

Método de Identificação de Ruído em Redes DSL usando Rede Neural Artificial

F. S. Farias, G. S. Borges, R. V. M. P. Pantoja, D. L. L. Silva, W. B. Monteiro, J. C. W. A. Costa

Federal University of Pará (UFPA), Rua Augusto Corrêa nº 01 – Guamá - P.O. Box: 8616, Belém-Pará, 66.075-900, Brazil, Tel: +559132017740, Email: {fabriciosf, gilvan, rvillar, jweyl}@ufpa.br; {daniel.levy, waldeir.monteiro}@itec.ufpa.br

Resumo— Este artigo propõe uma nova metodologia para identificação de ruídos em redes *Very-high-bit-rate Digital Subscriber Line* (VDSL). Uma Rede Neural Artificial (RNA) baseada no algoritmo *backpropagation* é usada para classificar, em tempo real, a ausência de ruído ou a predominância de ruído impulsivo, ou *crosstalk*. Os resultados demonstraram o potencial da utilização de algoritmos de inteligência computacional para a identificação de ruídos em redes VDSL. Dentre os testes, o melhor resultado alcançado foi de 99.3% de acerto.

Palavras-Chave—Redes Neurais, Identificação de Ruído, VDSL.

Abstract— This paper proposes a new methodology for noise identification in *Very-high-bit-rate Digital Subscriber Line* (VDSL). Artificial Neural Networks based on *backpropagation* algorithm is used for real time classification of absence or predominance of noise (*crosstalk* and impulsive). Results show the potential of applying computational intelligence algorithms in order to identify noise on VDSL networks. The best result achieved 99.3% accuracy.

Keywords—Artificial Neural Networks, Noise Identification, VDSL.

I. INTRODUÇÃO

TECNOLOGIAS baseadas na *Digital Subscriber Line* (DSL) transformam uma simples linha telefônica em um enlace de comunicação banda larga por meio de técnicas robustas de transmissão digital, que compensam as muitas interferências comuns em linhas telefônicas [1]. Dessa forma, as redes DSL revolucionaram a utilização de linhas telefônicas.

Apesar da crescente utilização de fibras ópticas e de arquiteturas tais como FTTC (*Fiber To The Cabinet*) e FTTH (*Fiber To The Home*) o uso de DSL para transmissão de dados ainda encontra nichos de utilização em termos de custo, em particular, em soluções de última milha tanto para usuários domésticos quanto usuários empresariais.

As tecnologias DSL mais recentes, em operação, são as *Very-high-bit-rate* DSL (VDSL) e VDSL2, as quais operam em enlaces curtos na ordem de centenas de metros e provém a comunicação, em geral, entre o armário da rede telefônica localizado na rua e o equipamento do usuário. Desta forma, percebe-se que as tecnologias DSL ainda terão uma grande participação no mercado de redes de acesso banda larga, seja pelo menor custo de instalação, ou por fornecer velocidades

compatíveis com os serviços de dados, os quais necessitam de alta taxa de transmissão tais como jogos online, videoconferência e VoIP.

Apesar dos avanços e melhorias na tecnologia DSL, seu principal problema continua sendo o impacto de ruído. Desta forma, o estudo na direção do desenvolvimento de formas da identificação, mitigação e de estimação de ruído é de grande importância [2]-[6]. Em [2] os autores apresentam um algoritmo de estimação de *crosstalk* em sistemas DSL considerando o impacto de tal ruído na Relação Sinal Ruído (*Signal-To-Noise Ratio* - SNR) de *downstream* do sistema. Em [3] os autores comparam três técnicas de mitigação de *crosstalk*, quando parâmetros como restrição de potência e frequência de operação são considerados. Em [4] o autor apresenta uma técnica de identificação e mitigação de *crosstalk* e ruído impulsivo em redes VDSL utilizando modulação discreta de múltiplos tons. Em [5] os autores apresentam uma estratégia para estimação de *crosstalk* utilizando informações de SNR obtidas do lado do usuário (*Customer Premises Equipment* - CPE). Em [6] é proposto um método para identificação de várias fontes de ruído *crosstalk* por meio do método dos mínimos quadrados. A densidade espectral de potência é utilizada como conhecimento a priori.

Ressalta-se que os artigos citados baseiam-se em modelos físicos do canal de comunicação DSL.

Dentre os diversos tipos de ruídos, o *crosstalk* é o ruído predominante em redes DSL [3][7]. O ruído impulsivo é o mais complexo de ser caracterizado e gera os maiores danos, como perda de conexão e diminuição do alcance do enlace [8]. Com o desenvolvimento de técnicas eficientes para o controle do *crosstalk*, o monitoramento de ruídos extrínsecos, como rádio frequência [9] e impulsivo [10], ganhou maior importância nos últimos anos.

Neste artigo é proposta uma estratégia de monitoramento de ruídos em redes DSL utilizando inteligência computacional. Foram efetuadas simulações utilizando o algoritmo baseado em Rede Neural Artificial (*backpropagation*).

Este estudo está estruturado da seguinte forma: a seção II apresenta os passos da Descoberta de Conhecimento em Base de Dados (*Knowledge Discovery in Databases* - KDD) aplicados neste artigo. A seção III apresenta os cenários DSL considerados. A seção IV apresenta os resultados obtidos com os algoritmos analisados. Finalmente, a última seção traz as conclusões finais deste artigo.

II. APLICAÇÃO DO KDD

O KDD (vide a Figura 1) consiste em um conjunto de técnicas organizadas em etapas, que iniciam desde a criação da base de dados até a extração de conhecimento através da aplicação de uma técnica de inteligência computacional ou estatística no processo de mineração de dados, também conhecido como *data mining*.

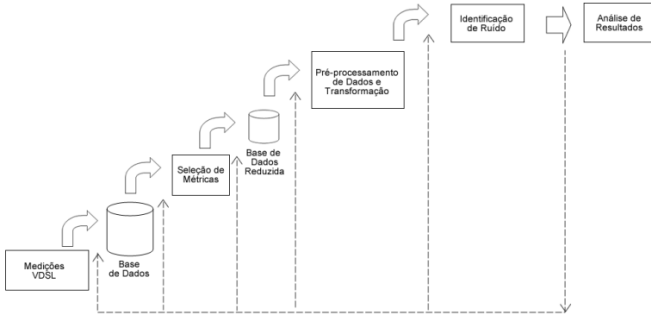


Figura 1: Fluxograma resumindo as etapas do KDD.

Na etapa de criação da base de dados, foram realizadas medições sem a injeção de ruído e com a injeção de ruídos de *crosstalk* e impulsivo.

As informações coletadas para composição da base de dados são medidas do *Management Information Base* (MIB) [11], as quais são um conjunto de dados retornados pelo modem e pelo *DSL Access Multiplexer* (DSLAM), os quais procuram abranger todas as informações necessárias para a gestão da rede. Tais métricas são gerenciadas pelo *Simple Network Management Protocol* (SNMP), o qual é um protocolo da camada de aplicação responsável em gerenciar dispositivos ligados em rede que permite a avaliação do desempenho e a alteração de determinadas configurações.

Dentre as métricas disponíveis no DSLAM, encontram-se algumas que medem a margem de SNR, o número de blocos corrigidos e a taxa de transmissão da central para o usuário, sendo algumas dessas métricas relevantes para a identificação do ruído.

A base foi composta a partir de medições realizadas com diferentes potências de ruído, de modo a garantir a diversidade dos dados.

Os registros coletados foram divididos em três classes, sendo 149 de *crosstalk*, 39 de ruído impulsivo e 94 sem injeção de ruído, totalizando 282 amostras, como apresentado na Tabela 1.

TABELA 1. DIVISÃO DOS REGISTROS

| Tipo de Ruído | Número de Amostras |
|------------------|--------------------|
| <i>Crosstalk</i> | 148 |
| Ruído impulsivo | 39 |
| Sem Ruído | 94 |

Durante a fase de medições VDSL foram coletados 60 atributos (métricas MIB) a cada 15 segundos, resultando em aproximadamente 11 horas de medição. Após as etapas iniciais do KDD, apenas 13 atributos foram usados como entrada no algoritmo simulado. A relevância de cada atributo foi estudada nas etapas do KDD [12].

O processo de seleção dos atributos relevantes ao modelo classificador foi realizado através de análise da variação dos registros de cada atributo de acordo com a presença ou a ausência de ruído. A base de dados reduzida foi normalizada na etapa de pré-processamento e transformação. Em seguida, os algoritmos de mineração de dados foram aplicados e uma análise dos resultados foi feita.

III. BACKPROPAGATION

A construção da Rede Neural Artificial (RNA) tem inspiração nos neurônios biológicos e nos sistemas nervosos humanos [13]. O principal objetivo da RNA é simular o aprendizado humano através de modelos matemáticos do neurônio.

De uma maneira geral, uma RNA pode ser vista como um sistema composto por unidades de processamento simples capazes de calcular algumas funções matemáticas dispostas em uma ou mais camadas interligadas por conexões. Na maioria dos modelos de RNAs, as conexões possuem pesos associados que armazenam o conhecimento adquirido e ponderam a entrada recebida por cada neurônio [14].

Uma RNA baseada no algoritmo *backpropagation* possui a etapa de propagação (*forward*) e retropropagação (*backward*).

O passo de propagação inicia pelo envio de sinais de entrada, representadas pelas métricas MIB. Esse processo se repete até a camada de saída, onde o cálculo de um valor de saída é realizado. O *backpropagation* calcula o erro comparando o valor calculado na saída da rede com a saída desejada apresentada. A correção do erro é baseada no método do gradiente descendente, através da retropropagação. Na retropropagação novos conjuntos de pesos são iterativamente calculados, através da modificação dos pesos existentes, buscando minimizar o erro até um erro mínimo global. O Erro Médio Quadrático (EMQ) é normalmente usado como uma medição do erro global.

A estrutura básica do funcionamento do algoritmo ocorre como é descrito em (1) - (6).

Primeiramente a matriz a ser treinada é normalizada na Eq. (1), em seguida é multiplicada pela matriz de pesos gerados aleatoriamente e a RNA *net* é criada em (2). A RNA *net* é apresentada à função de ativação do algoritmo em (3) e calculada em (5). Em (4) a saída gerada pelo treinamento está pronta para a comparação e realização do cálculo do erro da saída desejada em relação à saída encontrada.

$$x_j = \text{normaliza}(X_m) \quad (1)$$

$$\text{net}_k = \sum_{j=1}^{k-1} W_{kj} x_j \quad m \leq k \leq N + m \quad (2)$$

$$x_j = f(\text{net}_k) \quad m \leq k \leq N + m \quad (3)$$

$$O_{s-1} = x_{N+s} \quad 1 \leq s \leq n \quad (4)$$

$$f(\text{net}_k) = \frac{1}{1 + e^{-\text{net}_k}} \quad (5)$$

Em que X representa as entradas atuais para a rede; X_j representa a entrada normalizada; m é o número de entradas para a rede; n é o número de saídas da rede neural; W_{kj} são os

elementos da matriz de peso, nos quais, o tamanho depende do número de neurônios na camada adjacente correspondente da rede neural; N é o número de neurônios na rede neural, o qual é uma constante; O_s são saídas a partir da rede neural.

A função de ativação (5) é uma função *sigmoid*, em que k é um escalar positivo. A função *sigmoid* é uma função contínua que permite uma transição gradual e não linear entre os estados. Tal função foi utilizada neste problema devido ser a função mais utilizada na definição de uma RNA multicamadas [15].

O passo de retropropagação do aprendizado de uma RNA é um processo iterativo de ajuste dos pesos das conexões que armazenam o conhecimento que a rede adquiriu do ambiente externo. De um modo geral, o valor do vetor de pesos $\mathbf{w}(t+1)$ no instante $t+1$ pode ser escrito conforme (6), onde $\mathbf{w}(t)$ e $\mathbf{w}(t+1)$ representam os valores dos pesos nos instantes t e $t+1$, respectivamente, e $\Delta\mathbf{w}(t)$ é o ajuste aplicado.

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \Delta\mathbf{w}(t) \quad (6)$$

A retropropagação ocorre em várias iterações percorrendo os registros de entrada até que os parâmetros da rede estabilizem seu valor e o EMQ seja menor que um mínimo definido.

IV. ARQUITETURA DOS CENÁRIOS

Visando identificar os distúrbios no sistema VDSL, por meio dos algoritmos propostos, é necessário construir uma base de dados através de medições VDSL.

A arquitetura usada durante as medições é apresentada na Figura 2. As medições foram realizadas usando linhas DSL reais, cuja ligação aos outros equipamentos do cenário é realizada por cabos Ethernet indicados pelas linhas vermelhas. As linhas roxas representam os cabos Ethernet responsáveis por transmitir os dados a partir do gerador/analizador de tráfego para o DSLAM e do modem VDSL2 para o gerador/analizador de tráfego. A linha verde representa os cabos Ethernet utilizados para gerenciar o teste e coletar as métricas a partir do DSLAM. Além disso, ela conecta o Switch ao DSLAM, os quais foram usados para transmitir os dados de teste e gerenciar a comunicação. A descrição dos equipamentos que compõem a Figura 2 está disponível na Tabela 2.

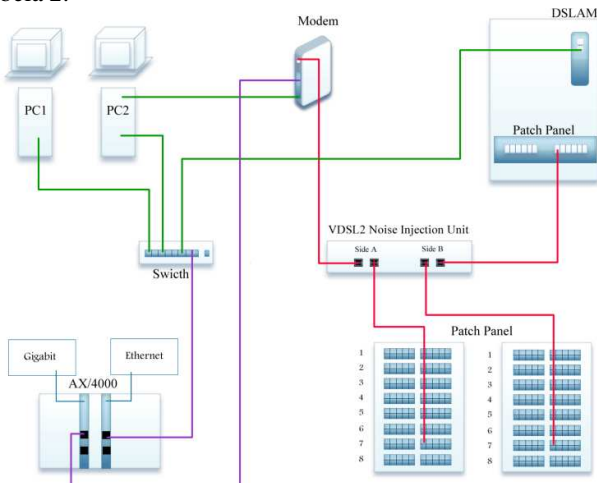


Figura 2: Cenário e seus equipamentos.

TABELA 2. EQUIPAMENTOS UTILIZADOS DURANTE OS TESTES NO LABORATÓRIO.

| Equipamento | Função |
|---|--|
| AX/4000 from Spirent Communications | Geração e análise de tráfegos no sistema. |
| VDSL2 Noise Injection Unit - DLS 5500 from Spirent Communications | Gerador de ruído impulsivo, branco e <i>crosstalk</i> . |
| Comtrend VDSL2 Modem | Equipamento do lado do cliente ou <i>Customer Premises Equipment (CPE)</i> . |
| Ericsson DSLAM EDN612p | Central telefônica ou <i>Central Office (CO)</i> . |
| Cabos reais | Loop DSL. |
| Computadores | Gerenciadores de teste. |

A recomendação utilizada para a escolha dos ruídos injetados foi baseada na *International Telecommunication Union (ITU) recommendation*. Dois tipos de ruídos foram aplicados no cenário projetado.

Objetivando validar a utilização de inteligência computacional para identificação de ruídos em redes DSL, foi utilizado um cenário com bitola 0.4 mm e comprimento 50 m como experimento inicial.

V. RESULTADOS

Nesta seção é apresentada uma descrição detalhada do algoritmo de classificação baseado na RNA. Durante as simulações os parâmetros de taxa de aprendizagem, número de neurônios na camada escondida, número de épocas, e base de dados de treinamento, validação e teste foram combinados e alterados de forma heurística durante 100 simulações.

Dentre as simulações realizadas a configuração da RNA mais adequada ao problema abordado possui treze neurônios na camada de entrada (representando as métricas selecionadas na etapa de seleção e normalizados na etapa de pré-processamento e transformação), dez neurônios na camada escondida e três neurônios na camada de saída (representando as três possíveis classes de saída – sem ruído, com *crosstalk*, ou com ruído impulsivo).

A base de dados foi dividida em treinamento, validação e teste, com 75%, 15% e 10% respectivamente. A função de ativação foi a *sigmoid*, como citado anteriormente. A taxa de aprendizagem foi de 0,1.

O algoritmo executou 55 épocas e obteve seu melhor resultado na época 49 alcançando o erro médio absoluto de 0.0095. Durante cada simulação foram calculados os EMQs e a taxa de registros classificados de forma incorreta (*misclassified*) em porcentagem (E%) para cada uma das três etapas, conforme pode ser visto Tabela 3. Nota-se que na etapa de validação a taxa de registros *misclassified* alcançada foi de 0%.

TABELA 3. RESULTADOS ALCANÇADOS PELA RNA.

| Etapas | Nº de Registros | EMQ | E% |
|-------------|-----------------|--------|------|
| Treinamento | 212 | 0.003 | 0.47 |
| Validação | 28 | 0.0002 | 0 |
| Teste | 42 | 0.008 | 2.4 |

A Figura 3 mostra o comportamento do EMQ graficamente para treino, validação e teste em todas as épocas percorridas pelo algoritmo. A porcentagem de exatidão alcançada pela RNA após as etapas de treinamento, validação e teste foi de 99.3%.

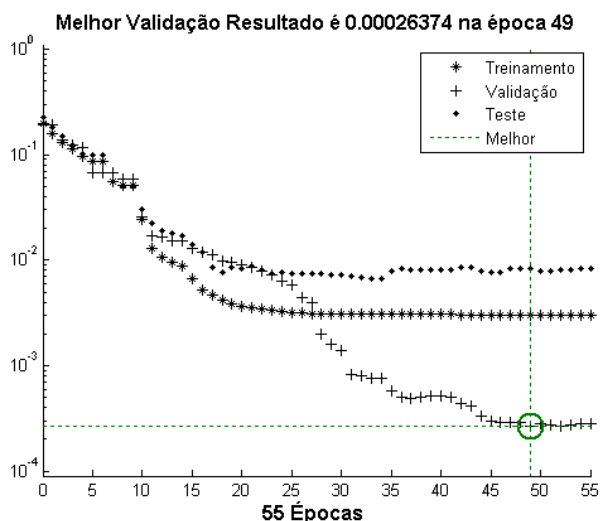


Figura 3: Representação gráfica dos resultados obtidos por época.

VI. CONCLUSÃO

Com base nos resultados alcançados, conclui-se que a proposta pode ser atrativa para as operadoras de telefonia e fornecedores do serviço VDSL, pois este tipo de metodologia não interfere no oferecimento do serviço de transmissão de dados e utilizam informações obrigatoriamente medidas pela central e pelo modem residencial.

Os resultados demonstram o elevado potencial da utilização de algoritmos de inteligência computacional, com o propósito de monitoramento em dados coletados da camada de aplicação (métricas MIB) para retirar informações da camada física (identificar ruído) em linha ativa, em tempo real, e sem necessidade de acréscimo de hardware.

A RNA obteve 99.3% de acerto na época 49 de 55. Confirmando o poder de generalização da RNA, durante a etapa de validação a porcentagem de acerto alcançada foi de 100%.

Com a utilização da RNA o *crosstalk* se configurou como o tipo de ruído mais difícil de ser identificado por um algoritmo de inteligência computacional devido ao seu baixo nível de potência. Em alguns casos, o *crosstalk* se torna confundível com uma medição sem a injeção de ruído.

Em trabalhos futuros são esperados a expansão da base de dados utilizada nas etapas de treinamento, e a aplicação de

outros algoritmos de inteligência computacional objetivando minimizar o erro.

AGRADECIMENTOS

Os autores agradecem o apoio financeiro recebido da Agência Sueca para Sistemas de Inovação, VINNOVA, através da Eureka - Celtic BANITS do projeto e do Centro de Pesquisa e Desenvolvimento, Ericsson Telecomunicações SA, Brasil.

Os Autores agradecem ao CNPQ – Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo apoio financeiro e logístico.

REFERÊNCIAS

- [1] Nedko H. Nedev. Analysis of the Impact of Impulse Noise in Digital Subscriber Line Systems. PhD thesis, The University of Edinburgh, 2003.
- [2] M. Guenach, J. Louveaux, L. Vandendorpe, P. Whiting, J. Maes, and M. Peeters. Performance analysis of the signal-to-noise ratio assisted crosstalk channel estimation for DSL systems, IEEE ICC 2009 proceedings, 2009.
- [3] D. Zanatta Filho, R. R. Lopes, R. Ferrari, M. B. Loiola, R. Suyama, G. C. C. P. Simões, and B. Dortschy. Achievable rates of DSL with crosstalk cancellation, European Transactions on Telecommunications, vol. 20, pp. 81-86, 2009.
- [4] T. A. Kumar. A Robust Multiuser Detection Based Scheme for Crosstalk Mitigation in DMT VDSL with Non-Gaussian Noise. 2009 International Conference on Signal Acquisition and Processing, pp. 234-238, 2009.
- [5] P. Whiting, A. Ashikhmin, G. Kramer, C. Nuzman, A. J. van Wijngaarden, M. Zivkovic, M. Peeters, M. Guenach, J. Maes, and J. Verlinden. DSL Crosstalk Coefficient Acquisition Using SNR Feedback, IEEE "GLOBECOM" 2008, pp. 1-5, 2008.
- [6] Ziqang Yang, Udayan Dasgupta, Arthur Redfern, and Murtaza Ali. Noise Identification in ADSL Modem. Texas Instrument DSP R&D Lab, Communications Systems Lab.
- [7] John W. Cook, Rob H. Kirkby, Martin G. Booth, Kevin T. Foster, Don E. A. Clarke, and Gavin Young. The Noise and Crosstalk Environment for ADSL and VDSL Systems. IEEE Communications Magazine. May 1999.
- [8] Tareq Y. Al-Naffouri, Ahmed A. Quadeer, and Giuseppe Caire. Impulsive Noise Estimation and Cancellation in DSL using Orthogonal Clustering. 2011 IEEE International Symposium on Information Theory Proceedings.
- [9] Tareq Y. Al-Naffouri, Ahmed A. Quadeer, and Giuseppe Caire. Impulsive Noise Estimation and Cancellation in DSL using Orthogonal Clustering. 2011 IEEE International Symposium on Information Theory Proceedings.
- [10] Vladimir Poulkov, Miglen Ovtcharov, Georgi Iliev, and Zlatka Nikolova. Radio Frequency Interference Mitigation in GDSL MIMO Systems Using Adaptive Complex Narrowband Filter Banks. Telsinks 2009.
- [11] MIB. [Online]. Available: <http://www.snmpLink.org/OnLineMIB/Standards/index.html#503>.
- [12] Fayyad, U. Shapiro, G. P. and Smyth, P. "From Data Mining to Knowledge Discovery in Databases". AI MAGAZINE. American Association for Artificial Intelligence, 1996.
- [13] F.M. de Azevedo. Uma proposta de modelos formais de neurônios e redes neurais artificiais. Em: Congresso Brasileiro de Redes (CBRM'97), Florianópolis, Brasil, 1997.
- [14] BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. *Redes Neurais Artificiais: Teoria e Aplicações*. 2ª Ed, 2007.
- [15] HAYKIN, Simon/ *Redes Neurais - Princípios e prática*; 2ª edição; editora Bookman.