aferições e otimizações são obtidas destas correlações, as quais foram realizadas por meio de métodos de Inteligência Computacional conhecidos como redes bayesianas e algoritmos genéticos.

Uma vez descobertas as correlações, via RB, das variáveis que influenciam diretamente o estudo desejado, pode-se estabelecer os cenários que permitem que os usuários de níveis decisórios possam prever quais condições (estados das medidas obtidas nas diversas camadas do protocolo TCP/IP) favorecem a obtenção de um valor meta, os quais foram estabelecidos por esses usuários.

6.2 – PRINCIPAIS CONTRIBUIÇÕES REALIZADAS

Desta forma, são entendidas como principais contribuições desta tese as seguintes:

- Elaboração de uma metodologia inédita de planejamento e avaliação de desempenho que mescla vantagens das técnicas de aferição e o poder de exploração de modelos de inteligência computacional;
- Levantamento do estado-da-arte no que se refere à aferição de redes de acesso DSL, do comportamento de aplicações de voz, vídeo e dados, e aplicabilidade de modelos de inteligência computacional para correlação e otimização das variáveis observadas;
- Projeto e implementação de *testbeds* para realização de testes, considerando parâmetros estabelecidos em normas, tidos como de extrema relevância para o estudo da tecnologia DSL;
- Realização de testes exaustivos e estatisticamente válidos para os cenários propostos, em situações associadas a problemas do mundo real, de maneira a constituir-se uma base de dados experimental bastante consistente, passível de ser tratada computacionalmente;
- Análise de cenários antecipatórios e estudos de correlação e dependência entre as variáveis aleatórias de interesse, utilizando-se para tanto as RBs. Adicionalmente, utilizou-se abordagem própria desenvolvida pelo grupo de pesquisa do LPRAD, em que incorporam características de otimização nas inferências bayesianas tradicionais, por intermédio do uso de algoritmos genéticos;

- Emprego técnicas híbridas de Inteligência Computacional com vistas à melhoria dos processos de descoberta de conhecimento em bases de dados; fato de bastante interesse para pesquisa em redes de acesso, já que não se deseja correlacionar parâmetros e sim, em alguns casos, também otimizá-los;
- Divulgação dos estudos realizados neste trabalho, junto às comunidades nacional e internacional, por meio da publicação de artigos em congressos e periódicos, o que corrobora a importância e contribuições das investigações realizadas. No Anexo A, são apresentadas as publicações realizadas no decorrer do desenvolvimento da tese e que possuem relação com as pesquisas que alicerçaram a mesma.
- Elaboração do documento de tese para disponibilização acadêmica dos modelos e técnicas desenvolvidos, e dos resultados alcançados.

6.3 - DIFICULDADES ENCONTRADAS

Cabe ressaltar, que ao longo do desenvolvimento da tese, foram encontradas dificuldades para a consecução do trabalho, as quais serão apresentadas aqui, com o objetivo de orientar pesquisadores e estudantes que venham a utilizar os conhecimentos disponíveis neste trabalho.

As dificuldades encontradas foram decorridas de diversos fatores, dentre os quais se destacam:

- Como pôde ser observada na literatura, uma das grandes dificuldades encontradas na execução de processo de Mineração de Dados, está relacionada à identificação das fontes e o pré-processamento dos dados. Neste trabalho não foi diferente, as bases de dados utilizadas não existiam, havendo a necessidade de implementações de *testbeds* complexos e utilização de aplicações reais;
- Compreensão do formalismo matemático próprio dos métodos que fundamentam as técnicas empregadas em Mineração de Dados;
- Pesquisa em diversas áreas da engenharia e computação (avaliação de desempenho, IA e redes de computadores);

- Durante o processo de aferição, houve preocupação em realizar medições com diferentes tamanhos de enlace e com diversas combinações de ruídos para não descaracterizar as adversidades existentes. Os resultados relacionados nesta tese foram obtidos a partir de médias extraídas das aferições realizadas em diversas situações. Constatou-se, por exemplo, que combinações não afetavam a comunicação, mesmo que confirmadas por referências na literatura.

É importante ressaltar que os obstáculos encontrados não estão limitados ao estudo das técnicas matemáticas, mas também, particularmente, na implementação dos *testbeds* e dos modelos computacionais.

Ao se optar por trabalhar com uma abordagem multidisciplinar para solução de um problema intrinsecamente de redes de acesso, de modo a definir propostas e objetivos em potencial a serem trabalhados, enfrentam-se de imediato os seguintes aspectos: o estudo exaustivo da literatura e trabalhos dos grupos de pesquisa existentes; a conseguinte análise das tendências futuras da área e do mercado, identificando as necessidades existentes; e a necessidade de se propor e introduzir contribuições que sejam significativas em uma área tão abrangente.

6.4 - TRABALHOS FUTUROS

Estima-se a ampliação e complementação da metodologia apresentada neste trabalho por meio das seguintes ações:

1. Realização de novos estudos e testes de desempenho em outras tecnologias de última milha, uma vez que a metodologia é abrangente;
2. Realização de novos estudos e testes de desempenho em outras técnicas de MD (e.g. lógica *fuzzy*, redes neurais artificiais) e técnicas de otimização (e.g. processo semi-markoviano de decisão, aprendizado por reforço, *particle swarm optimization*);
3. Incorporação do conhecimento obtido em modems e DSLANs, de modo a implementar técnicas de QoS e QoE através de *crosslayer* em tempo real. Embarcando este conhecimento, podem-se realizar ajustes em comportamentos e características dos fluxos;

4. Estudo de técnicas de DSM que permitam melhorias na comunicação. Estudo e desenvolvimento de algoritmos que viabilizem um melhor gerenciamento de espectros de linha DSL, através de ajustes de níveis de potência de sinal, viabilizando a diminuição de interferências que um enlace influi em outro;
5. Após estudo do DSM, realização de novos testes com estas informações, de modo a oferecer serviços diferenciados a usuários pertencentes ao mesmo enlace telefônico, contudo com diferentes distâncias entre a CO.
6. Aplicação de novas medidas de desempenho, principalmente para vídeo, com a incorporação de testes com ferramentas específicas tais como o Evalvid (LIE; KLAUE, 2007), que permitem a obtenção de parâmetros de qualidade tais como perda de *frames*, níveis de MOS de vídeo, PSNR (*Peak Signal Noise Ratio*), entre outros.

REFERÊNCIAS BIBLIOGRÁFICAS

- AL-HEZMI, A., REBAHI Y., MAGEDANZ T., ARBANOWSKI S. **Towards an Interactive IPTV for Mobile Subscribers**. In: ICDT/IEEE International Conference on Digital Telecommunications, (2006).
- AMBROSI, A., PEUGEOT V., PIMIENTA D. **Word Matters: Multicultural Perspectives On Information Societies**, C & F Editions (2005).
- ARAÚJO, J., RODRIGUES, J., FRAIHA, S., CAVALCANTE, G. P., FRANCÊS, C. R. L. VIJAYKUMAR, N. L. **The Influence of Interference Networks in QoS Parameters in a WLAN 802.11g Environment**. Lecture Notes in Computer Science, v. 4742, p. 932-945, (2007).
- BAYRAKTAR M. E., HASTAK M. A. **Decision Support System for Selecting the Optimal Contracting Strategy in Highway Work Zone Projects**. Automation in Construction, Elsevier Science Publishers, Vol. 18, Issue 6, October, pp. 834-843, (2009).
- BEASLEY, D., BULL, D. R., MARTIN, R.R. **An overview of Genetic Algorithms: part 1, fundamentals**. University Computing, 15(2): 58-69, (1993).
- BOUAM S., OTHMAN J. B., DIVERCHY B., NAÏT-ABDESSELAM F., **802.11 Crosslayer Protocol Based Propagations Conditions Adaptation**, In 32nd IEEE Conference on Local Computer Networks, (2007).
- BOUROUHA, M. Ci, S. BRAHIM, G.B. GUIZANI, M. **A cross-layer design for QoS support in the 3GPP2 wireless systems**. In: Global Telecommunications Conference Workshops, 2004. GlobeCom Workshops, (2004).
- CETIC, **Pesquisa sobre o uso das Tecnologias da Informação e da Comunicação no Brasil , 2008**. Disponível em <http://www.cetic.br/usuarios/tic/2008-total-brasil/index.htm>, acessado em 15/11/2009, (2008).
- CLAVENNA, S. **Telco Triple Play: the DSL Imperative**. New York: Light Reading, July, 47p. (Heavy Reading, V.2, n. 17) (2004).
- CHEN, Z., **Data Mining and Uncertain Reasoning - an Integrated Approach**, John Wiley Professional (2001).
- CHENG, J., BELL, D. A., LIU, W. **Learning Belief Networks From Data: An Information Theory Based Approach**, Proceedings of the sixth international conference on Information and knowledge management, (1): 325-331, (1997).
- CIOFFI J. M. **Very-High-Speed Digital Subscriber Lines**, IEEE Commun. Mag., pp. 72–79, April, (1999).

- COOPER, G., HERSKOVITZ, E. **A Bayesian method for the induction of probabilistic networks from data.** *Machine Learning*, (9): 309–347, (1992).
- DAWID, P. **Statistical theory – the frequential approach.** *Journal of the Royal Statistical Society*. (147): 178-292 (1984).
- EKLUND C., MARKS R. B., PONNUSWAMY S., STANWOOD K. L., **WirelessMAN: Inside the IEEE 802.16 Standard for Wireless Metropolitan Area Networks**, IEEE Press, (2006).
- ESHELMAN, L. J.; CARUANA, R. A., SCHAFFER, J. D. **Biases In The Crossover Landscape.** *Proceedings of the Third International Conference on Genetic Algorithms*, Morgan Kaufmann (1): 1009, (1989)
- FAUDON, V.; VLEESCHAUWER, D.; FESTRAETS, E.; ROSS, P.. **End-User Services for Broadband uptake in High-Growth Economies**, White Paper, Alcatel Telecommunications, September, (2006).
- FAYYAD, U. **Data Mining and Knowledge Discovery: making sense out of data.** *IEEE Expert*, 11(5): 20-25 (1996).
- FAYYAD, U., PIATESKY-SHAPIO, G., SMYTH, P., UTHURUSAMY, R., **Advances in Knowledge Discovery and Data Mining**, AAAI Press, 1996.
- GRFENSTETTE, J. **Optimization of control parameters for genetic algorithms.** *IEEE Transactions on Systems, Man and Cybernetics*. 16 (1): 122028, (1986).
- GOLDBERG, D.E. **Genetic Algorithms in search, optimization and machine learning.** Addison-Wesley (1989).
- GOLDSCHMIDT R., PASSOS E. **Data Mining: um guia prático.** Editora Campus, Rio de Janeiro: Elsevier, (2005).
- GONCALVES, G.E.; SANTOS, G.G.B.; FIDALGO, J.F.; KELNER, J.; SADOK, D.H.F., FERNANDES, S.F. **On the Use of an ADSL2+ Testbed for Video Quality Assessment.** *IEEE International Conference on Communications, ICC '09*, Page(s): 1-6, (2009).
- HECKERMAN, D.; GEIGER, D. e CHICKERING, D. **Learning Bayesian networks: the combination of knowledge and statistical data.** *Machine Learning*, 20: 197–243, (1995).
- HECKERMAN, D. **Bayesian networks for Data Mining.** *Data Mining and Knowledge Discovery*, Kluwer Academic Publishers, (1): 79-119, (1997).
- HOLLAND, J. H. **Adaptation in Natural and Artificial Systems.** University of Michigan Press: Ann Arbor, MI, (1975).

HUANG, H., SONG, H., TIAN, F., LU, Y., WANG, Q. **A comparatively research in incremental learning of Bayesian networks**, Intelligent Control and Automation. Fifth World Congress on, Volume 5, 4260-4264 (2004).

IWS - INTERNET WORLD STATUS, **Usage and Population Statistics, 2009**. Disponível em <http://www.internetworldstats.com/stats.htm>, acessado em 20/12/2009, (2009).

IPERF, Disponível em <http://www.noc.ucf.edu/Tools/Iperf/>, acesso em 16/12/2009, (2009).

ITU-T RECOMMENDATION G.992.5, **Asymmetric Digital Subscriber Line (ADSL) transceivers**, Extended bandwidth ADSL2 (ADSL2+), January, (2005).

ITU-TREC.H.264, **Advanced Video Coding for Generic Audio visual Services**, ITU-TRec.H.264 and ISO/IEC14496-10(MPEG-4AVC), ITU-Tand ISO/IEC JTC1, Version1:May2003, Version2:May2004, Version3:March2005, Version4:September2005, Versions5 and 6:June2006, Version7:April2007, Version8(includingSVCextension): consented in July, (2007).

KAMACI N.; ALTUNBASAK Y. **Performance Comparison Of The Emerging H.264 Video Coding Standard With The Existing Standards**. In: Proceedings of IEEE International Conference on Multimedia and Expo,Baltimore; p.345–8, (2003).

KERPEZ K., WARING D. L., GALLI S., DIXON J., MADON P. **Advanced DSL management**, IEEE Commun. Mag., vol. 41, no. 9, pp. 116–123, September, (2003).

KERPEZ K. J., KINNEY R., **Integrated DSL Test, Analysis, and Operations**, IEEE Transactions On Instrumentation And Measurement, Vol. 57, No. 4, April, (2008).

KIM, J.; CHUNG, K. **C-Snoop: Cross Layer Approach to Improving TCP Performance over Wired and Wireless Networks**.In: IJCSNS International Journal of Computer Science and Network Security, V.7, N.3, March, (2007).

KLIAZOVICH D. E GRANELLI F.; **A Cross-layer Scheme for TCP Performance Improvement in Wireless LANs**, IEEE Global Telecommunications Conference, GLOBECOM, vol. 2, pp. 840-844, (2004).

KORB K. B., NICHOLSON A. E. **Bayesian Artificial Intelligence**, CRC PRESS, (2003)

KUROSE, J. F., ROSS, K. W. **Redes de Computadores e a Internet - Uma abordagem top-down**. Addison Wesley. 3ª Edição. (2006).

LAM, W., BACCHUS, F. **Learning Bayesian Belief Networks: An Approach Based On The MDL Principle**. Computational Intelligence, 10(4): 269–293, (1994).

LIE A., KLAUE J., **Evalvid-RA: Trace Driven Simulation of Rate Adaptive MPEG-4 VBR Video**, Multimedia Systems, Volume 14, Number 1, 13. November, (2007).

- LIU C.; SHI L., LIU B. **Utility-Based Bandwidth Allocation for Triple-Play Services**. Em Fourth European Conference on Universal Multiservice Networks, ECUMN '07, Page(s): 327 – 336, (2007).
- MANNILA, H., **Data Mining: Machine Learning, Statistic and Databases**, Department of Computer Science, University of Helsinki, (1997).
- MICHALEWICZ, Z. **Genetic algorithms + data structures = evolution programs**. AI Series. Springer-Verlag (1994).
- MORROW, M.; VIJAYANADA, K. **Developing IP Based Service**. San Francisco: Morgan Kaufmann, 311 p. (2003).
- OLSEN, B.; KATSIANIS, D.; VAROUTAS, D.; STORDAHL, K.; HARNO, J.; ELNEGAARD, N.; WELLING I.; LOIZILLON, F.; MONATH, T.; CADRO, P. , **Technoeconomic Evaluation of the Major Telecommunication Investment Options for European Players**, IEEE Network, vol. 20, issue 4, pp.6-15, July/August, (2006).
- OPENH323, **Open Nam**. Disponível em <http://openh323.sourceforge.com/>, acessado em 10/10/2009, (2009).
- OTHMAN J. B., BOUAM S., DIVERCHY B., LAUQUIN N., Vanden-Abeelee F. **Facing 802.11 Anomaly and Improving 802.11 WLANs QoS Using a Crosslayer Design Based Unselfish Behavior**. ICN/ICONS/MCL 2006: 143, (2006).
- PAPAGIANNI C. A.; TSELIKAS N. D., KOSMATOS E. A.; PAPAPANAGIOTOU S., VENIERIS I. S.. **Performance Evaluation Study For Qos-Aware Triple Play Services Over Entry-Level Xdsl Connections**, Journal of Network and Computer Applications, 32, 215-225, (2009).
- PEARL, J., **Probabilistic Reasoning in Intelligent System**, Morgan Kaufmann Publishers (1988).
- PINILLA, I.; AREIZAGA E.; RIUS R. J.; BERG, M.; TROJER, E., OLSSON J. **Quality Multimedia Distribution Through Resource Management And Performance Monitoring Over DSL Networks**. Em IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, (2008).
- POINT TOPIC, **Broadband Forum Announces Broadband And IPTV Statistics For Q2-2009, 2009**. Disponível em <http://point-topic.com/content/dslanalysis/bbwfq209.html>, acessado em 22/12/2009, (2009).
- PRIMET P.; HAKALY R., BONNASSIEUX F. **Experiments of Network Throughput Measurement and Forecasting Using the Network Weather**, No 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid, 21-24 May, (2002).
- RADCOM, **The State of Art**, Disponível em <http://www.radcom.com>, acessado em 11/12/2009, (2009).
- RAMONI, M., SEBASTIANI, P. **Discovering Bayesian Networks In Incomplete Databases**. Knowledge Media Institute, The Open University, Kmi Technical Report, nº 46, (1997).

- RAO N. S. V.; POOLE S. W.; WING W. R.; CARTER S. M. **Experimental Analysis of Flow Optimization and Data Compression for TCP Enhancement**, In IEEE INFOCOM Workshops, IEEE 19-25, April, (2009).
- REZENDE, S. O. **Sistemas Inteligentes – Fundamentos e Aplicações**. Manole (2003).
- ROCHA C. A.; SANTANA A. L.; FRANCES C. R. L.; REGO L. **Sistema de Suporte à Decisão para Predição de Cargas e Modelagem de Dependência em Sistemas Elétricos de Potência**, Anais do XXVI Congresso da SBC, Campo grande, julho, (2007).
- RUSSEL, S., NORVIG, P. **Artificial Intelligence – a modern approach**, Prentice Hall. (2003).
- SADRI S. M. R.; HARANDI Y. N.; PIRHADI M.; WASKASI M. Y.; TABRIZIPOOR A. I., MIRZABAGHI M.. **Test strategy for DSL Broadband IP Access Services**, in High Capacity Optical Networks and Enabling Technologies, HONET, (2007).
- SALZBERG, S. L. **On comparing classifiers: pitfalls to avoid and a recommended approach**. Data Mining and Knowledge Discovery, 1 (3): 317-328 (1997).
- SPIRITES, P.; GLYMOUR, C., SCHEINES, R. **Causation, Prediction, And Search**. MIT Press, (2001)
- STARR T., CIOFFI J. M., SILVERMAN P. J., **Understanding Digital Subscriber Line Technology**, Prentice Hall, (1999).
- STARR T., SORBARA M., CIOFFI J. M., SILVERMAN P. J., **DSL Advances**, Prentice Hall, (2003).
- STOLLE R. **Electromagnetic Coupling of Twisted Pair Cables**, IEEE Journal on Selected Areas in Communications, vol. 20, pp. 883-892, June (2002).
- SYMEONIDIS, A. L., MITKAS, P. A., **Agent Intelligence Through Data Mining: Multiagent Systems, Artificial Societies, and Simulated Organizations**, Springer, (2006).
- TAN E., GUO L., CHEN S., ZHANG X. **CUBS: Coordinated Upload Bandwidth Sharing In Residential Networks**. Em 17th IEEE International Conference on Network Protocols, ICNP 2009, (2009).
- TELEBRÁS 225-540-788, **Sistema de Documentação Telebrás**, Abril de (1997).
- THOMPSON, R. **IPTV and The Future of Telecom Video Network Architectures**. New York: light Reading, January, 84 p. (Heavy Reading, V. 3, n. 9) (2005).
- TURBAN, E., ARONSON, J. E. **Decision Support Systems and Intelligent Systems**. Prentice-Hall (2001).
- TR-048, **ADSL Interoperability Test Plan**, DSL Forum Technical Report, April, (2002).
- TR-126, **Triple-play Services Quality of Experience (QoE) Requirements**, DSL Forum, December, (2006).
- VLC, **VLC Media PLayer**. Disponível em www.videolan.org/vlc, acessado em 10/10/2009, (2009).

XIAO Y., DU X., ZHANG J., HU F., GUIZANI S.. **Internet Protocol Television (IPTV): The Killer Application For The Next-Generation Internet**. IEEE Commun Mag 2007;45(11):126–34, (2007).

ZAPATE, M. N., BRESSAN, G. **A Proposed Approach for Quality of Experience Assurance of IPTV**. In: First International Conference on the Digital Society, 1, 2007. Proceedings of IEEE. Guadeloupe, p. 25-30 (2007).

WALLACE W.; HUMPHREY L.; KIRKBY R., PITT C.. **Enhanced DSL Algorithms - Deliverable number DB2.2**, MUSE, (Multi-Service Access Everywhere) Project, December (2005).

WILSONA, A. G., HUZURBAZAR, A. V. **Bayesian Networks For Multilevel System Reliability**. Reliability Engineering & System Safety, 92,(10), pp.1413-1420, (2007).

WIMAX FORUM, Disponível em <http://www.wimaxforum.org/>, acessado em 10/11/2009, (2009).

YONGLI, Z.; LIMIN, H. LIGUO, Z.; YAN, W. **Bayesian Network Based Time-Sequence Simulation for Power System Reliability Assessment**. Mexican International Conference on Artificial Intelligence, pp. 271-277. Seventh Mexican International Conference on Artificial Intelligence, (2008).

APÊNDICE A

PUBLICAÇÃO CIENTÍFICA

Publicações obtidas a partir desta tese.

Capítulos de livro:

1. Cardoso D. L., Santana A. L., Rocha C. A., Francês C. R. L. **A Novel Probabilistic Approach for Analysis and Planning of Large Capillarity Broadband Networks Based on ADSL2+ Technology.** In Bayesian Networks, SCIYO (2009)(Capítulo de livro aceito). Previsão de publicação: junho, 2009.

Periódicos:

1. CARDOSO D. L., SERUFFO M. C. R., FRANCÊS C. R. L., COSTA J. C. W. A., SILVA M. S., SOUZA L. V., RUSSILO D., CASTRO A. L. S., CAVALCANTE G. P. S., RIU J. R. I. **Impact of Non-Stationary Noise on xDSL Systems: an Experimental.** Proceedings of SPIE. , v.1, p.1 - , 2007.
2. CARDOSO D. L., COSTA J. C. W. A., FRANCÊS C. R. L., SOUZA J. A. M., SERUFFO M. C. R., RIU J. R. I., SOUZA, L. V. **Performance evaluation of IPTV traffic over pDSL.** Proceedings of SPIE. , v.6776, p.1 - 15, 2007.
3. CARDOSO D. L., COSTA J. C. W. A., FRANCÊS C. R. L., RIU, J. R. I., SERUFFO M. C. R., CASTRO, A. L. S., CAVALCANTE, G. P. S., SOUZA, L. V. **Triple play service under the impact of stationary noise in a DSL system: an Amazon approach.** Proceedings of SPIE. , v.6776, p.1 - 10, 2007.

Trabalhos publicados em anais de congressos:

1. CARDOSO D. L., COSTA J. C. W. A., FRANCÊS C. R. L., SERUFFO M. C. R., SILVA M. S., RIU J. R. I., SOUZA L. V. **Avaliação de Desempenho de Tráfego IPTV sobre pDSL - Uma Abordagem baseada em Aferição**, In: XXV Simpósio Brasileiro de Telecomunicações (SBTr'07), 2007, Recife. Anais do XXV Simpósio Brasileiro de Telecomunicações. , 2007.
2. CARDOSO D. L., COSTA J. C. W. A., FRANCÊS C. R. L., SERUFFO M. C. R., **Triple Play sobre ADSL2+ na Região Amazônica: Um Estudo de Caso envolvendo Experimentações e Simulações**, In: XXXIV Seminário Integrado de Software e Hardware, 2007, Rio de Janeiro. Anais do Congresso da Sociedade Brasileira de Computação. , 2007.
3. CARDOSO D. L., NASCIMENTO K., COUTINHO M. M., FRANCÊS C. R. L., COSTA J. C. W. A., **Proposta de Canal de Retorno para TVD Brasileira em um Cenário Típico da Região Amazônica**, In: Simpósio Brasileiro de Telecomunicações, 2005, Belém. Simpósio Brasileiro de Telecomunicações. , 2005.

ANEXO I

Cardoso D. L., Santana A. L., Rocha C. A., Francês C. R. L. **A Novel Probabilistic Approach for Analysis and Planning of Large Capillarity Broadband Networks Based on ADSL2+ Technology.** In Bayesian Networks, SCIYO (2009) (Capítulo de livro aceito). Previsão de publicação: junho, 2009.