



Universidade Federal do Pará
Instituto de Tecnologia
Faculdade de Engenharia da Computação

José Edicleiton Silva Wanzeler

Análise Bayesiana para Identificação do Crosstalk em Sistemas DSL

Belém
2011



Universidade Federal do Pará
Instituto de Tecnologia
Faculdade de Engenharia da Computação

José Edicleiton Silva Wanzeler

Análise Bayesiana para Identificação do Crosstalk em Sistemas DSL

Trabalho de Conclusão de Curso apresentado à Faculdade de Engenharia da Computação do Instituto de Tecnologia da Universidade Federal do Pará para obtenção do grau de Bacharel em Engenharia da Computação, sob orientação do Prof. Dr. João Crisóstomo Weyl Albuquerque Costa.

Belém

2011

Análise Bayesiana para Identificação de Crosstalk em Sistemas DSL

Trabalho de Conclusão de Curso submetido a avaliação da Faculdade de Engenharia da Computação do Instituto de Tecnologia da Universidade Federal do Pará e julgado adequado para obtenção do grau de Bacharel em Engenharia da Computação, sendo aprovado em sua forma final com conceito *EXCELENTE*.

Prof. Dr. João Crisóstomo Weyl Albuquerque Costa – Orientador
Faculdade de Engenharia da Computação – ITEC
Universidade Federal do Pará

Prof. Msc. Lamartine Vilar de Souza – Coorientador
Faculdade de Ciências Exatas e Tecnologia – FACET
Universidade Federal do Pará

Prof. Dr. Agostinho Luiz da Silva Castro – Membro
Faculdade de Engenharia da Computação – ITEC
Universidade Federal do Pará

Prof. Msc. Marcelino Silva da Silva – Membro
Faculdade de Engenharia da Computação – ITEC
Universidade Federal do Pará

Belém, 4 de julho de 2011.

Para minha mãe

Agradecimentos

Agradeço primeiramente a Deus; ao meu orientador Prof. João Crisóstomo; ao meu coorientador Prof. Lamartine; aos meus colaboradores Prof. Marcelino, Nayanne, Ramon e Rodrigo; aos integrantes do Laboratório de Eletromagnetismo Aplicado Liane, Roberto, Gilvan, Vinicius, Claudomiro, Lilian, Júnior, Amaury, Roberto Medeiros, Oswaldo, Gustavo, Ilan, Thiago, Ewerton, Renan Almeida, Renan, André e Lucas que propiciaram tão encorajador ambiente durante a produção deste trabalho; a fundação Ericsson Telecommunications que liberou verbas para meus estudos e pesquisas; aos meus professores Agostinho, Zampolo, Eurípedes, Leônidas e Rosana que contribuíram de forma ativa para minha formação; aos meus amigos e colegas de curso Fernanda, Elton, Natália, Tiago, Bruno Guedes, Alex, Rômeo, Bruno Haick, Débora, Cláudio, Antônio, Joarley, Cássio, Janize, Leonardo, Felipe Coelho, Felipe Freire, João, Antônio Bessa, Flávio, Camilo, Elton Silva, Francis, Jessé, Júnior Sá, Maylson, Marina, Jairo, Rafael, Renan Amador, Marcio Valerio, Lidinaldo, Hibernon, Sandro, Werson, Netinho; aos meus pais Maria e José Wanzeler; aos meus irmãos Cláudia, Erielson, Kátia, Rose, Áurea e Célia; e aos meus sobrinhos Fabricio, Tayane, Talyne, Taynara, Fabiane, Thiago, Rayssa, Beatriz, Ednor, Eliel e Carina.

*History teaches us to be skeptical of “theoretical limits”,
which are sometimes exceeded by creative persons
who break the rules by creating a new model.*

— Starr, Cioffi, and Silverman. *Understanding
Digital Subscriber Line Technology.*

Resumo

Recentemente, a tecnologia de linha digital do assinante (Digital Subscriber Line — DSL) ganhou popularidade como uma tecnologia de acesso de alta velocidade capaz de prover uma larga gama de serviços multimídia. Porém, um grande problema da tecnologia DSL é o *crosstalk* (diafonia) proveniente de outras linhas telefônicas presentes no mesmo cabo. Além disso, a natureza aleatória dessa interferência não permite de forma trivial sua identificação e, conseqüentemente, mitigação. Dessa forma, este trabalho pretende apresentar uma nova análise a fim de propor um modelo capaz de identificar a presença, assim como a margem desse ruído presente no canal. Essa análise será feita sobre a mais recente versão da tecnologia DSL, a segunda geração do DSL a taxas muito altas (Very-high bit rate DSL – VDSL) ou VDSL2.

O modelo proposto para descrever a margem de ruído *crosstalk* é baseado em regressão linear e a principal funcionalidade do modelo de regressão está assentada em seus parâmetros. Assim, será feita uma análise bayesiana sobre um conjunto de dados contendo variáveis que exprimem o estado do enlace VDSL2 no momento da transmissão com o objetivo de estimar os parâmetros do modelo. Desejando delimitar a análise, decidiu-se por fazer a identificação e a quantificação da margem de ruído apenas do lado da central telefônica, porém é possível também utilizar o método apresentado para produzir um modelo capaz de identificar o nível desse ruído do lado das instalações do cliente.

Palavras-chave: Bayes, Regressão, *crosstalk*, DSL

Abstract

Recently, the digital subscriber line (DSL) technology has gained popularity as a high-speed access technology capable of providing a wide range of multimedia services. However, one major problem in DSL technology is the crosstalk from other phone lines present in the same cable. In addition, the random nature of this interference does not allow so trivial its identification and, consequently, mitigation. Thus, this work intends to present a new analysis in order to propose a model able to identify the presence of this noise, as well the margin of this noise in the channel. This analysis will be made on the latest version of DSL technology, the second generation of Very-high bit rate DSL (VDSL) or VDSL2.

The model proposed to describe the crosstalk noise margin is based on linear regression and the main feature of the regression model is built on its parameters. Thus, it will be done a bayesian analysis on a data set containing variables that express VDSL2 link state in the moment of transmission in order to estimate the model parameters. Wishing to delimit the analysis, it was decided to do the identification and quantification of the noise margin just on the central office side, but it is also possible to use the method shown to produce a model capable of identifying the level of this noise on the customer premises.

Keywords: Bayes, Regressão, crosstalk, DSL

Estrutura do Trabalho

Este trabalho se estende por 6 capítulos:

Capítulo 1 introduz o contexto sobre o qual o trabalho foi desenvolvido, as principais motivações e os objetivos que se pretende alcançar ao término do mesmo.

Capítulo 2 trata dos sistemas DSL apontando as características do sistema de telefonia sobre o qual a tecnologia DSL foi instalada. Além de uma breve descrição sobre a família de tecnologias DSL que foram desenvolvidas e finaliza com as principais interferências presentes em enlaces DSL.

Capítulo 3 introduz conceitos sobre a utilização da inferência bayesiana, bem como sobre o modelo de regressão linear. Ambos necessários para que se atinja o objetivo de utilizar a análise bayesiana para estimar os parâmetros do modelo de regressão linear.

Capítulo 4 trata dos aspectos relacionados à coleta dos dados utilizados para análise, seleção das variáveis e estudo dessas variáveis. Além disso, faz-se a elicitação da informação *a priori* que será utilizada para a computação bayesiana.

Capítulo 5 contém informações sobre os resultados obtidos, como a análise dos parâmetros estimados para o modelo de regressão, os resultados calculados através do modelo de regressão e os erros de regressão associados ao modelo.

Capítulo 6 ressalta os principais objetivos alcançados por este trabalho, aponta certas limitações e sugere melhorias e atividades para pesquisas futuras.

Lista de Figuras

2.1	Enlace Local	6
2.2	Diagrama de blocos de um modelo de referência DSL	7
2.3	Linha temporal da sequência de ativação SHDSL	11
2.4	Modelo de referência ADSL	12
2.5	Seção de um cabo contendo fios de cobre	18
2.6	Acoplamento NEXT	22
2.7	Acoplamento FEXT	23
4.1	Configuração utilizada para as medições.	39
4.2	SNR do lado da central telefônica.	43
4.3	SNR do lado das instalações do cliente.	43
4.4	Taxa de transmissão real.	44
4.5	Taxa máxima de dados atingível do lado da central telefônica.	45
4.6	Taxa máxima de dados atingível do lado das instalações do cliente.	45
4.7	Potência total de saída transmitida pela central telefônica.	46
4.8	Potência total de saída transmitida pelas instalações do cliente.	47
4.9	Atraso de <i>interleave</i> do lado da central telefônica.	47
4.10	Real proteção ao ruído impulsivo em direção de downstream.	48
4.11	Comprimento do bloco de dados do canal com CRC.	49
4.12	Taxa de <i>bits</i> real da linha na direção de <i>downstream</i>	49
4.13	Gráfico de dispersão e histograma das variáveis SNR do lado das instalações do cliente e SNR do lado da central telefônica.	52
4.14	Gráfico de dispersão e histograma das variáveis taxa de transmissão real neste canal e SNR do lado da central telefônica.	54

4.15	Gráfico de dispersão e histograma das variáveis taxa de dados máxima atingível atualmente pela central telefônica e SNR do lado da central telefônica.	56
4.16	Gráfico de dispersão e histograma das variáveis taxa de dados máxima correntemente atingível pelas instalações do cliente e SNR do lado da central telefônica.	58
4.17	Gráfico de dispersão e histograma das variáveis potência total de saída medida transmitida pela central telefônica e SNR do lado da central telefônica.	60
4.18	Gráfico de dispersão e histograma das variáveis potência total de saída medida transmitida pelas instalações do usuário e SNR do lado da central telefônica.	62
4.19	Gráfico de dispersão e histograma das variáveis atraso de <i>interleave</i> do lado da central telefônica e SNR do lado da central telefônica.	64
4.20	Gráfico de dispersão e histograma das variáveis real proteção ao ruído impulsivo em direção de <i>downstream</i> e SNR do lado da central telefônica.	66
4.21	Gráfico de dispersão e histograma das variáveis comprimento do bloco de dados do canal no qual o CRC opera e SNR do lado da central telefônica.	68
4.22	Gráfico de dispersão e histograma das variáveis taxa de <i>bits</i> real da linha na direção de <i>downstream</i> e SNR do lado da central telefônica.	70
5.1	Curvas medidas e calculadas para a variável SNR do lado da central telefônica utilizando-se o modelo com <i>priori</i> não informativa.	77
5.2	Curvas medidas e calculadas para a variável SNR do lado da central telefônica utilizando-se o modelo com <i>priori</i> informativa.	77
5.3	Erro de regressão para a variável SNR do lado da central telefônica utilizando-se o modelo com <i>priori</i> não informativa.	79
5.4	Erro de regressão para a variável SNR do lado da central telefônica utilizando-se o modelo com <i>priori</i> informativa.	79
5.5	Histograma do erro de regressão para a variável SNR do lado da central telefônica utilizando-se o modelo com <i>priori</i> não informativa.	80

5.6	Histograma do erro de regressão para a variável SNR do lado da central telefônica utilizando-se o modelo com <i>priori</i> informativa. . .	80
-----	---------------------------------------------------------------------------------------------------------------------------------------------	----

Lista de Tabelas

4.1	Média dos parâmetros <i>a priori</i> e <i>a posteriori</i>	72
4.2	Desvio padrão dos parâmetros <i>a priori</i> e <i>a posteriori</i>	72
4.3	Média e desvio padrão do erro de precisão <i>a priori</i> e <i>a posteriori</i>	73
5.1	Parâmetros <i>a posteriori</i> do modelo utilizando-se <i>priori</i> informativa e não informativa.	75

Sumário

1	Introdução	1
1.1	Motivação	1
1.2	Objetivos	2
2	Sistemas DSL	4
2.1	O Sistema de Telefonia	5
2.2	Visão Geral sobre a Tecnologia DSL	7
2.3	A Família de Tecnologias DSL	8
2.3.1	ISDN	9
2.3.2	HDSL	10
2.3.3	SHDSL	11
2.3.4	ADSL	12
2.3.5	VDSL	14
2.4	Características Físicas dos Fios de Cobre	18
2.4.1	Isolamento, Composição do Núcleo e Blindagem dos Pares Trançados	18
2.5	Interferências em Enlaces DSL	19
2.5.1	Ruído Impulsivo	20
2.5.2	Interferência de Rádio Frequência	20
2.5.3	Crosstalk	21
3	Inferência Bayesiana	24
3.1	Introdução	24
3.1.1	Teoria Bayesiana	24
3.1.2	Computação Bayesiana	28

3.1.3	Programas para Computação Bayesiana	29
3.2	Modelo de Regressão Linear	29
3.3	Função de Verossimilhança	31
3.4	Densidade <i>a priori</i>	33
3.5	Densidade <i>a posteriori</i>	34
3.6	Predição	36
4	Análise Bayesiana para Identificação de Crosstalk em Sistemas DSL	38
4.1	Configuração Utilizada para as Medições	39
4.2	Seleção das Variáveis Explanatórias	41
4.3	Características das Variáveis Presentes na Base de Dados	42
4.3.1	adslAtucCurrSnrMgn	42
4.3.2	adslAturCurrSnrMgn	42
4.3.3	adslAtucChanCurrTxRate	44
4.3.4	adslAtucCurrAttainableRate	44
4.3.5	adslAturCurrAttainableRate	44
4.3.6	adslAtucCurrOutputPwr	46
4.3.7	adslAturCurrOutputPwr	46
4.3.8	adslAtucChanInterleaveDelay	46
4.3.9	adslAtucProprietaryChanActualInp	48
4.3.10	adslAtucChanCrcBlockLength	48
4.3.11	adslAtucProprietaryPhysXActualLineBitRate	48
4.4	Elicitação da Informação <i>a priori</i>	50
4.4.1	Obtenção dos Hiperparâmetros para a Densidade <i>a priori</i>	50
4.5	Computação Bayesiana	71
5	Análise dos Resultados	74
5.1	Análise dos Parâmetros Estimados para o Modelo de Regressão	74
5.2	Resultados Calculados através do Modelo de Regressão	76
5.3	Erros de Regressão	78
6	Considerações Finais	81
6.1	Objetivos Alcançados	81

6.2	Limitações deste Trabalho e Pesquisas Futuras	84
-----	---------------------------------------------------------	----

Capítulo 1

Introdução

Este trabalho trata da identificação do ruído *crosstalk* no enlace local de sistemas DSL (Digital Subscriber Line). DSL é uma tecnologia que provê transportes a altas taxas de informação digital sobre linhas telefônicas do assinante [1]. As linhas telefônicas são uma estrutura de tal abrangência que é capaz de conectar residências e locais de trabalho ao redor do mundo todo. Esse nível de maturidade que a rede telefônica possui, aliado ao fato de se poder transportar dados digitais em alta velocidade, a um custo relativamente baixo sobre as modificações necessárias faz com que essa tecnologia seja bastante atrativa. Além disso, a crescente demanda por taxas de transmissão mais altas gerada por serviços de comunicação digital que transportam voz, vídeo e dados a altas taxas de transmissão torna os sistemas DSL ainda mais populares como ferramenta de comunicação entre as pessoas.

1.1 Motivação

O *crosstalk* juntamente com outras interferências como o ruído impulsivo e sinais de radiofrequência são os maiores limitantes para que se atinjam taxas de transmissão de dados mais altas. No canal DSL, o tipo de ruído dominante é o *crosstalk* proveniente de outros sistemas DSL. É ele que tipicamente determina a taxa máxima de dados que se pode esperar atingir. *Crosstalk* é a interferência dentro de um canal proveniente da potência de transmissão de um sinal de um

outro canal [2].

Dessa forma, o *crosstalk* é o acoplamento eletromagnético entre pares trançados de fio de cobre no mesmo cabo. Apesar de os picos das funções de acoplamento parecerem seguir uma tendência, essas funções são na verdade aleatórias. Essas complicadas e imprevisíveis formas das funções de acoplamento fazem com que a maioria dos trabalhos práticos utilize modelos simplificados de tendências. O acoplamento segue diferentes tendências dependendo das configurações da rede [2].

De principal interesse para este estudo é o DSL a taxas muito altas (Very-high bit rate DSL – VDSL), mais especificamente a segunda geração de VDSL, por se tratar da mais recente tecnologia DSL desenvolvida e capaz de atingir as mais altas taxas de transmissão. Além disso, a primeira geração do DSL a taxas muito altas (VDSL) surgiu como o “próximo passo” além do DSL assimétrico (ADSL) [3] que é a tecnologia mais largamente empregada no mundo. E essa segunda geração, o VDSL2, buscava incorporar as melhores partes do ADSL e do VDSL em uma única especificação. Essa tecnologia aplicada a enlaces curtos pode suportar taxas de dados simétricas de até 100 Mb/s utilizando um espectro de até 30 MHz [3].

1.2 Objetivos

O objetivo deste trabalho é identificar o nível de ruído *crosstalk* do lado da central telefônica em enlaces DSL, mais especificamente em enlaces VDSL2 e verificar a influência que as características medidas do estado do enlace têm sobre o nível desse ruído. Para se atingir este objetivo, será utilizada uma abordagem bayesiana que estima os parâmetros de um modelo de regressão linear com múltiplas variáveis explanatórias [4] que são valores correntes de características da linha telefônica durante a transmissão DSL, buscando explicar o nível de ruído *crosstalk* do lado da central telefônica. Para realizar-se esta análise, medições foram feitas para diferentes situações: na ausência de ruído, com ruído *crosstalk* inserido no enlace na direção de *upstream* e com ruído *crosstalk* inserido na direção de *downstream*.

Os resultados das medições mostram que nem todas as variáveis que foram

coletadas influenciam no nível de ruído *crosstalk* do lado da central telefônica. As variáveis que possuem influência no nível de ruído *crosstalk* do lado da central telefônica são: a taxa de transmissão real no canal; a taxa de dados máxima atingível pela central telefônica; a taxa de dados máxima atingível pelas instalações do cliente; a potência total de saída transmitida pela central telefônica; a potência total de saída transmitida pelas instalações do cliente; o espaçamento relativo entre *bytes* de entrada na entrada intercaladora do canal e a sua localização no fluxo de *bits* na saída intercaladora; a real proteção ao ruído impulsivo; comprimento do bloco de dados do canal no qual a verificação de redundância cíclica (Cyclic Redundancy Check — CRC) opera; a taxa de bits da linha na direção de *downstream*.

Os resultados alcançados descritos neste trabalho são bastante significativos. Principalmente, porque a abordagem bayesiana para a estimação dos parâmetros do modelo de regressão linear que explica o nível de ruído do lado da central telefônica permite não apenas identificar a presença do ruído, como também é capaz de quantificar o nível desse ruído e descrever a influência que cada variável em particular possui sobre a variável nível de *crosstalk* do lado da central telefônica. Além disso, outra significância deste trabalho está no fato de a metodologia utilizada para formular o modelo apresentado, bem como estimar seus parâmetros pode ser aplicada a outros tipos de sistemas de telecomunicações para o mesmo fim. Dessa forma, pode-se verificar a presença do ruído, o nível do mesmo e as características do enlace que influenciam para o nível dessa interferência do lado da central telefônica a fim de se aplicar as medidas adequadas para sua mitigação.

Capítulo 2

Sistemas DSL

A tecnologia DSL transforma uma simples linha telefônica em um enlace de comunicação banda larga através sofisticadas técnicas de transmissão digital que compensam as muitas interferências comuns em linhas telefônicas como atenuação, *crosstalk* — sinais presentes em outros fios no mesmo cabo, reflexão, interferências de rádio frequência e ruído impulsivo. Dessa forma, a tecnologia DSL revolucionou a utilização de linhas telefônicas que utilizam pares trançados.

Embora existam vários outros meios de acesso banda larga para assinantes residenciais e corporativos, nenhum outro possui o nível de maturidade desenvolvido pela rede telefônica. São centenas de milhões de linhas telefônicas de pares trançados instalados mundialmente. Como citado por [5], a infraestrutura de pares trançados conecta quase todas as residências e locais de trabalho ao redor do mundo.

Porém as linhas digitais têm suas limitações. Apesar de tal abrangência a nível mundial, uma significativa quantidade de linhas telefônicas necessitam de medidas a fim de permitir operações DSL a altas velocidades. Portanto, medidas corretivas para longos enlaces, como repetidores de meia distância, por exemplo, precisam ser instaladas, além de melhorias nas técnicas de identificação e mitigação das interferências nos enlaces telefônicos.

Nas seções que seguem serão apresentados conceitos básicos sobre o funcionamento da tecnologia DSL desde o sistema telefônico sobre o qual foi implementada passando pelos diferentes tipos de DSL desenvolvidas e chegando às interferências presentes nos enlaces telefônicos. Também, serão introduzidas as principais características físicas dos fios de cobre empregados na rede telefônica e conceitos sobre qualificação de enlaces para serviços DSL. Com maior nível de detalhamento serão tratados a tecnologia VDSL e o ruído *crosstalk*.

2.1 O Sistema de Telefonia

O sistema de telefonia era um serviço ponto a ponto de curta distância que veio a se transformar em uma rede de altíssima maturidade interligando quase o mundo todo e possibilitando comunicação extremamente rápida mesmo a grandes distâncias [2].

Essa transformação se deu por meio do desenvolvimento de multiplexadores locais, *hubs* físicos situados nas extremidades das múltiplas linhas telefônicas. Esses multiplexadores compõem os nós da rede de telefonia comutada pública (Public Switched Telephone Network — PSTN) usualmente chamada de antigo serviço de telefonia simples (Plain Old Telephone Service — POTS).

Outro reconhecido fator que propiciou tal alcance da rede de telefonia foi o desenvolvimento de *backbones* — esquemas de conexões centrais de um sistema mais amplo — para interligar vários multiplexadores locais permitindo a conexão interurbana e em seguida a internacional. Originalmente, os *backbones* utilizavam os tradicionais cabos telefônicos feitos de cobre ou alumínio, porém hoje em dia são compostos por fibra óptica e enlaces de satélite.

Como se observa na Figura 2.1, do lado direito do multiplexador encontram-se as linhas telefônicas compostas de tradicionais fios de cobre que fazem a conexão com as instalações do cliente. Essa parte da rede telefônica é conhecida como “enlace local” e também como “última milha” (last mile).

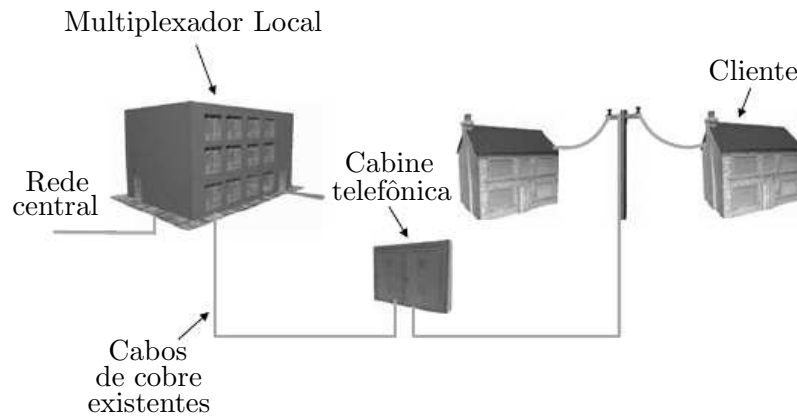


Figura 2.1: Enlace Local [2].

É sempre desejado minimizar o comprimento do enlace local [2], dado que é a parte mais lenta e densa do sistema telefônico. Para isso, faz-se uso de multiplexadores remotos — multiplexadores menores mais próximos às instalações do cliente — conectados aos multiplexadores tradicionais através de um enlace de alta velocidade, geralmente fibra óptica. Esses multiplexadores remotos são conhecidos como transportadoras de enlace digitais (Digital Loop Carrier – DLC).

Do ponto de vista da tecnologia DSL, o percurso de maior interesse é o enlace local que compreende a estrutura de fios de cobre entre o usuário e o multiplexador. Pois, por razões econômicas, mantém-se os cabos com pares trançados entre o cliente e o multiplexador [2], uma vez que o custo de substituir essa infraestrutura por outra que permita maior velocidade — fibra óptica, por exemplo, seria elevado demais.

2.2 Visão Geral sobre a Tecnologia DSL

A Figura 2.2 mostra o diagrama de blocos de uma configuração DSL habitual. Nesse sentido, a “linha digital do assinante” se refere ao enlace analógico local entre as instalações de cada cliente e sua respectiva central telefônica (Central Office – CO). Além disso, é necessário um modem DSL em cada extremidade do enlace.

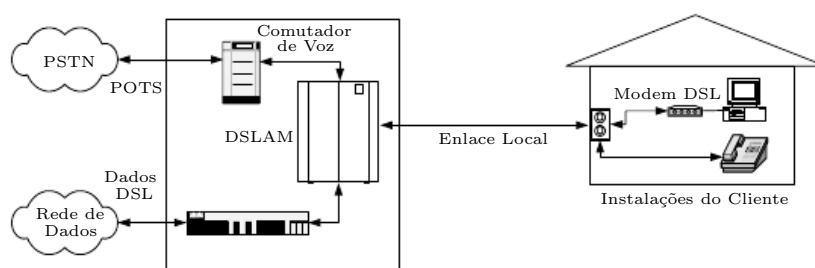


Figura 2.2: Diagrama de blocos de um modelo de referência DSL [2].

O serviço DSL pode ser considerado como sendo fornecido por meio de uma rede sobreposta que não faz parte da rede telefônica comutada normal, ou seja, as tecnologias DSL e de telefonia são aplicadas sobre a mesma infraestrutura de pares trançados. Dessa forma, o provedor de serviços precisa separar o serviço DSL do serviço POTS, de modo que os dados transportados via DSL devem ser enviados para uma rede de dados independente da rede de voz comutada.

O CO provê serviço DSL para as instalações do cliente usando multiplexadores de acesso DSL (DSL Access Multiplexer – DSLAM). Os DSLAMs geralmente contêm muitos modems servindo os múltiplos clientes.

Uma “linha digital do assinante” existe em um único enlace local entre as instalações do cliente e a central telefônica. Enquanto a conexão entre modems *voiceband* — que realizam modulação na faixa de frequência de voz entre computadores presentes nas extremidades da rede — inclui dois enlaces locais além dos elementos de rede entre eles.

Antes que modems DSL sejam usados para a transmissão de dados das aplicações, eles precisam passar por uma fase de ativação. Vários elementos receptores e transmissores precisam ser inicializados para que uma comunicação confiável seja realizada naquele enlace em particular sob as condições de interferência daquele momento.

A fase de ativação é também utilizada para permitir aos modems de cada extremidade da linha estipular qual configuração eles utilizarão, através dessa sessão chamada de *handshaking*.

2.3 A Família de Tecnologias DSL

A quantidade de tecnologias DSL é bem grande. Por conseguinte, esta seção pretende descrever brevemente apenas as principais tecnologias desenvolvidas e que estão atualmente em desenvolvimento.

De um ponto de vista simplista, [2] divide as tecnologias DSL em duas amplas classes:

- **Simétricas.** As taxas de dados transmitidos tanto na direção da central telefônica para o usuário — *downstream* — quanto na direção usuário-central telefônica — *upstream* — são as mesmas. Essa configuração é preferida por clientes corporativos visto que realizam transferências de dados em ambas as direções na mesma proporção.
- **Assimétricas.** Neste caso, ocorre assimetria dado que a taxa de *downstream* é mais alta que a taxa de *upstream*. Essa solução se apresenta melhor para clientes residenciais que fazem uso de aplicações como navegação Web, por exemplo.

Algumas das várias tecnologias são capazes de operar tanto de forma simétrica quanto assimétrica. Além disso, há casos de sistemas DSL que são capazes de operação a taxas variáveis.

Tecnologias DSL simétricas geralmente fazem uso modulação de banda base, cuja largura de banda do sinal transmitido se prolonga desde o início do espectro de frequências (0 Hz). Enquanto tecnologias assimétricas geralmente usam modulação passa faixas a qual evita as frequências mais baixas que poderiam ser usadas para serviços na faixa de voz como telefonia analógica por exemplo. Dessa forma, clientes residenciais que fazem uso de DSLs assimétricas poderiam continuar utilizando os serviços tradicionais da rede telefônica. Para usuários corporativos a utilização dos serviços telefônicos não constitui um problema, pois o seu tráfego de voz pode ser transportado através do enlace DSL.

As subseções que seguem descreverão brevemente os vários tipos de tecnologia DSL em ordem cronológica de desenvolvimento classificando-as entre as classes descritas anteriormente quando necessário.

2.3.1 ISDN

Considerada como a “tecnologia DSL original”, a rede digital de serviços integrados (Integrated Services Digital Network – ISDN) objetiva uma rede digital global para a transmissão de sinais de voz e sinais de dados de baixa velocidade.

Como descrito por [2], essa é uma tecnologia simétrica com taxa de transmissão de até 160 kbit/s a distâncias de aproximadamente 5,5 km podendo atingir maiores distâncias com o uso de repetidores. A taxa de transmissão é dividida entre dois canais de 64 kbit/s cada, um canal de 16 kbit/s para transmissão de dados e 16 kbit/s para controle. A maior parte da rede ISDN instalada utiliza modulação em amplitude por pulso (Pulse Amplitude Modulation – PAM) a uma frequência de 80 kHz.

2.3.2 HDSL

DSL a altas taxa de bits (High-bitrate Digital Subscriber Line – HDSL) [6] é o termo usualmente aplicado ao suprimento de taxa simétricas de até 2,048 Mbit/s sobre um, dois ou três pares trançados. A tecnologia HDSL faz uso de modulação PAM assim como o ISDN.

O HDSL geralmente usa dois pares, onde metade dos dados transmitidos é enviada por cada par. Porém o uso de um único par é bastante comum entre pequenos clientes corporativos em que o volume de tráfego não justifica o uso de uma taxa completa [2].

O uso de dois pares HDSL permite a operação sobre um percurso de até 3,7 km de fio de 0,5 mm, mas pode atingir distâncias superiores através de repetidores. O uso de dois pares ou mais ajuda a assegurar uma transmissão mais robusta a longas distâncias, uma vez que a largura de banda que é usada em cada um dos par é menor que aquela que seria usada em apenas um par, além de diminuir a atenuação do sinal. Porém, o uso de mais de um par por cliente significa que um número menor pares estará disponível para os demais clientes.

Como dito em [2], o desenvolvimento de uma versão do HDSL de alto desempenho foi executado visando o uso em um único par, porém mantendo a compatibilidade espectral com os sistemas existentes. Essa segunda geração da tecnologia HDSL (HDSL2) utiliza modulação PAM em 16 níveis discretos (16-PAM) a 517,33 kHz, permite a transmissão assimétrica de dados e atinge até 2,7 km com fio de 0,4 mm ou 3,7 km com fios de 0,5 mm. Outro desenvolvimento permitiu uma versão do HDSL2 que opera sobre dois pares trançados como o HDSL e é referida como HDSL4. O HDSL4 utiliza basicamente a mesma tecnologia do HDSL2, porém atinge distâncias de aproximadamente 3,4 km usando fios de 0,4 mm.

2.3.3 SHDSL

DSL de alta velocidade em um único par (Single-pair High-speed Digital Subscriber Line) é uma tecnologia DSL multitaxa definida pela norma G.991.2 [7]. A taxa de vazão de bits vai de 192 kbit/s até 2,304 Mbit/s com incrementos de 8 kbit/s.

A tecnologia SHDSL incorpora muitas características do HDSL2/4 como, por exemplo, taxas simétricas e operações multitaxa. Além disso, os elementos operacionais do SHDSL são específicos para cada região.

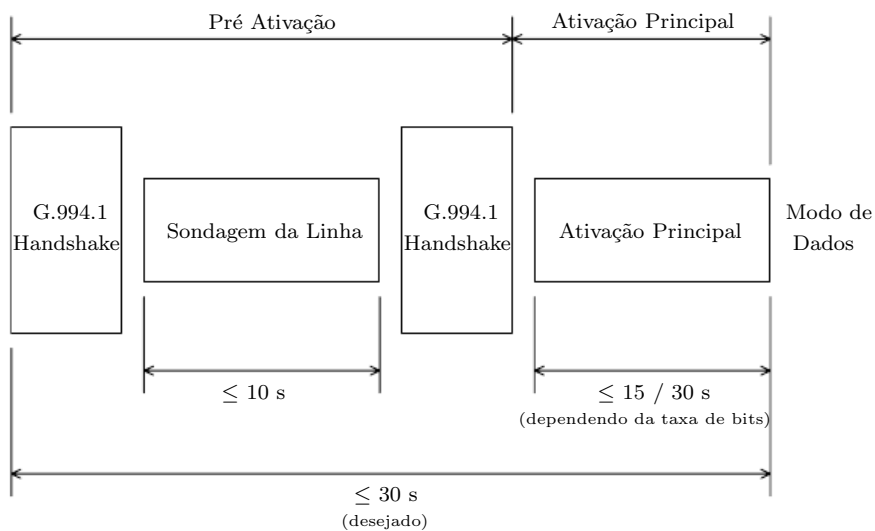


Figura 2.3: Linha temporal da sequência de ativação SHDSL [2].

Como pode ser verificado na Figura 2.3, a sequência de inicialização se dá em duas fases: pré ativação e ativação. Na pré ativação os modems de cada extremidade da linha trocam informações sobre suas capacidades em uma subseção conhecida como *handshake* como definido pela norma G.994.1 [8] e têm a opção de fazer uma sondagem da linha através do envio de sinais entre ambos para verificar quais taxas podem ser suportadas considerando-se o enlace em particular assim como as condições de ruído. Enquanto que durante a ativação principal os blocos funcionais do receptor e do transmissor são adaptados para que uma comunicação confiável seja estabelecida na taxa de bits definida previamente.

Desenvolvimentos adicionais foram anexados ao padrão original [7] para atingir melhorias na transmissão como, por exemplo: modo de operação estendido para múltiplos pares; provisões que permitem taxas de até 5,696 Mbit/s; e uma inicialização mais rápida.

2.3.4 ADSL

Como referido em [2], a definição original do DSL assimétrico (ADSL) [9] é a mais amplamente empregada em todo o mundo. As taxas de transmissão dessa tecnologia chegam a até 8 Mbit/s para *downstream* e até 896 kbit/s para *upstream* utilizando um par trançado. Também em [2] é citado que a motivação inicial para o desenvolvimento dessa tecnologia foi a aplicação de vídeo por demanda e em seguida o acesso à *Internet* de alta velocidade.

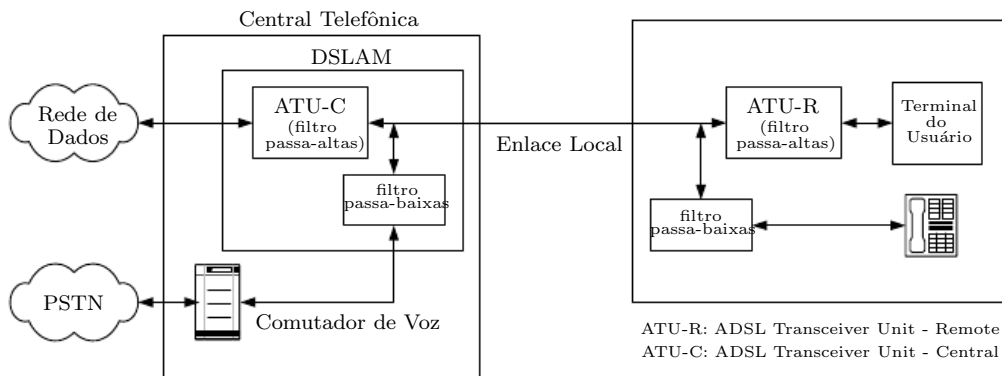


Figura 2.4: Modelo de referência ADSL [2].

Como mostrado nesta seção, a maioria das tecnologias DSL assimétricas utilizam modulação passa-faixa para evitar as frequências mais baixas utilizadas pelo serviço de telefonia. Assim, o espectro do sinal ADSL começa em aproximadamente 25 kHz. Para separar os dois serviços, um par de filtros divisores passa-altas e passa-baixas são necessários como pode ser visto no modelo de referência mostrado na Figura 2.4.

A tecnologia ADSL é baseada em modulação multi-tono discreta (Discrete Multi-tone – DMT) — trata-se do envio de múltiplos sinais em diferentes frequências, uma espécie de multiplexação por divisão de frequência — e usa constelações — representações de sinais modulados digitalmente — de tamanhos variáveis além de sofisticados algoritmos de alocação de bits para determinar onde melhor distribuir a energia na banda usável.

A largura de banda necessária para *upstream* é muito menor que a necessária para *downstream* dado que a taxa de *downstream* é muito superior à de *upstream*. A faixa de frequências para *upstream* é de 25 kHz a 138 kHz, enquanto a faixa para *downstream* utilizando sobreposição entre as bandas de *downstream* e *upstream* é de 25kHz a 1,104 MHz.

Como uma largura de banda maior implica uma taxa de transmissão também maior, é possível se utilizar a sobreposição de bandas através de sofisticadas técnicas de cancelamento de eco a fim de se obter maiores taxas de transmissão, porém essa sobreposição acarreta um aumento do nível de ruído *crosstalk* no canal de *upstream*. O uso do modo de transmissão baseado na divisão em frequência (Frequency Division Duplexing – FDD) evita a sobreposição fazendo com que a faixa de frequência de *upstream* inicie em 138 kHz [2].

Desde o desenvolvimento do ADSL, uma série de melhorias foram adicionadas de forma a proporcionar um aumento de desempenho em um padrão conhecido como ADSL2 [10]. Entre as melhorias destacam-se: a adição de uma constelação de um único bit para um desempenho mais robusto permitindo a operação sobre enlaces mais longos que antes não eram usados; a inclusão de um modo opcional totalmente digital para permitir a transmissão ADSL sobre a banda reservada para o serviço telefônico; e melhorias nos procedimentos para permitir uma inicialização mais rápida.

Além disso, desenvolveu-se uma especificação adicional para o ADSL2 chamada ADSL2+ [11] que teve como principal melhoria a extensão do limite superior da banda de *downstream* para 2,2 MHz que resulta em taxas de *downstream* mais altas em enlaces de comprimentos curto e médio .

2.3.5 VDSL

A DSL a taxas muito altas (Very-high bit rate DSL – VDSL) é a variante DSL de mais alta velocidade atualmente. De certa forma, o VDSL é uma extensão da tecnologia ADSL existente, porém essas taxas mais altas só podem ser alcançadas sobre enlaces mais curtos.

Para possibilitar o uso da tecnologia VDSL, uma transmissão de fibra óptica é usada para conectar a central telefônica a uma unidade de rede óptica (Optical Network Unit – ONU) remota. E a distância entre a ONU e as instalações do cliente é coberta usando transmissão VDSL sobre o par trançado comum de cobre. Essa arquitetura é conhecida como fibra-cobre (*fiber-copper*).

O VDSL pode suportar taxas de bits simétricas ou assimétricas. A primeira geração de VDSL (VDSL1), em particular, atinge taxas simétricas de 13 ou 26 Mbit/s, enquanto que pode atingir assimetricamente taxas de até 52 Mbit/s para *downstream* e 6,4 Mbit/s para *upstream*. Para que o VDSL alcance estas altas taxas de transmissão de bits, a tecnologia faz uso de larguras de banda que vão até 13 MHz. Através da capacidade de transmissão tanto simétrica quanto assimétrica, o VDSL torna-se conveniente tanto para a categoria de clientes residenciais quanto corporativos.

O emprego de VDSL geralmente concentra-se no uso do modo de transmissão FDD. A primeira geração de tecnologia faz uso tanto de modulação (Quadrature Amplitude Modulation / Carrierless Amplitude Phase – QAM/CAP) quanto DMT. Porém, a segunda geração (VDSL2) especifica apenas o uso de modulação DMT.

VDSL1

A tecnologia VDSL foi proposta para ser o “próximo passo” além do ADSL e objetivava um sistema capaz de prover taxas de transmissão mais elevadas que o ADSL em enlaces mais curtos. Sabendo-se que em enlaces mais curtos, as frequências acima daquelas ocupadas pela tecnologia ADSL eram úteis para a transmissão de dados, a tecnologia VDSL foi especificada sob tais condições de comprimento de enlace e faixa de frequência [12]. Inicialmente, o comprimento máximo para enlaces VDSL foi assumido ser de 1500 m, porém os comprimentos empregados geralmente encontram-se na faixa de 150 a 2000 m [2].

O advento do VDSL trouxe uma nova gama de serviços sobre a infraestrutura de cobre existente. A largura de banda ampliada e vazão do VDSL davam a possibilidade de oferecer o serviço de acesso à *Internet* de alta velocidade que era fornecido pela tecnologia ADSL instalada, porém era necessário o emprego de cenários diferentes do ADSL.

Esperava-se que a tecnologia VDSL oferecesse serviços tanto simétricos quanto assimétricos. Essa necessidade de suportar os dois tipos de serviços impacta sobre o projeto do sistema, especialmente sobre o plano de frequência apropriado.

Para clientes corporativos as taxas simétricas eram preferíveis, pois eles utilizam serviços como , por exemplo, acesso a redes locais (Local Area Network – LAN) que permitem ao usuário conectar-se remotamente à LAN da empresa a fim de fazer tanto *download* quanto o *upload* de dados. Além disso, videoconferências de alta qualidade, bastante comuns no ambiente corporativo, necessitam da mesma largura de banda para *downstream* e *upstream*.

Por outro lado, as taxas assimétricas eram direcionadas aos clientes residenciais cujo objetivo final era receber o serviço chamado *triple-play*, que consiste no serviço combinado de telefonia, transmissão de TV e acesso à *Internet*.

Porém, mesmo com tantos novos horizontes que a tecnologia VDSL podia proporcionar, o emprego da mesma resulta em uma cobertura limitada em que poucos

clientes poderiam ser beneficiados dadas as restrições de comprimento do enlace. Para aumentar a cobertura, o VDSL precisa ser instalado mais próximo do usuário final.

Além do problema de cobertura, o VDSL apresenta outros desafios mesmo para aqueles clientes que estão aptos a receber esse serviço. Como os enlaces VDSL são curtos, o *crosstalk* do tipo FEXT¹ é a fonte dominante de ruído principalmente nas altas frequências onde o VDSL opera.

É bastante provável que em um sistema VDSL empregado em um enlace curto que venha a dividir um cabo contendo pares de cobre chamado *binder* com um sistema VDSL empregado em um enlace longo – aproximadamente 1500 m, o FEXT causado pelo enlace mais curto reduziria significativamente o desempenho do *upstream* no enlace mais longo.

Como o VDSL1 transmite a frequências de até 12 MHz e o VDSL2 usa frequências ainda mais altas, pode ocorrer interferências provenientes dos espectros de rádio.

VDSL2

Diferentemente do VDSL1, não foram desenvolvidas versões regionais do VDSL2. As comissões de normalização contribuíram juntas diretamente para a construção de um único documento contendo os requisitos individuais: a Recomendação G.993.2 [14].

Além disso, os problemas quanto ao tipo de modulação adotado que ocorreram durante a padronização do VDSL1 convergiram para que a modulação DMT se tornasse padrão para o VDSL2 [3]. Dessa forma, uma recomendação VDSL bastante melhorada foi gerada.

Um dos objetivos estabelecidos para o VDSL2 era facilitar implementações ADSL2/VDSL2 multimodo, ou seja, ambas as tecnologias deveriam ser similares

¹*Far-end crosstalk* ocorre de um transmissor dentro de um receptor distante [13].

em termos de características, gerência e assim pudessem ser implementadas em uma única plataforma. No final das contas o VDSL2 era uma tentativa de incorporar as melhores partes dos padrões VDSL e ADSL existentes em uma única recomendação.

Desde o início, era evidente que os operadores tinham em mente uma larga gama de aplicações para esta nova tecnologia. Alguns operadores objetivavam uma banda bastante larga e taxa de dados extremamente alta para emprego em ambientes como Unidades com Múltiplas Habitações ou Unidades com Múltiplos Inquilinos (Multiple Dwelling Units / Multiple Tenant Units — MDU/MTU). Estes ambientes possuem enlaces curtos que podem suportar potencialmente taxas de dados simétricas de até 100 Mbit/s ao utilizar o espectro a até 30 MHz [3].

Enquanto isso, outros operadores desejavam alcances mais longos, taxas de bits mais baixas para o VDSL2 e estavam interessados em estender o alcance do VDSL2 para além daquele alcançado pelo VDSL1. Já outros operadores mostravam interesse em sistemas de baixa complexidade a fim de possibilitar seu fechamento dentro de cabines telefônicas, por exemplo.

Obviamente, os requisitos eram diversos para serem compilados em um único sistema de forma otimizada. Assim, um sistema que tentasse atender a todos os requisitos seria subotimizado e repleto de complexidades desnecessárias na maioria das aplicações.

Assim, implementou-se o conceito de perfis – *profiles*, a fim de lidar com essa diversidade de aplicações objetivadas para o VDSL2. Basicamente, um perfil é um subconjunto do conjunto completo de configurações definidos no padrão VDSL2. Dessa forma, um determinado perfil é tem o objetivo de produzir um sistema otimizado para determinada aplicação.

Para que um sistema seja considerado compatível com o padrão VDSL2, não é necessário que ele suporte todos os perfis definidos pela norma, mas é necessário que ele suporte pelo menos um perfil definido no padrão.

2.4 Características Físicas dos Fios de Cobre

As primitivas instalações de fios de cobre para a telefonia utilizavam um único fio condutor e o terra como retorno em uma transmissão chamada modo comum, resultando em um sistema telefônico de péssima qualidade.

A partir de 1881, com a utilização das recomendações que constavam na patente de Alexander Graham Bell, realmente, significativas melhorias foram alcançadas no sistema de telefonia [2]. Dentre elas o uso de pares blindados.

2.4.1 Isolamento, Composição do Núcleo e Blindagem dos Pares Trançados

Uma das medidas para melhor aproveitamento dos pares de cobre consiste em envolver o par de cobre com um escudo metálico a fim de isolá-lo de fontes de interferência eletromagnética. Os pares de cobre são agrupados dentro de um *binder* para prevenir que os pares toquem-se uns com os outros ou com agentes externos.

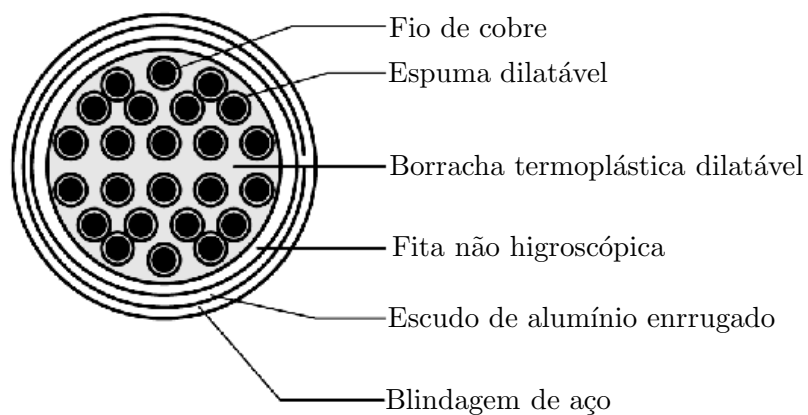


Figura 2.5: Seção de um cabo contendo fios de cobre [2].

A Figura 2.5 mostra a seção de um *binder*. Os fios condutores utilizam um isolamento plástico de poliolefina. Geralmente, utiliza-se espuma dilatável (polietileno dilatável, cloreto de polivinila (polyvinyl chloride — PVC) por proporcionar fios mais flexíveis e de menor custo em relação a outros isolantes sólidos.

O núcleo do cabo é preenchido com um composto de borracha termoplástica dilatável com ótimas propriedades dielétricas e resistente a oxidações para evitar espaçamentos entre os pares. Além disso, uma fita não higroscópica, ou seja, que não absorve nem propaga umidade, com uma sobrecamada envolve o núcleo do cabo.

Em seguida, uma fita de alumínio enrugada descoberta é aplicada para formar o escudo que é aterrado nas extremidades para reduzir a interferência de fontes externas. Para dar a resistência a corrosão, o alumínio é vinculado a uma fina camada plástica.

Por fim, uma armadura obtida através de uma fita de aço protege o cabo. Essa armadura é coberta com uma capa de polietileno de baixa densidade. Geralmente, os cabos possuem uma forma cilíndrica devido o posicionamento dos pares no núcleo.

2.5 Interferências em Enlaces DSL

O enlace local por onde os dados são transmitidos em alta velocidade através da tecnologia DSL está exposto a uma grande quantidade de interferências que podem chegar até mesmo ao ponto de interromper o serviço. Entre os ruídos presentes nos sistemas DSL, o dominante é o *crosstalk*, a ação desse ruído é tal que impõe a taxa máxima de dados que se deve esperar transmitir. Outro ruído a ser tratado é o ruído impulsivo responsável pela maioria dos erros de bit durante a transmissão. Além das interferências de rádio frequência recebidas pelo par trançado. Nas seções seguintes serão abordados os principais tipos interferência presentes em enlaces DSL.

2.5.1 Ruído Impulsivo

O ruído impulsivo é uma interferência eletromagnética estocástica não estacionária que é caracterizada pela ocorrência aleatória de curtos pulsos com significativa energia e amplitude e conteúdo espectral aleatórios [1]. Esse ruído está altamente correlacionado com a atividade humana devido a comutação de transientes [2].

As fontes desse ruído são comutações de natureza elétrica próximas às instalações do cliente. Algumas das conhecidas fontes desse ruído são a soldagem de vigas metálicas, descargas elétricas, cercas elétricas e o acionamento de lâmpadas fluorescentes.

Como os pulsos variam em forma, duração, amplitude e espaçamento não conseguiu-se chegar a modelos aceitáveis, embora as várias tentativas de fazê-lo baseado em observações de redes reais [2].

2.5.2 Interferência de Rádio Frequência

Interferência de rádio frequência é a interferência de sinais de transmissão sem fio, como transmissões de rádio amador e radiodifusão (*broadcast*) AM em que determinada informação é enviada para muitos receptores ao mesmo tempo por meio de ondas AM.

Sinais de rádio frequência interferem principalmente em linhas telefônicas aéreas que, sendo compostas de cobre, sob incidência de ondas eletromagnéticas tornam-se antenas com fluxo de carga induzida em relação à terra.

DSLs com frequências na faixa de 560 kHz a 30 MHz podem sobrepor bandas de rádio e conseqüentemente receber alguma forma de rádio interferência [13]. Assim, as bandas de rádio amador sobrepõem a banda de sistemas VDSL acarretando um problema para sua transmissão.

A interferência de rádio é tão significativa em sistemas que operam em elevadas taxas de frequência que pode atingir níveis de ruído superiores ao *crosstalk*. Porém, como sinais de rádio amador são banda estreita, métodos de transmissão que se baseiam em filtragem no receptor tentam evitar as poucas bandas ocupadas por este ruído.

2.5.3 Crosstalk

Crosstalk é a interferência em um canal de um sinal proveniente de um outro canal. Os sinais elétricos em um par trançado geram um pequeno campo eletromagnético que circunda o par e induz um sinal elétrico nos pares vizinho conduzindo a um sinal indesejado nesses outros pares. O trançamento dos pares reduz esse acoplamento indutivo.

O *crosstalk* de outros sistemas de transmissão no mesmo cabo é o principal fator limitante da taxa de transmissão e do alcance do serviço a enlaces mais longos [5, 15]. Além disso, o *crosstalk* aumenta com a frequência e com o comprimento do cabo. Os acoplamentos possuem todos a mesma causa física, mas diferentes configurações produzem diferentes misturas de vazamentos de crosstalk e atenuação [2]. A seguir são apresentadas duas configurações de como esse ruído se apresenta nos enlaces.

NEXT

Near-end crosstalk – NEXT é a maior interferência para sistemas que compartilham a mesma banda de frequência para *upstream* e *downstream*. O NEXT é percebido pelo receptor localizado na mesma extremidade do cabo que o transmissor como mostrado na Figura 2.6. Para evitar o NEXT proveniente de outros fios que utilizam a mesma tecnologia — *self-NEXT*, utiliza-se diferentes bandas de frequência para *upstream* e *downstream*. Porém, os sistemas geralmente ainda precisam lidar com o NEXT proveniente de outros tipos de sistemas que transmitem na mesma banda de frequência [3].

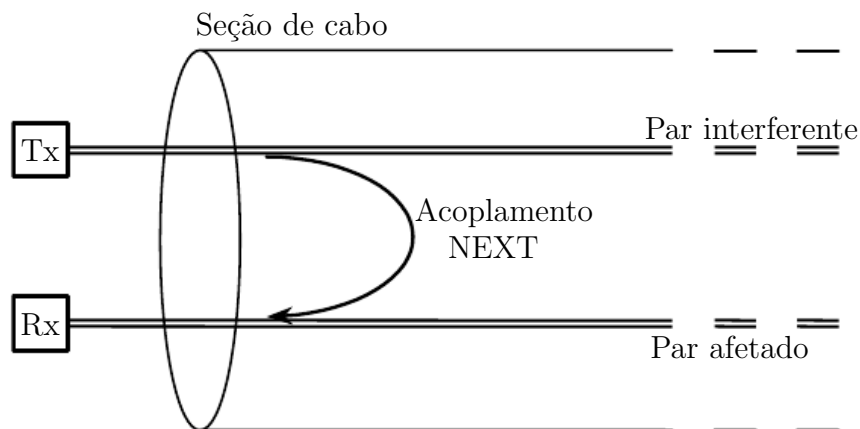


Figura 2.6: Acoplamento NEXT [2].

FEXT

Far-end crosstalk – FEXT é o ruído detectado pelo receptor localizado na extremidade distante do cabo transmissor que é a fonte do ruído como mostrado na Figura 2.7. O FEXT é atenuado ao atravessar totalmente o comprimento do cabo, porém como a tecnologia VDSL utiliza enlaces muito curtos o impacto desse ruído pode significar um entrave para a transmissão [2].

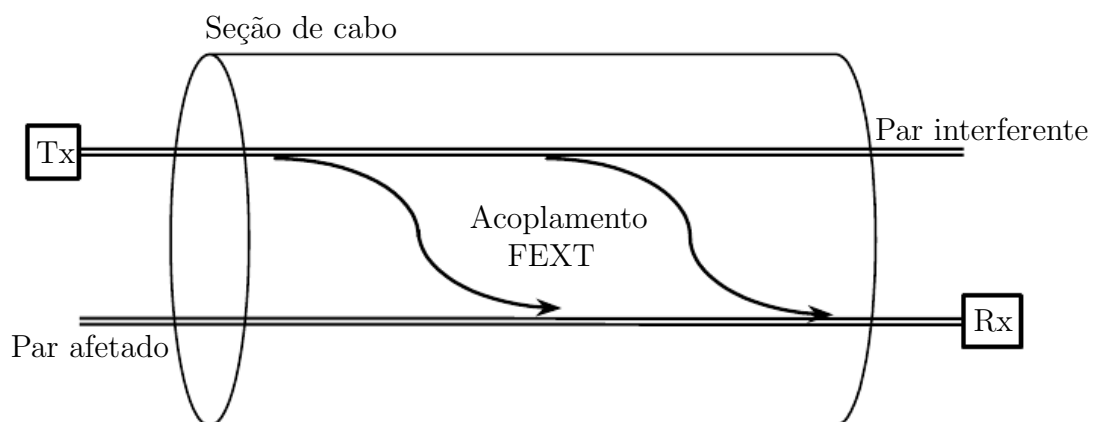


Figura 2.7: Acoplamento FEXT [2].

No capítulo seguinte serão introduzidos os conceitos mais importantes da abordagem bayesiana. Esse capítulo, terá a finalidade de apresentar os métodos e procedimentos que serão utilizados para se obter estimativas para os parâmetros do modelo de regressão linear que relaciona as variáveis explicativas com a variável dependente no modelo de ruído *crosstalk*.

Capítulo 3

Inferência Bayesiana

3.1 Introdução

A principal vantagem da abordagem bayesiana resume-se ao fato de envolver apenas os conceitos básicos de probabilidade. Os blocos básicos dessa abordagem são a função de verossimilhança e a probabilidade prévia (ou mais comumente chamada *a priori*), uma vez que do produto de ambas resulta a probabilidade posterior (ou *a posteriori*) que é a base para inferências a respeito dos parâmetros desconhecidos de um modelo [4].

Outra grande importância que a probabilidade *a posteriori* possui é dar suporte para se comparar diferentes modelos. Além disso, a predição é baseada na densidade preditiva. Porém, uma desvantagem da abordagem bayesiana é, para a maioria dos casos, a impossibilidade de se trabalhar analiticamente com todos esses blocos. Dessa forma, um tópico de relevante importância é a computação bayesiana, onde o método predominante utilizado é a simulação *a posteriori*.

3.1.1 Teoria Bayesiana

A teoria bayesiana é baseada em algumas poucas regras de probabilidade. Como dito anteriormente, esta é uma das principais vantagens da abordagem bayesiana. Assim, tudo o que se deseja extrair frente a um problema como es-

timar parâmetros de um modelo, comparar diferentes modelos e obter predições de um modelo envolve as mesmas regras da probabilidade [4].

A fim de mostrar de onde deriva a abordagem bayesiana, considere-se duas variáveis aleatórias A e B . A probabilidade conjunta destas duas variáveis aleatórias pode ser expressa pela equação 3.1.

$$p(A, B) = p(A|B)p(B) \quad (3.1)$$

onde $p(A|B)$ é a probabilidade condicional de A ocorrer dado que B tenha ocorrido e $p(B)$ é a probabilidade marginal de B , ou seja, a probabilidade de B ocorrer. Outra forma de se representar a mesma probabilidade conjunta para essas duas variáveis pode ser dada pela equação 3.2

$$p(A, B) = p(B|A)p(A) \quad (3.2)$$

Igualando-se as equações 3.1 e 3.2 e reordenando-as obtém-se a expressão do teorema de Bayes apresentada na equação 3.3, que governa toda a análise que se deseja fazer sobre os dados disponíveis.

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} \quad (3.3)$$

O objetivo de utilizar a inferência bayesiana neste trabalho consiste em analisar os dados a respeito de uma variável, os quais são obtidos através de observações ou medições para aprender sobre o fenômeno da presença do ruído na linha telefônica. Assim, trabalha-se com modelos que dependem de parâmetros. Um modelo de grande valia para esta análise é o de regressão, uma vez que este procura descrever o comportamento dos dados e a relação entre a variável sobre qual se deseja aprender e as demais variáveis. Neste modelo, o principal interesse está centrado

nos coeficientes da regressão que ponderam a influência das demais variáveis na variável principal e em estimar estes coeficientes.

Neste caso, os coeficientes são os parâmetros em estudo, considerando y como um vetor ou matriz de dados e θ como um vetor ou matriz que contém os parâmetros para um modelo que busca explicar y . Assim, deseja-se aprender sobre θ baseado nos dados, y . Onde, se aplica o teorema de Bayes expresso pela equação 3.3 como mostrado na equação 3.4.

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (3.4)$$

O termo $p(\theta|y)$ é de fundamental interesse para a análise bayesiana, tanto que [4] resume-o na questão “Com base nos dados, o que nós sabemos sobre θ ?” — “*Given the data, what do we know about θ ?*”. Dessa forma, a análise bayesiana gira em torno de aprender sobre algo desconhecido (coeficientes em uma regressão) com base em algo conhecido (dados). Assim, a probabilidade condicional dos coeficientes com base nos dados é a melhor forma de resumir o que se aprendeu.

Além disso, há outras utilizações para o termo $p(\theta|y)$. Em [16], onde pode-se encontrar uma excelente introdução à teoria bayesiana, o termo é utilizado para tomadas de decisão entre grupos correlatos com o objetivo de classificação.

Sabendo que o termo $p(\theta|y)$ da equação 3.4 é de fundamental importância para este estudo e o interesse está em se aprender apenas a respeito de θ , pode-se ignorar o termo $p(y)$ por não envolver θ :

$$p(\theta|y) \propto p(y|\theta)p(\theta) \quad (3.5)$$

O termo $p(\theta)$ é estabelecido como a densidade *a priori*. A densidade *a priori* $p(\theta)$ não depende dos dados. Em suma, ela contém qualquer informação disponível não baseada nos dados, ou seja, ela resume o que se sabe sobre θ antes de se analisar

suscintamente os dados.

O termo $p(y|\theta)$ é chamado função de verossimilhança. A função de verossimilhança $p(y|\theta)$ é a densidade de probabilidade dos dados y com base em θ , o vetor contendo os parâmetros do modelo.

O termo $p(\theta|y)$ é referido como a densidade *a posteriori*. A densidade *a posteriori* é de fundamental interesse no estudo, pois resume tudo o que se sabe a respeito de θ depois de se analisar os dados. Assim, a densidade *a posteriori* combina tanto as informações baseadas nos dados quanto as informações baseadas apenas no conhecimento *a priori*.

Porém, além de se desejar aprender a respeito dos parâmetros de um modelo, comparar diferentes modelos é uma tarefa interessante, porque se deseja obter o modelo que melhor represente o fenômeno em estudo. Segundo [4], um modelo é definido por uma função de verossimilhança e uma densidade *a priori*. Supondo que tenham-se m diferentes modelos, M_i , para $i = 1, \dots, m$, que tentam explicar y . M_i depende dos parâmetros θ^i . Assim, a densidade *a posteriori* para os parâmetros calculados usando o modelo M_i é escrita como:

$$p(\theta^i|y, M_i) = \frac{p(y|\theta^i, M_i)p(\theta^i|M_i)}{p(y|M_i)} \quad (3.6)$$

A probabilidade *a posteriori* do modelo pode ser usada para avaliar o grau de suporte de M_i , uma vez que se pode utilizar os dados para dizer se um modelo é o correto ou não. Portanto, utilizando a equação 3.3 tem-se:

$$p(M_i|y) = \frac{p(y|M_i)p(M_i)}{p(y)} \quad (3.7)$$

Onde, $p(M_i)$ é a probabilidade *a priori* do modelo e tem o efeito de o quanto acredita-se que o modelo M_i está correto antes de se analisar os dados. O termo $p(y|M_i)$ é a verossimilhança marginal e é obtida integrando-se a equação 3.6 em

relação a θ^i e o fato de $\int p(\theta^i|y, M_i)d\theta^i = 1$, uma vez que a integral de uma função densidade de probabilidade é igual a 1, tem-se:

$$p(y|M_i) = \int p(y|\theta^i, M_i)p(\theta^i|M_i)d\theta^i \quad (3.8)$$

Porém, o que realmente se deseja é fazer predições, isto é, com os dados observados y , fazer a predição de alguns dados futuros não observados y^* . Com base na abordagem bayesiana, deve-se sintetizar a informação desconhecida y^* a partir da informação disponível y . Assim, a predição se baseia na densidade preditiva $p(y^*|y)$. Fazendo uso de algumas regras de probabilidade, [4] expressa a densidade preditiva da seguinte forma:

$$p(y^*|y) = \int p(y^*|y, \theta)p(\theta|y)d\theta \quad (3.9)$$

O procedimento descrito nesta subseção resume de forma estruturada os conceitos teóricos básicos do uso da abordagem bayesiana para estimar parâmetros, comparar modelos e fazer predições.

3.1.2 Computação Bayesiana

Basicamente, a abordagem bayesiana é simples, uma vez que envolve apenas algumas regras simples de probabilidade, porém, na prática, essa abordagem necessita de uma grande quantidade de cálculos matemáticos.

Um exemplo do quão importante é computação bayesiana pode ser observado pelas equações relativas a comparação entre modelos e a predição que direta ou indiretamente envolvem integrais. Em geral, é necessária a utilização de ferramentas computacionais para avaliar essas integrais e muitos algoritmos destinados a este fim foram desenvolvidos.

Mesmo quando não ocorre o uso de integrais, como para o cálculo da probabilidade *a posteriori* $p(\theta|y)$, a quantidade de informação presente em y torna difícil sua análise sem a utilização de técnicas computacionais [4].

O método predominante para a avaliação das integrais envolvidas nos cálculos é a simulação *a posteriori*. A simulação *a posteriori* possui uma grande quantidade de variantes, mas que, segundo [4], geralmente se baseiam na lei dos grandes números, cuja implicação resulta na integração de Monte Carlo.

Da forma mais básica possível, a integração de Monte Carlo, como descrito em [4], consiste em se utilizar o computador para receber amostras aleatórias da probabilidade *a posteriori* e aproximar a função desejada.

3.1.3 Programas para Computação Bayesiana

Com a grande quantidade de probabilidades *a priori* e verossimilhanças torna-se difícil construir um pacote de programas que seja capaz de cobrir a necessidade de todos os pesquisadores que utilizam a abordagem bayesiana. Dessa forma, a prática de criar programas próprios para a computação dos dados é uma boa alternativa e, como tal, o modelo a ser proposto no próximo capítulo será implementado em código fonte, ou seja, não será utilizado nenhum conjunto de programas projetado para análise bayesiana como o *toolbox* do MATLAB.

3.2 Modelo de Regressão Linear

Um modelo de regressão linear relaciona uma variável dependente y_i com k variáveis explanatórias x_{i1}, \dots, x_{ik} para $i = 1, \dots, N$ através da equação 3.10.

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (3.10)$$

Assim, todo o processo descrito sucintamente na seção anterior visa justamente obter através da abordagem bayesiana estimativas a respeito dos parâmetros β_1, \dots, β_k de um modelo de regressão linear. Onde, ε_i é o erro de regressão e à variável x_{i1} é atribuído o valor 1 a fim de permitir uma interceptação em y_i quando todas as outras variáveis são nulas.

Como pretende-se analisar uma base de dados proveniente de uma transmissão em sistema DSL a fim de avaliar os fatores que influenciam para a margem de ruído *crosstalk* as variáveis explanatórias serão: a taxa de transmissão real no canal; a taxa de dados máxima atingível pela central telefônica; a taxa de dados máxima atingível pelas instalações do cliente; a potência total de saída transmitida pela central telefônica; a potência total de saída transmitida pelas instalações do cliente; o espaçamento relativo entre *bytes* de entrada na entrada intercaladora do canal e a sua localização no fluxo de *bits* na saída intercaladora; a real proteção ao ruído impulsivo; comprimento do bloco de dados do canal no qual a verificação de redundância cíclica (Cyclic Redundancy Check — CRC) opera; a taxa de *bits* da linha na direção de *downstream*. Dessa forma, trabalhar com essa base de dados envolve muitas variáveis explanatórias e conseqüentemente muitos parâmetros. A fim de se evitar uma notação demasiada complexa será feito o uso de álgebra matricial para simplificar a forma de representar a equação 3.10.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & x_{13} & \dots & x_{1k} \\ 1 & x_{22} & x_{23} & \dots & x_{2k} \\ 1 & x_{32} & x_{33} & \dots & x_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N2} & x_{N3} & \dots & x_{Nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

$$y = X\beta + \varepsilon \tag{3.11}$$

3.3 Função de Verossimilhança

Como descrito em [4], a forma da função de verossimilhança será determinada por algumas suposições acerca de ε e X que visam simplificar a formulação do modelo. Estas suposições são enumeradas a seguir:

1. ε possui distribuição Normal multivariada com média 0_N e matriz de covariância $\sigma^2 I_N$. Assim, ε é independente e identicamente distribuída – i.i.d. como $N(0_N, h^{-1} I_N)$, onde h é o erro de precisão definido por $h = \frac{1}{\sigma^2}$.
2. Os elementos de X são variáveis aleatórias independentes de ε com uma função densidade de probabilidade $p(X|\lambda)$, onde λ é um vetor de parâmetros que não inclui β nem h .

A função de verossimilhança é definida como uma função densidade de probabilidade de todos os dados com base nos parâmetros desconhecidos. Então, a função de verossimilhança torna-se $p(y, X|\beta, h, \lambda)$ e a segunda suposição implica:

$$p(y, X|\beta, h, \lambda) = p(y|X, \beta, h)p(X|\lambda)$$

Como a distribuição de X não é de interesse, pode-se utilizar a função de verossimilhança condicional sobre X – $p(y|X, \beta, h)$, ou seja, a distribuição condicional de y dado X e não a distribuição conjunta das duas variáveis. Para simplificar a notação, X não será explicitado.

Como a primeira suposição diz que ε possui distribuição Normal multivariada e a equação 3.11 é uma combinação linear de ε , tem-se que $p(y|\beta, h)$ é também Normal multivariada definida por:

$$p(y|\beta, h) = \frac{h^{\frac{N}{2}}}{2\pi^{\frac{N}{2}}} e^{-\frac{h}{2}(y-X\beta)'(y-X\beta)} \quad (3.12)$$

Pode ser conveniente escrever a função de verossimilhança em termos de mínimos quadrados ordinários (Ordinary Least Squares – OLS), pois são parâmetros comumente presentes na abordagem estatística clássica.

$$v = N - k \quad (3.13)$$

$$\hat{\beta} = (X'X)^{-1} X'y \quad (3.14)$$

$$s^2 = \frac{(y - X\hat{\beta})' (y - X\hat{\beta})}{v} \quad (3.15)$$

Nestes termos, a função de verossimilhança pode ser escrita como:

$$p(y|\beta, h) = \frac{1}{2\pi^{\frac{N}{2}}} h^{\frac{1}{2}} e^{-\frac{h}{2}(\beta - \hat{\beta})' X'X(\beta - \hat{\beta})} h^{\frac{v}{2}} e^{-\frac{hv}{2s^2}} \quad (3.16)$$

É válido notar que o argumento da primeira exponencial da equação 3.16 lembra o núcleo de uma densidade Normal mostrada na equação 3.17 para β e o argumento da segunda exponencial parece bastante com uma densidade Gamma mostrada na equação 3.18 para h [4].

$$f_N(y|\mu, \Sigma) = \frac{1}{2\pi^{\frac{k}{2}}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(y-\mu)' \Sigma^{-1}(y-\mu)} \quad (3.17)$$

$$f_G(y|\mu, v) = \begin{cases} c_G^{-1} y^{\frac{v-2}{2}} e^{-\frac{yv}{2\mu}} & \text{se } 0 < y < \infty \\ 0 & \text{caso contrário} \end{cases} \quad (3.18)$$

onde $c_G = \frac{2\mu}{v} \Gamma(\frac{v}{2})$ e $\Gamma(\cdot)$ é a função Gamma.

3.4 Densidade *a priori*

A densidade *a priori* consiste em qualquer informação que se tenha antes de analisar os dados. Porém, dada a grande quantidade de possíveis densidades *a priori*, é razoável a escolha por uma densidade que seja fácil de interpretar e fácil de computar, ou seja, que não seja necessária a aplicação de métodos iterativos de computação.

Em [4], sugere-se a escolha de uma densidade *a priori* conjugada natural. A propriedade de ser natural faz com que a densidade *a priori* tenha a mesma forma funcional da função de verossimilhança. Enquanto que a propriedade de ser conjugada faz com que a combinação entre a densidade *a priori* e a verossimilhança resulte em uma densidade *a posteriori* da mesma classe que as anteriores. Portanto, a densidade *a priori* é obtida de um conjunto de dados fictício gerado pelo mesmo processo que os dados reais.

A densidade *a priori* para β e h é representada por $p(\beta, h)$, podendo ser escrita como $p(\beta, h) = p(\beta|h)p(h)$, ou seja, a densidade *a priori* é dada em termos de $p(\beta|h)$ e $p(h)$. Como a equação 3.16 sugere, a densidade *a priori* é dada por uma densidade Normal para β dado h e uma densidade Gamma para h , cujo produto resulta em uma Normal-Gamma:

$$p(\beta|h) = N(\underline{\beta}, h^{-1}\underline{V})$$

$$p(h) = G(\underline{s}^{-2}, \underline{v})$$

$$p(\beta, h) = NG(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{v}) \tag{3.19}$$

Os chamados hiperparâmetros $\underline{\beta}$, \underline{V} , \underline{s}^{-2} e \underline{v} são escolhidos de forma tal que reflitam a informação *a priori*. Neste trabalho, será adotada a mesma notação utilizada em [4], ou seja, parâmetros sublinhados indicam parâmetros de uma densidade *a priori*, enquanto parâmetros sobrelinhados representam parâmetros de uma densidade *a posteriori*.

3.5 Densidade *a posteriori*

A densidade *a posteriori* reúne todas as informações disponíveis a respeito dos parâmetros β e h , ou seja, combina a função de verossimilhança com a densidade *a priori*. Como explicado na seção anterior, ao utilizar uma densidade *a priori* conjugada natural resulta em uma densidade *a posteriori* também da forma Normal-Gamma [4]:

$$p(\beta, h|y) = NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{v}) \quad (3.20)$$

Sendo que:

$$\bar{V} = (V^{-1} + X'X)^{-1} \quad (3.21)$$

$$\bar{\beta} = \bar{V} \left(\underline{V}^{-1} \underline{\beta} + X'X \hat{\beta} \right) \quad (3.22)$$

$$\bar{v} = \underline{v} + N \quad (3.23)$$

$$\bar{v} \bar{s}^2 = \underline{v} \underline{s}^2 + v s^2 + \left(\hat{\beta} - \underline{\beta} \right)' \left[\underline{V} + (X'X)^{-1} \right]^{-1} \left(\hat{\beta} - \underline{\beta} \right) \quad (3.24)$$

Como ao se utilizar modelo de regressão, o maior interesse se concentra em β , uma vez que este parâmetro representa a influência da variável explanatória sobre a variável dependente, é razoável que se deseje obter a densidade *a posteriori*

marginal $p(\beta|y)$. A densidade marginal de β pode ser obtida integrando-se a densidade *a posteriori* conjunta $p(\beta, h|y)$ em relação a h . O resultado é uma densidade *t* multivariada como segue [4]:

$$p(\beta|y) = t(\bar{\beta}, \bar{s}^2 \bar{V}, \bar{v}) \quad (3.25)$$

Segundo [4] a média *a posteriori* $E(\beta|y)$ é uma estimativa comumente utilizada e a variância *a posteriori* $var(\beta|y)$ é uma métrica comumente utilizada para medir a incerteza associada a essa estimativa. Pela definição da densidade *t* tem-se:

$$E(\beta|y) = \bar{\beta} \quad (3.26)$$

$$var(\beta|y) = \frac{\bar{v} \bar{s}^2}{\bar{v} - 2} \bar{V} \quad (3.27)$$

O erro de precisão h possui menor importância em relação a β [4], uma vez que apenas traduz a incerteza relacionada à estimativa dos parâmetros, enquanto que β reflete a influência das variáveis explanatórias sobre a variável dependente. Porém as seguintes informações podem ser explicitadas pelas propriedades da Normal-Gamma:

$$p(h|y) = G(\bar{s}^{-2}, \bar{v}) \quad (3.28)$$

$$E(h|y) = \bar{s}^{-2} \quad (3.29)$$

$$var(h|y) = \frac{2\bar{s}^{-2}}{\bar{v}} \quad (3.30)$$

3.6 Predição

Supondo-se que após os métodos apresentados para se estimar os parâmetros β e h , deseja-se fazer predições de T valores não observados da variável dependente que foram gerados pelo mesmo modelo. Assim, tem-se para a variável não observada $y^* = (y_1^*, y_2^*, y_3^*, \dots, y_T^*)'$:

$$y^* = X^* \beta + \varepsilon^* \quad (3.31)$$

Além disso, as suposições para a equação 3.31 são as mesmas apresentadas na seção 3.3, ou seja, ε^* é independente de ε e possui distribuição do tipo $N(0, h^{-1} I_T)$ e X^* é uma matriz $T \times k$, contendo as k variáveis explanatórias para os T valores não observados.

Como descrito pela equação genérica 3.9, a predição bayesiana é produzida pelo cálculo de :

$$p(y^*|y) = \int \int p(y^*|y, \beta, h) p(\beta, h|y) d\beta dh \quad (3.32)$$

O fato de ε^* ser independente de ε implica que y e y^* também são independentes e consequentemente $p(y^*|y, \beta, h) = p(y^*|\beta, h)$. Utilizando a mesma metodologia que produziu a equação 3.12, tem-se:

$$p(y^*|\beta, h) = \frac{h^{\frac{S}{2}}}{2\pi^{\frac{S}{2}}} e^{-\frac{h}{2}(y^* - X^* \beta)'(y^* - X^* \beta)} \quad (3.33)$$

Integrando-se o produto das equações 3.20 e 3.33 como descrito pela equação 3.32 resulta em uma densidade preditiva t multivariada da forma:

$$p(y^*|y) = t(X^* \bar{\beta}, \bar{s}^2(I_T + X^* \bar{V} X^{*'}), \bar{v}) \quad (3.34)$$

Nesse capítulo, foram abordados os aspectos teóricos de uma análise bayesiana completa para um modelo de regressão linear com densidade *a priori* conjugada natural para diversas variáveis aleatórias. No Capítulo 4, essa abordagem será utilizada para implementar a análise para um modelo de regressão linear que descreve as variáveis de transmissão que influenciam no nível do ruído *crosstalk*.

Capítulo 4

Análise Bayesiana para Identificação de Crosstalk em Sistemas DSL

Neste Capítulo, será realizada uma análise completa dos dados referentes à transmissão VDSL2 através de um modelo de regressão linear com múltiplas variáveis explanatórias. As medições de tráfego no sistema VDSL2 resultaram em uma base de dados com 50 variáveis explanatórias e uma variável dependente que retrata o nível do ruído *crosstalk* na linha em 270 observações das variáveis disponíveis durante a transmissão. Sendo que dessas, 90 foram observações quando não havia *crosstalk*, 90 quando havia *crosstalk* na direção de *upstream* e 90 com *crosstalk* na direção de *downstream*.

Porém, desse conjunto de dados apenas uma parte é realmente de interesse para a análise. Assim, foram selecionadas variáveis que possuem informações relacionadas ao modelo de regressão linear que descreve o ruído *crosstalk* como será tratado na Seção 4.2.

4.1 Configuração Utilizada para as Medições

A Figura 4.1 mostra a configuração utilizada para as medições em VDSL2 a fim de se coletar o conjunto de dados para a análise bayesiana.

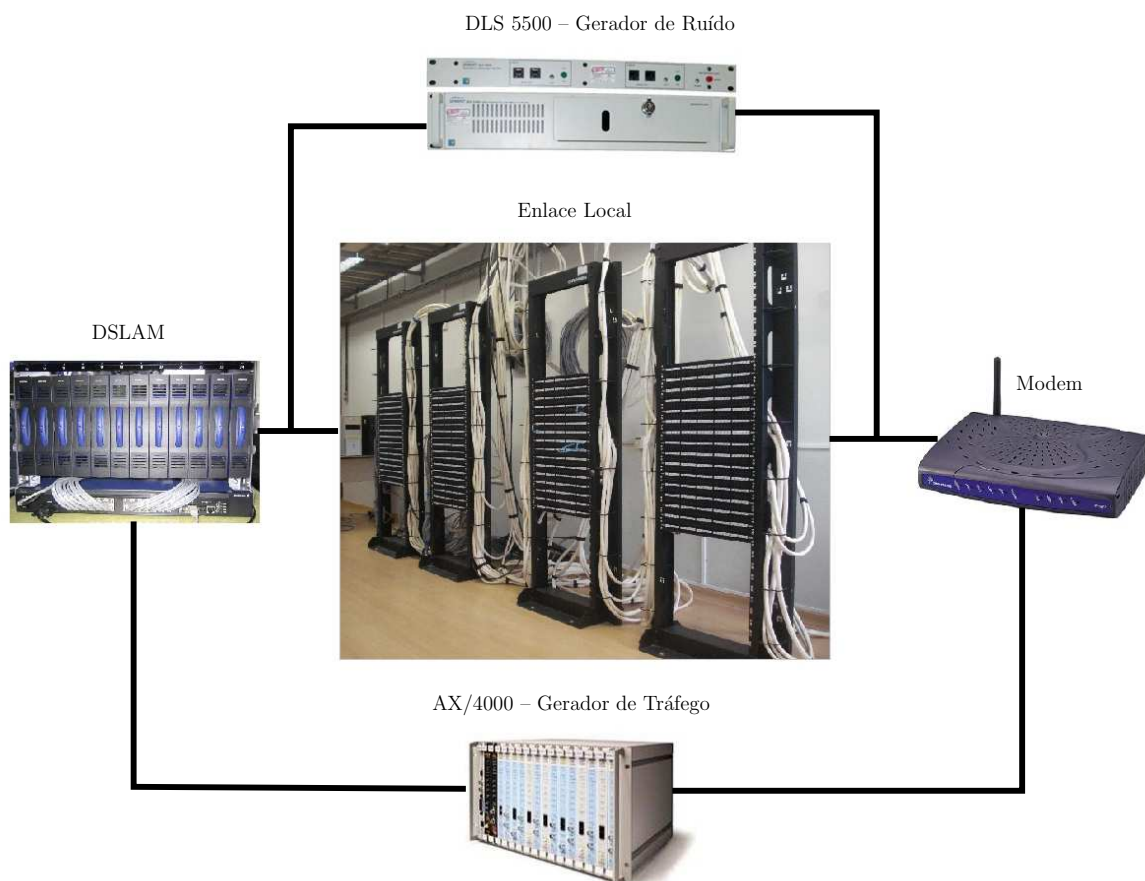


Figura 4.1: Configuração utilizada para as medições.

Os equipamentos gerador de tráfego AX/4000 e gerador de ruído DLS 5500 são controlados por um computador com os respectivos programas de controle e configuração. O gerador de tráfego AX/4000 é utilizado para simular o envio de pacotes de voz, vídeo e dados através do enlace DSL. Este gerador de tráfego simula um aplicativo que esteja funcionando em um computador do lado do cliente

e/ou do outro lado do enlace.

O computador do lado do cliente está conectado a um modem que por sua vez modulariza os sinais de dados para envio através do enlace DSL local ou demodulariza os sinais recebidos pelo enlace. O enlace local é constituído por cabos telefônicos, estendidos em esteiras metálicas, que contornam o prédio de engenharia elétrica de forma que ambas as pontas dos cabos encontrem-se dentro da sala do Laboratório de Eletromagnetismo Aplicado. Esta disposição dos cabos telefônicos permite que se utilize diferentes comprimentos para os enlaces de teste. Para este enlace em particular, o comprimento utilizado foi de 50 m, pois faz parte de um projeto maior que visa uma campanha de medições com comprimentos de 750 e 2600 m.

Do outro lado do enlace DSL, encontra-se um DSLAM que faz a multiplexação dos sinais provenientes das instalações do cliente, assim como o envio de sinais modularizados em direção ao mesmo.

Porém, enlaces DSL sofrem de vários tipos de interferência provenientes de variados tipos de fontes. Para simular o ruído *crossstalk* de interesse para esta pesquisa, utiliza-se o equipamento DLS 5500 que gera sinais aleatórios com distribuição de probabilidade similares às fontes reais de ruído que são injetados nas extremidades do enlace DSL.

Durante a transmissão VDSL2, utilizou-se um código escrito em MATLAB para a coleta de dados referentes ao estado do enlace. Esses dados são, por fim, armazenados para a posterior análise.

4.2 Seleção das Variáveis Explanatórias

A base de dados foi reduzida para apenas 10 variáveis explanatórias e 1 variável dependente a fim de se fazer uma análise mais consistente do conjunto de dados. Essa seleção foi realizada em duas etapas.

A primeira etapa consistiu em traçar os gráficos de todas as 50 variáveis disponíveis e eliminar aquelas que permaneciam constantes frente a variação no nível de ruído presente no enlace.

Porém, ainda restaram muitas variáveis que não poderiam ser utilizadas por se tratarem de variáveis que armazenam a contagem de todos os dados acumulados durante o dia. Como as medições foram realizadas em dias diferentes, estas variáveis não se alteravam em razão do nível de ruído, mas sim por motivo da alteração do dia.

Assim, foi possível gerar um novo conjunto de dados com apenas 10 variáveis explanatórias e uma variável dependente para o nível de ruído.

Contudo, há duas variáveis que expressam o nível de ruído na linha: uma variável que exhibe a Relação Sinal-Ruído (Signal-to-Noise Ratio — SNR) do lado da central telefônica e outra que mostra a SNR do lado das instalações do cliente.

Como está fora do escopo deste trabalho formular modelos para as duas variáveis, a base de dados será utilizada para estimar apenas os parâmetros do modelo que procura explicar o nível de ruído do lado da central telefônica.

4.3 Características das Variáveis Presentes na Base de Dados

Como explicado na seção anterior, este trabalho busca formular um modelo de regressão linear para avaliar as variáveis coletadas durante a transmissão que afetam o nível de ruído *crosstalk* no enlace VDSL2 do lado da central telefônica. Assim, para este modelo a variável dependente é a margem de ruído pelo lado da central telefônica.

Nas subseções seguintes será possível visualizar todas as variáveis selecionadas para a base de dados, assim como uma breve descrição a respeito das mesmas e gráficos que denotam as faixas de valores e o comportamento destas variáveis. Mais informações sobre estas variáveis e outras podem ser encontradas em [17], [18] e em [19]. Vale ressaltar que as 90 primeiras medições foram realizadas em transmissões com ausência de ruído, enquanto que as 90 medições seguintes foram realizadas quando havia ruído *crosstalk* na direção de *upstream* e, por fim, as últimas 90 medições foram executadas com *crosstalk* na direção de *downstream*.

4.3.1 `adslAtucCurrSnrMgn`

Esta variável é a margem de ruído do lado da central telefônica com relação ao respectivo sinal recebido em décimos de dB (tenth dB). Na Figura 4.2 pode ser visualizado o gráfico desta variável em função do número de medições.

4.3.2 `adslAturCurrSnrMgn`

É a margem de ruído do lado das instalações do cliente com relação ao respectivo sinal recebido em décimos de dB. A Figura 4.3 exibe o gráfico desta variável em função do número de medições.

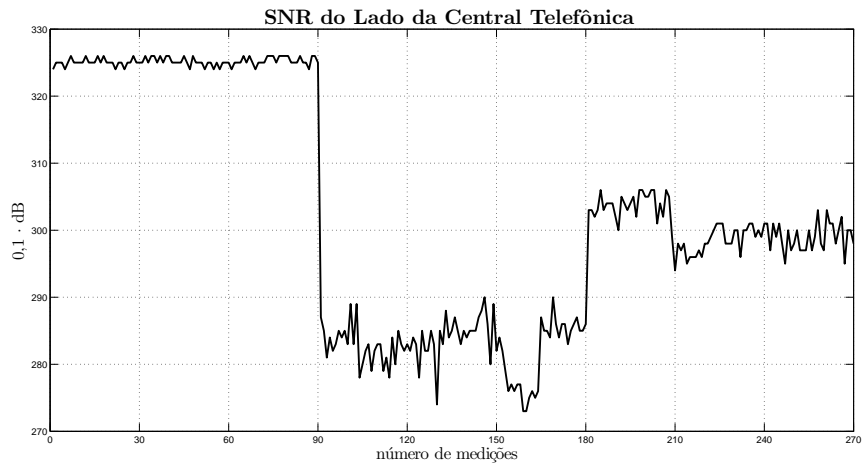


Figura 4.2: SNR do lado da central telefônica.

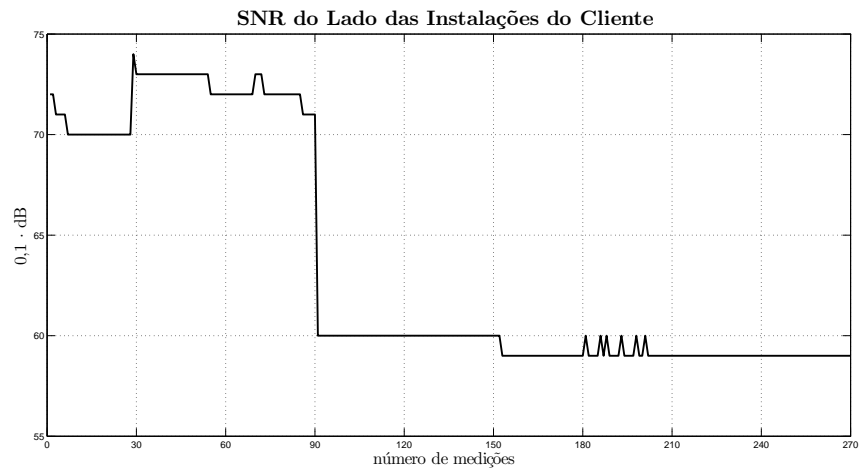


Figura 4.3: SNR do lado das instalações do cliente.

4.3.3 adslAtucChanCurrTxRate

É a taxa de transmissão real do lado da central telefônica em *bits/s*. Pode-se visualizar na Figura 4.4 o gráfico do comportamento desta variável.

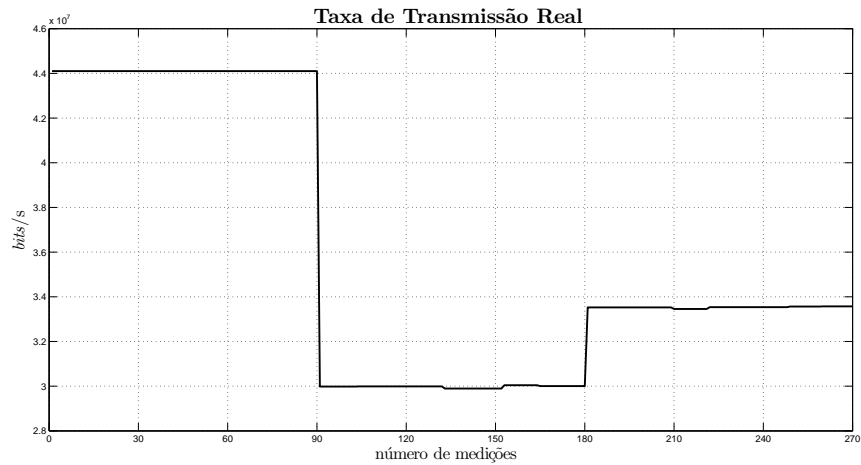


Figura 4.4: Taxa de transmissão real.

4.3.4 adslAtucCurrAttainableRate

Indica a taxa de dados máxima atingível atualmente pela central telefônica em *bits/s*. A variável pode ser vista na Figura 4.5.

4.3.5 adslAturCurrAttainableRate

Indica a taxa de dados máxima correntemente atingível pelas instalações do cliente em *bits/s*. A Figura 4.6 mostra o gráfico de medições para essa variável.

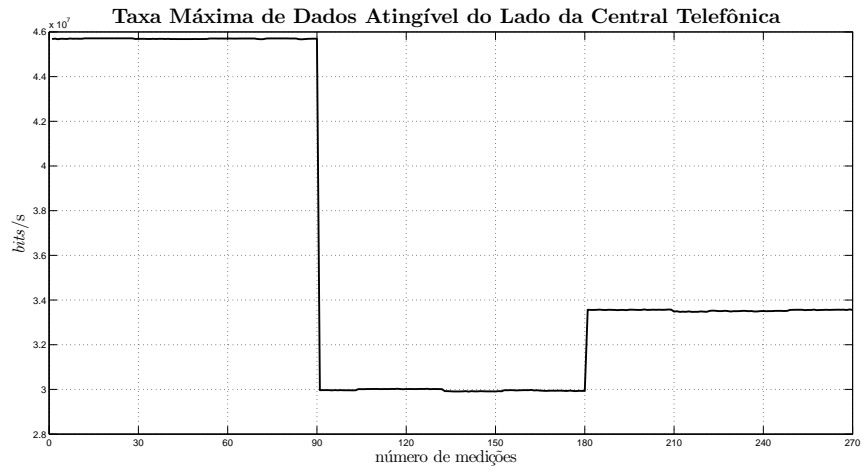


Figura 4.5: Taxa máxima de dados atingível do lado da central telefônica.

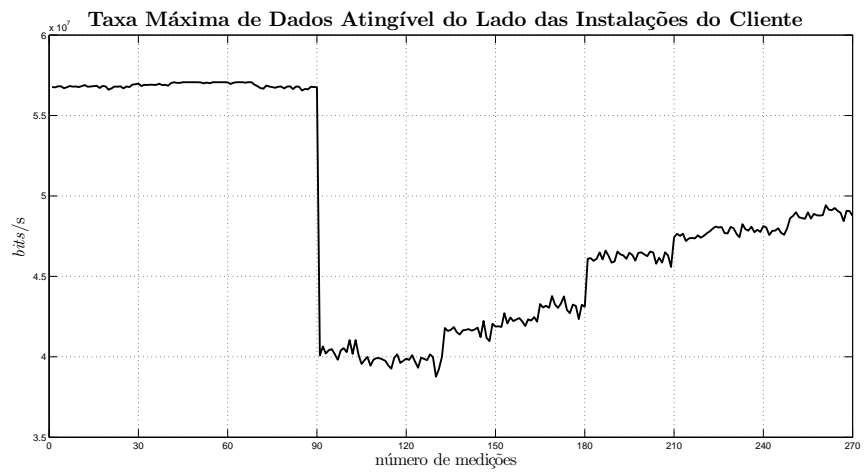


Figura 4.6: Taxa máxima de dados atingível do lado das instalações do cliente.

4.3.6 adslAtucCurrOutputPwr

Potência total de saída medida transmitida pela central telefônica em décimos de dBm. Esta variável é mostrada na Figura 4.7.

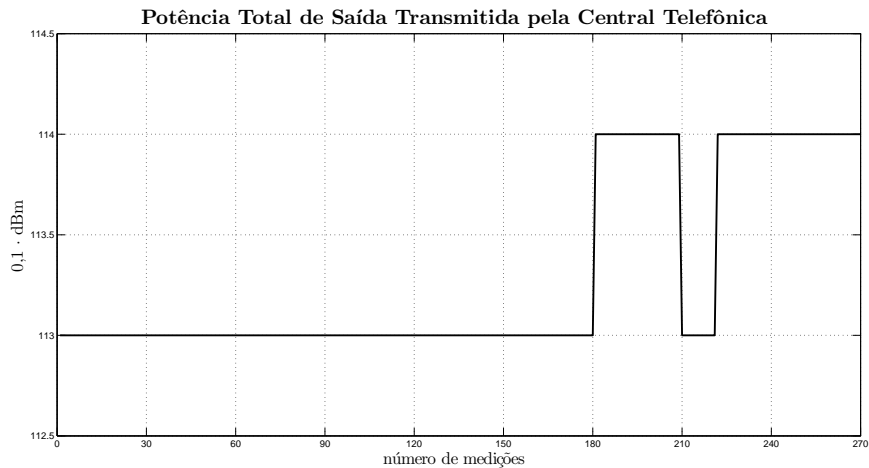


Figura 4.7: Potência total de saída transmitida pela central telefônica.

4.3.7 adslAturCurrOutputPwr

Potência total de saída medida transmitida pelas instalações do usuário em décimos de dBm. Na Figura 4.8 são exibidos os valores para essa variável durante as medições.

4.3.8 adslAtucChanInterleaveDelay

Define o espaçamento relativo entre *bytes* de entrada na entrada do canal e a sua localização no fluxo de *bits* na saída em ms. A Figura 4.9 exibe o comportamento desta variável.

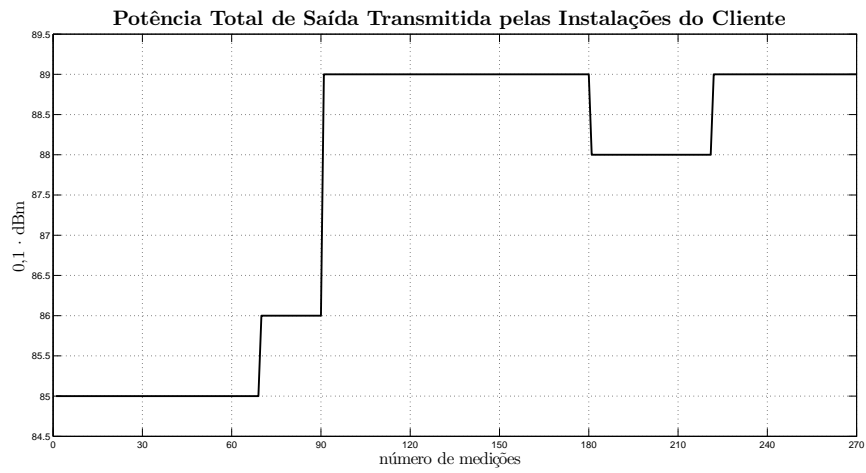


Figura 4.8: Potência total de saída transmitida pelas instalações do cliente.

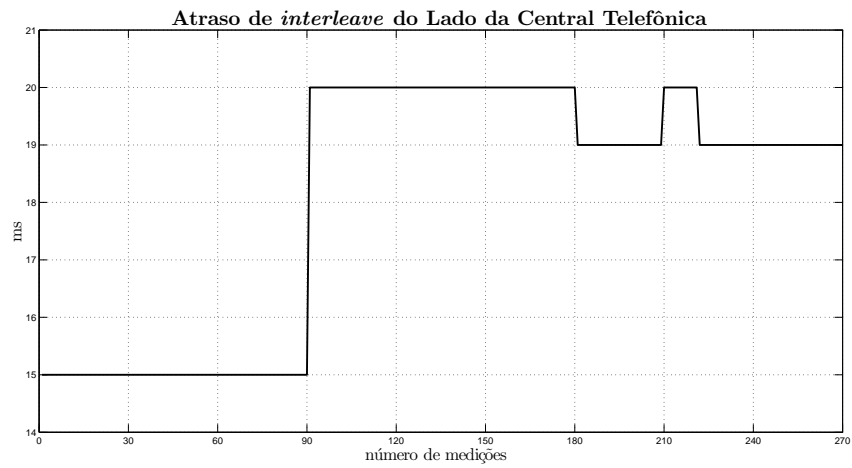


Figura 4.9: Atraso de *interleave* do lado da central telefônica.

4.3.9 adslAtucProprietaryChanActualInp

Real proteção ao ruído impulsivo (Impulse Noise Protection — INP) em símbolos DMT multiplicada por 100 para o canal em direção de *downstream*. Essa variável pode ter seus valores visualizados na Figura 4.10.

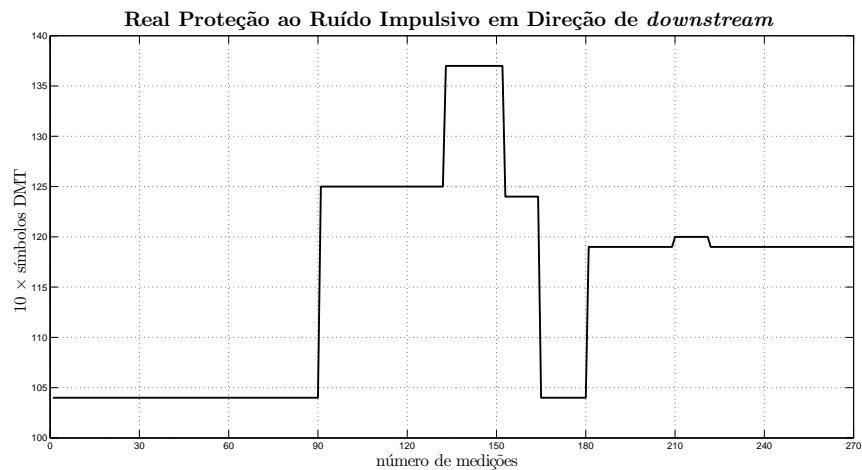


Figura 4.10: Real proteção ao ruído impulsivo em direção de downstream.

4.3.10 adslAtucChanCrcBlockLength

Indica o comprimento do bloco de dados do canal no qual a verificação de redundância cíclica (Cyclic Redundancy Check — CRC) opera. CRC é um método de detecção de erros em transmissão de dados. A Figura 4.11 mostra o comportamento desta variável.

4.3.11 adslAtucProprietaryPhysXActualLineBitRate

Especifica a taxa de *bits* real da linha na direção de *downstream* em *bits/s*. Pode-se visualizar na Figura 4.12 os valores medidos para esta variável.

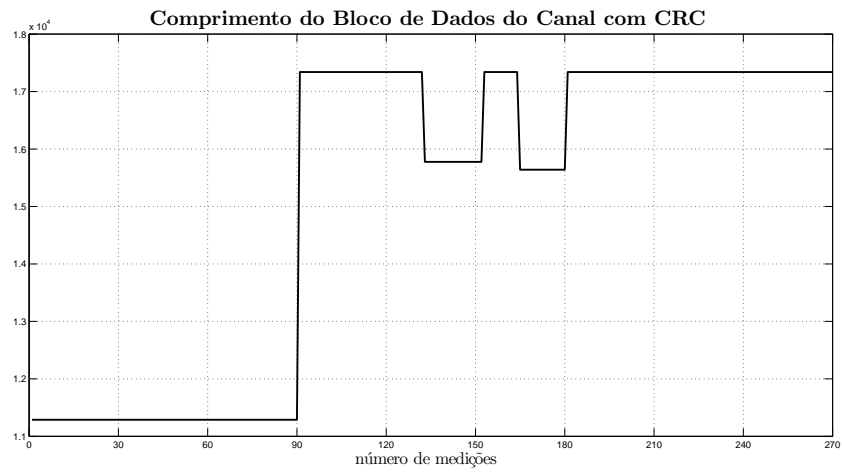


Figura 4.11: Comprimento do bloco de dados do canal com CRC.

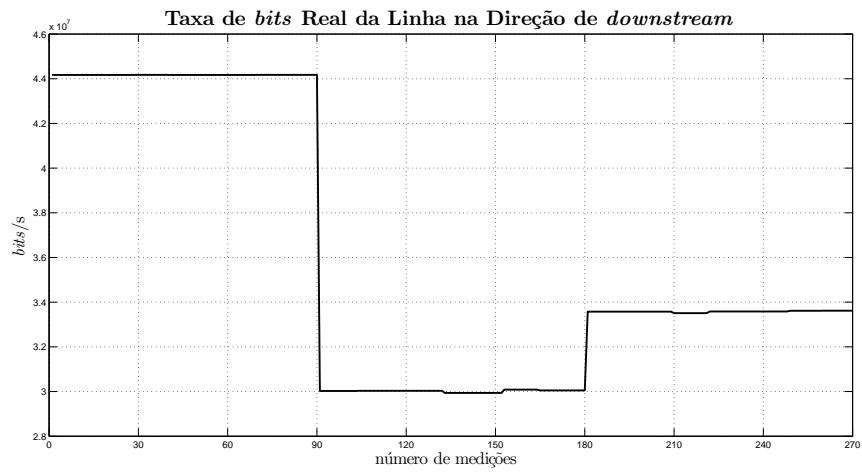


Figura 4.12: Taxa de *bits* real da linha na direção de *downstream*.

4.4 Elicitação da Informação *a priori*

As características das variáveis presentes na base de dados tratadas na seção anterior são de extrema importância para que se proponha uma quantidade de informação *a priori* bastante razoável.

Na subseção 4.4.1, serão abordados os aspectos de provimento da informação *a priori* para a identificação do nível de *crosstalk* do lado da central telefônica.

4.4.1 Obtenção dos Hiperparâmetros para a Densidade *a priori*

Para dar maior importância a informação proveniente da base de dados, será atribuído um menor grau de confiança a informação *a priori*, ou seja, as variâncias do hiperparâmetro *a priori* serão grandes.

As margens de ruído *crosstalk* do lado da central telefônica apresentam uma grande variação, porém estão distribuídas na faixa de 275 a 325 décimos de dB. Assim, um modelo de regressão capaz de descrever bem o nível desse ruído deve ter erros na ordem de no máximo 5 décimos de dB [4]. Essa suposição implica que σ seja grosseiramente 2,5 décimos de dB. Como o erro de precisão é obtido por $h = \frac{1}{\sigma^2}$, o que sugere que h seja dado por 0,16 e conseqüentemente $\underline{g}^{-2} = 0,16$.

Como esta suposição possui grande grau de incerteza associado, será atribuído pouco peso a ela fazendo \underline{v} igual a um valor muito pequeno em relação ao número de medições N . Assim, dado que $N = 270$ e deseja-se atribuir um peso bastante pequeno (da ordem de 1% [4]) a informação *a priori* sobre h , fazer $\underline{v} = 3$ é bastante razoável.

O passo seguinte é atribuir valores para os parâmetros *a priori* do modelo de regressão assim como as variâncias associadas aos mesmos. Isso será realizado com base na visualização dos gráficos de dispersão e histograma obtidos para cada variável explanatória individualmente e para a variável dependente.

Parâmetro *a priori* $\underline{\beta}_1$

Via de regra será associado o valor zero para o parâmetro de interceptação $\underline{\beta}_1$, porém, esta informação é bastante incerta, fazendo com que seja necessário a atribuição de um valor bastante alto para a variância deste parâmetro [4]. Neste caso, pode-se atribuir $var(\beta_1) = 18000^2$ resultando em um desvio-padrão de 18000 e, assim, é dizer que β_1 é um valor que possui 95% de probabilidade de estar entre -36000 e 36000 .

Parâmetro *a priori* $\underline{\beta}_2$

O parâmetro β_2 associado à variável explanatória SNR do lado das instalações do cliente é obtido com base na variação que esta variável provoca na variável dependente SNR do lado da central telefônica com base no gráfico de dispersão e histograma da Figura 4.13. Assim, o parâmetro β_2 é estimado da seguinte forma:

$$\begin{array}{l} adslAturCurrSnrMgn(0,1 \cdot dB) : \quad 59 \quad \xrightarrow{1} \quad 60 \quad \xrightarrow{10} \quad 70 \\ adslAtucCurrSnrMgn(0,1 \cdot dB) : \quad 300 \quad \xrightarrow{-15} \quad 285 \quad \xrightarrow{40} \quad 325 \end{array}$$

$$\begin{aligned} \underline{\beta}_2 &= \frac{-15 + 40}{1 + 10} \\ \underline{\beta}_2 &\approx 2,27 \end{aligned}$$

Como existe grande probabilidade de que o efeito da SNR do lado das instalações do cliente esteja entre -1 e 5 , pode-se escolher $var(\beta_2) = 9$ resultando em um desvio padrão igual a 3 , ou seja, como trata-se de uma distribuição gaussiana há uma probabilidade de 95% que o valor de β_2 esteja dentro da faixa que é o dobro do desvio padrão [4]. Portanto, entre -1 e 5 .

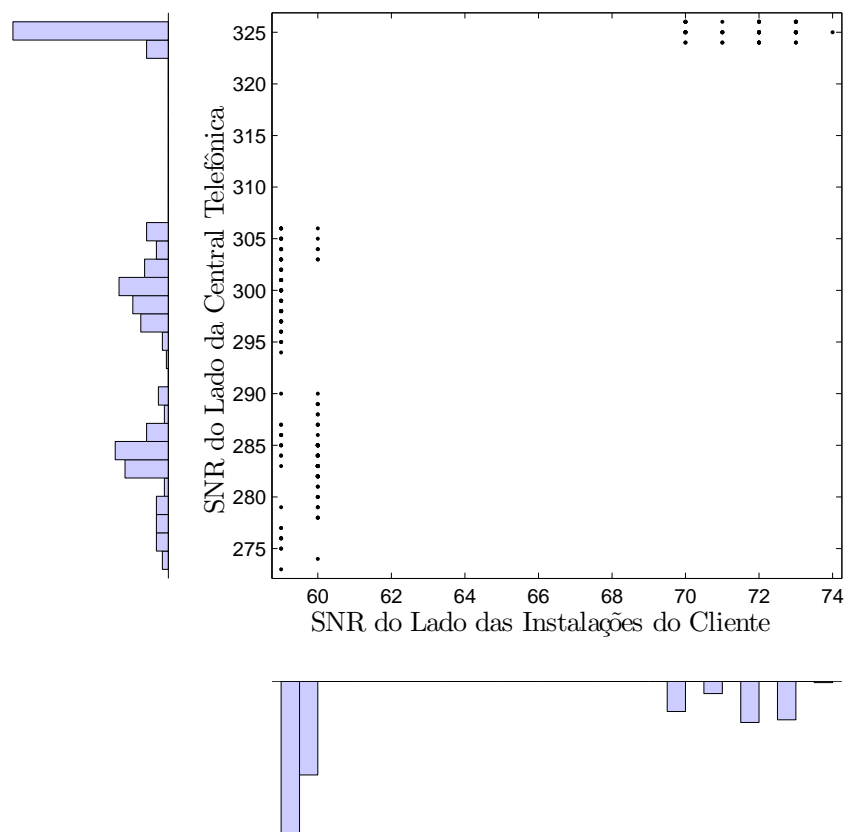


Figura 4.13: Gráfico de dispersão e histograma das variáveis SNR do lado das instalações do cliente e SNR do lado da central telefônica.

Parâmetro *a priori* $\underline{\beta}_3$

O parâmetro β_3 associado à variável explanatória taxa de transmissão real neste canal é obtido com base na variação que esta variável provoca na variável dependente SNR do lado da central telefônica com base no gráfico de dispersão e histograma da Figura 4.14. Assim, o parâmetro β_3 é estimado da seguinte forma:

$$\begin{array}{l} \text{adslAtucChanCurrTxRate(bits/s)} : 3 \cdot 10^7 \xrightarrow{0,4 \cdot 10^7} 3,4 \cdot 10^7 \xrightarrow{1 \cdot 10^7} 4,4 \cdot 10^7 \\ \text{adslAtucCurrSnrMgn}(0,1 \cdot \text{dB}) : 280 \xrightarrow{20} 300 \xrightarrow{25} 325 \end{array}$$

$$\begin{aligned} \underline{\beta}_3 &= \frac{20 + 25}{0,4 \cdot 10^7 + 1 \cdot 10^7} \\ \underline{\beta}_3 &\approx 3,21 \cdot 10^{-6} \end{aligned}$$

Como existe grande probabilidade de que o efeito da taxa de transmissão real neste canal esteja entre -1 e 1 , pode-se escolher $var(\beta_3) = 1$ resultando em um desvio padrão igual a 1 , ou seja, como trata-se de uma distribuição gaussiana há uma probabilidade de 95% que o valor de β_3 esteja dentro da faixa que é o dobro do desvio padrão [4]. Portanto, entre -1 e 1 .

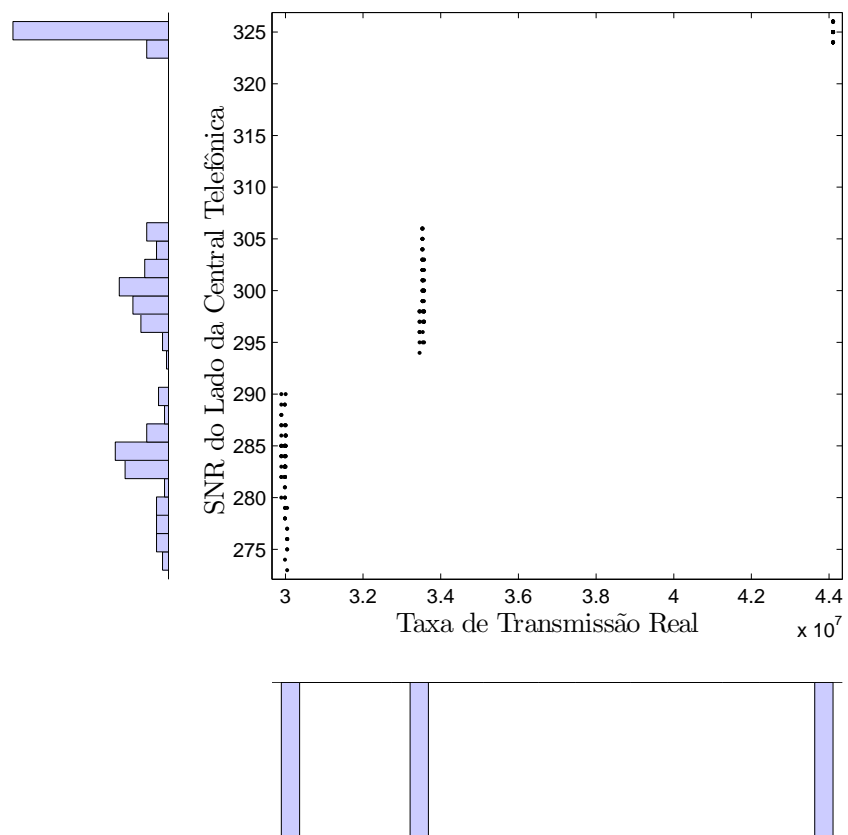


Figura 4.14: Gráfico de dispersão e histograma das variáveis taxa de transmissão real neste canal e SNR do lado da central telefônica.

Parâmetro *a priori* $\underline{\beta}_4$

O parâmetro β_4 associado à variável explanatória taxa de dados máxima atingível atualmente pela central telefônica é obtido com base na variação que esta variável provoca na variável dependente SNR do lado da central telefônica com base no gráfico de dispersão e histograma da Figura 4.15. Assim, o parâmetro β_4 é estimado da seguinte forma:

$$\begin{array}{l} \text{adslAtucCurrAttainableRate(bits/s)} : \quad 3 \cdot 10^7 \quad \xrightarrow{0,4 \cdot 10^7} \quad 3,4 \cdot 10^7 \quad \xrightarrow{1,2 \cdot 10^7} \quad 4,6 \cdot 10^7 \\ \text{adslAtucCurrSnrMgn}(0,1 \cdot \text{dB}) : \quad 280 \quad \xrightarrow{20} \quad 300 \quad \xrightarrow{25} \quad 325 \end{array}$$

$$\begin{aligned} \underline{\beta}_4 &= \frac{20 + 25}{0,4 \cdot 10^7 + 1,2 \cdot 10^7} \\ \underline{\beta}_4 &\approx 2,81 \cdot 10^{-6} \end{aligned}$$

Existe grande probabilidade de que o efeito da taxa de dados máxima atingível atualmente pela central telefônica esteja entre -1 e 1 . Assim, pode-se escolher $\text{var}(\beta_4) = 1$ resultando em um desvio padrão igual a 1 , ou seja, como trata-se de uma distribuição gaussiana há uma probabilidade de 95% que o valor de β_4 esteja dentro da faixa que é o dobro do desvio padrão [4]. Portanto, entre -1 e 1 .

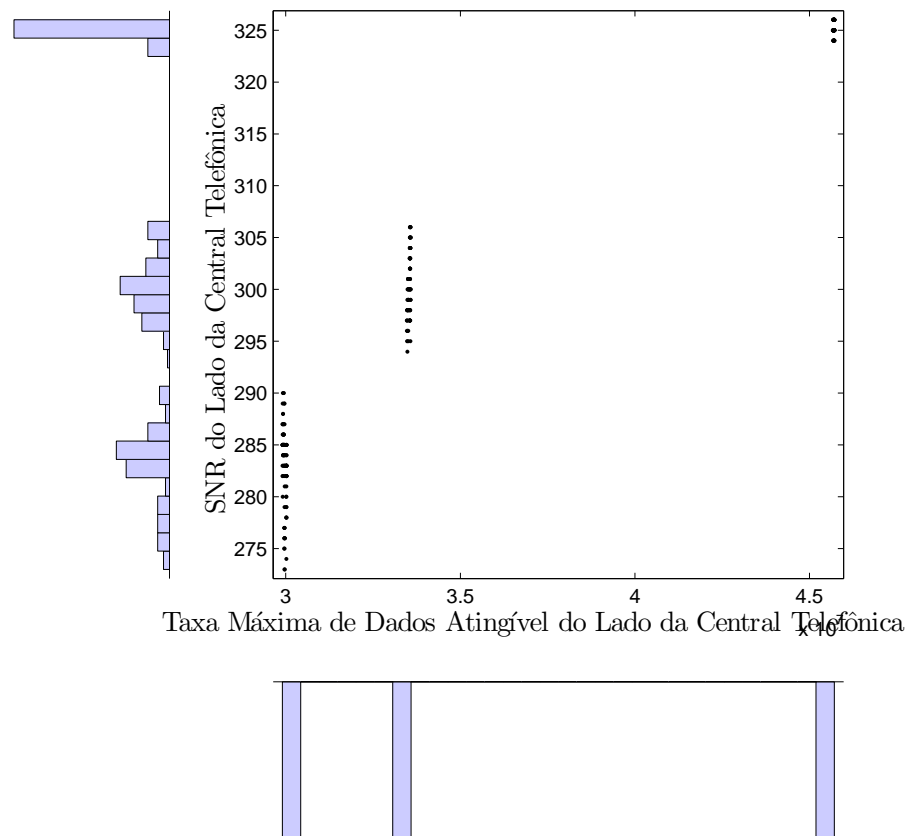


Figura 4.15: Gráfico de dispersão e histograma das variáveis taxa de dados máxima atingível atualmente pela central telefônica e SNR do lado da central telefônica.

Parâmetro *a priori* $\underline{\beta}_5$

O parâmetro β_5 associado à variável explanatória taxa de dados máxima correntemente atingível pelas instalações do cliente é obtido com base na variação que esta variável provoca na variável dependente SNR do lado da central telefônica com base no gráfico de dispersão e histograma da Figura 4.16. Assim, o parâmetro β_5 é estimado da seguinte forma:

$$\begin{array}{l} \text{adslAturCurrAttainableRate(bits/s)} : \quad 4,1 \cdot 10^7 \quad \xrightarrow{0,7 \cdot 10^7} \quad 4,8 \cdot 10^7 \quad \xrightarrow{0,9 \cdot 10^7} \quad 5,7 \cdot 10^7 \\ \text{adslAtucCurrSnrMgn}(0,1 \cdot \text{dB}) : \quad 280 \quad \xrightarrow{20} \quad 300 \quad \xrightarrow{25} \quad 325 \end{array}$$

$$\begin{aligned} \underline{\beta}_5 &= \frac{20 + 25}{0,7 \cdot 10^7 + 0,9 \cdot 10^7} \\ \underline{\beta}_5 &\approx 2,81 \cdot 10^{-6} \end{aligned}$$

Existe grande probabilidade de que o efeito da taxa de dados máxima correntemente atingível pelas instalações do cliente esteja entre -1 e 1 . Assim, pode-se escolher $\text{var}(\beta_5) = 1$ resultando em um desvio padrão igual a 1 , ou seja, como trata-se de uma distribuição gaussiana há uma probabilidade de 95% que o valor de β_5 esteja dentro da faixa que é o dobro do desvio padrão [4]. Portanto, entre -1 e 1 .

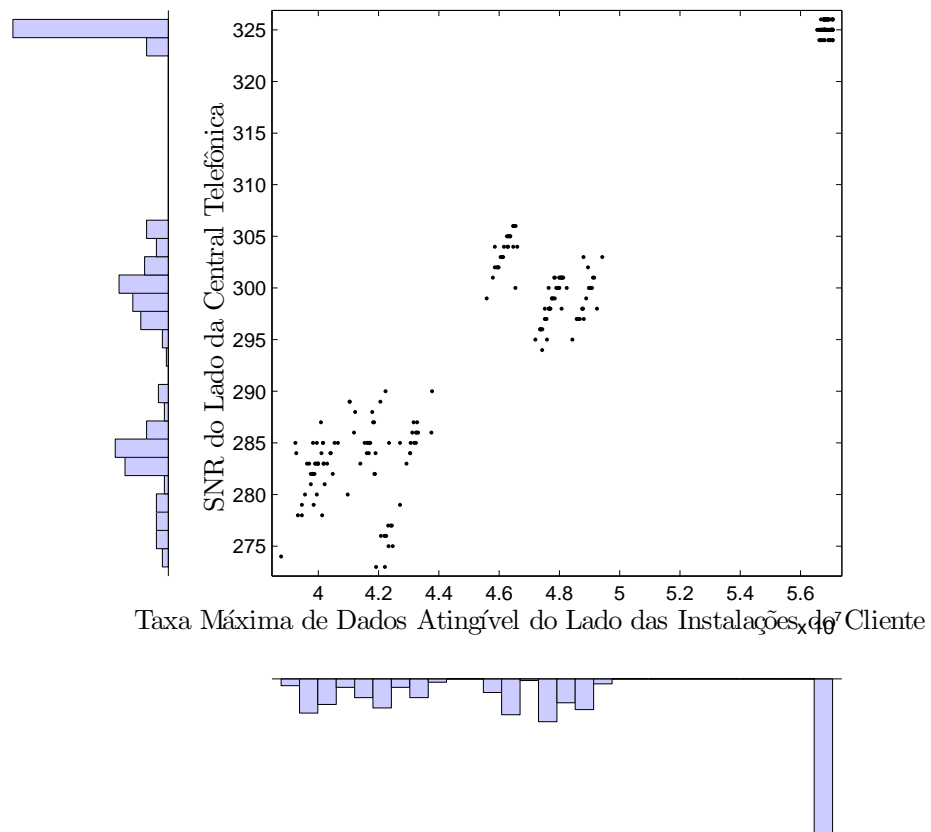


Figura 4.16: Gráfico de dispersão e histograma das variáveis taxa de dados máxima correntemente atingível pelas instalações do cliente e SNR do lado da central telefônica.

Parâmetro *a priori* $\underline{\beta}_6$

O parâmetro β_6 associado à variável explanatória potência total de saída medida transmitida pela central telefônica é obtido com base na variação que esta variável provoca na variável dependente SNR do lado da central telefônica com base no gráfico de dispersão e histograma da Figura 4.17. Assim, o parâmetro β_6 é estimado da seguinte forma:

$$\begin{array}{l} \text{adslAtucCurrOutputPwr(dBm)} : 113 \xrightarrow{1} 114 \\ \text{adslAtucCurrSnrMgn}(0, 1 \cdot \text{dB}) : 325 \xrightarrow{-25} 300 \end{array}$$

$$\begin{aligned} \underline{\beta}_6 &= \frac{-25}{1} \\ \underline{\beta}_6 &= -25 \end{aligned}$$

Existe grande probabilidade de que o efeito da potência total de saída medida transmitida pela central telefônica esteja entre -175 e -125 . Assim, pode-se escolher $var(\beta_6) = 22500$ resultando em um desvio padrão igual a 150 , ou seja, como trata-se de uma distribuição gaussiana há uma probabilidade de 95% que o valor de β_6 esteja dentro da faixa que é o dobro do desvio padrão [4]. Portanto, entre -175 e -125 .

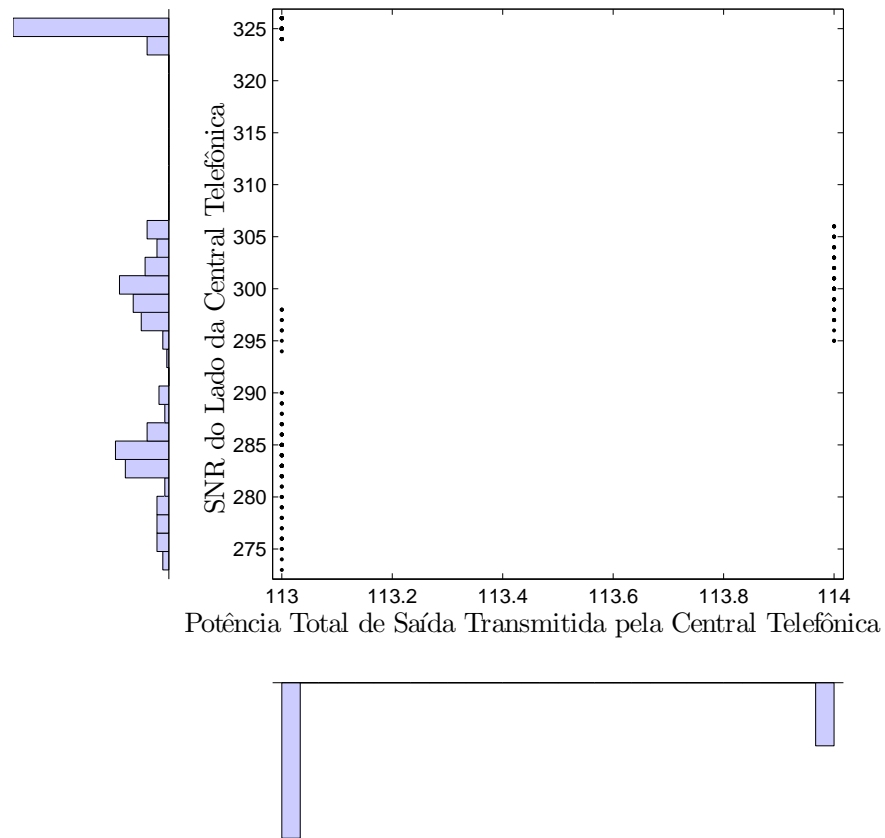


Figura 4.17: Gráfico de dispersão e histograma das variáveis potência total de saída medida transmitida pela central telefônica e SNR do lado da central telefônica.

Parâmetro *a priori* $\underline{\beta}_7$

O parâmetro β_7 associado à variável explanatória potência total de saída medida transmitida pelas instalações do usuário é obtido com base na variação que esta variável provoca na variável dependente SNR do lado da central telefônica com base no gráfico de dispersão e histograma da Figura 4.18. Assim, o parâmetro β_7 é estimado da seguinte forma:

$$\begin{array}{rccccccc} \text{adslAturCurrOutputPwr(dBm)} : & 85 & \xrightarrow{3} & 88 & \xrightarrow{1} & 89 & \\ \text{adslAtucCurrSnrMgn}(0,1 \cdot \text{dB}) : & 325 & \xrightarrow{-25} & 300 & \xrightarrow{-20} & 280 & \end{array}$$

$$\begin{aligned} \underline{\beta}_7 &= \frac{-25 - 20}{3 + 1} \\ \underline{\beta}_7 &\approx -11,25 \end{aligned}$$

Existe grande probabilidade de que o efeito da potência total de saída medida transmitida pelas instalações do usuário esteja entre -22 e 0 . Assim, pode-se escolher $\text{var}(\beta_7) = 121$ resultando em um desvio padrão igual a 11 , ou seja, como trata-se de uma distribuição gaussiana há uma probabilidade de 95% que o valor de β_7 esteja dentro da faixa que é o dobro do desvio padrão [4]. Portanto, entre -22 e 0 .

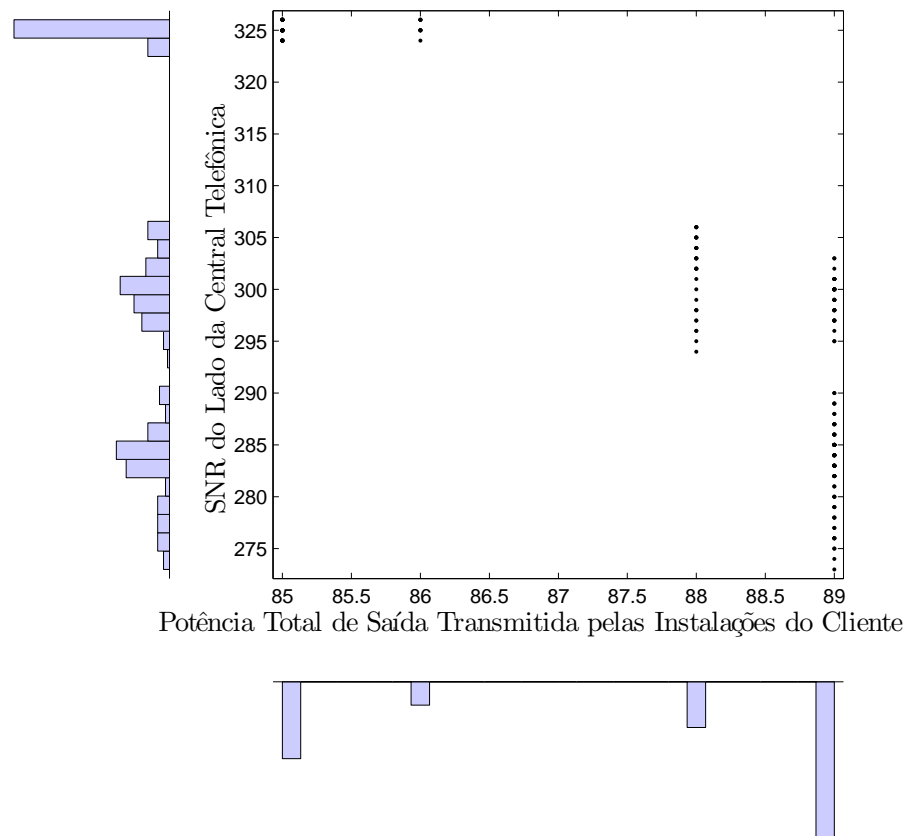


Figura 4.18: Gráfico de dispersão e histograma das variáveis potência total de saída medida transmitida pelas instalações do usuário e SNR do lado da central telefônica.

Parâmetro *a priori* $\underline{\beta}_8$

O parâmetro β_8 associado à variável explanatória atraso de *interleave* do lado da central telefônica é obtido com base na variação que esta variável provoca na variável dependente SNR do lado da central telefônica com base no gráfico de dispersão e histograma da Figura 4.19. Assim, o parâmetro β_8 é estimado da seguinte forma:

$$\begin{array}{l} \text{adslAtucChanInterleaveDelay}(ms) : \quad 15 \quad \xrightarrow{4} \quad 19 \quad \xrightarrow{1} \quad 20 \\ \text{adslAtucCurrSnrMgn}(0, 1 \cdot dB) : \quad 325 \quad \xrightarrow{-25} \quad 300 \quad \xrightarrow{-20} \quad 280 \end{array}$$

$$\begin{aligned} \underline{\beta}_8 &= \frac{-25 - 20}{4 + 1} \\ \underline{\beta}_8 &= -9 \end{aligned}$$

Existe grande probabilidade de que o efeito do atraso de *interleave* do lado da central telefônica esteja entre -40 e 22 . Assim, pode-se escolher $var(\beta_8) = 961$ resultando em um desvio padrão igual a 31 , ou seja, como trata-se de uma distribuição gaussiana há uma probabilidade de 95% que o valor de β_8 esteja dentro da faixa que é o dobro do desvio padrão [4]. Portanto, entre -40 e 22 .

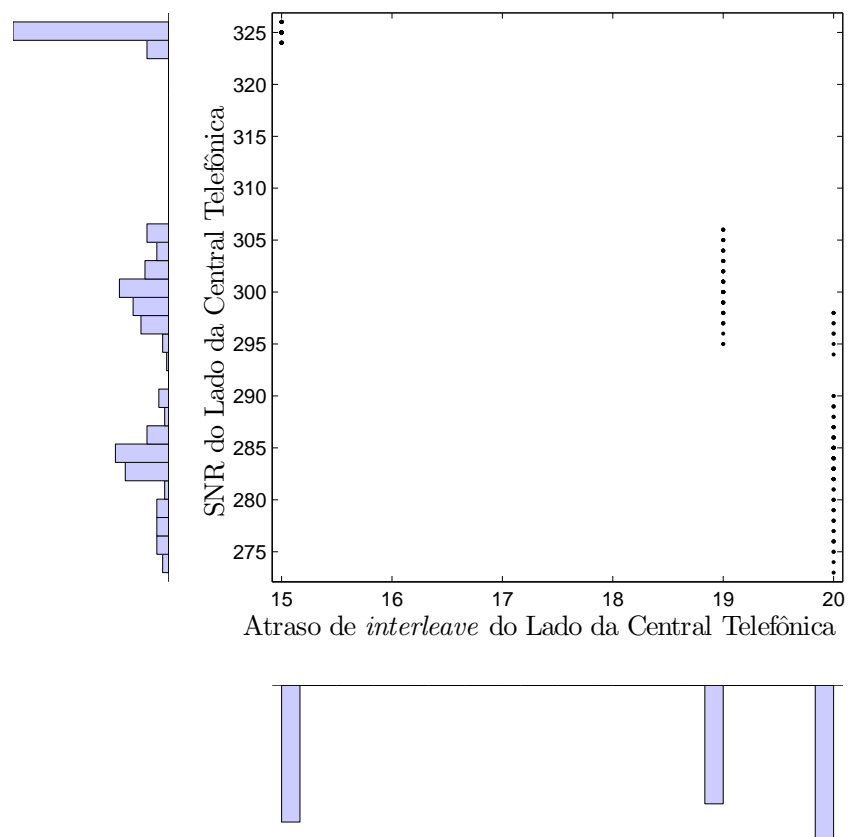


Figura 4.19: Gráfico de dispersão e histograma das variáveis atraso de *interleave* do lado da central telefônica e SNR do lado da central telefônica.

Parâmetro *a priori* $\underline{\beta}_9$

O parâmetro β_9 associado à variável explanatória real proteção ao ruído impulsivo em direção de *downstream* é obtido com base na variação que esta variável provoca na variável dependente SNR do lado da central telefônica com base no gráfico de dispersão e histograma da Figura 4.20. Assim, o parâmetro β_9 é estimado da seguinte forma:

$$\begin{array}{ccccccccc} \text{adslAtuc...ActualInp(simb. DMT)} : & 105 & \xrightarrow{15} & 120 & \xrightarrow{5} & 125 & \xrightarrow{15} & 140 & \\ \text{adslAtucCurrSnrMgn}(0, 1 \cdot \text{dB}) : & 325 & \xrightarrow{-25} & 300 & \xrightarrow{-20} & 280 & \xrightarrow{5} & 285 & \end{array}$$

$$\begin{aligned} \underline{\beta}_9 &= \frac{-25 - 20 + 5}{15 + 5 + 15} \\ \underline{\beta}_9 &\approx -1,14 \end{aligned}$$

Existe grande probabilidade de que o efeito da real proteção ao ruído impulsivo em direção de *downstream* esteja entre -4 e 2 . Assim, pode-se escolher $\text{var}(\beta_9) = 9$ resultando em um desvio padrão igual a 3 , ou seja, como trata-se de uma distribuição gaussiana há uma probabilidade de 95% que o valor de β_9 esteja dentro da faixa que é o dobro do desvio padrão [4]. Portanto, entre -4 e 2 .

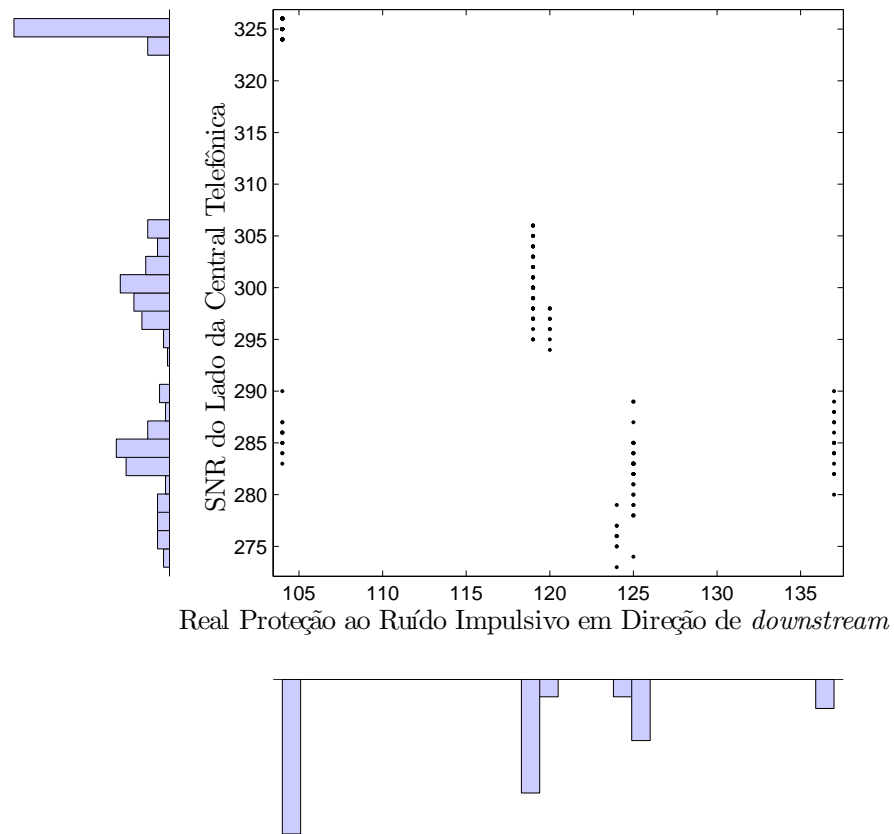


Figura 4.20: Gráfico de dispersão e histograma das variáveis real proteção ao ruído impulsivo em direção de *downstream* e SNR do lado da central telefônica.

Parâmetro *a priori* $\underline{\beta}_{10}$

O parâmetro β_{10} associado à variável explanatória comprimento do bloco de dados do canal no qual o CRC opera é obtido com base na variação que esta variável provoca na variável dependente SNR do lado da central telefônica com base no gráfico de dispersão e histograma da Figura 4.21. Assim, o parâmetro β_{10} é estimado da seguinte forma:

$$\begin{array}{l} \text{adslAtucChanCrcBlockLength} : \quad 1,1 \cdot 10^4 \quad \xrightarrow{0,5 \cdot 10^4} \quad 1,6 \cdot 10^4 \quad \xrightarrow{0,2 \cdot 10^4} \quad 1,8 \cdot 10^4 \\ \text{adslAtucCurrSnrMgn}(0,1 \cdot \text{dB}) : \quad 325 \quad \xrightarrow{-40} \quad 285 \quad \xrightarrow{15} \quad 300 \end{array}$$

$$\begin{aligned} \underline{\beta}_{10} &= \frac{-40 + 15}{0,5 \cdot 10^4 + 0,2 \cdot 10^4} \\ \underline{\beta}_{10} &\approx -3,57 \cdot 10^{-4} \end{aligned}$$

Existe grande probabilidade de que o efeito do comprimento do bloco de dados do canal no qual o CRC opera esteja entre -1 e 1 . Assim, pode-se escolher $\text{var}(\beta_{10}) = 1$ resultando em um desvio padrão igual a 1 , ou seja, como trata-se de uma distribuição gaussiana há uma probabilidade de 95% que o valor de β_{10} esteja dentro da faixa que é o dobro do desvio padrão [4]. Portanto, entre -1 e 1 .

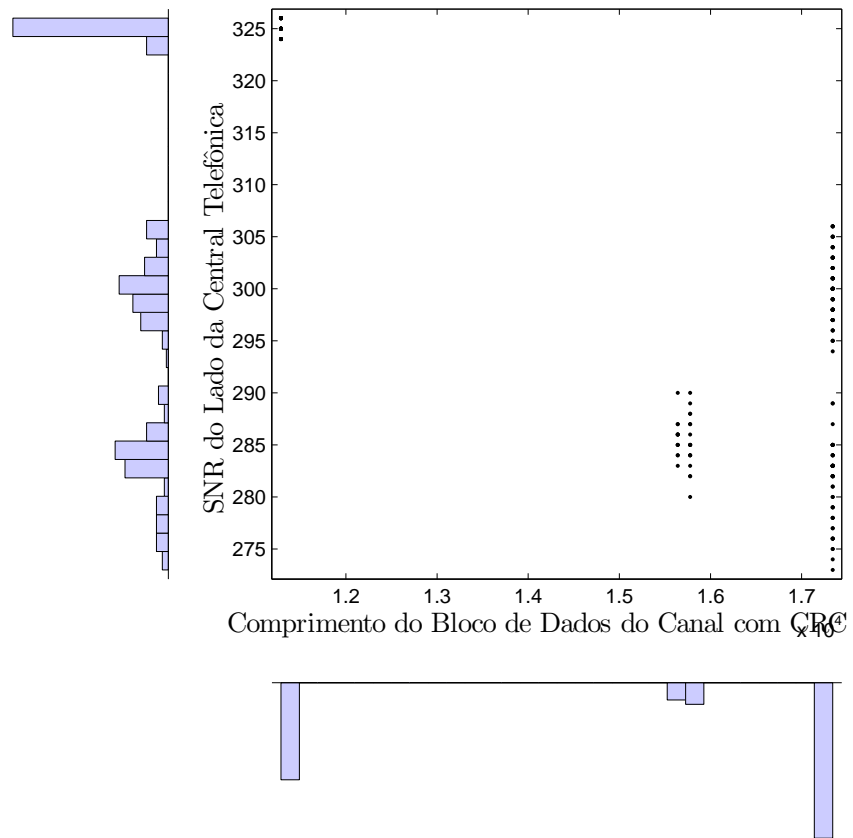


Figura 4.21: Gráfico de dispersão e histograma das variáveis comprimento do bloco de dados do canal no qual o CRC opera e SNR do lado da central telefônica.

Parâmetro *a priori* $\underline{\beta}_{11}$

O parâmetro β_{11} associado à variável explanatória taxa de *bits* real da linha na direção de *downstream* é obtido com base na variação que esta variável provoca na variável dependente SNR do lado da central telefônica com base no gráfico de dispersão e histograma da Figura 4.22. Assim, o parâmetro β_{11} é estimado da seguinte forma:

$$\begin{array}{l} \text{adslAtuc...ActualLineBitRate(bits/s)} : \quad 3 \cdot 10^7 \quad \xrightarrow{0,4 \cdot 10^7} \quad 3,4 \cdot 10^7 \quad \xrightarrow{1 \cdot 10^7} \quad 4,4 \cdot 10^7 \\ \text{adslAtucCurrSnrMgn}(0,1 \cdot \text{dB}) : \quad 280 \quad \xrightarrow{20} \quad 300 \quad \xrightarrow{25} \quad 325 \end{array}$$

$$\begin{aligned} \underline{\beta}_{11} &= \frac{20 + 25}{0,4 \cdot 10^7 + 1 \cdot 10^7} \\ \underline{\beta}_{11} &\approx 3,21 \cdot 10^{-6} \end{aligned}$$

Existe grande probabilidade de que o efeito da taxa de *bits* real da linha na direção de *downstream* esteja entre -1 e 1 . Assim, pode-se escolher $\text{var}(\beta_{11}) = 1$ resultando em um desvio padrão igual a 1 , ou seja, como trata-se de uma distribuição gaussiana há uma probabilidade de 95% que o valor de β_{11} esteja dentro da faixa que é o dobro do desvio padrão [4]. Portanto, entre -1 e 1 .

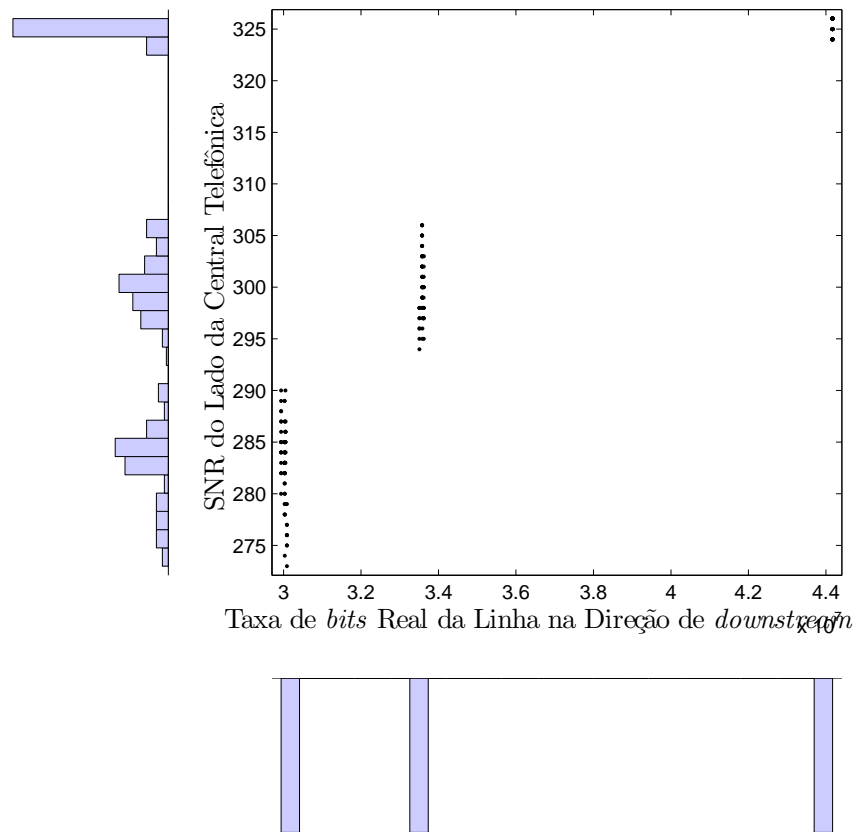


Figura 4.22: Gráfico de dispersão e histograma das variáveis taxa de *bits* real da linha na direção de *downstream* e SNR do lado da central telefônica.

De posse dessas informações *a priori* é possível obter o hiperparâmetro \underline{V} que, através das propriedades da distribuição Normal-Gamma [4], possui a forma:

$$var(\beta) = \frac{vs^2}{v-2}\underline{V}$$

Como $\frac{vs^2}{v-2} = 18,75$, o hiperparâmetro \underline{V} é dado por:

$$\underline{V} = \begin{bmatrix} 4157^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,48 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,0533 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,0533 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,0533 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1200 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6,4533 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 51,2533 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,48 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,0533 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,0533 \end{bmatrix}$$

4.5 Computação Bayesiana

Com as informações *a priori* elicitadas na seção anterior, alimentou-se um *script* escrito em MATLAB que faz a computação dos dados através das equações 3.20 a 3.30 e gera como saída os parâmetros *a posteriori* do modelo tanto para o caso em que os parâmetros são baseados apenas no dados — *priori* não informativa, quanto para o caso em que os parâmetros são baseados na combinação entre informação *a priori* e dados — *priori* informativa.

As Tabelas 4.1 e 4.2, mostram respectivamente os parâmetros *a posteriori* informativos e não informativos estimados e os desvios padrões associados a estes parâmetros. Os resultados apresentados na Tabela 4.1 reforçam o peso maior que se deu a informação baseada nos dados, uma vez que os resultados para o caso da densidade *a priori* informativa são similares ao caso em que a densidade *a priori* é não informativa.

	<i>priori</i>	<i>posteriori</i> com <i>priori</i> não informativa	<i>posteriori</i> com <i>priori</i> informativa
β_1	0,00	-8798,4261	-8057,5209
β_2	2,27	0,2904	0,3633
β_3	$3,21 \cdot 10^{-6}$	-0,0564	-0,0485
β_4	$2,81 \cdot 10^{-6}$	$3,4629 \cdot 10^{-5}$	$3,6274 \cdot 10^{-5}$
β_5	$2,81 \cdot 10^{-6}$	$1,4929 \cdot 10^{-6}$	$1,2633 \cdot 10^{-6}$
β_6	-25,00	91,6166	82,8656
β_7	-11,25	-1,2370	-1,2110
β_8	-9,00	-37,5052	-28,7850
β_9	-1,14	1,2125	1,0212
β_{10}	$-3,57 \cdot 10^{-4}$	0,0162	0,0135
β_{11}	$3,21 \cdot 10^{-6}$	0,0563	0,0483

Tabela 4.1: Média dos parâmetros *a priori* e *a posteriori*.

	<i>priori</i>	<i>posteriori</i> com <i>priori</i> não informativa	<i>posteriori</i> com <i>priori</i> informativa
β_1	18000	667,4649	619,2433
β_2	3	0,1616	0,1620
β_3	1	0,0063	0,0056
β_4	1	$6,8175 \cdot 10^{-6}$	$6,8602 \cdot 10^{-6}$
β_5	1	$3,1940 \cdot 10^{-7}$	$3,1123 \cdot 10^{-7}$
β_6	150	7,4604	6,8116
β_7	11	0,3644	0,3674
β_8	31	6,9791	6,1937
β_9	3	0,1528	0,1362
β_{10}	1	0,0022	0,0020
β_{11}	1	0,0063	0,0056

Tabela 4.2: Desvio padrão dos parâmetros *a priori* e *a posteriori*.

Além disso, pode-se notar claramente pela Tabela 4.2 que a maioria dos desvios padrões associados à *priori* informativa são menores que aqueles resultantes da *priori* não informativa. Portanto, a elicitación de informação *a priori* melhora a precisão dos parâmetros estimados.

	<i>priori</i>	<i>posteriori</i> com <i>priori</i> não informativa	<i>posteriori</i> com <i>priori</i> informativa
Média	0,16	0,3000	0,2899
Desvio padrão	0,40	0,0471	0,0461

Tabela 4.3: Média e desvio padrão do erro de precisão *a priori* e *a posteriori*.

Outro resultado que pode ser usado para notar os benefícios da abordagem bayesiana em se utilizar uma densidade *a priori* informativa é mostrado na Tabela 4.3. Pode-se ver que o erro de precisão é menor para o caso em que a *priori* é informativa do que para o caso em que os cálculos baseiam-se apenas nos dados.

No Capítulo 5, será realizada a análise dos resultados e testes para a validação do modelo proposto para a identificação do nível de ruído do lado da central telefônica.

Capítulo 5

Análise dos Resultados

Após a computação bayesiana realizada para se estimar os parâmetros do modelo de regressão que foi proposto para explicar o nível de ruído *crosstalk* do lado da central telefônica, este capítulo pretende realizar a análise dos resultados obtidos.

Nas próximas seções, pretende-se realizar a interpretação dos parâmetros estimados em termos da variável dependente SNR do lado da central telefônica; validar o modelo de regressão através da comparação entre o valor medido da variável dependente e o valor obtido pelo modelo de regressão; além disso, será feita a uma análise sobre o erro de regressão para o caso em que se utiliza uma densidade *a priori* informativa e para o caso em que a densidade *a priori* é não informativa.

5.1 Análise dos Parâmetros Estimados para o Modelo de Regressão

A Tabela 5.1 mostra os parâmetros *a posteriori* para os casos em que se utilizou densidades *a priori* informativas e não informativas. A fim de se interpretar esses parâmetros do modelo de regressão, pode-se utilizar como exemplos os parâmetros β_2 e β_7 .

	<i>posteriori com priori não informativa</i>	<i>posteriori com priori informativa</i>
β_1	-8798,4261	-8057,5209
β_2	0,2904	0,3633
β_3	-0,0564	-0,0485
β_4	$3,4629 \cdot 10^{-5}$	$3,6274 \cdot 10^{-5}$
β_5	$1,4929 \cdot 10^{-6}$	$1,2633 \cdot 10^{-6}$
β_6	91,6166	82,8656
β_7	-1,2370	-1,2110
β_8	-37,5052	-28,7850
β_9	1,2125	1,0212
β_{10}	0,0162	0,0135
β_{11}	0,0563	0,0483

Tabela 5.1: Parâmetros *a posteriori* do modelo utilizando-se *priori* informativa e não informativa.

O parâmetro β_2 exprime a influência que a variável explanatória SNR do lado das instalações do cliente possui sobre a variável dependente SNR do lado da central telefônica. Como pode-se notar, o parâmetro β_2 possui valor 0,2904 para a *priori* não informativa e 0,3633 para a *priori* informativa. Os valores positivos desse parâmetro significam que o aumento de 1 unidade de SNR do lado das instalações do cliente reflete o aumento de 0,2904 unidades de SNR do lado da central telefônica quando se utiliza uma *priori* não informativa e 0,3633 unidades quando se utiliza uma *priori* informativa [4].

Já o parâmetro β_7 exprime a influência que a variável explanatória potência total de saída medida transmitida pelas instalações do usuário possui sobre a variável dependente SNR do lado da central telefônica. Como pode-se notar, o parâmetro β_7 possui valor $-1,2370$ para a *priori* não informativa e $-1,2110$ para a *priori* informativa. Nesse caso, os valores negativos desse parâmetro significam que o aumento de 1 unidade da potência total de saída medida transmitida pelas instalações do usuário reflete a redução de 1,2370 unidades de SNR do lado da central telefônica quando se utiliza uma *priori* não informativa e 1,2110 unidades quando se utiliza uma *priori* informativa.

Além disso, como o módulo do parâmetro β_7 é maior que o módulo do parâmetro β_2 , isso significa que a variável explanatória potência total de saída medida transmitida pelas instalações do usuário possui maior influência sobre a SNR do lado da central telefônica do que a variável explanatória SNR do lado das instalações do cliente.

5.2 Resultados Calculados através do Modelo de Regressão

A fim de se validar o modelo de regressão linear proposto neste trabalho, utilizou-se os parâmetros encontrados no capítulo anterior juntamente com as variáveis explanatórias da base de dados para se calcular a variável dependente SNR do lado da central telefônica obtida pelo modelo de regressão.

As curvas obtidas utilizando o modelo de regressão linear e as variáveis explanatórias medidas resultaram nas curvas das Figuras 5.1 e 5.2 que utilizam respectivamente densidades *a priori* não informativas e informativas. Pode-se observar que o modelo de regressão obtido com *prioris* informativas ou não informativas são bastante similares. E, ainda mais importante que isso, ambas as curvas calculadas possuem uma excelente aproximação para as curvas medidas em laboratório mostrando que o modelo proposto pode descrever com bastante acurácia o comportamento da variável dependente SNR do lado da central telefônica em função das demais variáveis explanatórias.

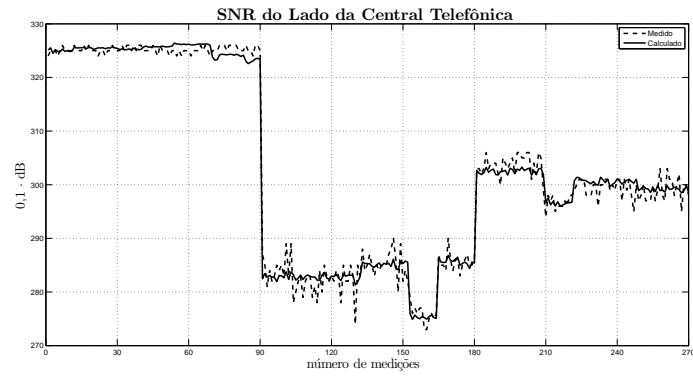


Figura 5.1: Curvas medidas e calculadas para a variável SNR do lado da central telefônica utilizando-se o modelo com *priori* não informativa.

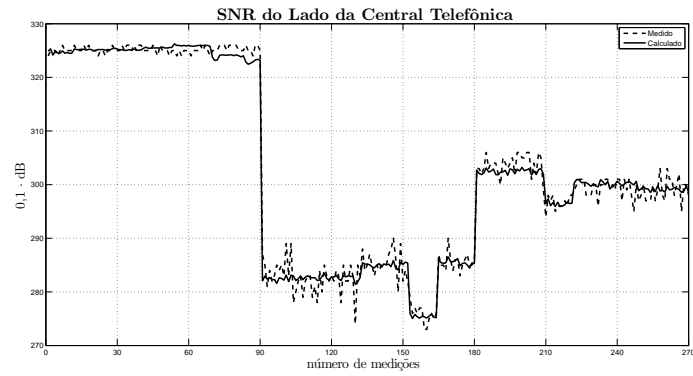


Figura 5.2: Curvas medidas e calculadas para a variável SNR do lado da central telefônica utilizando-se o modelo com *priori* informativa.

5.3 Erros de Regressão

Nesta seção, será realizada a análise dos erros de regressão que estão associados ao modelo proposto neste trabalho. Os erros de regressão ε podem ser calculados como sendo a diferença entre a variável dependente medida e a variável dependente calculada pelo modelo, como sintetizado pela equação 5.1.

$$\varepsilon = y - X\beta \quad (5.1)$$

Ou de outra forma, como mostrado pela equação 5.2.

$$\varepsilon = y_{medido} - y_{calculado} \quad (5.2)$$

As Figuras 5.3 e 5.4 mostram respectivamente os gráficos dos erros associados aos modelos de regressão que utilizam *priori* não informativa e informativa. Pode-se notar que os erros associados são bastante pequenos e estão distribuídos entre -8 e 6 décimos de dB. Os erros de regressão médios tanto para o caso de *priori* não informativa quanto para o caso de *priori* informativa são da ordem de 0,1835 décimos de dB, ou seja, bastante aceitáveis [4].

Além disso, outra característica do erro de regressão linear que pode ser notada pelos histogramas das Figuras 5.5 e 5.6 é a distribuição destes valores que parece bastante com uma distribuição Normal com média em zero, como foi suposto. Dessa forma, pode-se ver que a maioria dos erros associados são próximos de zero reforçando a adequação do modelo proposto [4].

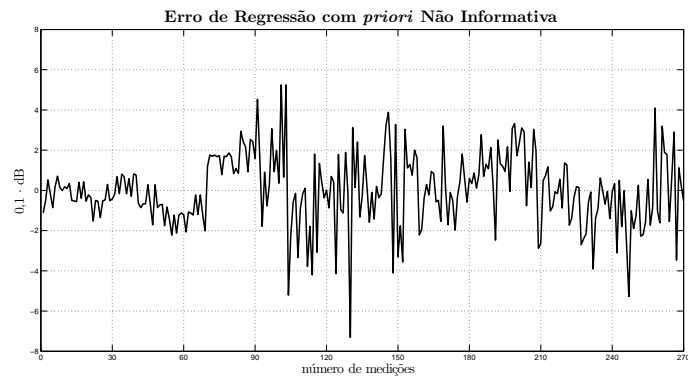


Figura 5.3: Erro de regressão para a variável SNR do lado da central telefônica utilizando-se o modelo com *priori* não informativa.

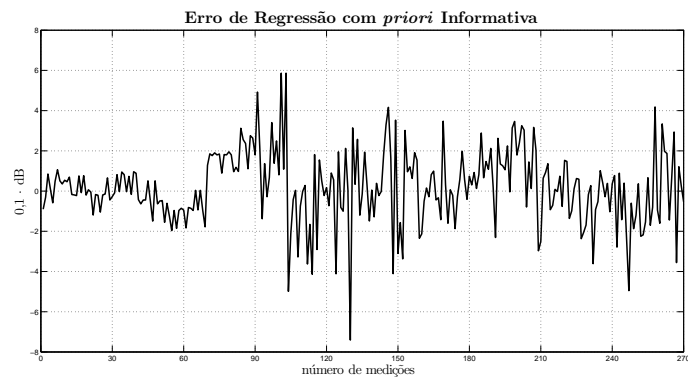


Figura 5.4: Erro de regressão para a variável SNR do lado da central telefônica utilizando-se o modelo com *priori* informativa.

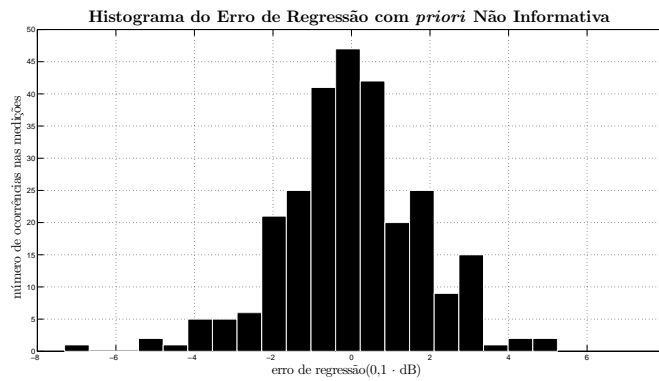


Figura 5.5: Histograma do erro de regressão para a variável SNR do lado da central telefônica utilizando-se o modelo com *priori* não informativa.

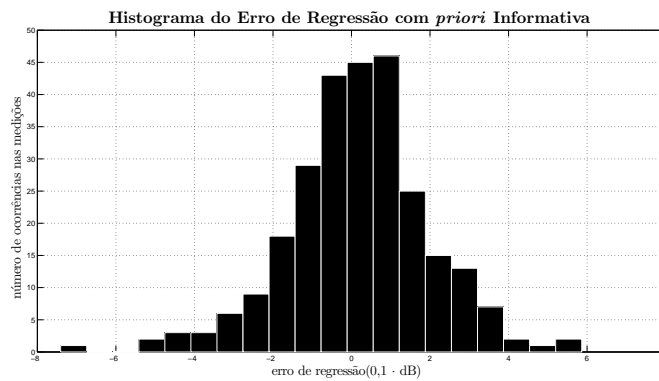


Figura 5.6: Histograma do erro de regressão para a variável SNR do lado da central telefônica utilizando-se o modelo com *priori* informativa.

Capítulo 6

Considerações Finais

O objetivo principal do trabalho descrito neste documento foi identificar o nível de ruído *crosstalk* do lado da central telefônica em enlaces DSL, utilizando como exemplo o caso específico de um enlace VDSL2 e verificar a influência que as variáveis explanatórias medidas têm sobre o nível desse ruído. Isso foi feito utilizando-se a abordagem bayesiana para estimar os parâmetros de um modelo de regressão linear com múltiplas variáveis que busca explicar o nível de ruído *crosstalk* do lado da central telefônica. Neste capítulo, as principais conclusões do trabalho como os objetivos alcançados e sugestões para trabalhos futuros são apresentados.

6.1 Objetivos Alcançados

Uma das principais contribuições deste trabalho foi a obtenção de um modelo de regressão linear capaz de relacionar as variáveis selecionadas para serem explanatórias para o cálculo do nível de ruído *crosstalk* do lado da central telefônica.

Além disso, um ponto importante deste trabalho foi a avaliação da influência que as variáveis explanatórias possuem sobre a variável dependente margem de ruído *crosstalk* do lado da central telefônica. A seguir são feitos comentários baseados nos resultados obtidos com relação à influência das variáveis explanatórias sobre a variável dependente:

- Margem de ruído do lado das instalações do cliente: possui pouca influência sobre a variável dependente, porém seu aumento resulta um aumento de 0,3633 décimos de dB na variável dependente;
- Taxa de transmissão real neste canal: possui uma influência menor ainda sobre a variável dependente. Um aumento desta variável implica diminuição de 0,0485 décimos de dB na variável dependente;
- Taxa de dados máxima atingível atualmente pela central telefônica: possui influência positiva sobre a variável dependente, ou seja, o aumento dessa variável explanatória provoca um aumento na variável dependente, porém muito pequena (da ordem de $3,6274 \cdot 10^{-5}$ décimos de dB);
- Taxa de dados máxima correntemente atingível pelas instalações do cliente: também possui influência positiva sobre a variável dependente, porém muito pequena (cerca de $1,2633 \cdot 10^{-6}$ décimos de dB);
- Potência total de saída medida transmitida pela central telefônica: é a variável que possui maior influência sobre a variável dependente. Um aumento nesta variável provoca grande aumento na variável dependente (em torno de 82,8656 décimos de dB);
- Potência total de saída medida transmitida pelas instalações do usuário: possui alguma influência sobre a variável dependente (cerca de 1,211 décimos de dB), porém de forma negativa, ou seja, o aumento desta variável resulta na diminuição do valor da variável dependente.
- Atraso de *interleave*: possui grande influência sobre a variável dependente (em torno de 28,785 décimos de dB), porém de forma negativa;
- Real proteção ao ruído impulsivo: possui alguma influência positiva sobre a variável dependente (da ordem de 1,0212 décimos de dB).
- Comprimento do bloco de dados do canal no qual o CRC opera: possui pouca influência sobre a variável dependente e de forma positiva (cerca de 0,0135 décimos de dB);

- Taxa de *bits* real da linha na direção de *downstream*: seu aumento provoca um aumento pequeno na variável dependente (aproximadamente 0,0483 décimos de dB).

6.2 Limitações deste Trabalho e Pesquisas Futuras

As contribuições feitas por este estudo mostraram que embora o ruído *crosstalk* possua natureza aleatória, a técnica apresentada neste trabalho mostrou-se bastante promissora em detectar a presença, bem como o nível desse ruído em um enlace DSL. Porém, algumas suposições adotadas neste estudo limitam as conclusões deste trabalho de dar uma estimativa mais precisa dos parâmetros da regressão.

Em primeiro lugar, este estudo é baseado na suposição de que a influência das variáveis explanatórias sobre a variável dependente pode ser resumida através de uma relação linear, uma vez que foi proposto um modelo linear. Embora os resultados deste trabalho tenham sido satisfatórios com erros bastante próximos de zero, seria interessante em pesquisas futuras que fossem propostos modelos de regressão do tipo não lineares.

Outro ponto de interesse está em estender o número de medições para que se obtenha uma base de dados mais completa para ser analisada a fim de se estimar parâmetros mais precisos para o modelo de regressão. Dessa forma, o modelo seria capaz de responder adequadamente para uma gama maior de situações que possam ocorrer.

Finalmente, seria de grande valia a identificação do nível de ruído impulsivo e interferência de rádio frequência em sistemas VDSL2. Uma vez que o impacto desses ruídos em sistemas VDSL2 é bastante prejudicial e sua identificação proporcionaria a aplicação de técnicas mais adequadas e precisas de mitigação.

Referências Bibliográficas

- [1] Nedko H. Nedev. *Analysis of the Impact of Impulse Noise in Digital Subscriber Line Systems*. PhD thesis, The University of Edinburgh, 2003.
- [2] Philip Golden, Hervé Dedieu, and Krista Jacobsen, editors. *Fundamentals of DSL Technology*. Auerbach Publications, 472 pg., 2006.
- [3] Philip Golden, Hervé Dedieu, and Krista Jacobsen, editors. *Implementation and Applications of DSL Technology*. Auerbach Publications, 1^o edition, 816 pg., 2008.
- [4] Gary Koop. *Bayesian Econometrics*. John Wiley & Sons Ltd, 1^o edition, 374pg., 2003.
- [5] Thomas Starr, John M. Cioffi, and Peter J. Silverman. *Understanding Digital Subscriber Line Technology*. Prentice Hall, 1^o edition, 480 pg., 1999.
- [6] ITU-T Recommendation G.991.1. *High Bit ate Digital Subscriber Line (HDSL) Transceivers*, 1998.
- [7] ITU-T Recommendation G.991.2. *Single-pair High-speed Digital Subscriber Line (SHDSL) Transceivers*, 2003.
- [8] ITU-T Recommendation G.994.1. *Handshake Procedures for Digital Subscriber Line (DSL) Transceivers*, 2007.
- [9] ITU-T Recommendation G.992.1. *Asymmetric Digital Subscriber Line (ADSL) Transceivers*, 1999.
- [10] ITU-T Recommendation G.992.3. *Asymmetric Digital Subscriber Line Transceivers 2 (ADSL2)*, 2009.

- [11] ITU-T Recommendation G.992.5. *Asymmetric Digital Subscriber Line Transceivers 2 (ADSL2) – Extended Bandwidth (ADSL2plus)*, 2009.
- [12] ITU-T Recommendation G.993.1. *Very High Speed Digital Subscriber Line Transceivers*, 2004.
- [13] John Cioffi, Thomas Starr, and Massimo Sorbara. *DSL Advances*. Prentice Hall, 1^o edition, 576 pg., 2002.
- [14] ITU-T Recommendation G.993.2. *Very High Speed Digital Subscriber Line Transceivers 2 (VDSL2)*, 2006.
- [15] Eduardo L. Medeiros. *Um Protótipo para Desenvolvimento e Aplicação de Estratégias de Otimização em Sistemas xDSL Práticos*. Master's thesis, Universidade Federal do Pará, 2010.
- [16] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley & Sons Inc, 2^o edition, 654 pg., 2001.
- [17] Ericsson Telecommunications. *EDN612 Standard MIB Description*, 2009.
- [18] Ericsson Telecommunications. *EDN312x, EDN312, EDN110 Proprietary MIB Description*, 2009.
- [19] ITU-T Recommendation G.997.1. *Physical Layer Management for Digital Subscriber Line (DSL) Transceivers*, 2009.