STEPHAN KOVACH

DETECÇÃO DE FRAUDES EM TRANSAÇÕES FINANCEIRAS VIA INTERNET EM TEMPO REAL

São Paulo
2011

STEPHAN KOVACH

DETECÇÃO DE FRAUDES EM TRANSAÇÕES FINANCEIRAS VIA INTERNET EM TEMPO REAL

Tese apresentada a Escola Politécnica da Universidade de São Paulo para obtenção do Título de Doutor em Engenharia Elétrica

São Paulo
2011

STEPHAN KOVACH

DETECÇÃO DE FRAUDES EM TRANSAÇÕES FINANCEIRAS VIA INTERNET EM TEMPO REAL

Tese apresentada a Escola Politécnica da Universidade de São Paulo para obtenção do Título de Doutor em Engenharia Elétrica

Área de Concentração:
Sistemas Digitais

Orientador:
Prof. Dr. Wilson Vicente Ruggiero

São Paulo
2011

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 15 de junho de 2011.

Assinatura do autor _____

Assinatura do orientador _____

FICHA CATALOGRÁFICA

Kovach, Stephan
Detecção de fraudes em transações financeiras via internet em tempo real / S. Kovach. -- ed. rev.-- São Paulo, 2011.
134 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1. Fraude 2. Internet 3. Teoria de Dempster-Shafer I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II. t.

À memória dos meus pais e da minha irmã

AGRADECIMENTOS

Ao meu orientador Prof. Dr. Wilson Vicente Ruggiero, pela amizade, pelos ensinamentos e pela confiança depositada em mim. As discussões e reflexões realizadas durante estes anos foram primordiais para a conclusão deste trabalho.

Um agradecimento especial a Profa. Dra. Graça Bressan pelo incentivo, pelas sugestões e contribuições feitas durante o desenvolvimento deste trabalho.

Aos Professores Dr. Paulo Sergio Licciardi Messeder Barreto e Dr. Leonardo Augusto Martucci pelas sugestões e contribuições feitas durante a qualificação deste trabalho.

Ao meu colega Eng. Daniel Soriano pelo incentivo e contribuições valiosas para a realização deste trabalho.

A todos os colegas do LARC pelo incentivo e sugestões.

Aos professores do PCS pelo incentivo.

Aos meus colegas da Scopus pelas discussões sobre segurança e fraudes que muito enriqueceram meus conhecimentos sobre o assunto

A minha esposa Sandra e as minhas filhas Bianca, Camila e Giulia pelo apoio, carinho, paciência, compreensão e, sobretudo por me incentivar a não desistir do trabalho.

A vida é como uma caixa de chocolates, você nunca sabe o que vai encontrar.

(Forrest Gump)

RESUMO

KOVACH, S. **Detecção de Fraudes em Transações Financeiras Via Internet em Tempo Real**. 2011. 134 p. Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo, São Paulo, 2011.

Um dos objetivos mais importantes de qualquer sistema de detecção de fraudes, independente de seu domínio de operação, é detectar o maior número de fraudes com menor número de alarmes falsos, também denominados de falsos positivos. A existência de falsos positivos é um fato inerente a qualquer sistema de detecção fraudes. O primeiro passo para alcançar esse objetivo é identificar os atributos que podem ser usados para diferenciar atividades legítimas das fraudulentas. O próximo passo consiste em identificar um método para cada atributo escolhido para efetuar essa distinção. A escolha adequada dos atributos e dos métodos correspondentes determina em grande parte o desempenho de um detector de fraudes tanto em termos da relação entre o número de fraudes detectadas e o número de falsos positivos, quanto em termos de tempo de processamento. O desafio desta escolha é maior ao se tratar de um detector de fraudes em tempo real, isto é, fazer a detecção antes que a fraude seja concretizada. O objetivo deste trabalho é apresentar a proposta de uma arquitetura de um sistema de detecção de fraudes em tempo real em transações bancárias via Internet, baseando-se em observações do comportamento local e global de usuários. O método estatístico baseado em análise diferencial é usado para obter a evidência local de uma fraude. Neste caso, a evidência de fraude é baseada na diferença entre os perfis de comportamento atual e histórico do usuário. A evidência local de fraude é fortalecida ou enfraquecida pelo comportamento global do usuário. Neste caso, a evidência de fraude é baseada no número de acessos efetuados em contas diferentes feitos pelo dispositivo utilizado pelo usuário, e por um valor probabilístico que varia com o tempo. A teoria matemática de evidências de Dempster-Shafer é utilizada para combinar estas evidências e obter um escore final. Este escore é então comparado com um limiar para disparar um alarme indicando a fraude. A principal inovação e contribuição deste trabalho estão na definição e exploração dos métodos de detecção baseados em atributos globais que são de natureza específica do domínio de transações financeiras. Os resultados da avaliação utilizando uma base de dados com registros de transações correspondentes a perfis reais de uso demonstraram que a integração de um detector baseado em atributos globais fez aumentar a capacidade do sistema de detectar fraudes em 20%.

Palavras-chaves: Fraude. Detecção de Fraude. Perfil de Comportamento. Análise Diferencial e Global. Teoria de Dempster-Shafer. Transação Financeira. Internet.

ABSTRACT

KOVACH, S. **Frauds Detections in Financial Transactions Via Internet in Real Time**. 2011. 134 p. Thesis (Doctoral) - Escola Politécnica da Universidade de São Paulo, São Paulo, 2011.

One of the most important goals of any fraud detection system, whichever is the domain where it characterizes the possibility for fraud, is to detect the largest number of frauds with fewer false alarms, also denominated false positives. The existence of false positives is a fact inherent to any fraud detection system. The first step in achieving this goal is to identify the attributes that can be used to differentiate between legitimate and fraudulent activities. The next step is to identify a method for each attribute chosen to make this distinction. The proper choice of the attributes and corresponding methods largely determines the performance of a fraud detector, not only in terms of the rate between the number of detected frauds and the number of false positives, but in terms of processing time. The challenge of this choice is higher when dealing with fraud detection in real time, that is, making the detection before the fraud is carried out. The aim of this work is to present the proposal of an architecture of a real time fraud detection system for Internet banking transactions, based on local and global observations of user's behavior. The statistical method based on differential analysis is used to obtain the local evidence of fraud. In this case, the evidence of fraud is based on the difference between the current and historical behavior of the user. The fraud's local evidence is strengthened or weakened by the user's global behavior. In this case, the evidence of fraud is based on the number of accesses performed on different accounts made by the device used by the user and by a probability value that varies over time. The Dempster-Shafer's mathematical theory of evidence is applied in order to combine these evidences for final suspicion score of fraud. This score is then compared with a threshold to trigger an alarm indicating the fraud. The main innovation and contribution of this work are the definition and exploration of detection methods based on global attributes which are domain specific of financial transactions. The evaluation results using a database with records of transactions corresponding to actual usage profiles showed that the integration of a detector based on global attributes improves the system capacity to detect frauds in 20%.

Keywords: Fraud. Fraud Detection. Behavior Profile. Differential and Global Analysis. Dempster-Shafer Theory. Financial Transaction. Internet.

LISTA DE FIGURAS

Figura 1 - Espaço de curvas ROC ilustrando sete classificadores discretos (Baseado em HAMEL, L, 2011 e FAWCETT, T., 2006)..... 30
Figura 2 - Gráfico ROC de um classificador parametrizado 32
Figura 3 - Análise absoluta e diferencial sob visão probabilística 37
Figura 4 - Subconjuntos possíveis no ambiente de aviões (GIARRATANO J., RILEY G., 1998) 41
Figura 5 - Relações entre as grandezas de confiança 49
Figura 6 - A arquitetura geral do sistema de detecção proposta 72
Figura 7 - Exemplo de alguns atributos contidos nas mensagens de uma transação bancária 74
Figura 8 - A construção do perfil de comportamento atual de um usuário 78
Figura 9 - Cálculo da distância probabilística entre PA e PH usando teoria de Dempster-Shafer..... 82
Figura 10 - Atualização de PH 85
Figura 11 - Curvas exponencialmente decrescentes..... 92
Figura 12 - A probabilidade de fraude pela função exponencial decrescente 93
Figura 13 - Componentes básicos da análise global 93
Figura 14 - Módulo de combinação de evidências..... 96
Figura 15 - Campos de uma transação..... 101
Figura 16 - Registro simplificado de uma transação..... 103
Figura 17 - Trecho de uma sequência de transações em uma conta corrente 109
Figura 18 - Curva ROC e AUC aplicando z-score sem análise global, em apenas contas fraudadas111
Figura 19 - Curva ROC e AUC aplicando z-score e análise global com Nmax=5, em apenas contas fraudadas..... 112
Figura 20 - Curva ROC e AUC aplicando z-score e análise global com Nmax=7, em apenas contas fraudadas..... 113
Figura 21 - Curva Curva ROC e AUC aplicando média ponderada sem análise global, em apenas contas fraudadas 114
Figura 22 - Curva ROC e AUC aplicando média ponderada e análise global com Nmax=5, em apenas contas fraudadas 115
Figura 23 - Curva ROC e AUC aplicando média ponderada e análise global com Nmax=7, em apenas contas fraudadas 115
Figura 24 - Curva ROC e AUC aplicando z-score sem análise global, em todas as contas 116
Figura 25 - Curva ROC e AUC aplicando z-score e análise global com Nmax=5, em todas as contas 116
Figura 26 - Curva ROC e AUC aplicando z-score e análise global com Nmax=7, em todas as contas 117
Figura 27 - Curva ROC e AUC aplicando média ponderada sem análise global, em todas as contas 118

Figura 28 - Curva ROC e AUC aplicando média ponderada e análise global com Nmax=5, em todas as contas 118
Figura 29 - Curva ROC e AUC aplicando média ponderada e análise global com Nmax=7, em todas as contas 119

LISTA DE TABELAS

Tabela 1 – Matriz de confusão ou tabela de contingência 27
Tabela 2 - Comparação entre massa da teoria de Dempster-Shafer e a teoria de probabilidade (GIARRATANO J., RILEY G., 1998) 43
Tabela 3 - Tabela de intersecções e produtos de m_1 e m_2 45
Tabela 4 - Tabela de intersecções e produtos entre m_1 , m_2 e m_3 49
Tabela 5 - Tabela comparativa dos resultados obtidos em cada um dos cenários 120

LISTA DE ABREVIATURAS E SIGLAS

e-mail	Correio eletrônico
Framework	Arcabouço de uma solução a ser completado para se tornar operacional
Hardware	Computador ou um dispositivo
IP	Internet Protocol
log	Registro de atividades geradas por um programa de computador
login	Identificação inicial de um usuário
MAC	Medim Access Control (Controle de Acesso ao Meio)
online	Conectado a um computador
ROC	Receiver Operating Characteristics
AUC	Area Under Curve
Software	Programas que executam em um computador

SUMÁRIO

1. INTRODUÇÃO 16
1.1. Motivação 18
1.2. Objetivo 19
1.3. Metodologia 20
1.3.1. Revisão bibliográfica 20
1.3.2. Delimitação do escopo 21
1.3.3. Elaboração da especificação da solução proposta 22
1.3.4. Validação e avaliação de desempenho da solução proposta 22
1.4. Contribuições Originais 22
1.5. Estrutura da Tese 22
2. CONCEITOS 24
2.1. Prevenção e Detecção de Fraudes 24
2.2. Perfil de Comportamento 25
2.3. Métricas de Desempenho 26
2.3.1. Curva ROC (*Receiver Operating Characteristics*) 28
2.3.2. AUC (*Area Under Curve*) 33
2.4. Métodos de Detecção de Fraudes 34
2.4.1. Métodos de detecção supervisionada e não supervisionada 35
2.4.2. Métodos estatísticos 35
2.4.3. Métodos baseados em regras 38
2.4.4. Métodos baseados em redes neurais artificiais 38
2.5. Teoria Matemática de Evidências de Dempster-Shafer 39
2.5.1. Função de massa e ignorância 42
2.5.2. Regra de combinação de Dempster 44
2.5.3. Função de Confiança 46
2.5.4. Intervalo de confiança 47
3. REVISÃO DA LITERATURA 52
3.1. Considerações Iniciais 52
3.2. Fraudes em Cartões de Crédito 53
3.3. Intrusão em Computadores 54
3.4. Fraudes em Comunicação Móvel 57
3.5. Considerações Finais 65
4. PROPOSTA DE ARQUITETURA PARA DETECÇÃO DE FRAUDES EM TRANSAÇÕES FINANCEIRA VIA INTERNET 67
4.1. Considerações Iniciais 67
4.2. Atributos de uma Transação Financeira 69
4.3. Características de Fraudes de Transações Bancárias *Online* 70
4.4. Descrição Geral do Sistema de Detecção 71

4.5. Identificação dos Dispositivos de Acesso 72
4.6. Atributos Locais de uma Transação 74
4.7. Perfil de Atividade Local 75
4.8. Perfil de Comportamento 77
4.9. Análise Diferencial 79
4.9.1. Distâncias individuais entre os elementos de PA e PH 83
4.9.2. Atualização do perfil histórico 84
4.9.3. Inicialização do perfil histórico 86
4.10. Atributos Globais e o Comportamento Global de Usuários 87
4.11. Análise Global 88
4.11.1. A lista de suspeitos e a função exponencial decrescente 89
4.12. Combinação de Evidências 94
4.13. Integração com Outros Modelos de Detecção 97
4.14. Considerações Finais 98
5. VALIDAÇÃO DA PROPOSTA 100
5.1. Considerações Iniciais 100
5.2. metodologia para a validação da proposta 100
5.3. dados para avaliação de Desempenho 101
5.4. Determinação de atributos 104
5.5. Avaliação de Desempenho do Detector de Fraudes 105
5.5.1. Análise Diferencial 105
5.5.2. Análise Global 108
5.5.3. Combinação de Dempster-Shafer 109
5.5.4. Avaliação de desempenho e os seus resultados 110
5.5.4.1. Cenário 1- Considerando apenas contas correntes que foram fraudadas .. 111
5.5.4.2. Cenário 2- Considerando todas as contas correntes da base de dados 115
5.5.4.3. Análise dos resultados 120
5.6. Considerações a respeito de Detecção em Tempo Real 121
5.7. Considerações Finais 122
6. CONCLUSÕES E TRABALHOS FUTUROS 125
6.1. Considerações Finais 125
6.2. Trabalhos futuros 127
REFERÊNCIAS 129

1. INTRODUÇÃO

O ambiente de transação financeira via Internet se caracteriza por alguns atributos de natureza variável associados aos seus usuários e as instituições bancárias que fornecem os serviços aos seus clientes via Internet.

Pelo lado dos clientes, a confiança depositada por eles quanto a segurança fornecida pelos bancos na utilização dos seus serviços é um dos atributos. Pelo lado dos bancos, a confiança que estes depositam quanto a legitimidade dos usuários que estão utilizando os seus serviços é um exemplo de outro atributo.

Os mecanismos de segurança utilizados pelas instituições como autenticação de usuários e infra-estrutura de chaves públicas, constituem alguns dos mecanismos do modelo tradicional de segurança com o qual os clientes e as instituições bancárias depositam suas confianças e impõem as regras de suas políticas de segurança.

Por meio destes mecanismos, as instituições procuram impedir que fraudadores acessem o sistema fazendo-se passar por usuários legítimos.

Fraude é definida no escopo desta tese como sendo qualquer acesso não autorizado ou uma transação não autorizada efetuada em uma conta corrente através da Internet.

Segundo Bolton e Hand (BOLTON, R. J.; HAND, D. J, 2002), prevenção de fraudes descreve as medidas de segurança para evitar indivíduos não autorizados a iniciarem transações em contas nas quais eles não foram autorizados

Apesar de vários mecanismos para prevenção de fraudes disponíveis para aplicações bancárias online, eles podem falhar, pois eles não protegem contra “as falhas de segurança dos humanos”.

A engenharia social é ainda, uma das armadilhas mais utilizadas para a obtenção de informações sigilosas e importantes, especialmente dos usuários

domésticos. Através disso, junto com a natureza aberta da Internet, os fraudadores têm conseguido transpor a barreira imposta pelo modelo tradicional de segurança.

Entre as abordagens de engenharia social usadas pelos fraudadores, *phishing* é uma das formas mais comuns para roubar dos usuários, os detalhes para autenticação das contas.

Ele vem normalmente na forma de *e-mail* tentando convencer usuários a abrirem anexos ou então direcioná-los para algum site fraudulento, e na maioria das vezes eles são tão bem elaborados que muitos usuários são induzidos a informar os detalhes de suas contas.

Detecção de fraudes consiste em identificar atividades não autorizadas uma vez que a prevenção de fraudes falhou. Na prática, detecção de fraudes deve ser aplicada constantemente, pois o sistema não tem ciência de quando a prevenção falha (BOLTON, R. J.; HAND, D. J, 2002).

Os detectores de fraude são ferramentas cada vez mais utilizadas para a identificação de transações ilegítimas que conseguiram passar pelas barreiras impostas pelos mecanismos de segurança.

Vários métodos são propostos para detecção de fraudes, sendo uns mais adequados que os outros dependendo do domínio de cada aplicação, como de cartões de crédito, de intrusão de computadores, de telefonia móvel, assim como de transações financeiras *online*.

Como exemplos de métodos de detecção, existem os métodos supervisionados baseados em regras, os métodos não supervisionados baseados em modelos estatísticos que detectam evidências de fraude através de desvios do comportamento normal de usuário. Existem também os métodos baseados em redes neurais, supervisionados ou não, e métodos híbridos, entre outros.

Uma questão associada a detectores de fraude é a relação entre a detecção correta de fraudes e a detecção incorreta de fraudes (falsos alarmes). O aumento de

um sempre vem acompanhado pelo aumento do outro. Portanto, o objetivo de todos os detectores de fraude é maximizar a predição correta de fraudes mantendo as predições incorretas em um nível aceitável. (KOU, Y., et al., 2004)

Esta tese apresenta uma abordagem inovadora para detectar fraudes em transações financeiras via Internet em tempo real baseando-se em uma abordagem híbrida para evidenciar fraudes: a primeira baseada em um método estatístico similar aos usados em outros domínios, e a segunda baseada na observação do comportamento global de usuários. As evidências de fraude determinadas pelas duas abordagens são combinadas usando a teoria matemática de evidências de Dempster-Shafer (SHAFER, G., 1976).

1.1. MOTIVAÇÃO

Um dos objetivos mais importantes de um sistema de detecção de fraudes é identificar transações fraudulentas com menor número de alarmes falsos.

Uma transação legítima que é sinalizada como uma fraude caracteriza um alarme falso (ou falso positivo).

No caso de transações financeiras, por exemplo, o custo de não detectar uma fraude pode ser bem alta. Por outro lado, disparar alarmes mediante qualquer suspeita pode gerar uma taxa elevada de falsos alarmes, o que pode gerar insatisfação aos clientes legítimos.

Tom Fawcett e Foster Provost, no seu artigo sobre detecção de fraudes de clonagens de celulares (FAWCETT, T.; PROVOST, F, 1997), considera dois problemas que tornam a tarefa de detecção de fraudes muito difícil. O primeiro problema ocorre em situações onde uma transação não normal para um usuário pode ser típica para outro.

O segundo problema ocorre quando um usuário legítimo pode ocasionalmente fazer uma transação que pode parecer suspeita.

Os fraudadores por outro lado, podem também adotar novas técnicas de fraude resultando em novos padrões de comportamento, dificultando mais a tarefa de detecção além dos dois problemas apresentados.

Surge, portanto, a necessidade de um sistema de detecção capaz de identificar estas evidências e combiná-las gerando um escore total de suspeita com menor taxa de alarmes falsos.

Além disso, é desejável que o sistema possa aprender os padrões de comportamento mais recentes dos usuários, e se adapte as suas mudanças.

Em se tratando de transações financeiras via Internet, existe ainda outro grande desafio que consiste em detectar fraudes em tempo real, isto é, antes de terminar a execução das transações, pois o custo associado com uma fraude numa transação financeira pode ser elevado. (EDGE, M. E.; SAMPAIO, P. R., 2009)

Vários trabalhos foram publicados sobre detecção de fraudes nas áreas de telefonia celular, cartões de crédito, e sobre detecção de intrusões em computadores e redes, mas poucos sobre detecção de fraudes em transações financeiras.

1.2. OBJETIVO

O objetivo desta tese é propor uma arquitetura para detectar fraudes em transações financeiras via Internet em tempo real, utilizando uma abordagem híbrida baseada em dois métodos: o primeiro baseado em um método estatístico similar aos adotados em outros domínios, como telefonia celular, cartões de crédito e intrusão em sistemas de computação, e o segundo baseado em um método inovador utilizando atributos globais específicos do domínio de aplicações bancárias via Internet.

Esse objetivo é alcançado através de quatro metas. A primeira meta consiste em determinar os atributos mais adequados para evidenciar fraudes dentro cada um dos métodos. Estes atributos são utilizados para definir o perfil de comportamento de usuários que acessam suas contas bancárias através da Internet.

A segunda meta consiste em determinar os modelos estatísticos mais adequados para cada um dos atributos selecionados na meta anterior para caracterizar evidências de fraude.

A terceira meta consiste em determinar um método para combinar estas evidências e produzir um escore de suspeita total.

A quarta meta é especificar uma arquitetura para detectar fraudes em transações financeiras *online* em tempo real, contemplando as três primeiras metas.

1.3. METODOLOGIA

Esta seção visa apresentar a metodologia adotada na elaboração do trabalho para atingir as metas do objetivo proposto nesta tese.

As subseções a seguir descrevem as etapas do trabalho:

1.3.1. Revisão bibliográfica

Abrangeu o levantamento de informações referentes aos assuntos e tópicos em geral nos quais este trabalho de pesquisa está inserido. Foi feito uma análise das fontes consultadas na literatura para proporcionar um embasamento teórico ao trabalho. Também fez parte desta etapa, realizar uma revisão de pesquisas similares com o objetivo de posicionar o trabalho proposto em relação aos demais.

Foi possível identificar algumas tarefas importantes neste levantamento:

- Estudo de requisitos e de ambientes onde se aplicam o uso de detectores tanto de fraudes como de intrusão de computadores;
- Estudo dos métodos utilizados para detecção de fraudes e de intrusão;
- Estudo de técnicas usadas em mineração de dados para aprendizado de máquinas;
- Estudo de métodos estatísticos para caracterização de perfis de comportamento de usuários e para determinação de seus desvios;
- Estudo de técnicas para o tratamento de incertezas, em particular a teoria matemática de evidências de Dempster-Shafer;
- Busca e análise de trabalhos correlatos.
- Publicação do artigo técnico em evento internacional versando sobre a contribuição original da tese.

1.3.2. Delimitação do escopo

Com base nos estudos realizados, o escopo deste trabalho consistiu na:

1. Definição de atributos mais adequados para detectar desvios de comportamento que possam caracterizar fraude;
2. Definição de métricas e modelos estatísticos para cada um dos atributos definidos para a determinação de desvios;
3. Especificação de um mecanismo de detecção baseado em desvios individuais de atributos de uma transação, e utilizar a teoria de evidências de Dempster-Shafer para a determinação de escores de suspeita de fraude;
4. Especificação de uma arquitetura para combinar todos os itens anteriores;
5. Apresentação dos resultados quanto a eficácia e desempenho da arquitetura por meio de simulação.

1.3.3. Elaboração da especificação da solução proposta

Consistiu na especificação da arquitetura proposta nesta tese, com a definição de seus respectivos componentes.

1.3.4. Validação e avaliação de desempenho da solução proposta

Consistiu na realização de ensaios de simulação com a finalidade de verificar e validar as soluções propostas;

1.4. CONTRIBUIÇÕES ORIGINAIS

Diante do objetivo e de suas metas, as principais contribuições desta tese são:

- Definição de atributos globais para determinação de anomalias para reforçar evidências de fraude determinada por meio de atributos locais
- Especificação de métodos probabilísticos para determinar o grau de evidência de fraude a partir de atributos globais;
- Especificação de uma arquitetura para detectar fraudes em transações financeiras *online* em tempo real utilizando atributos locais e globais; e
- A especificação de uma forma de combinar evidências de fraude utilizando cálculos baseados na teoria matemática de evidências de Dempster-Shafer;

1.5. ESTRUTURA DA TESE

Esta tese está organizada como segue:

Capítulo 1: Introdução. Este capítulo tem por finalidade situar o leitor sobre a tese apresentando a motivação do trabalho, seu objetivo, escopo e organização do documento.

Capítulo 2: Conceitos. Este capítulo apresenta o embasamento teórico e os principais termos utilizados no trabalho.

Capítulo 3: Revisão da Literatura: Este capítulo contém um resumo de algumas publicações mais relevantes para esta tese.

Capítulo 4: Arquitetura para detecção de fraudes em um ambiente de transação financeira via Internet. Este capítulo descreve a arquitetura proposta nesta tese.

Capítulo 5: Validação da Proposta. Este capítulo avalia a eficácia e o desempenho do sistema de detecção de fraude proposto no capítulo 4 através de uma série de simulações com dados obtidos dos registros de transações reais de uma instituição financeira.

Capítulo 6: Conclusão. Finaliza a discussão sobre a proposta discutindo as contribuições obtidas e sugestões para trabalhos futuros.

2. CONCEITOS

Neste capítulo são apresentados conceitos e terminologias presentes nas literaturas com objetivo de dar melhor fundamentação na pesquisa realizada e para que sirvam de subsídios para os demais capítulos da tese.

Alguns conceitos são descritos com mais detalhes de acordo com a sua relevância no trabalho

2.1. PREVENÇÃO E DETECÇÃO DE FRAUDES

Prevenção de fraude consiste em tomar medidas para evitar que ocorram fraudes antes do termino de uma transação. A prevenção é feita normalmente durante a fase de autenticação de um usuário tradicionalmente utilizando senhas, frases secretas, dispositivos de geração de códigos secretos *(tokens)*, etc.

A detecção de fraudes entra em ação quando a prevenção não consegue evitar a fraude. Ela consiste em identificar uma fraude o mais rápido possível assim que ela ocorrer.

Tanto prevenção quanto detecção de fraudes são disciplinas que estão em constante evolução, pois sempre que um novo método é implantado, os criminosos adaptam suas estratégias de ataque ou tentam outras. E em consequência disso, novos métodos para detecção de fraudes são elaborados e o ciclo se repete. (BOLTON, R. J.; HAND, D. J.,2002).

Existe uma grande dificuldade em desenvolver novos métodos de detecção em virtude da pouca troca de informações nesta área. Isto vem do fato de que a divulgação detalhada das técnicas de detecção fornece ao mesmo tempo, informações que os fraudadores precisam para burlar a detecção. (BOLTON, R. J.; HAND, D. J.,2002).

2.2. PERFIL DE COMPORTAMENTO

Existem várias definições na literatura para caracterizar o perfil de comportamento de usuários durante a sua interação com o sistema

Denning (DENNING, D. E., 1987) utiliza o termo *perfil de atividade* para caracterizar o comportamento de um *sujeito (ou conjunto de sujeitos)* com relação a um *objeto (ou conjunto de objetos),* servindo de assinatura ou descrição de atividade normal para os seus respectivos *sujeito(s) e objetos(s).*

Sujeitos são os iniciadores de ações. Normalmente são os próprios usuários do sistema, mas podem ser quaisquer entidades que atuam em nome de usuários.

Objetos são os receptores de ações, e incluem entidades como, mensagens, programas, arquivos, registros, estruturas de dados, *e-mails*, etc.

No caso de uma transação financeira, cada perfil de atividade corresponde a um dos atributos da transação selecionados para distinguir comportamento legítimo de um comportamento fraudulento.

Um comportamento observado pode ser caracterizado em termos de uma métrica e de um modelo estatístico.

Métrica representa uma medida quantitativa de uma variável aleatória acumulada durante um período. Este período pode ser um intervalo de tempo fixo ou entre dois eventos.

Dado um conjunto de observações sobre uma variável aleatória, o objetivo de um modelo estatístico é determinar se uma nova observação é anormal com relação as observações anteriores.

Outras definições são encontradas para perfil de comportamento. Em (LUNT T. F., TAMARU A., 1991), ao contrário de (DENNING, D. E., 1987), o perfil de

comportamento é definido como sendo um vetor composto de variáveis aleatórias. Em várias publicações (CORTES, C.; PREGIBON, D., 2001; FERREIRA, P. et al., 2006) este vetor recebe o nome de assinatura.

2.3. MÉTRICAS DE DESEMPENHO

As métricas, normalmente utilizadas para avaliar o desempenho de detector de fraudes são as seguintes (KOU, Y., et al., 2004):

Taxa de verdadeiro positivo (Tvp) ou sensibilidade, é a fração de transações fraudulentas que foram corretamente classificadas como fraudulentas.

Taxa de falso positivo (Tfp) ou taxa de falsos alarmes é a fração de transações legítimas que foram incorretamente classificados como fraudulentas.

Taxa de verdadeiro negativo (Tvn) é a fração de transações legítimas que foram corretamente classificados como legítimas.

Taxa de falso negativo (Tfn) é a fração de transações fraudulentas que foram incorretamente classificados como legítimas.

Exatidão (Ex) é a fração do número total de transações, legítimas e fraudulentas, corretamente classificadas.

Precisão (Pr) é a fração das transações classificadas como fraudulentas que estavam corretas

Estas métricas podem ser derivadas a partir de uma tabela conhecida como *matriz de confusão ou tabela de contingência*, cuja explicação segue abaixo. (FAWCETT, T., 2006)

Considerando que um detector de fraudes é um classificador de duas classes, P (Positiva ou Fraude) e N (Negativa ou Legítima), existem quatro possíveis

resultados ao classificar uma transação, que pode ser apresentada na matriz de confusão, como mostra a tabela 1.

Tabela 1 – Matriz de confusão ou tabela de contingência

		Classe correta	
		F	**L**
Resultado da detecção	**F**	Verdadeiro Positivo (VP)	Falso Positivo (FP) (*Legítimo como fraude*)
	L	Falso Negativo (FN) (*Fraude como legítimo*)	Verdadeiro Negativo (VN)

onde,

F = Fraude

L = Legítima

VP = O número de positivos (fraudes) classificados corretamente;

FP = O número de negativos (legítimos) classificados incorretamente;

VN = O número de negativos (legítimos) classificados corretamente;

FN = O número de positivos (fraudes) classificados incorretamente;

As classificações dispostas na diagonal principal (verdadeiras positivas e verdadeiras negativas) da matriz de confusão são as classificações corretas. Os demais campos significam classificações erradas.

Com base nestes campos, as métricas de desempenho são derivadas como segue:

Taxa de verdadeiro positivo ou sensibilidade: $Tvp = VP/P$

Taxa de falso positivo: $Tfp = FP/\ N$

Taxa de verdadeiro negativo: $Tvn = VN/\ N$

Taxa de falso negativo: $Tfn = FN/\ P$

Exatidão: $Ex = VP + VN/\ (P + N)$

Precisão: $Pr = VP/\ (VP + FP)$

onde,

$P = VP + FN$ é igual ao número total de positivos (fraudes);

$N = VN + FP$ é igual ao número total de negativos (legítimos).

Exatidão, que representa a proporção de transações identificadas corretamente em relação ao total de transações, é uma métrica muito comum utilizada para avaliar o desempenho de um classificador em geral, como detector de fraudes.

2.3.1. Curva ROC (*Receiver Operating Characteristics*)

O uso de exatidão (Ex) para avaliar o desempenho de um classificador (detector de fraudes) a partir da matriz de confusão é uma métrica muitas vezes considerada inadequada, pois ela é altamente dependente da distribuição de classes e os custos decorrentes de erros de classificação não são uniformes.

O uso de exatidão como métrica de avaliação supõe que a distribuição de classes entre os dados é constante e relativamente balanceada. Isso raramente ocorre no mundo real, em especial no domínio de detecção de fraudes, cuja distribuição de classes é extremamente assimétrica.

Por exemplo, num domínio onde as classes aparecem na proporção de 999:1, uma regra que classifica sempre a classe de maior proporção terá uma exatidão de 99,9%, o que não é razoável na prática.

Outra suposição ao se utilizar exatidão como métrica para avaliar um classificador é que os custos decorrentes de erros de classificação são iguais, isto é, o custo de erro nos casos de falso positivo é igual ao custo de erro nos casos de falso negativo. Raramente este é o caso no mundo real.

Provost e Fawcett (PROVOST, F.; FAWCETT, T., 1997) propõem o uso de curvas ROC (*Receiver Operating Characteristics*) para avaliar o desempenho de um classificador.

Os gráficos ROC são gráficos bidimensionais onde o eixo vertical representa a taxa de verdadeiro positivo (Tvp) e o eixo horizontal representa a taxa de falso positivo (Tfp).

A motivação pelo uso de curvas ROC vem do fato de desacoplar o desempenho do classificador tanto da distribuição de classes quanto dos custos causados por erros. (PROVOST, F.; FAWCETT, T., 1997).

Uma curva ROC descreve um compromisso relativo entre benefício (Verdadeiro Positivo) e custo (Falso Positivo).

Um classificador discreto (binário) produz um par (Tfp,Tvp) correspondendo a um ponto espaço ROC.

A figura 1 apresenta 7 classificadores discretos representados no espaço das curvas ROC.

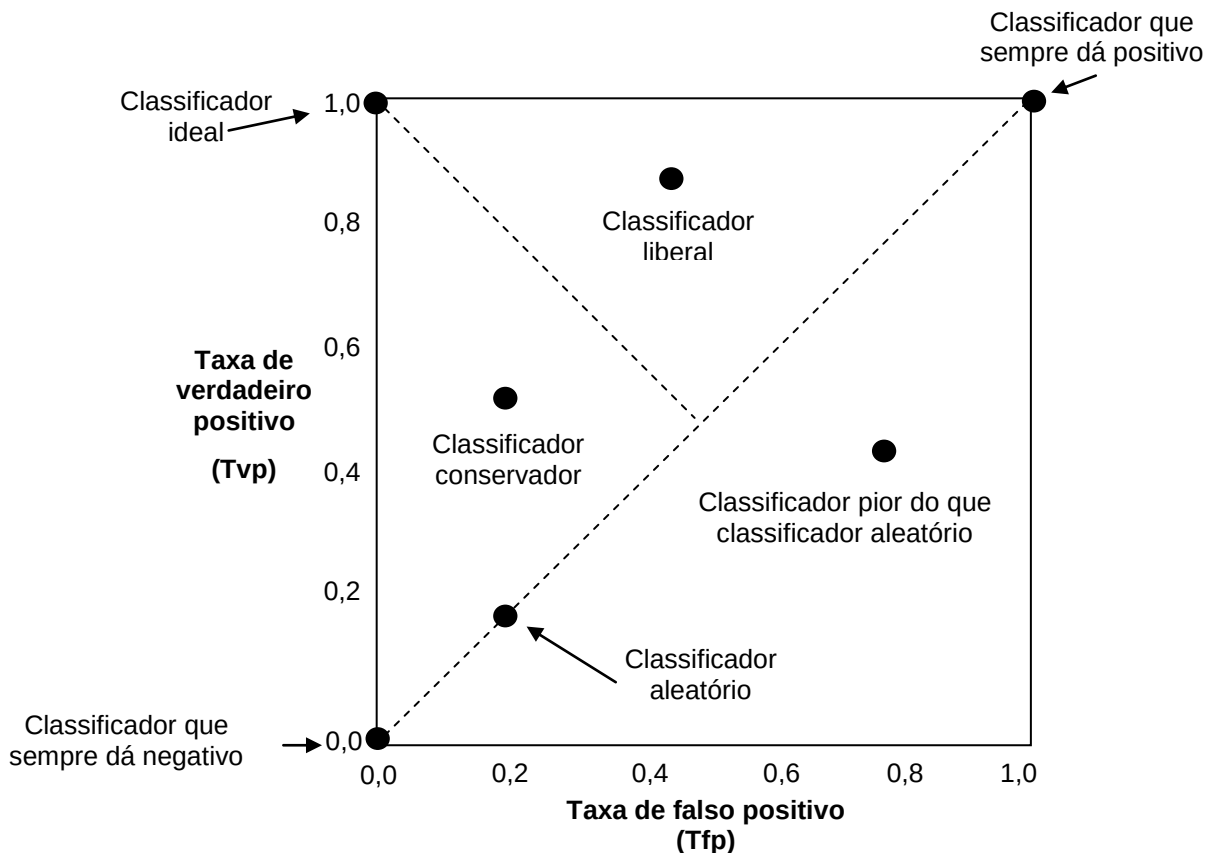

Figura 1 - Espaço de curvas ROC ilustrando sete classificadores discretos (Baseada em HAMEL, L, 2011 e FAWCETT, T., 2006)

Algumas regiões de interesse no espaço ROC são comentadas a seguir:

A linha diagonal (0,0; 1,1) denota o desempenho de um classificador aleatório, isto é, um classificador mapeado nesta linha produz respostas tanto falsos positivos quanto verdadeiro positivos de forma aleatória.

O ponto (0,0) é um caso extremo que corresponde a um classificador conservador que classifica todas as instâncias como negativas. Isto é, não produz nenhum falso positivo, como também nenhum verdadeiro positivo.

Os classificadores na região liberal produzem muitos verdadeiros positivos, mas também muitos falsos positivos.

O ponto (1,1) é outro caso extremo que classifica todas as instâncias como positivas, sejam verdadeiras ou falsas.

Classificadores que caem na região a direita da linha diagonal (0,0; 1,1) têm um desempenho pior do que um classificador aleatório, pois produzem mais falsos positivos do que verdadeiros positivos.

O ponto (0,1) denota um classificador perfeito, fornecendo 100% de verdadeiros positivos e 0% de falsos positivos.

Classificadores mapeados no espaço ROC podem ser avaliados conforme as suas distâncias ao ponto de desempenho perfeito (0,1), isto é, quanto mais perto do ponto (0,1) melhor é o desempenho do classificador.

Um elemento fundamental para a construção das curvas ROC é a noção de probabilidade ou escore de uma instância, isto é, um valor numérico que indica a probabilidade com que uma instância é membro de uma classe (Fraude (P) ou Legítimo (N)).

Curvas ROC podem ser usadas em qualquer modelo de classificação que anexa um valor probabilístico, *ranking*, ou um valor de confiança para cada predição.

Muitos modelos de classificação produzem tais valores probabilísticos como parte de seus algoritmos. Dentre eles incluem os classificadores baseados em *Naïve Bayes* e Redes Neurais Artificiais.

Entretanto, existem técnicas que calculam valores probabilísticos para modelos de classificação que não produzem tais valores, como por exemplo, árvores de decisão. (BREIMAN, L ET AL., 1984)

Os classificadores que geram valores probabilísticos podem ser usados com um limiar (*threshold*) para produzir um classificador binário. Isto é, para cada instância se a saída do classificador estiver acima do limiar, o classificador produz 1 (Positivo), caso contrário 0 (Negativo).

Cada valor de limiar produz uma matriz de confusão diferente e, portanto, um ponto diferente no espaço ROC, representando o desempenho do classificador em termos de taxas de verdadeiro positivo e falso positivo para este valor de limiar.

Cada valor de limiar produz um ponto no espaço ROC. Portanto, variando-se o valor de limiar constrói-se a curva ROC.

As curvas ROC representam a porcentagem de fraudes reais detectados contra a porcentagem de falsos alarmes para diversos valores de limiares.

A figura 2 ilustra uma curva no espaço ROC de um classificador parametrizado, isto é, dotado de limiar variável.

A partir desta figura, pode-se determinar o melhor valor de limiar para este modelo de classificador. Ele ocorre no ponto de Tvp máximo e Tfp mínimo (também denominado, ponto ótimo de operação), isto é, aproximadamente nos pontos onde Tvp = 0,80 e Tfp= 0,20.

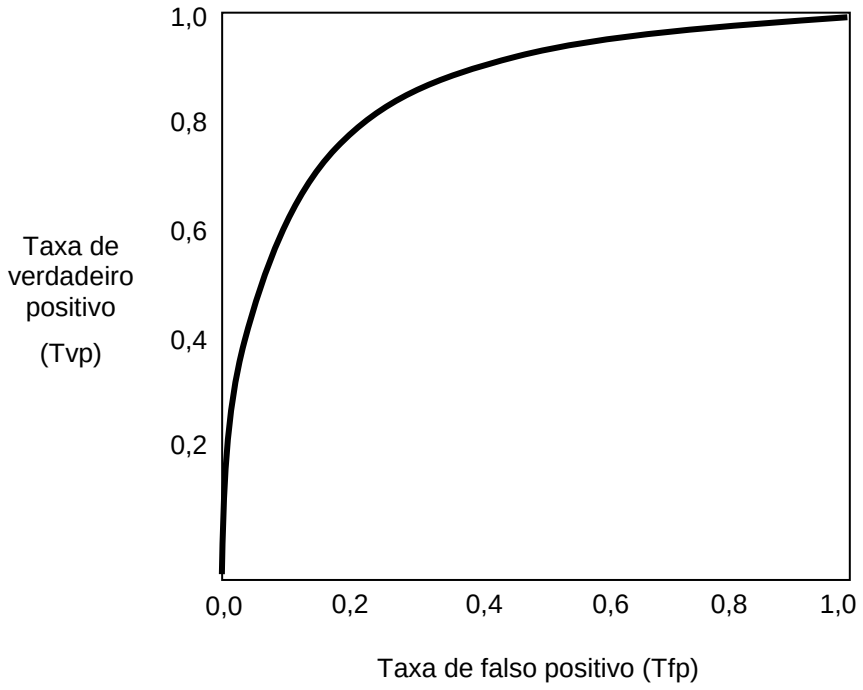

Figura 2 - Gráfico ROC de um classificador parametrizado

Como observação final a respeito das curvas ROC, deve-se dizer que na prática elas são construídas para determinar o melhor valor do limiar que deve ser utilizado para se obter a melhor relação entre taxa de verdadeiro positivo (Tvp) e a taxa de falso positivo (Tfp) a partir da base real de dados.

Entretanto, o ponto ótimo de operação determinado pela curva, nem sempre significa o melhor ponto num caso real. Normalmente, os custos relacionados com a taxa de falsos positivos e a taxa de falsos negativos são levados em conta para a escolha do melhor ponto.

Em todos os sistemas de classificação parametrizada, o aumento de taxa de verdadeiro positivo é sempre acompanhado do aumento da taxa de falso positivo. Diante disso, existe situações que é preferível reduzir o número de detecções de fraudes legítimas para se ter um número menor de falsos positivos devido ao seu custo. (MOREAU, Y., Vandewalle, J., 1997)

2.3.2. AUC (*Area Under Curve*)

A curva ROC descreve o desempenho de um classificador utilizando duas dimensões. Ela permite verificar os comportamentos das taxas de verdadeiros positivos (Tvp) e falsos positivos (Tfp) com relação a vários níveis de limiar.

A curva ROC ajuda na determinação do melhor ponto de operação de um detector. Entretanto, ela não é muito prática quando se quer comparar desempenhos de detectores pelo fato de ser baseado em duas dimensões. Para estes casos, a métrica muito utilizada é um índice, denominado AUC (*Area Under Curve*) que é determinado pela área sob a curva ROC. (FAWCETT, T., 2006)

AUC equivale a probabilidade de uma transação fraudulenta sorteada aleatoriamente receber um escore da evidência de fraude maior do que uma transação legítima sorteada aleatoriamente. (FAWCETT, T., 2006)

Em outras palavras, AUC mede a capacidade de um detector classificar corretamente transações fraudulentas e transações legítimas, isto é, mede a capacidade de discriminação de um detector.

AUC igual a 1,0 representa um classificador ideal e AUC igual a 0,5 representa um classificador “aleatório”, isto é, sem nenhuma utilidade.

Um detector com AUC na faixa de 0,8 a 0,9 pode ser considerado um bom detector, enquanto que um detector com AUC na faixa de 0,5 e 0,6, um detector fraco.

2.4. MÉTODOS DE DETECÇÃO DE FRAUDES

Nesta seção são apresentados alguns dos métodos mais utilizados para detecção de fraudes.

Alguns detectores de fraudes utilizam métodos para classificar o comportamento de uma transação como fraude ou legítima baseando-se nos padrões de perfis já conhecidos, isto é, o detector mantém uma base de dados com os perfis de fraudes conhecidos. Se o perfil de comportamento de uma transação coincidir com um destes perfis, um alarme é gerado.

Outros detectores de fraude utilizam métodos estatísticos para caracterizar uma fraude baseando-se em desvios significativos do comportamento usual.

Em qualquer um dos casos, eles procuram classificar uma transação ou um evento como sendo legítimo ou fraudulento, baseando-se no seu comportamento. Se o comportamento observado não estiver dentro das expectativas esperadas é sinalizado pelo detector.

Normalmente, os métodos de detecção geram um escore que é comparado com um limiar determinado manualmente por um especialista em fraudes ou por um

processo adaptativo utilizando técnicas de inteligência artificial (FAWCETT, T.; PROVOST, F, 1997).

2.4.1. Métodos de detecção supervisionada e não supervisionada

Os métodos de detecção de fraudes podem ser classificados como supervisionados e não supervisionados. (HILAS, C. S. e SAHALOS, J. N., 2005)

Métodos supervisionados são aqueles em que amostras de comportamentos normais e fraudulentos são usadas para construir modelos que permitem o sistema classificar as novas observações em uma destas duas classes.

Uma característica do método supervisionado é que ele é capaz de identificar apenas atividades fraudulentas conhecidas.

Os métodos não supervisionados apenas procuram observações que são diferentes do comportamento usual.

2.4.2. Métodos estatísticos

Métodos estatísticos utilizam métricas e modelos estatísticos para determinar as variações de comportamento dos usuários.

Existem duas abordagens baseadas em métodos estatísticos: (BURGE et al., 1997; HOLLMÉN, J., 2000)

- Análise absoluta; e
- Análise diferencial

Na abordagem baseada em análise absoluta, a detecção é feita por meio de algum critério de comparação de um ou mais campos de uma transação com valores fixos preestabelecidos, denominados limiares.

A análise absoluta é útil para detectar atividades fraudulentas extremas, como número elevado de erros de senha.

Entretanto, alguns padrões comportamentais podem ser indicativos de fraude para alguns tipos de usuários enquanto que para outros eles podem ser considerados aceitáveis. É o caso de número de transações de pagamento em um determinado período de tempo através da Internet.

Na abordagem baseada em análise diferencial, o padrão comportamental dos acessos as contas bancárias são monitorados, comparando suas atividades mais recentes com o histórico de sua utilização. Alarmes são gerados quando o padrão de utilização muda de forma significativamente em um curto período de tempo.

Um exemplo típico de sistema de detecção que usa uma abordagem baseada em análise diferencial é o detector de intrusão em computadores descrito em (LUNT T. F., TAMARU A., 1991). Neste exemplo, um detector observa as atividades dos componentes de um sistema e gera um perfil que representa o comportamento do usuário sobre os mesmos.

Quando uma nova atividade for observada, o sistema gera um escore que determina o grau da anormalidade do comportamento desta observação. Este escore é gerado como resultado da comparação do perfil da atividade observada com o perfil de comportamento anterior.

Caso não haja anormalidade, o perfil da atividade observada é normalmente fundido com o perfil anterior para se adaptar as variações do comportamento de um usuário com o tempo.

A vantagem de utilizar métodos estatísticos é que eles são baseados em teorias bem conhecidas.

Como algumas das desvantagens podemos citar as seguintes:

- Medidas estatísticas não levam em conta a ordem de ocorrência de eventos;
- Os detectores podem ser treinados gradualmente até que eles passam a considerar um comportamento anormal como normal;
- A determinação do limiar para que um comportamento seja considerado anormal é difícil de ser estabelecida.

As duas abordagens são ilustradas na figura 3, usando uma visão probabilística.

Na análise absoluta, representada pela figura (a), os modelos de comportamento legítimo (L) e comportamento fraudulento (F) são apresentados por duas curvas normais.

Na análise diferencial, representada pela figura (b), o modelo assumindo comportamento normal (L) é representado por uma curva normal.

As linhas verticais (*t e -t*) indicam limiares arbitrários de decisão e as áreas sombreadas denotam regiões classificadas como fraudulentas.

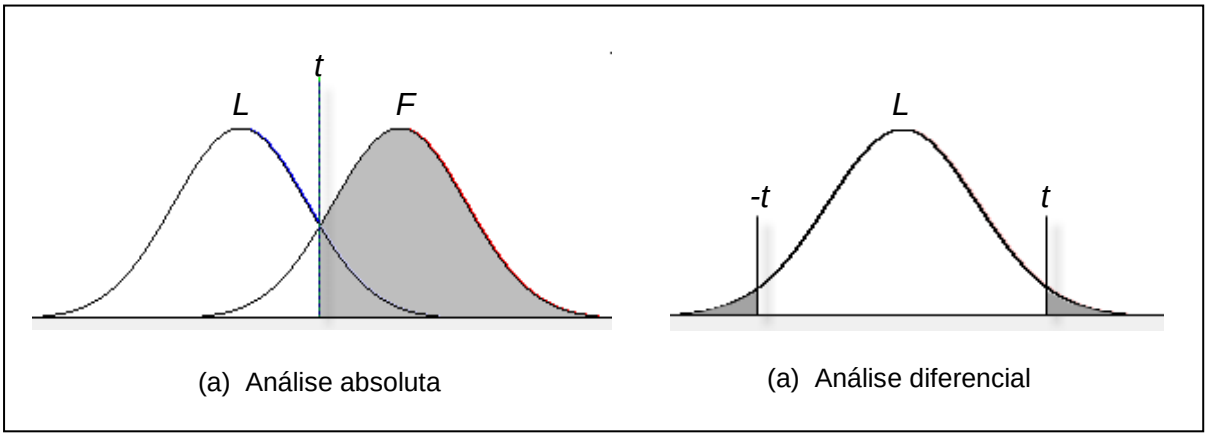

Figura 3 - Análise absoluta e diferencial sob visão probabilística (Baseada em HOLLMEN, J., 2000)

2.4.3. Métodos baseados em regras

Métodos baseados em regras são algoritmos de aprendizado supervisionado que produzem classificadores usando regras da forma *if {condições) then {ação}*. Quando as condições de uma regra forem satisfeitas, normalmente, um alarme é gerado. (BOLTON, R. J.; HAND, D. J., 2002).

Como exemplos de alguns algoritmos podem ser citados os seguintes: RIPPER (COHEN, W., 1995), C4.5 (QUINLAN, J. R.,1993) e BAYES (CLARK, P.; NIBLETT, T., 1989).

A grande desvantagem de usar este método é que o sistema só é capaz de detectar fraudes identificadas pelas regras.

Portanto, um requisito importante para sistemas que usam métodos baseados em regras é que eles precisam ser atualizados com novas regras a medida que novas fraudes são descobertas.

Uma das consequências desta abordagem pode ser a necessidade de uma grande quantidade de memória de armazenamento o que pode acarretar atraso no processamento.

2.4.4. Métodos baseados em redes neurais artificiais

Uma rede neural artificial é composta de vários elementos de processamento (neurônios) interconectados para resolverem um problema.

Cada neurônio pode ser visto como um elemento de soma ponderada seguida de uma função sigmoidal *f,* como expressa abaixo:

$$f = \frac{1}{1+e^{-(\sum_i w_i A_i)}} \qquad (1)$$

onde, w_i representa o peso aplicado ao valor da entrada A_i do somador.

A saída de cada neurônio alimenta os próximos neurônios.

O processo de aprendizado de uma rede neural é um processo de otimização para determinar o melhor conjunto de valores dos pesos para resolver um problema.

Alguns detectores de fraude utilizam redes neurais artificiais como técnica de reconhecimento de padrões.

Durante a fase de treinamento, os parâmetros da rede neural são otimizados para associar a saída com uma das classes (legítima ou fraude) de acordo com cada padrão de entrada aplicada.

Durante a fase de utilização, ao identificar um padrão de entrada, ele gera uma saída com a classe correspondente.

Se a rede não tiver nenhuma classe associada com uma determinada entrada, a rede neural fornece uma saída que corresponde ao padrão que mais se aproxima do padrão fornecido na entrada. (MORANDI, M.; ZULKERNINE, M., 2004; BEBAR, H.ET AL., 2002)

Uma desvantagem nos métodos baseados em redes neurais é que a topologia da rede e os pesos assinalados para cada elemento da rede são determinados apenas depois de considerável tentativa e erro.

A vantagem desta abordagem é que ela não depende da natureza dos dados.

2.5. TEORIA MATEMÁTICA DE EVIDÊNCIAS DE DEMPSTER-SHAFER

A teoria de Dempster-Shafer (DS) (1976, apud GIARRATANO J., RILEY G., 1998) oferece uma alternativa a teoria tradicional de probabilidades para a representação matemática de incerteza ou ignorância.

Na teoria de probabilidades, a probabilidade deve ser distribuída igualmente mesmo havendo incerteza. Por exemplo, se não existir nenhum conhecimento a priori, deve-se assumir probabilidade P igual a 1/N para cada uma das N possibilidade pelo *princípio da indiferença*.

O extremo caso da aplicação deste princípio ocorre quando existe apenas duas possibilidades, como por exemplo, ter ou não ter petróleo sob seu terreno, simbolizado por H e H'. Neste caso, a probabilidade de cada uma das possibilidades é de 50 por cento mesmo não havendo nenhum conhecimento, pois a teoria de probabilidade diz que P(H) + P(H') = 1. Seguindo esta linha de raciocínio, podemos concluir que também pode ter uma chance de 50 por cento de encontrar, diamantes, tesouro, pastel, ou qualquer coisa sob seu terreno (GIARRATANO J., RILEY G., 1998).

Mesmo que o princípio da indiferença não seja usado, a condição de contorno P(H) + P(H') = 1 da teoria de probabilidades diz que qualquer evidência que não dê suporte a uma hipótese deve refutar esta hipótese, mesmo não existindo nenhuma evidência para isto.

Na teoria de DS, ao contrário, é possível atribuir medidas de incerteza a um conjunto de hipóteses disjuntas.

A abordagem seguida pela teoria de DS para tratar de incerteza começa com um conjunto inicial de todas as hipóteses no domínio de um problema, conhecido como *conjunto ou quadro de discernimento (frame of discernment)*.

Domínio de um problema, ou ambiente, é um conjunto de elementos (Θ) mutuamente exclusivos e exaustivos. Exemplos de dois ambientes:

$\Theta = \{\text{Fraude}, \neg\ \text{Fraude}\}$; $\Theta = \{\text{Avião comercial}, \text{Bombardeiro}, \text{Caça}\}$;

O quadro de discernimento é um conjunto de elementos dentro de um ambiente que podem ser interpretados como possíveis respostas e apenas uma é correta.

Ele é constituído por todos os subconjuntos deste ambiente incluindo ele próprio, isto é, pelo seu conjunto de partes *(power set),* simbolizado por 2^{Θ} ou $P(\Theta)$.

No caso de $\Theta = \{$Fraude, $\neg$ Fraude}, por exemplo, o seu $P(\Theta)$ é constituído de $\varnothing$, {Fraude}, {$\neg$ Fraude} e o próprio ambiente Θ.

No caso de $\Theta = \{$ $A_{\text{vião comercial}}$,$B_{\text{ombardeiro}}$,$C_{\text{aça}}$ }, o seu $P(\Theta)$ é constituído de $\varnothing$, {A}, {B}, {C}, {A, B}, {A, C}, {B, C}, e o próprio ambiente Θ. A figura 4 ilustra os subconjuntos possíveis de Θ (GIARRATANO J., RILEY G., 1998).

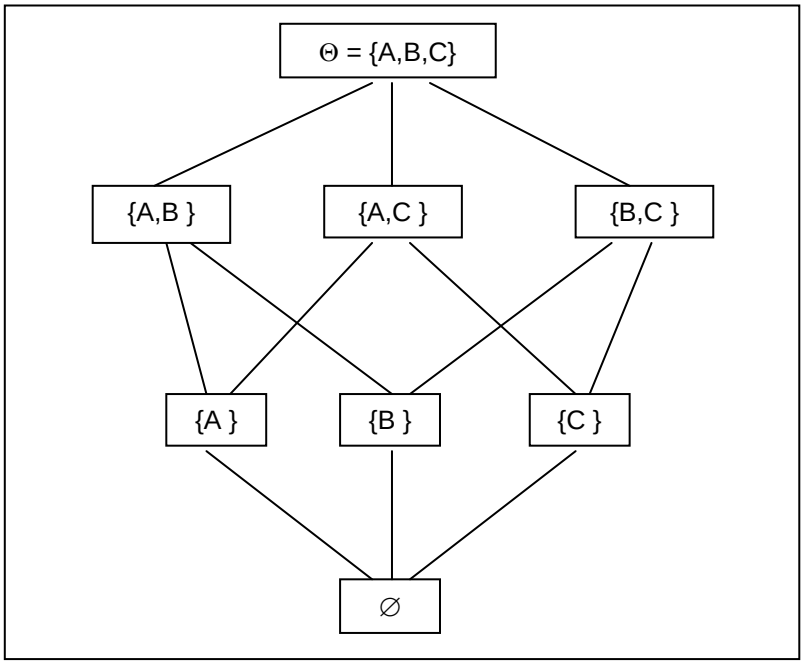

Figura 4 - Subconjuntos possíveis no ambiente de aviões (GIARRATANO J., RILEY G., 1998)

2.5.1. Função de massa e ignorância

Na teoria de DS, costuma-se considerar o grau de confiança (*belief*) sobre uma evidência como uma *massa* de um objeto físico. Isto é, a medida de confiança, simbolizada pela letra *m*, é análoga a quantidade de massa.

A razão desta analogia é que a confiança é considerada como uma quantidade que pode ser movida, dividida e combinada.

Outro termo utilizado para esta massa é *atribuição de probabilidade básica (bpa-basic probability assignment) ou simplesmente probabilidade básica*.

A massa é atribuída apenas aos subconjuntos do ambiente aos quais se deseja atribuir confiança.

Se uma confiança não estiver atribuída a um subconjunto específico, este subconjunto é considerado *sem nenhuma confiança (nonbelief ou no belief)* e esta confiança fica associada ao ambiente Θ.

Confiança que refuta uma hipótese é denominada *desconfiança (disbelief),* e não tem o mesmo significado de *sem nenhuma confiança.*

Todo conjunto num conjunto de partes *(power set)* que tiver uma massa > 0 é denominado *elemento focal*.

Para ilustrar estes conceitos, consideremos como exemplo, o domínio $\Theta = \{$Fraude, $\neg$ Fraude$\}$.

Neste ambiente, suponha que um detector de fraudes indique uma confiança de 0,7 na evidência de que uma transação seja uma fraude.

De acordo com a teoria de DS, a atribuição de massa é para o conjunto {Fraude}, expressa por m ({ Fraude }) = 0,7.

O resto da confiança é deixado para o ambiente Θ como *sem nenhuma confiança (nonbelief)*, expressa por m (Θ) = 1- 0,7 = 0,3.

Na teoria de probabilidade, este 0,3 representa uma desconfiança na fraude (isto é, considera que não é uma fraude). Na teoria de DS significa: *"confiamos que a transação seja uma fraude com um grau de 0,7 e estamos reservando um parecer de 0,3 tanto na desconfiança como na confiança adicional de que seja uma fraude"* (GIARRATANO J., RILEY G., 1998).

É importante notar que a atribuição de 0,3 no ambiente Θ não atribui nenhum valor aos seus subconjuntos, mesmo que estes subconjuntos incluam {Fraude}, e {¬Fraude}.

Outro exemplo que pode ser citado para ilustrar o conceito seria:

Se m(obter nota 10 e se graduar) for igual a 0,7, não implica que m(obter nota 10 e não se graduar) seja igual a 0,3, a menos que ambos os valores sejam assinalados explicitamente (GIARRATANO J., RILEY G., 1998).

A tabela 2 apresenta uma comparação entre massa e probabilidade:

Tabela 2 - Comparação entre massa da teoria de Dempster-Shafer e a teoria de probabilidade (GIARRATANO J., RILEY G., 1998)

Teoria de Dempster-Shafer	***Teoria de Probabilidade***
m(Θ) não tem que ser 1	$\sum P_i = 1$
Se X $\subseteq$ Y, não significa que m(X) $\leq$ m(Y)	P(X) $\leq$ P(Y)
Nenhum requisito de relacionamento entre m(X) e m(X')	P(X) + P(X') = 1

A massa na teoria de DS pode ser expressa formalmente como uma função que mapeia cada elemento do seu conjunto de partes, P(Θ) ou 2^{Θ}, em um número real entre 0 e 1. Este mapeamento é expresso formalmente por

m: $2^{\Theta} \rightarrow [0,1]$,

com as seguintes propriedades,

m (∅) = 0, isto é a massa de um conjunto vazio é zero,

e a soma de todas as massas de todos os subconjuntos de 2^{Θ} é igual a 1

$$\sum_{X \in P(\Theta)} m(X) = 1$$

lembrando que $P(\Theta) = 2^{\Theta}$.

No exemplo de fraudes acima,

$$\sum_{X \in P(\Theta)} m(X) = m(\{Fraude\}) + m\{\Theta\} = 0{,}7 + 0{,}3 = 1$$

2.5.2. Regra de combinação de Dempster

A regra de combinação de Dempster fornece um procedimento numérico para combinar evidências obtidas de diversas fontes sobre um mesmo quadro de discernimento e produzir uma estimativa agregada de confiança na evidência.

A teoria de Dempster-Shafer assume que estas fontes são independentes.

Suponha duas massas de evidência, $m_1(Z)$ e $m_2(Z)$, fornecidas por duas fontes independentes sobre o mesmo quadro de discernimento.

As massas de evidência podem ser combinadas usando a seguinte regra de combinação de Dempster:

$$m_3(Z) = m_1(Z) \oplus m_2(Z) = \frac{\sum_{X \cap Y = Z} m_1(X) \,.\, m_2(Y)}{1 - K} \qquad (2)$$

onde, o operador $\oplus$ representa uma soma ortogonal ou soma direta definida pela soma dos produtos das massas da intersecção $X \cap Y = Z$.

O denominador 1 – K, é um fator de normalização onde K é definida pela soma dos produtos das massas da intersecção $X \cap Y = \varnothing$, isto é,

$$K = \sum_{X \cap Y = \varnothing} m_1(X) \,.\, m_2(Y)$$

Os exemplos a seguir, ilustram o procedimento para combinar evidências segundo regra de Dempster:

Suponha que no exemplo de fraudes acima, um segundo detector identificou a transação como uma fraude com uma confiança de 0,9.

As massas de confiança dos dois detectores, m_1 e m_2, são as seguintes:

m_1 ({ Fraude }) = 0,7 e m_1 (Θ) = 0,3
m_2 ({ Fraude }) = 0,9 e m_2 (Θ) = 0,1

A tabela 3 apresentada a seguir mostra as intersecções e os produtos resultantes. Os elementos da tabela são calculados fazendo a intersecção das linhas e colunas para obter os conjuntos resultantes e multiplicando as suas respectivas massas:

Tabela 3 - Tabela de intersecções e produtos de m_1 e m_2

	m_2 ({ Fraude }) = 0,9	m_2 (Θ) = 0,1
m_1 ({ Fraude }) = 0,7	{ Fraude } = 0,63	{ Fraude } = 0,07
m_1 (Θ) = 0,3	{ Fraude } = 0,27	Θ = 0,03

De acordo com a regra de Dempster, os conjuntos comuns resultantes devem ser somados:

m_3 ({ Fraude }) = $m_1 \oplus m_2$ ({ Fraude }) = 0,63 + 0,27 + 0,07 = 0,97

m_3 (Θ) = $m_1 \oplus m_2$ (Θ) = 0,03.

O valor de K da função de combinação é igual a zero, pois as intersecções não resultaram em nenhum conjunto vazio.

m_3 ({ Fraude }) representa a confiança na evidência combinada de fraude.

m_3 (Θ) implica numa informação adicional, pois como ele inclui {Fraude}, é *plausível* que contribua na confiança da evidência de fraude.

Portanto, a sua massa, 0,03, pode ser adicionada a confiança de 0,97 no conjunto {Fraude} para produzir a *máxima confiança* de que ele seja uma fraude.

Isto significa que existe um *intervalo de confiança* na evidência de fraude entre 0,97 e 1,0, representado por [0.97, 1.0].

O limite inferior é conhecido como *Confiança (Bel – Belief)* e o limite superior como *Potencial de Confiança ou Confiança Plausível (Pls – Plausibility)*.

2.5.3. Função de Confiança

A função de confiança (*Bel*), fornece a confiança total de um conjunto e de todos os seus subconjuntos, isto é, Bel é a soma de todas as massas que dão suporte a um conjunto, e é definido em termos de massa:

$$Bel\ (\ X\ ) = \sum_{Y \subseteq X} m(Y)$$

No caso de $\Theta = \{$A,B,C}, por exemplo,

Bel_1 ({B,C}) = m_1 ({B,C}) + m_1 ({B}) + m_1 ({C})

Como as funções de confiança podem ser definidas em termos de massa, a combinação de duas funções de confiança pode ser expressa em termos de soma ortogonal das massas de um conjunto e de todos os seus subconjuntos. Por exemplo:

$Bel_1 \oplus Bel_2$ ({B,C}) = $m_1 \oplus m_2$({B,C}) + $m_1 \oplus m_2$({B}) + $m_1 \oplus m_2$({C})

No exemplo de $\Theta = \{$Fraude, $\neg$ Fraude} anterior, tem-se o seguinte:

$Bel_1 \oplus Bel_2$ ({ Fraude }) = $m_1 \oplus m_2$ ({ Fraude }) = 0,97

Neste mesmo exemplo, a função de confiança para Θ é dada por

$Bel_1 \oplus Bel_2$ (Θ) = $m_1 \oplus m_2$ (Θ) + $m_1 \oplus m_2$ ({ Fraude }) = 0,03 + 0,97 = 1.

De fato, Bel (Θ) deve ser igual a 1, pois a soma de todas as massas é igual a um.

2.5.4. Intervalo de confiança

Bel (S) representa o grau com que a evidência suporta a hipótese S, isto é, fornece um limite inferior de confiança.

Bel (S') representa o grau com que a hipótese S é refutada (desconfiada).

Pls (S) = 1 – Bel (S') representa a confiança total não atribuída a S', de forma que fornece um limite superior de confiança a S.

A diferença Pls (S) – Bel (S) expressa o grau de incerteza (ou ignorância) com relação a S.

Intervalo de confiança (ou intervalo de evidência) de um conjunto S é definido por:

$$IC(S) = [\, Bel(S), Pls(S) \,]$$

No exemplo de $\Theta = \{A, B, C\}$, se $S = \{B\}$, teremos $S' = \{A,C\}$

No exemplo de fraudes, se $S = \{Fraude\}$, $S' = \{\neg Fraude\}$, teremos

$$Bel(\{\neg Fraude\}) = Bel_1 \oplus Bel_2(\{\neg Fraude\}) = 0,$$

pois, ele não é um elemento focal, isto é, não lhe foi atribuída nenhuma massa. Portanto,

$$Pls(\{Fraude\}) = 1 - 0 = 1,$$

e portanto,

$$IC(\{Fraude\}) = [\, 0.97, 1 \,].$$

A seguir estão ilustrados alguns exemplos de intervalos de confiança (ESMAILI, M., 1997):

- Se [Bel (x), Pls (x)] = [0, 1], nenhuma informação a respeito da hipótese x é disponível.
- Se [Bel (x), Pls (x)] = [0, 0], a hipótese x é totalmente negada.
- Se [Bel (x), Pls (x)] = [1, 1], a hipótese x é totalmente confirmada.
- Se [Bel (x), Pls (x)] = [0, 0.8], existe alguma evidência contra a hipótese x.
- Se [Bel (x), Pls (x)] = [0.3, 1], existe alguma evidência a favor da hipótese x.
- Se [Bel (x), Pls (x)] = [0.15, 0.75], existe alguma evidência a favor ,assim como, contra a hipótese x.

Se Pls (x) – Bel (x) = 0, para todos os $x \subseteq \Theta$, a teoria de DS é igual a teoria convencional de probabilidade.

A figura 5 a seguir ilustra as relações entre as grandezas de confiança.

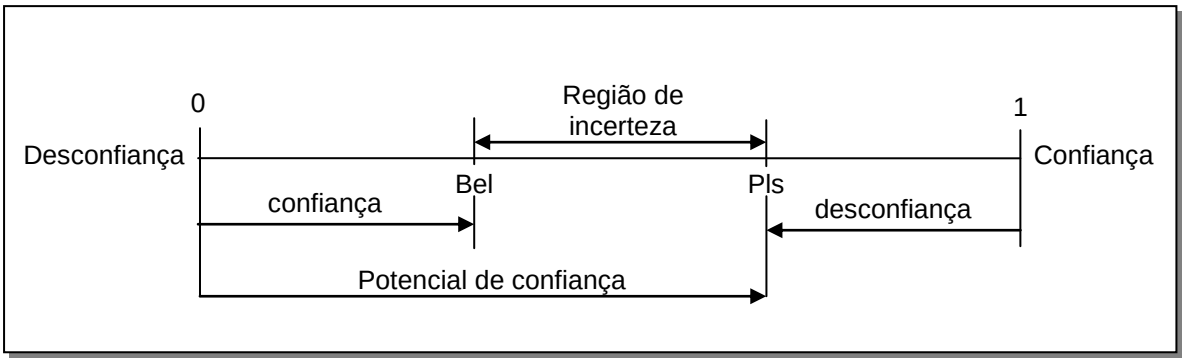

Figura 5 - Relações entre as grandezas de confiança

Suponha agora que um terceiro detector reporta uma evidência conflitante de 0,95 de que a transação não é uma fraude, isto é,

$$m_3 (\{\neg Fraude\}) = 0{,}95 \quad e \quad m_3 (\Theta) = 0{,}05$$

A tabela 4 abaixo mostra como os produtos cruzados são calculados.

Tabela 4 - Tabela de intersecções e produtos entre m_1 , m_2 e m_3

	$m_1 \oplus m_2 (\{Fraude\}) = 0{,}97$	$m_1 \oplus m_2 (\Theta) = 0{,}03$
$m_3 (\{\neg Fraude\}) = 0{,}95$	$\{ \varnothing \}$ 0,9215	$\{ \neg Fraude \}$ 0,0285
$m_3 (\Theta) = 0{,}05$	{ Fraude } 0,0485	Θ 0,0015

O conjunto vazio $\varnothing$ ocorre porque {Fraude} e $\{\neg Fraude\}$ não tem nenhum elemento em comum.

O fator K é igual a soma das massas dos conjuntos vazios que resultaram da intersecção, ou seja, K = 0,9215.

Portanto, 1 – K = 1 – 0,9215 = 0,0785

Aplicando a função de combinação em cada um dos conjuntos resultantes da intersecção, temos:

$m_1 \oplus m_2 \oplus m_3$ ({¬Fraude}) = 0,0285 / 0,0785= 0,363

$m_1 \oplus m_2 \oplus m_3$ ({Fraude}) = 0,0485 / 0,0785= 0,617

$m_1 \oplus m_2 \oplus m_3$ (Θ) = 0,0015 / 0,0785= 0,020

A confiança total no subconjunto { Fraude } é agora,

Bel ({Fraude}) = $m_1 \oplus m_2 \oplus m_3$ ({Fraude}) = 0,617, e

Bel ({¬Fraude}) = $m_1 \oplus m_2 \oplus m_3$ ({¬Fraude}) = 0,363

Pls ({Fraude}) = 1 - Bel ({¬Fraude }) = 1 – 0,363 = 0,637

Portanto,

IC ({ Fraude }) = [0,617, 0,637].

O suporte a hipótese (Bel) e o potencial de confiança (Pls) para {Fraude} foram reduzidos pela evidência conflitante de {¬Fraude}.

Este exemplo teve como objetivo ilustrar a teoria de evidências de Dempster-Shafer como uma forma de combinar evidências de várias fontes de detecção, assim como quantificar o nível de confiança nos resultados obtidos.

A grande vantagem da teoria de DS é que nenhum conhecimento a priori é requerido para a combinação de evidências, tornando-a adequada para detecção de

fraudes ou de qualquer tipo de anomalia antes não observadas. (CHEN, Q.; AICKELIN U., 2006)

Outra vantagem é que a teoria de DS permite expressar a ignorância, dando informação sobre a incerteza da situação.

A desvantagem da teoria de DS é a sua complexidade computacional, pois depende do número de elementos no quadro de discernimento (Θ).

Se existir *n* elementos em Θ, haverá até 2^{n-1} elementos focais para a função de massa. Isso significa que a combinação de duas massas precisa de uma computação de até 2^{n} intersecções. (CHEN, Q.; AICKELIN U., 2006)

Entretanto, quando o quadro de discernimento tiver apenas dois elementos (Fraude, $\neg$Fraude), a complexidade computacional é pequena, pois existem no máximo três elementos focais: {Fraude}, {$\neg$Fraude} e { Fraude, $\neg$Fraude} (isto é, a incerteza), resultando portanto em baixa complexidade computacional.

3. REVISÃO DA LITERATURA

3.1. CONSIDERAÇÕES INICIAIS

Este capítulo descreve e comenta alguns dos trabalhos mais relevantes encontrados na literatura técnica especializada relacionado com o tema da tese.

Existem poucas publicações sobre detecção de fraudes no domínio de aplicações de transações bancárias *online*. É provável que este fato seja mais em virtude da privacidade, do sigilo e dos interesses comerciais neste domínio, do que pela ausência de pesquisa.

Em virtude da pouca troca de informações, o desenvolvimento de novos métodos de detecção na área de transações bancárias tem sido difícil. (KARSEN K. N.; KILLINGBERG T. G, 2008).

A maioria dos trabalhos publicados está relacionada a fraudes nos domínios de cartões de crédito, intrusão de computadores e comunicação móvel.

Considerando-se que os princípios conceituais utilizados em sistemas de detecção de fraudes são os mesmos em todas as áreas, os resultados obtidos a partir desta pesquisa serão aplicados para os objetivos desta tese.

Alguns dos trabalhos são apresentados aqui com mais detalhes do que os outros em virtude da relevância que os mesmos tiveram para o desenvolvimento da tese.

3.2. FRAUDES EM CARTÕES DE CRÉDITO

A maioria dos trabalhos encontrados na literatura técnica-científica sobre prevenção e detecção de fraudes em cartões de crédito foi conduzida dando especial ênfase em *data mining* e redes neurais.

Ghosh e Reilly (GHOSH, S.; REILLY D. L.,1994) apresentam um detector de fraudes baseado em redes neurais. Esta rede foi treinada com um grande número de amostras de transações pré-classificadas. O estudo de viabilidade demonstrou que este método permitiu reduzir de 20% a 40% do total de perdas em fraudes.

Aleskerov, Frisleben e Rao (ALESKEROV, E.; FREISLEBEN, B.; RAO, B.,1997) descrevem um sistema baseado em *data mining* junto com uma rede neural. A rede neural é treinada com os dados específicos de consumo de clientes gerando então um modelo que é usado para detectar fraudes.

Bolton e Hand (BOLTON, R. J.; HAND, D. J., 2001) propõem uma técnica de detecção não supervisionada utilizando análise de *breakpoint* (ponto de quebra) para identificar mudanças em comportamentos de gastos.

Breakpoint é uma observação ou um instante onde um comportamento anormal é detectado. Um exemplo disso é o aumento repentino no número de transações que pode indicar um comportamento fraudulento.

Guo e Li (GUO T.; LI G. Y.,2008) apresentam uma forma de modelar uma sequência de operações no processamento de transações em cartão de crédito usando uma rede neural combinado com nível de confiança. Isto é, se uma transação de cartão de crédito não for aceita pelo modelo de rede neural, inicialmente treinado, com um mínimo de confiança, ela é considerada fraudulenta.

A técnica de análise baseada em curvas ROC (*Receiver Operating Characteristic*) foi utilizada para assegurar a exatidão e a eficácia da detecção.

A curva ROC permitiu determinar o melhor valor do *limiar* para se obter o melhor compromisso entre as taxas de verdadeiro positivo e falso positivo com a introdução de custos na ocorrência de falsos positivos.

Uma rede neural é inicialmente treinada com dados sintéticos. Se uma transação de cartão de crédito não for aceita pelo modelo de rede neural com um nível mínimo de confiança, a transação é considerada fraudulenta.

Este artigo mostra como a introdução de nível de confiança e ROC podem ser combinados para se obter bons resultados na detecção de fraudes.

3.3. INTRUSÃO EM COMPUTADORES

Abordagens de detecção de intrusões em computadores são normalmente classificadas em duas categorias conforme o modelo de intrusão: detecção de *abuso*[1] e detecção de *anomalia*.

Detecção de *abuso* procura reconhecer os ataques de intrusões previamente observadas na forma de padrões ou assinaturas e então monitorar tais ocorrências.

Detecção de *anomalia* procura estabelecer um perfil histórico normal para cada usuário e então usar desvios suficientemente grandes deste perfil para indicar possíveis intrusões (KOU, Y., et al., 2004).

Ghosh e Schwrtzbard (GHOSH, A. K.; SCHWRTZBARD, A.,1999) descrevem uma abordagem que emprega redes artificiais neurais tanto para detectar anomalias quanto para detectar abusos.

Denning (DENNING, D. E., 1987) apresenta um modelo estatístico para detecção de intrusões em tempo real baseado em detecção de anomalias.

---

[1] O termo *abuso* foi utilizado aqui como tradução de *misuse* por ser mais adequado do que o termo *incorreto*.

De acordo com a autora, um objetivo importante da pesquisa foi determinar quais atividades e medidas estáticas fornecem o maior poder de discriminação, isto é, aqueles que oferecem uma alta taxa de detecção com uma baixa taxa de alarmes falsos. Isso se deve ao fato de que o próprio usuário legítimo pode ter um comportamento anormal.

Denning dedica uma boa parte deste trabalho as métricas e modelos estatísticos para a caracterização de perfis.

A métrica é definida como sendo uma variável aleatória x representando uma medida quantitativa acumulada durante um período. São definidos três tipos de métricas:

- *Contador de eventos;*
- *Temporizador de intervalo;*
- *Medidor de consumo de recurso*.

Dada uma métrica para uma variável aleatória x e n observações, $x_1, x_2, \ldots, x_n$, a finalidade de um modelo estatístico é determinar se uma nova observação x_{n+1} é anormal com relação as observações anteriores. Os seguintes modelos são definidos neste trabalho:

- *Modelo operacional* – baseado na suposição de que uma anomalia pode ser determinada comparando-se a nova observação de x com limites prefixados.
- *Modelo de média e desvio padrão* – baseado na suposição de que conhecemos a média (μ) e o desvio padrão (σ) das n observações. Uma nova observação x_{t+1} é definida como anormal se ele cair fora de um intervalo de confiança, isto é, d desvios padrões da média. Uma variação do modelo de média e desvio padrão é aplicar pesos nos valores observados.

- *Modelo multivariado* – modelo similar ao modelo de média e desvio padrão, com a diferença de que ele é baseado na correlação entre duas ou mais métricas.

Outros modelos são considerados, como por exemplo, o modelo que usa processos de Markov, séries temporais, etc.

Entretanto, a autora cita dois problemas em potencial que são também citados em outros artigos. Estes dois problemas ocorrem quando novos usuários são introduzidos no sistema.

O primeiro problema é causado pela falta de informação sobre o comportamento ou pela inexperiência do próprio usuário com o sistema, podendo gerar um número excessivo de alarmes.

Uma sugestão para este problema, segundo Denning, seria ignorar alarmes de anomalias para novos usuários, mas isso leva ao segundo problema que consiste em não detectar uma intrusão pensando ser um usuário novo. O que se deseja é uma solução que minimiza a ocorrência de falsos alarmes sem ignorar intrusões reais.

A abordagem proposta pela autora para minimizar falsos alarmes é através do uso de modelos estatísticos apropriados para as atividades que causam alarmes e por uma escolha apropriada de perfis.

Por exemplo, no modelo de média e desvio padrão, o intervalo de confiança utilizado poderia ser maior no início, e ir diminuindo a medida que são coletados mais dados sobre o comportamento do usuário.

Isso pode reduzir o número de falsos alarmes causados por um perfil individual de usuário, mas não protege o sistema contra novos usuários (ou usuários infrequentes) cujo comportamento é duvidoso, ou contra usuários que tem um comportamento não normal desde o início.

Contra isso, Denning sugere que a atividade corrente seja comparada com um conjunto de perfis comuns para todos os usuários ou de usuários pertencentes a um mesmo grupo.

O trabalho de Denning mostra que as técnicas e os métodos utilizados para detectar intrusões em computadores são basicamente os mesmos de detecção de fraudes que utilizam métodos estatísticos. Neste sentido, este artigo tem sido uma referência para quase todos os artigos sobre sistemas de detecção de fraudes e de intrusões.

3.4. FRAUDES EM COMUNICAÇÃO MÓVEL

Fraude em redes de comunicação móvel se refere ao acesso ilegal a rede, assim como no uso ilegal de seus serviços.

Cortes e Pregibon (CORTES, C.; PREGIBON, D., 2001) apresentam um modelo de detecção baseado em sumários estatísticos, nomeados por eles de *assinaturas.*

Assinatura é um vetor de variáveis correspondentes aos atributos de uma transação telefônica utilizados para definir o comportamento de um usuário. São exemplos de atributos, o número de chamadas efetuadas num período, tempo médio de cada chamada, localização da origem das chamadas, entre outros.

São definidas duas janelas de tempo onde as assinaturas são determinadas. Uma relativa a atividade atual na rede e outra relativa a informação histórica do usuário. A atividade atual na rede é comparada com a sua atividade histórica e a existência de alguma diferença é verificada.

Este trabalho discute aspectos estatísticos e computacionais dos métodos baseados em assinatura para capturar o desvio de comportamento de usuário diretamente entre de dados de uma transação.

Para detecção de fraudes, as assinaturas podem ser usadas de duas formas:

- Métodos de detecção baseados em perfil, onde uma biblioteca de perfis de ataques é armazenada numa base de dados. As assinaturas obtidas do tráfego são comparadas com estes perfis para detectar comportamento fraudulento ou detectar intrusão. É um método supervisionado.
- Métodos de detecção baseados em anomalia, onde a própria assinatura do usuário serve de base para a comparação. Assinaturas obtidas do tráfego são comparadas com a assinatura do usuário para determinar se houve desvio de comportamento. Um desvio significativo da assinatura é usado como sinal de fraude.

Um dos pontos positivos deste trabalho é a discussão exemplificada das distribuições de probabilidades dos componentes da assinatura e a maneira como estas distribuições são combinadas. Na prática, a assinatura é uma estimativa da distribuição de probabilidade conjunta de todos os componentes.

Um dos desafios na utilização de métodos baseados em assinaturas é a necessidade de se adaptar as mudanças nos padrões de utilização dos usuários.

Neste trabalho, os autores escolheram não atualizar uma assinatura com os dados da nova transação se a assinatura desta transação desviar significativamente da assinatura histórica do usuário.

Dois modelos de processamento são definidos para atualização de assinaturas:

- Processamento *orientado a tempo*, onde os registros são coletados e armazenados por um certo período de tempo, e no final deste período, os registros são sumarizados e as assinaturas atualizadas.
- Processamento *orientado a evento (ou ação)*, onde as assinaturas são atualizadas a medida que chegam os novos registros.

Nos sistemas onde os custos associados com fraude são elevados e requer detecção em tempo real, o modelo mais adequado de processamento é o *orientado a eventos*, embora o custo de processamento seja mais elevado, pois existem certos tipos de variáveis que são difíceis de serem calculados em tempo hábil.

Tanto um modelo de processamento como outro seguem o mesmo modelo computacional para atualização de assinaturas:

- Sejam S_t, a assinatura do usuário no instante t e Pc, a assinatura extraída da transação no instante t+1, respectivamente.
- A nova assinatura S_{t+1} no instante t+1, é formada a partir da assinatura do instante t e por Pc, variável por variável.

$$S_{t+1} = \begin{cases} S_t & \text{se } |S_t - Pc| > \varepsilon \\ \beta.S_t + (1-\beta).Pc, & \text{caso contrário} \\ \text{onde,} \quad 0 < \varepsilon < \infty \quad \text{e} \quad 0 \leq \beta \leq 1 & \end{cases}$$

O parâmetro ε serve como um limiar para evitar que uma assinatura seja contaminada por um padrão de comportamento de um elemento estranho.

O parâmetro β é um fator heurístico que determina o peso a ser aplicado nos dados novos e nos dados velhos.

Para atualização usando modelo *orientado a tempo*, β é normalmente constante. O seu valor é baseado na escala de tempo em que os dados mais antigos ficam irrelevantes.

Por exemplo, para uma atualização diária com β = 0,85, permite manter as informações dos últimos 30 dias. Com, β = 0,5, são mantidas informações dos últimos 7 dias.

Para atualização usando modelo *orientado a evento*, os autores sugerem que a escolha do valor de β seja em função do tempo entre as chegadas das transações ou da taxa de chegada.

A importância deste trabalho reside no fato de propor uma solução para detectar anomalias diretamente sobre transações por meio de assinaturas e, portanto, em tempo real.

Entretanto, esta metodologia pode ser enganada por fraudes que se introduzem lentamente no tráfego de uma conta.

FAWCETT, T. e PROVOST, F. (1997) apresentam métodos baseados em regras e em redes neurais para gerar detectores de fraude de clonagem de celulares.

Um dos pontos importantes destacados pelos próprios autores é que os sistemas de detecção de fraudes precisam ser adaptativos. Isto é, que se adaptem as novas condições de uso, que sejam capazes de detectar novos padrões de fraude, e que também sejam capazes de modificar o comportamento do gerador de alarmes em função das mudanças do nível de fraude e dos custos associados com os falsos positivos.

O sistema abordado neste artigo obtém esta adaptabilidade através de um processo de geração automática de detectores de fraude utilizando técnicas de *data mining* e de *redes neurais.*

A construção do detector é feita em três estágios.

No primeiro estágio são gerados padrões (indicadores) de fraude na forma de regras de classificação utilizando um programa de aprendizado de regras em base de dados contendo registros das transações mais recentes.

No segundo estágio são construídos monitores de perfil a partir desta lista de regras e um conjunto de gabaritos pré-definidos que determinam o modelo estatístico a ser utilizado para o cálculo de desvios.

Cada monitor é responsável por uma regra e tem duas etapas de operação.

Na primeira etapa, conhecida como etapa de treinamento, o monitor é utilizado para medir as atividades normais, sem fraudes, de um usuário aplicado durante um período de tempo. Como resultado desta etapa tem-se um valor estatístico que representa o perfil de comportamento normal do usuário com relação a regra pela qual o monitor é responsável.

Na segunda etapa, conhecida como etapa de uso, o monitor é usado para processar as atividades diárias de uma conta. Nesta etapa o monitor gera um valor numérico para cada atividade indicando o seu desvio com relação ao valor do perfil de comportamento normal do usuário

Dois gabaritos de monitores são utilizados: *Monitor de limiar e Monitor de desvio padrão*

O terceiro estágio da construção do detector consiste em determinar a forma de combinar as evidências geradas pelos monitores.

Esta combinação é feita por meio de um LTU[2] (*Linear Threshold Unit*), isto é, por um modelo de neurônio artificial que é constituído de um somador ponderado e de um detector de limiar a ser aplicado no resultado da soma.

Durante a etapa de treinamento, as saídas dos monitores referentes aos dados de uma conta diária do usuário já classificados são apresentados ao LTU junto com os resultados desejados (fraude, não fraude). São então calculados os pesos a serem aplicados nas saídas dos monitores, assim como, o valor de limiar do

[2] LTU é também conhecido como *percéptron.* (MITCHELL T. M.,1997)

resultado da soma de forma que os alarmes possam ser disparados com grande confiança

Um dos pontos forte deste artigo é a descrição dos processos de treinamento tanto dos monitores quanto do estágio de combinação de evidências.

Além disso, se destaca também a discussão sobre a importância dos custos associados com falsos positivos e falsos negativos. A solução adotada para minimizar estes custos foi a incorporação de uma segunda etapa no treinamento que consiste em um ajuste manual no limiar do LTU.

O sistema de detecção de fraudes proposto pelos autores não serve para aplicações que exigem uma detecção em tempo real, pois o processamento é feito sobre registros de chamadas acumuladas em um dia.

Panigrahi, Kundu, Sural, and Majumdar (PANIGRAHI, S.; KUNDU, A.; SURAL, S.; MAJUMBAR, A. K., 2007) descrevem um *framework* para detecção de fraudes em redes de comunicação móvel usando a teoria matemática de evidências de Dempster-Shefer (DS) para combinar múltiplas evidências fornecidas pelo componente detector de desvio baseado em regras e determinar um escore total de suspeita.

O detector de desvio determina o nível de suspeita de cada chamada baseando-se na extensão do desvio em relação aos padrões de comportamento esperados.

Uma chamada é classificada como normal, ou suspeita dependendo do escore de suspeita.

Se essa chamada for considerada suspeita, a sua confiança (*belief*) é reforçada ou enfraquecida baseando-se na similaridade com os dados históricos de chamadas, utilizando a teoria Bayesiana.

Este *framework* é constituído de quatro componentes:

- Componente de detecção de desvio baseado em regras
- Componente de combinação de Dempster-Shafer
- Componente de base de dados com históricos de chamadas
- Componente de aprendizado Bayesiano

O componente de detecção de desvio baseado em regras consiste de regras genéricas e específicas que classificam uma chamada como sendo genuína ou fraudulenta com certa probabilidade.

O valor desta probabilidade é usado como escore de suspeita que indica o quanto uma chamada se desvia do padrão de utilização normal do assinante.

Duas técnicas baseadas em regras são apresentadas no trabalho: Análise de *breakpoint* (Regra 1) e detecção de desvio de frequência (Regra 2).

Segundo os autores, o termo *breakpoint* significa uma observação ou um instante em que é detectado um comportamento anômalo.

Na análise de *breakpoint,* as chamadas recentes são comparadas com padrões de utilização anterior para detectar mudanças bruscas no comportamento.

Uma janela deslizante (*sliding window*) de tamanho fixo é usada de tal forma que ao chegar uma nova chamada, ela entra na janela e a mais antiga é removida.

As chamadas mais recentes da janela são comparadas com as chamadas da parte mais antiga. Métodos estatísticos utilizados nesta comparação determinam o grau de variação das chamadas com relação ao perfil das chamadas normais.

Na detecção de desvio de frequência, a regra utilizada consiste em detectar excesso de atividade em uma conta em termos de frequência de chamadas e o correspondente tipo de chamada (local, internacional, ...).

Para detectar um aumento súbito na frequência de chamadas de um determinado tipo por um telefone móvel, o sistema compara a sua atividade mais recente com o perfil histórico de sua utilização. Da mesma forma que na regra anterior, o desvio de frequência é calculado usando métodos estatísticos.

O componente de combinação de Dempster-Shafer tem o papel de combinar as evidências obtidas pelas regras 1 e 2, e calcular o escore de suspeita de uma chamada. É utilizada a teoria matemática de evidências de Dempster-Shafer (SHAFER, G, 1976 apud GIARRATANO J., RILEY G. 1998) no cálculo destes escores.

Baseado nos escores de suspeita, uma chamada pode ser detectado como normal, anormal ou simplesmente suspeita.

As chamadas consideradas suspeitas (isto é, aquelas que não tiveram evidências suficientes para serem consideradas normais ou anormais) são passadas para o componente de base de dados com históricos de chamadas para reforçar ou enfraquecer a confiança baseando-se na similaridade com o histórico das chamadas fraudulentas ou genuínas.

O componente de aprendizado Bayesiano é usado para atualizar o escore de suspeita obtido pelo componente de combinação de Dempster-Shafer sob luz da nova evidência obtida pelo componente de base de dados com históricos de chamadas.

Para demonstrar a eficácia do sistema, foi desenvolvido um simulador para modelar o comportamento de assinantes genuínos assim como de fraudadores em diferentes situações.

Segundo os autores, foi conseguido até 96% de taxa de verdadeiros positivos e menos de 10% de taxa de falsos positivos.

3.5. CONSIDERAÇÕES FINAIS

Este capítulo descreveu alguns dos aspectos mais relevantes dos sistemas de detecção de fraudes e intrusões em computadores.

De uma forma geral, os pontos comuns destes artigos são a adaptação dos detectores de fraude as novas condições de uso, a capacidade de detectarem novos padrões de fraude e minimizar os falsos positivos.

A principal diferença entre as abordagens decorre em virtude de alguns terem como requisito, a detecção em tempo real. Observa-se que nestes sistemas, a sua capacidade de detecção é menor do que aqueles que não possuem este requisito.

O último artigo aborda um sistema que utiliza a teoria de Dempster-Shafer para combinar evidências de fraude determinadas por duas regras no domínio de telefonia móvel.

O sistema proposto nesta tese para a detecção de fraudes em transações financeiras via Internet tem como requisito fundamental a detecção em tempo real. Para alcançar este objetivo e, além disso, obter um bom desempenho (relação Tvp e Tfp) sem utilizar métodos que exigem grande capacidade de processamento, ele utiliza duas abordagens diferentes que são combinadas usando teoria de Dempster-Shafer.

O grande diferencial deste sistema com relação aos sistemas abordados neste capítulo está na concepção modular de sua arquitetura e na identificação de atributos de natureza específica do domínio de fraudes de transações financeiras via Internet.

A arquitetura básica do sistema é constituída de dois módulos de detecção que utilizam métodos diferentes e os resultados obtidos por estes módulos são combinados por um terceiro módulo.

O primeiro módulo utiliza uma abordagem baseada em análise diferencial que avalia a probabilidade de uma transação ser uma fraude baseando-se no seu comportamento local.

O segundo módulo utiliza uma abordagem inovadora que avalia a probabilidade de uma transação ser uma fraude baseando-se no seu comportamento global.

A distinção entre os comportamentos, local e global, é baseada na identificação de duas classes de atributos específicos do domínio de transação financeira via Internet.

A teoria de Dempster-Shafer é usada para combinar evidências de fraude determinados pelos demais módulos;

Apesar de o sistema proposto ter sido definido utilizando dois módulos independentes de detecção, a arquitetura proposta na tese permite que outros módulos sejam incorporados e combinados, respeitando os requisitos de detecção em tempo real.

4. PROPOSTA DE ARQUITETURA PARA DETECÇÃO DE FRAUDES EM TRANSAÇÕES FINANCEIRA VIA INTERNET

4.1. CONSIDERAÇÕES INICIAIS

Este capítulo apresenta a proposta de uma arquitetura para detectar fraudes em tempo real em transações financeiras na Internet abordando quatro itens que sintetizam o objetivo do trabalho:

- A determinação dos atributos de uma transação financeira para definir o perfil de comportamento de um usuário e de seus dispositivos de acesso;
- A escolha dos modelos estatísticos mais adequados para cada um dos atributos para caracterizar evidências de fraude;
- Um método para combinar estas evidências e determinar um escore de suspeita de fraude; e
- a especificação da arquitetura de um sistema para detecção de fraudes em tempo real contemplando os três aspectos acima.

Quanto a arquitetura do sistema proposto para a detecção de fraudes, ela se baseia na proposta de utilizar várias abordagens de detecção, cada uma baseada em uma técnica diferente, e então combinar os resultados individuais para se obter uma avaliação global da fraude.

A idéia de combinar várias abordagens diferentes vem do fato de que uma das características mais marcantes em detecção de fraudes é a importância do compromisso entre a detecção das fraudes verdadeiras e a produção de falsos alarmes (ou falsos positivos).

A utilização de múltiplas abordagens (MOREAU, Y., et al., 1999) possibilita um desempenho melhor que as abordagens isoladas

O sistema proposto nesta tese usa duas abordagens complementares para detecção de fraudes:

- *Abordagem global* que é centrada nos dispositivos de acesso operando num conjunto de contas correntes pertencentes a diferentes usuários; e
- *Abordagem local* que é centrada no cliente, que embora possa acessar por meio de diferentes dispositivos de acesso, manipula contas correntes de sua titularidade

Na *abordagem local*, ou *abordagem baseada em análise diferencial*, os padrões de utilização da conta corrente são monitorados e comparados com o histórico de sua utilização, que representa o comportamento normal do usuário. Qualquer desvio significativo do comportamento normal indica uma potencial fraude.

Na *abordagem global*, ou *abordagem baseada em análise global*, cada dispositivo é monitorado e classificado como legítimo ou fraudulento com certa probabilidade baseado em informações globais.

A abordagem de análise global é baseada em três suposições, a saber:

- A primeira assume que cada dispositivo usado em *online banking* tem uma identificação única.
- A segunda suposição é baseada no fato de que a probabilidade de uma transação ser uma fraude cresce com o número de contas correntes acessadas pela mesma fonte que solicita a transação atual.
- A terceira suposição vem do fato de que a única forma de ter certeza de que uma fraude foi perpetrada é quando o próprio usuário reporta a fraude na sua conta corrente

A maior contribuição deste trabalho é a constatação, através da análise empírica de um conjunto de transações do mundo real, de que a identificação eficaz dos dispositivos de acesso e monitoramento das contas acessadas por

cada dispositivo é um suplemento que pode ajudar outros métodos na detecção de comportamentos fraudulentos em aplicações bancárias online.

4.2. ATRIBUTOS DE UMA TRANSAÇÃO FINANCEIRA

Nos domínios onde se caracterizam a existência de fraudes, existem alguns atributos de natureza específica de cada domínio que podem ser utilizados para detectar fraudes com grande poder de discriminação.

No domínio de telefonia celular, por exemplo, um usuário (conta) só acessa outra conta (usuário), enquanto que no domínio de transações financeiras um usuário (conta) só acessa a sua própria conta.

Portanto, o primeiro passo para a escolha de atributos que serão usados na construção de indicadores de fraude deve ser a identificação das características específicas de seu domínio.

No caso do domínio de transações financeiras onde se aplica a proposta da arquitetura, foram identificadas duas classes de atributos a partir de observações empíricas em dados reais:

- Uma, centrada no usuário legítimo, isto é, que acessa apenas a sua própria conta; e
- Outra, centrada no fraudador, isto é, que acessa múltiplas contas pertencentes a outros usuários.

Com base nesta classificação, os atributos centrados no usuário legítimo foram denominados de *atributos locais*, em virtude de estarem relacionados apenas a sua conta, ou seja, localmente a sua conta.

Os atributos centrados no fraudador, por outro lado, foram denominados de *atributos globais* em virtude da sua característica global, isto é, de acessar várias contas não autorizadas.

4.3. CARACTERÍSTICAS DE FRAUDES DE TRANSAÇÕES BANCÁRIAS *ONLINE*

Uma análise empírica efetuada em um conjunto de transações do mundo real revelou que a maioria das fraudes tem algumas características de comportamento semelhantes, sendo algumas descritas abaixo:

- Grande número de contas correntes diferentes acessadas por um único dispositivo;
- Transações originadas de um mesmo dispositivo envolvendo pequenas quantias em muitas contas correntes;
- Mais transações de pagamento que o usual em uma única conta corrente;
- Grande número de erros de senha por tentativas provenientes de um mesmo dispositivo de acesso;

Enquanto que as duas últimas características podem ser detectadas através da análise diferencial usando atributos locais, as duas primeiras características precisam de informações sobre ataques similares em outras contas.

O sistema de detecção de fraude proposto nesta tese leva em conta estas características para a determinação de atributos que possam ser utilizados na determinação de indicadores de fraude.

4.4. DESCRIÇÃO GERAL DO SISTEMA DE DETECÇÃO

A idéia básica do sistema proposto tem como objetivo obter o maior número de evidências quanto a veracidade da fraude considerando que algumas formas de ataque não são detectadas por um único método, mas possivelmente por um conjunto de métodos diferentes.

Neste sentido, considerando-se as características de fraude em transações financeiras *online* em que a observação das atividades globais desempenha um papel importante na detecção de fraudes, a arquitetura proposta possui pelo menos um método de detecção baseado no comportamento global dos usuários.

Nesta tese serão considerados apenas dois métodos de detecção, sendo um baseado em atributos globais, pois o objetivo aqui é demonstrar a importância do método de detecção baseado no comportamento global de usuário e seus dispositivos de acesso em transações financeiras *online* como complemento de outros métodos.

A arquitetura do sistema proposto é constituída por mais de um método de detecção operando em paralelo:

- Um baseado em análise diferencial (centrado em usuários legítimos) que detecta mudanças significativas nos padrões de transações em uma conta; e
- Outro baseado em análise global (centrado em dispositivos de acesso), baseado na observação do comportamento global de usuários, utilizando um conjunto de contadores relacionados a parâmetros globais.

As evidências de fraude determinadas pelas duas abordagens são então combinadas para produzir um escore (ou valor) final de suspeita que pode disparar um alarme dependendo deste valor estar acima ou abaixo de um limiar predeterminado.

A arquitetura geral do detector de fraude proposto está ilustrada na figura 6.

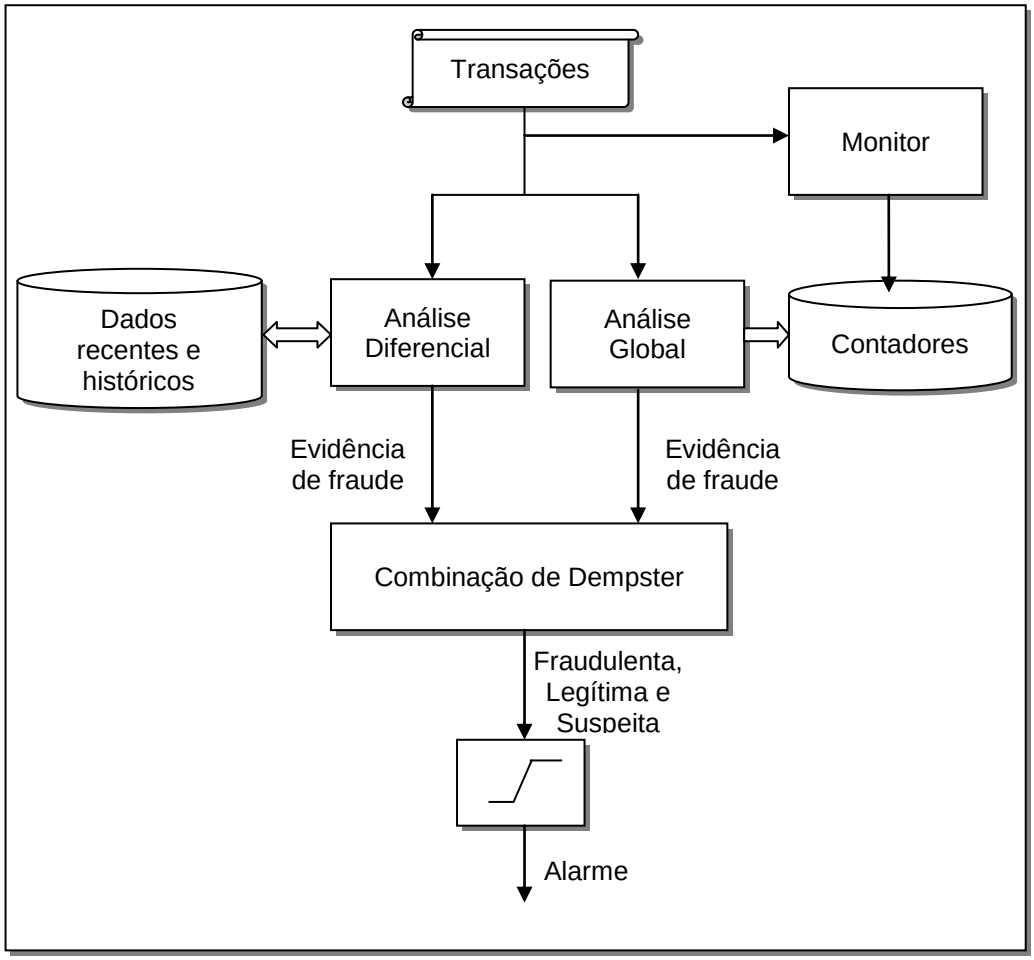

Figura 6 - A arquitetura geral do sistema de detecção proposta

Os detalhes e as principais questões relacionadas com a arquitetura estão descritas nas próximas seções.

4.5. IDENTIFICAÇÃO DOS DISPOSITIVOS DE ACESSO

A técnica de detecção baseada em análise global proposta tem como principal conceito, a noção de identidade do dispositivo de acesso.

Os dispositivos de acesso são utilizados pelo usuário para acessar as contas bancárias via Internet.

No domínio de transações bancárias *online*, onde acessos são feitos através da Internet, a identificação do dispositivo baseado apenas no endereço IP não é tão eficaz, pois ele pode mudar periodicamente.

Na abordagem proposta, a identificação do dispositivo de acesso é feita por um componente que deve ser baixado e instalado no dispositivo do cliente. Um dispositivo sem essa identificação não conseguirá acessar nenhuma conta corrente.

Este componente gera uma impressão digital do dispositivo de acesso e envia esta informação para o *site* do banco como parte dos dados das transações.

A impressão digital é calculada aplicando funções criptográficas nas informações de hardware e de software, como número de série do processador e do sistema operacional, endereço MAC, e outros detalhes de configuração, garantindo que a identificação seja difícil de ser replicada.

Os detalhes de implementação deste componente estão fora do escopo desta tese. Ela apenas assume que a implementação do componente segue três requisitos fundamentais:

- Ele gera uma impressão digital única para cada dispositivo diferente de acesso;
- Ele introduz aleatoriedade durante a geração da impressão digital para dificultar a sua imitação por outros dispositivos;
- Ele informa a nova impressão digital sempre que a configuração do dispositivo tiver alguma alteração, correlacionando-a com a impressão digital anterior antes da alteração da configuração.

Na análise global, estas identidades são usadas junto com um conjunto de contadores para monitorar o perfil de contas correntes diferentes acessadas por cada dispositivo.

4.6. ATRIBUTOS LOCAIS DE UMA TRANSAÇÃO

Para fins de detecção de fraudes, uma transação é caracterizada por um conjunto de atributos (ou características). Estes atributos descrevem as atividades de uma transação numa conta corrente pertencente a um usuário.

Alguns atributos são obtidos diretamente a partir dos dados contidos nas mensagens de transação, como tipo de transação (pagamento, transferência, etc.), o valor envolvido, data e hora, identidade do dispositivo de acesso, entre outros.

A figura 7 abaixo ilustra alguns atributos que podem ser obtidos diretamente dos dados de uma transação bancária *online*

Agência	Conta corrente	Data	Hora	Identidade do dispositivo	Endereço IP	Tipo de transação	Valor

Figura 7 - Exemplo de alguns atributos contidos nas mensagens de uma transação bancária

Eventualmente outros atributos poderiam ser obtidos de outras fontes, como contadores e temporizadores externos.

A escolha destes atributos tem um impacto fundamental na detecção de fraudes, pois é a partir deles que os perfis de atividades individuais são construídos.

4.7. PERFIL DE ATIVIDADE LOCAL

Perfil de atividade local, ou simplesmente perfil de atividade, descreve um aspecto observável do comportamento de um usuário.

Os perfis de atividades são construídos a partir dos atributos de uma transação efetuada em uma cota corrente.

Eles são monitorados para distinguir comportamento legítimo de um comportamento fraudulento.

Exemplos de alguns perfis de atividade individual que são monitorados estão descritas abaixo:

Frequência de transações de pagamento

Este perfil é monitorado para detectar o aumento repentino das transações de pagamento que não é normal ao usuário legítimo.

Erros de senha

Este perfil é monitorado para detectar o número de falhas de senha no instante de *login* que pode caracterizar tentativa de invasão.

Frequência de login

Este perfil é monitorado para detectar tentativas de *login* durante períodos de tempo em que os usuários legítimos não costumam acessar as suas contas.

Cada perfil de atividade individual é caracterizado por meio de uma métrica e de um modelo estatístico.

A métrica estatística representa uma medida quantitativa de uma variável aleatória acumulada durante um período de observações.

As métricas estatísticas são normalmente obtidas por meios de contagens, temporizações e medições de recursos.

Um modelo estatístico determina se uma nova observação é normal ou anormal com relação as observações anteriores.

Dentre os modelos estatísticos, os modelos mais comuns utilizados em detecção são os modelos baseado em um limiar fixo e os modelos baseado em média e desvio padrão.

As seguintes métricas e modelos estatísticos são propostos para os perfis de atividade apresentados acima:

Frequência de transações de pagamento

Métrica: Contagem do número de transações de pagamento efetuadas em uma conta corrente durante um determinado período de tempo;

Modelo estatístico: média e desvio padrão.

Erros de senha

Métrica: Contagem do número de tentativas mal sucedidas de senhas em uma conta corrente durante um período curto de tempo. Requer um contador externo.

Modelo estatístico: limiar fixo.

Frequência de login

Métrica: Contagem do número de *logins* por hora, por período dia, ou por dias da semana;

Modelo estatístico: média e desvio padrão.

Para determinar as frequências de ocorrências de eventos, como frequência de transações de pagamento e frequência de *login*, é necessário manter a data e a hora da transação anterior.

Isso pode ser feito armazenando os atributos mais relevantes da última transação efetuada em cada conta corrente, e recuperando os atributos correspondentes a uma determinada conta quando chegar uma transação nesta conta.

4.8. PERFIL DE COMPORTAMENTO

O perfil de comportamento (PC) de um usuário, ou simplesmente perfil de usuário, é definido como sendo um vetor de variáveis aleatórias, onde cada variável corresponde a um perfil de atividade individual.

A expressão a seguir descreve o comportamento de um usuário.

$$PC = \{ pa_0, pa_1, ..., pa_{n-1} \} \qquad (2)$$

onde, pa_i = perfil de atividade individual correspondente ao elemento i do vetor.

A idéia de construir o perfil de usuário vem do fato de que os comportamentos passados de um usuário podem ser acumulados para construir o perfil normal, isto é, um vetor contendo valores esperados do comportamento normal (ou histórico) do usuário. (HILAS, C. S., SAHALOS, J. N., 2005)

O perfil de comportamento atual de um usuário pode então, ser comparado com o seu perfil normal, para verificar se existe uma consistência com ele, ou se existe um desvio significativo que pode implicar em fraude.

A construção do perfil atual assim como a verificação de sua consistência com o perfil histórico de comportamento de um usuário é feitos dentro do módulo de análise diferencial.

A figura 8 ilustra os principais componentes envolvidos na construção do perfil atual de um usuário.

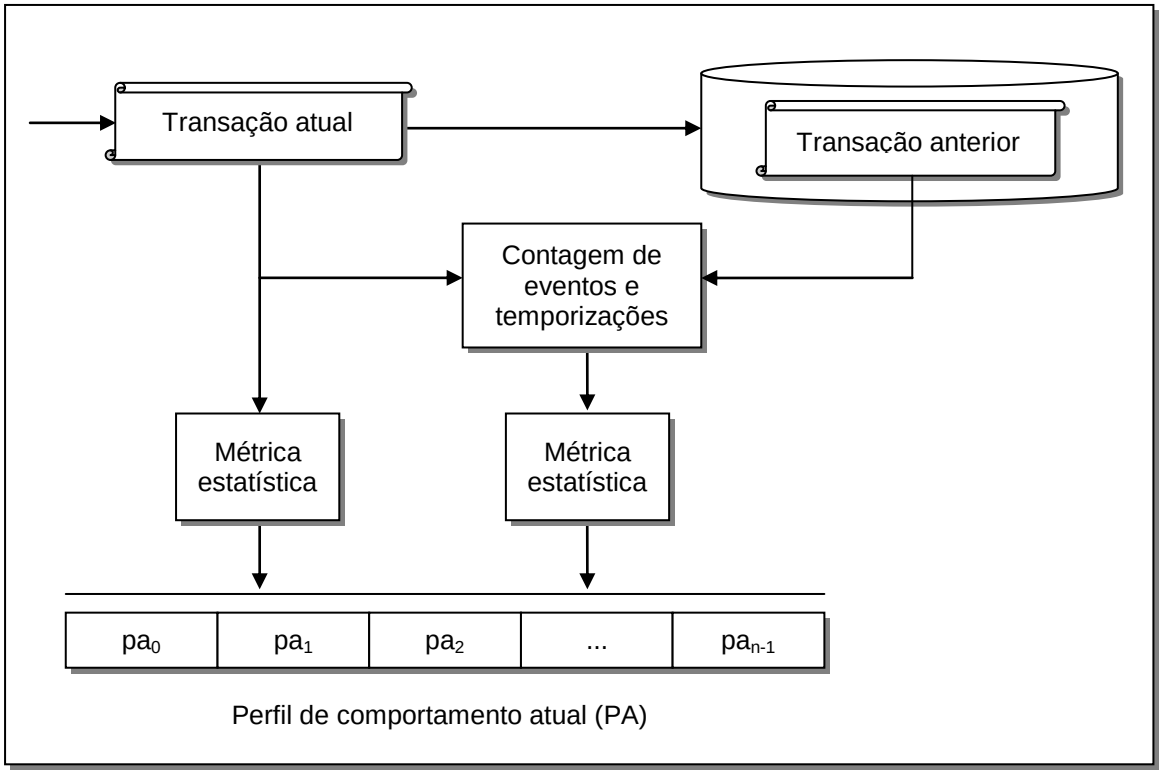

Figura 8 - A construção do perfil de comportamento atual de um usuário

4.9. ANÁLISE DIFERENCIAL

Na abordagem utilizando análise diferencial, o perfil de comportamento correspondente a transação atual é comparada com o perfil de comportamento que caracteriza o padrão de uso normal do usuário legítimo. Se o padrão de utilização atual desviar significativamente do padrão médio de uso do usuário, pode ser uma indicação de uma fraude.

De uma forma geral, para uma análise diferencial é necessário dispor de informações sobre o comportamento histórico das transações do usuário além de uma amostra da atividade mais recente.

(BURGE, P. and SHAWE-TAYLOR, J., 1997; BRITOS, P. et al., 2006; MURAD, 1999; CORTES, C.; PREGIBON, D., 2001) propõem alternativas semelhantes para a análise diferencial em ambientes de comunicação móvel e de intrusão de computadores.

Em uma das propostas, o sistema mantém duas janelas de transações executadas para cada usuário: uma que descreve os padrões de atividades mais recentes e outro que descreve os padrões de atividades passadas.

As duas janelas são concatenadas de forma que, ao chegar uma nova transação, ela é inserida na janela dos mais recentes, e a transação mais antiga desta janela é inserida na janela das transações antigas, sendo a mais antiga desta janela, excluída.

A partir dos conteúdos destas duas janelas são calculados o perfil médio de comportamento mais recente e o perfil de comportamento histórico, e então, eles são comparados.

Essa abordagem, entretanto, não representa uma boa opção para aplicações bancárias em tempo real, pois exige a recuperação dos conteúdos das duas janelas

da base de dados antes do cálculo de desvio de comportamento de cada nova transação que chega.

A abordagem proposta nesta tese utiliza dois perfis:

- Um que descreve o comportamento da transação atual do usuário sendo submetida (PA - Perfil Atual); e
- Outro que descreve o perfil médio histórico do usuário (PH - Perfil Histórico).

O perfil de comportamento da transação atual do usuário (PA - Perfil Atual) é calculado a partir das informações obtidas da transação recém chegada, conforme descrito na seção 4.6 e 4.7.

Este perfil é comparado com o perfil de comportamento médio histórico (PH), que é mantido pré-calculado, para determinar a sua distância probabilística com este último.

Esta distancia é calculada por meio de modelos estatísticos aplicados em cada um dos elementos do vetor,

Cada modelo estatístico gera um valor probabilístico que é encarado como um escore individual de suspeita com relação a um perfil individual de atividade.

A distância total entre os dois perfis é determinada através da união de todas as distâncias individuais.

Uma das alternativas para obter a distância entre os dois perfis é através da soma ponderada das distâncias individuais de cada elemento (FERREIRA et al., 2006), como expressa abaixo:

$$dist\ (PA, PH) = \alpha_0.f_0\ (pa_0, ph_0) + \alpha_1.f_1\ (pa_1, ph_1) + ... + \alpha_{n-1}.f_{n-1}\ (pa_{n-1}, ph_{n-1}) \qquad (3)$$

onde,

α_i são fatores de peso que exprimem a importância de cada perfil de atividade na determinação da distância total;

f_i é a função específica que usa um modelo específico para calcular a distância probabilística entre os perfis de atividade correspondente ao *i-ésimo* elemento, considerando que cada elemento possui uma distribuição de probabilidade diferente.

A *dist (PA, PH)* determina o nível de evidência (suspeita) de que uma transação seja fraudulenta baseando-se na extensão do desvio em relação ao perfil histórico.

O sistema proposto nesta tese usa outra abordagem para determinar a distância total entre os dois perfis. A abordagem utilizada é a própria teoria de Dempster-Shafer (DS) para combinar as distâncias individuais e determinar a distância total entre os dois perfis, PA e PH.

A razão da escolha desta abordagem em vez da soma ponderada vem da dificuldade de determinar os valores mais adequados dos pesos.

Dependendo dos pesos escolhidos, uma evidência individual pode não influenciar significativamente na soma final em virtude do peso associado a ele ser pequeno em relação aos pesos associados aos demais elementos.

A teoria de Dempster-Shafer tem sido usada com bons resultados na fusão de evidências determinadas individualmente. (Singh R. ET AL., 2006; CHEN, Q.; AICKELIN U., 2006).

A figura 9 ilustra os elementos envolvidos no cálculo da distância entre os perfis PA e PH usando regra de Combinação de Dempster.

Nesta figura, os blocos denominados f_0, f_1, ..., f_{n-1} correspondem aos métodos estatísticos que são aplicados a cada par de perfil de atividade individual dos perfis de comportamento, PA e PH. As distâncias probabilísticas, m_0, m_1, ..., m_{n-1},

determinados por cada um dos métodos estatísticos são combinados pela regra de Dempster para se obter a distância total, dist (PA, PH).

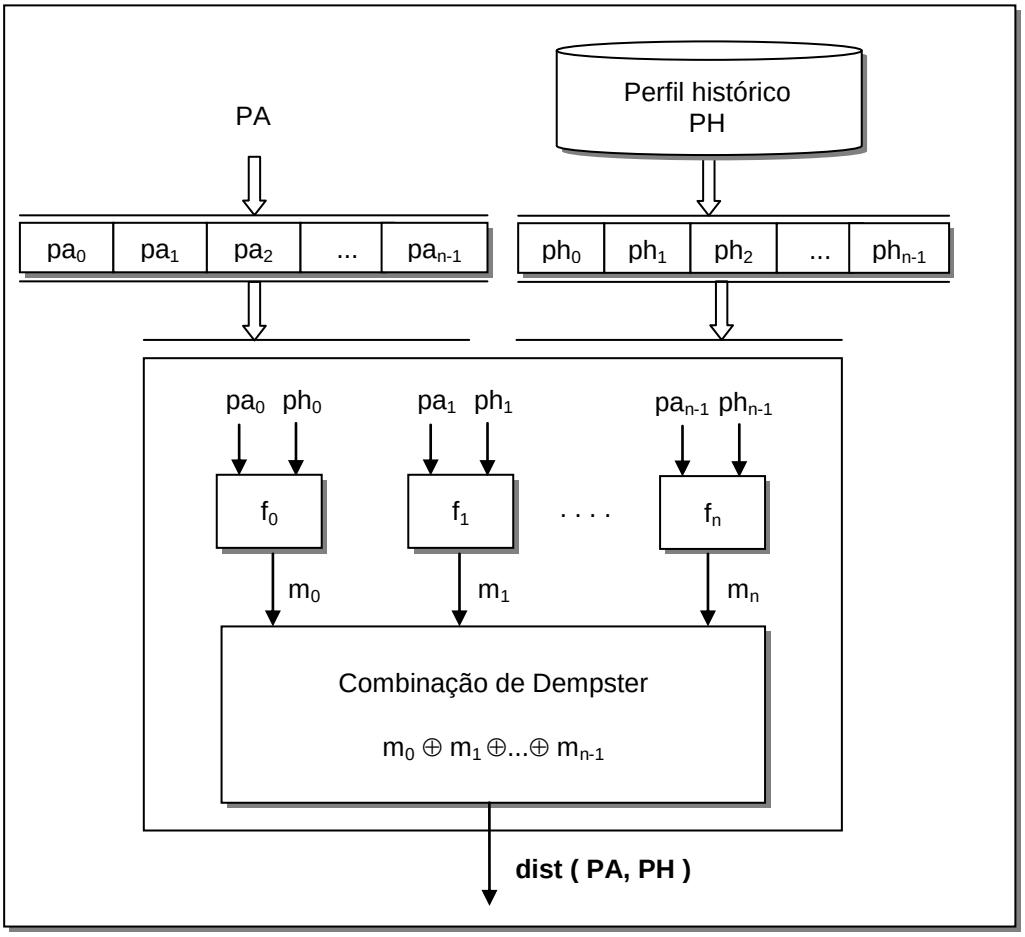

Figura 9 - Cálculo da distância probabilística entre PA e PH usando teoria de Dempster-Shafer

O valor determinado pelo *dist (PA, PH)* é combinado com as evidências determinadas pelos demais módulos de detecção para produzir um escore total de suspeita no modulo de combinação de Dempster-Shafer, descrita na seção 4.10.

Se o escore gerado não determinar uma fraude, o perfil histórico (PH) é atualizado com o perfil da transação recente (PA).

4.9.1. Distâncias individuais entre os elementos de PA e PH

Um modelo estatístico específico deve ser usado para calcular a distância probabilística entre cada um dos perfis de atividade individual de PA e PH, considerando que cada um possui uma distribuição de probabilidade diferente.

Os modelos estatísticos geram um valor probabilístico (entre 0 e 1) que pode ser utilizado como um escore individual de suspeita com relação a um perfil individual de atividade.

Entretanto, sob ponto de vista estatístico, muitas vezes as variáveis possuem uma distribuição de probabilidade que pode ser descrita por uma distribuição normal, se a média μ e o desvio padrão σ forem especificados. (WALPOLE R. et al., 2009, apud FERREIRA, P. et al., 2006)

A distribuição normal nos dá uma aproximação razoável para muitas variáveis que ocorrem nas situações do mundo real, em especial na área de detecção de fraude. Sendo assim, a função de distribuição normal pode ser adaptada para medir a distância entre dois elementos cujo modelo estatístico utilizado seja baseado em média e desvio padrão. (FERREIRA, P. et al., 2006)

Se uma variável aleatória X tem uma distribuição normal com média μ e variância σ^2, a função *z-escore* pode fornecer a probabilidade de X ter um valor igual a *x,* isto é,

$$P(X = x) = P\left(\frac{X-\mu}{\sigma} = \frac{x-\mu}{\sigma}\right) = P(Z = z) \qquad (4)$$

Z é conhecido como variável aleatória padrão normal e $z = (x-\mu)/\sigma$ é o z-escore correspondente a x.

P(Z=z) pode ser usada na análise diferencial como uma medida de quão longe ou quão perto o perfil de atividade atual está em relação ao perfil de atividade normal de um usuário, em termos de probabilidade.

A função *z–escore* fornece um valor de probabilidade menor a medida que o valor de X tende a μ, atingindo o mínimo quando X = μ, isto é, 0,5.

A distância probabilística pode ser dada através do processo de normalização para dar resultado entre 0 e 1, como segue:

$$Distância = 2 * (P(Z = z) - 0{,}5) \quad (5)$$

Existem outros modelos estatísticos que podem ser utilizados para a determinação de distâncias entre duas variáveis aleatórias dependendo da distribuição de probabilidade adotada. (WALPOLE R. et al., 2009, apud FERREIRA, P. et al., 2006).

4.9.2. Atualização do perfil histórico

O perfil de comportamento histórico (PH) do usuário é atualizado para incorporar o perfil das transações mais recentes.

A idéia por trás disso é que o perfil de comportamento histórico do usuário se adapte as variações de comportamento desde que ela seja lenta.

A atualização é feita através de uma função de média ponderada que determina a taxa com que os valores antigos são descartados em cada reavaliação do perfil histórico.

O valor de PH é atualizado com o valor de PA, decaindo os valores de cada elemento de PH por um fator α *(0 <* $\alpha \leq$ *1),* através da expressão:.

$$ph_i = \alpha.\ ph_i + (1-\alpha).pa_i \qquad (6)$$

onde,

ph_i *e* pa_i representam o *i-ésimo* elemento de PH e PA respectivamente; e o valor de α determina a taxa com que os valores antigos se tornam irrelevantes, sendo determinado pelos especialistas.

A figura 10 ilustra em blocos a atualização de PH com os valores de PA

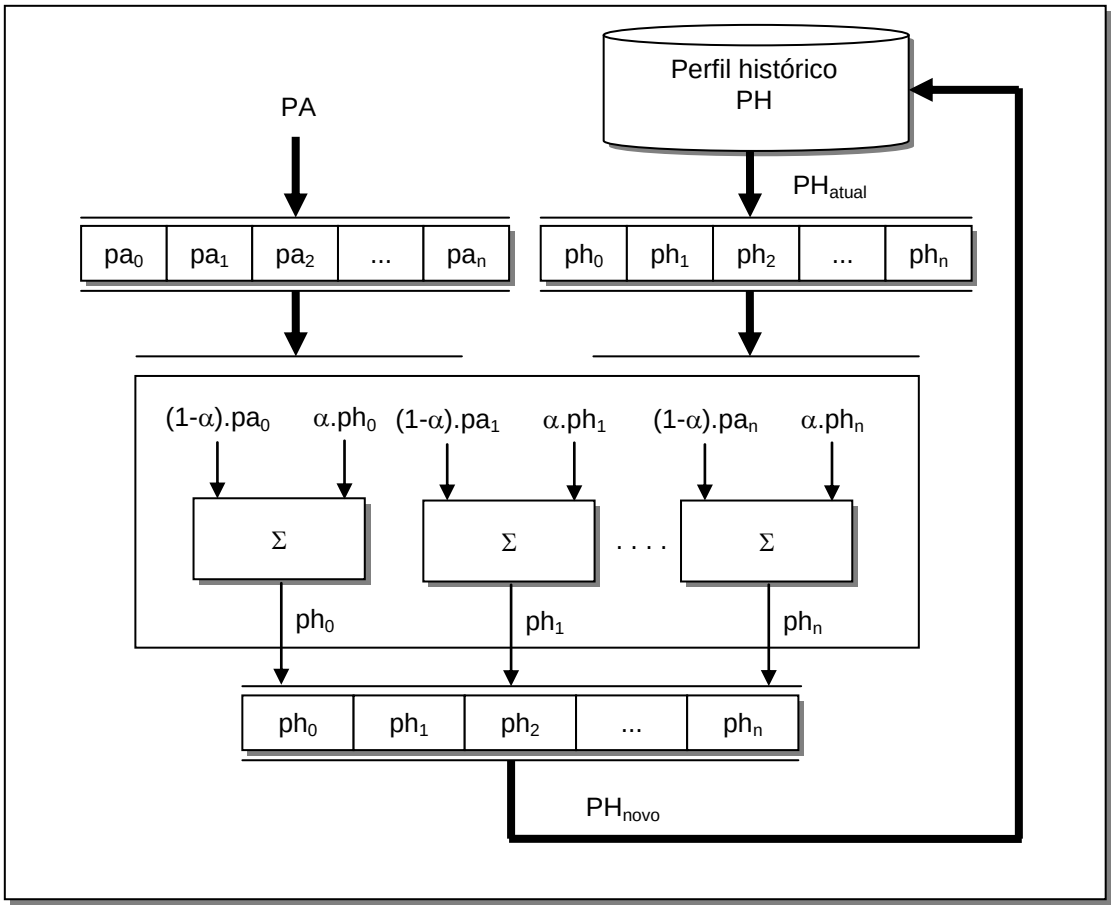

Figura 10 - Atualização de PH

4.9.3. Inicialização do perfil histórico

O valor inicial do perfil histórico representa um problema em virtude da falta de transações para estabelecer o histórico de comportamento do usuário podendo como consequência gerar um número excessivo de falsos alarmes.

(FERREIRA, P. et al., 2006) apresenta uma abordagem para resolver este problema em que abre uma janela pequena de tempo no início das atividades sem detectar fraudes. O problema desta solução consiste em não detectar um fraudador dentro dessa janela de tempo, pois seria considerado um usuário novo.

O que se deseja é uma solução que minimize a ocorrência de falsos alarmes sem ignorar intrusões reais.

A abordagem para minimizar falsos alarmes é usar modelos estatísticos apropriados ou então, fazer uma escolha apropriada de perfil inicial.

Um exemplo do primeiro caso seria o modelo de média e desvio padrão, onde o intervalo de confiança utilizado seria maior no início, e ir diminuindo a medida que são coletados mais dados sobre o comportamento do usuário.

Isso pode reduzir o número de falsos alarmes causados por um perfil individual de usuário, mas não protege o sistema contra usuários novos cujo comportamento é duvidoso, ou contra usuários que tem um comportamento não normal desde o início.

.

Cortes e Pregibon (CORTES, C.; PREGIBON, D., 2001) sugerem a criação de classes de equivalência de contas, associando um perfil inicial para cada classe.

Quando uma nova conta é criada, os atributos das primeiras transações são utilizados para mapear o recém chegado numa classe de equivalência.

Cortes e Pregibon observam que a inicialização não precisa ser exata, pois considerando que um perfil inicial adequado tenha sido estabelecido, logo o perfil histórico será atualizado com os dados individuais reais.

O sistema proposto neste trabalho sugere uma abordagem semelhante ao do Ferreira et al., em que as primeiras β (valor arbitrariamente escolhido) transações são efetuadas sem detectar fraude.

A razão desta sugestão é a sua simplicidade na implementação e também pelo fato de que o problema de inicialização ocorre apenas no módulo da análise diferencial e não no módulo de análise global. Outro fato considerado foi que o perfil histórico é logo atualizado com os dados reais.

Considerando $\beta = 2$, as determinações dos valores de PH nos instantes t=0 e t=1 são apresentadas abaixo:

$$PH_{t=0} = PA_{t=0}$$

$$PH_{t=1} = \alpha . PH_{t=0} + (1-\alpha). PA_{t=1}$$

4.10. ATRIBUTOS GLOBAIS E O COMPORTAMENTO GLOBAL DE USUÁRIOS

Através de uma análise empírica efetuada em um conjunto de transações reais verificou-se que a observação do comportamento global de usuários, e seus dispositivos de acesso, desempenham um papel importante no sistema de detecção de fraudes proposto nesta tese.

Um exemplo de comportamento global que pode evidenciar uma fraude é o número grande de contas diferentes acessadas por um mesmo dispositivo.

Embora outros atributos globais possam ser monitorados, como ocorrência de falhas de *login* em diversas contas usando as mesmas senhas, será considerada nesta tese apenas o número de contas diferentes acessadas por cada dispositivo de acesso como atributo global.

Na figura 6, o monitor e o conjunto de contadores são usados para acompanhar o comportamento global de usuários.

4.11. ANÁLISE GLOBAL

O objetivo do módulo de análise global é fortalecer ou enfraquecer as evidências de fraude determinadas pelo módulo de análise diferencial.

Esta análise é feita pela avaliação de evidências de fraude observando o comportamento global dos dispositivos de acesso utilizados para acessar as contas correntes.

A evidência de fraude, dada por um valor probabilístico, é determinada por meio de três listas e por uma função exponencial decrescente

As três listas estão relacionadas a seguir:

- ***Lista Negra***, que contem a identidade dos dispositivos associados a transações que já foram classificadas como fraudulentas.

- ***Lista Branca***, que contém as identidades dos dispositivos, assim como dos números das contas acessadas por eles, associadas a transações classificadas como legítimas.

- ***Lista de Suspeitos*** que contem as identidades dos dispositivos cujas transações ainda não foram classificadas.

O assinalamento dos dispositivos em uma destas listas e a determinação de sua probabilidade de fraude são conduzidos por meio de regras descritas a seguir:

Para cada transação que chega,

- Se o dispositivo estiver na Lista Negra, a probabilidade de fraude será assinalada com um, significando que a transação é fraudulenta com alto grau de evidência;
- Se o dispositivo e o número da conta acessada por este dispositivo estiverem na Lista Branca, a probabilidade de fraude é assinalada com zero denotando que a transação é legítima com alto nível de evidência. Note que a identidade do dispositivo pode estar associada com um ou mais contas na Lista Branca. Este é o caso em que um único usuário tem acesso a diversas contas diferentes;
- Se o dispositivo não estiver em nenhuma destas listas, ele e o número da conta acessada são incluídos na Lista de Suspeitos. Enquanto ele estiver nesta lista, a probabilidade de fraude desta transação é determinada por uma função exponencial decrescente descrita na próxima seção.

4.11.1. A lista de suspeitos e a função exponencial decrescente

Se o dispositivo que está enviando uma transação estiver inserido na Lista de Suspeitos, ele permanecerá ali até que a transação seja classificada explicitamente como fraudulenta ou legítima, quando a identidade do dispositivo associado e o número da conta associada serão inseridos na Lista Negra ou Branca, respectivamente.

A idéia por traz desta regra vem do fato de que uma transação só pode ser assegurada como fraudulenta pelo próprio cliente.

Se nenhuma fraude for reportada até o final de um período de tempo prefixado, nada pode ser dito a respeito da legitimidade deste dispositivo.

Neste caso, o dispositivo será movido para a Lista Branca, pois é mais provável ser legítimo baseado nas análises efetuadas em transações no mundo real.

Entretanto, um *sinal* será ativado indicando que este dispositivo foi movido para a Lista Branca no fim do período predefinido e não classificado explicitamente como legítimo.

Este *sinal* pode ser usado pelos analistas se uma possível fraude for detectada a partir deste dispositivo mais tarde.

Como o dispositivo foi movido para a Lista Branca, as próximas transações a partir deste dispositivo serão consideradas como legítimas por este módulo.

O tempo decorrido desde a ocorrência de uma fraude e a sua detecção pelo cliente pode levar mais de um mês e de acordo com as informações colhidas de relatórios de fraudes do mundo real, foi observado que em alguns casos a notificação de fraude pelo cliente levou até dois meses para ser reportada.

A provável razão desta demora é o fato de muitas transações fraudulentas envolverem quantias muito pequenas.

Quando um dispositivo é incluído na Lista de Suspeitos, um valor inicial é assinalado a probabilidade de fraude para este dispositivo.

Este valor varia de acordo com uma função exponencial decrescente que depende do número de contas diferentes que foram acessadas por este dispositivo cujas transações ainda não foram classificadas. (KOVACH, S.; RUGGIERO, W. V., 2011)

Se uma fraude em qualquer uma destas contas for reportada por um cliente, a identidade do dispositivo associado será movida para a Lista Negra.

A função exponencial decrescente foi escolhida empiricamente devido ao fato de que a maioria das fraudes são reportadas assim que elas são cometidas e muito poucas no fim de um período de tempo, como por exemplo, um ou dois meses depois.

Em outras palavras, a probabilidade de ser uma fraude é maior no início de uma transação, decaindo rapidamente ao longo do tempo.

A função exponencial decrescente é expressa como segue

$$P(t) = P_{max} \cdot e^{-\lambda t} \qquad (7)$$

onde,

P_{max} é o valor máximo de probabilidade assinalado ao dispositivo quando for incluído na lista.

Este valor depende do número de contas diferentes (N) acessadas pelo dispositivo, pois a probabilidade de ser uma fraude cresce com este número.

Como exemplo da escolha de P_{max}, considere que N maiores de 10 seja uma situação de fraude, isto é, $N_{max} = 10$.

Neste caso, o valor de $P_{max} = 1,0$ para $N \geq 10$.

Quando a probabilidade $P(t)$ atingir 1,0 o dispositivo é incluído automaticamente na Lista Negra e a transação é considerada fraudulenta.

Para valores de probabilidade abaixo de 100%, serão decididos pelo módulo combinador de Dempster-Shafer junto com outras evidências.

λ é calculada de tal forma que no final do período (t_{end}), o valor da probabilidade tenha alcançado um valor arbitrariamente pequeno.

Assumindo t_{end} = 60 dias e $P(t_{end})$ = 0.01, temos

$$\lambda = -(1/60).ln(0.01/P_{max}) \quad (8)$$

A Figura 11 ilustra algumas curvas da função exponencial decrescente para valores de N até 8.

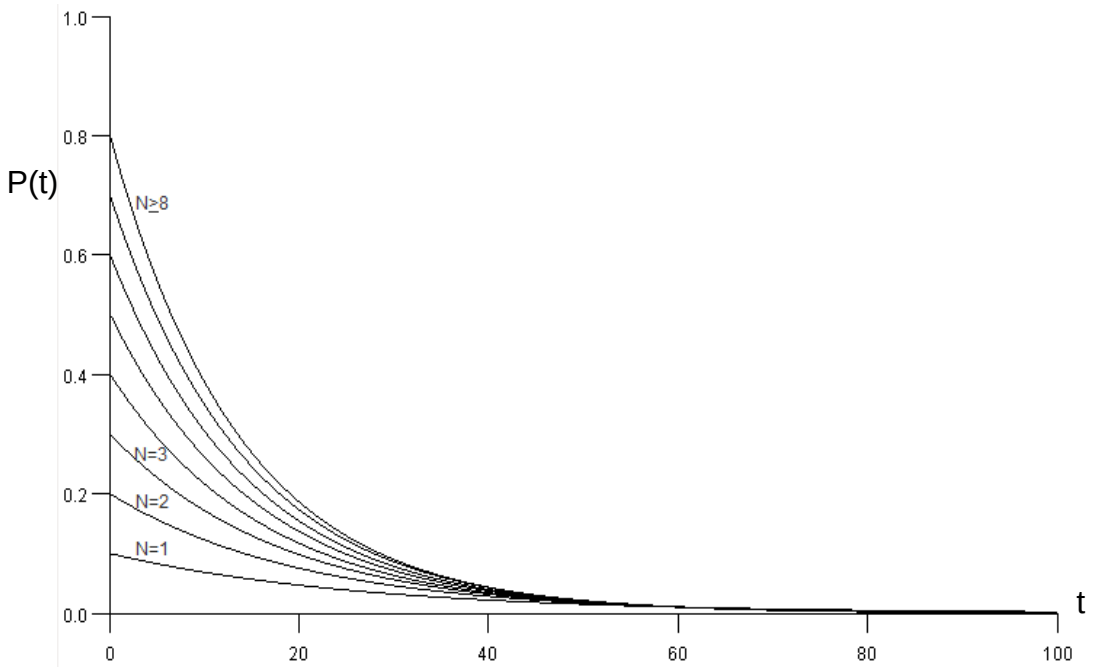

Figura 11 - Curvas exponencialmente decrescentes

A linha tracejada da figura 12 ilustra um exemplo dos valores de probabilidade assinalados a um dispositivo variando com o tempo. Pode-se notar que quando ocorre um novo acesso a uma conta corrente (diferente), o valor de P_{max} salta par um novo patamar decaindo, então, exponencialmente a partir deste ponto.

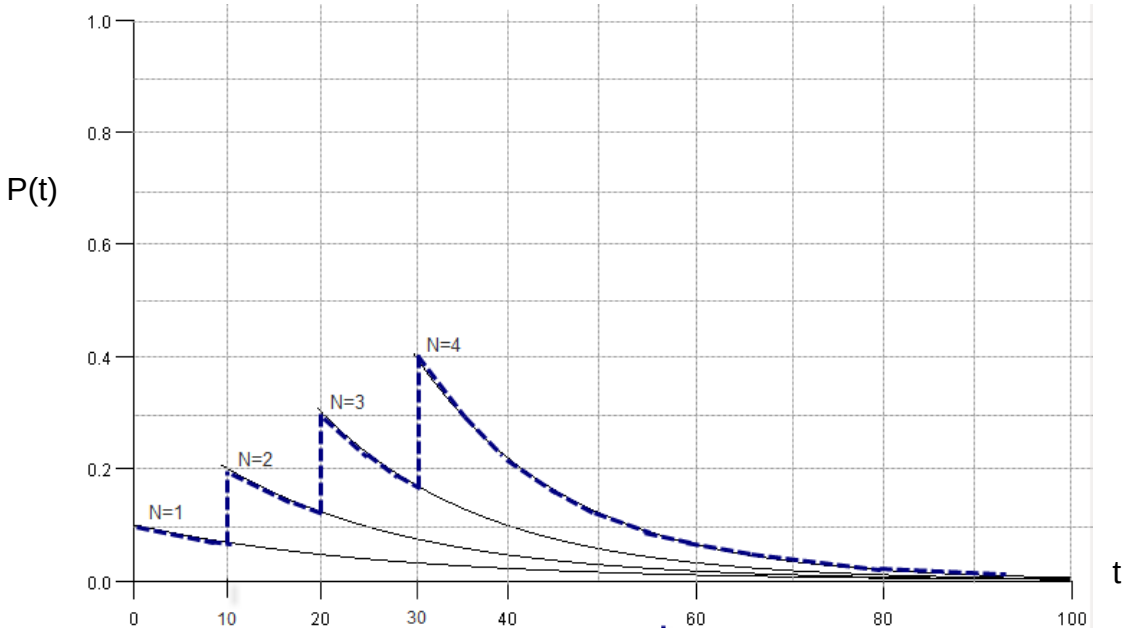

Figura 12 - A probabilidade de fraude pela função exponencial decrescente

A figura 13 ilustra os componentes básicos do módulo de análise global e os seus relacionamentos.

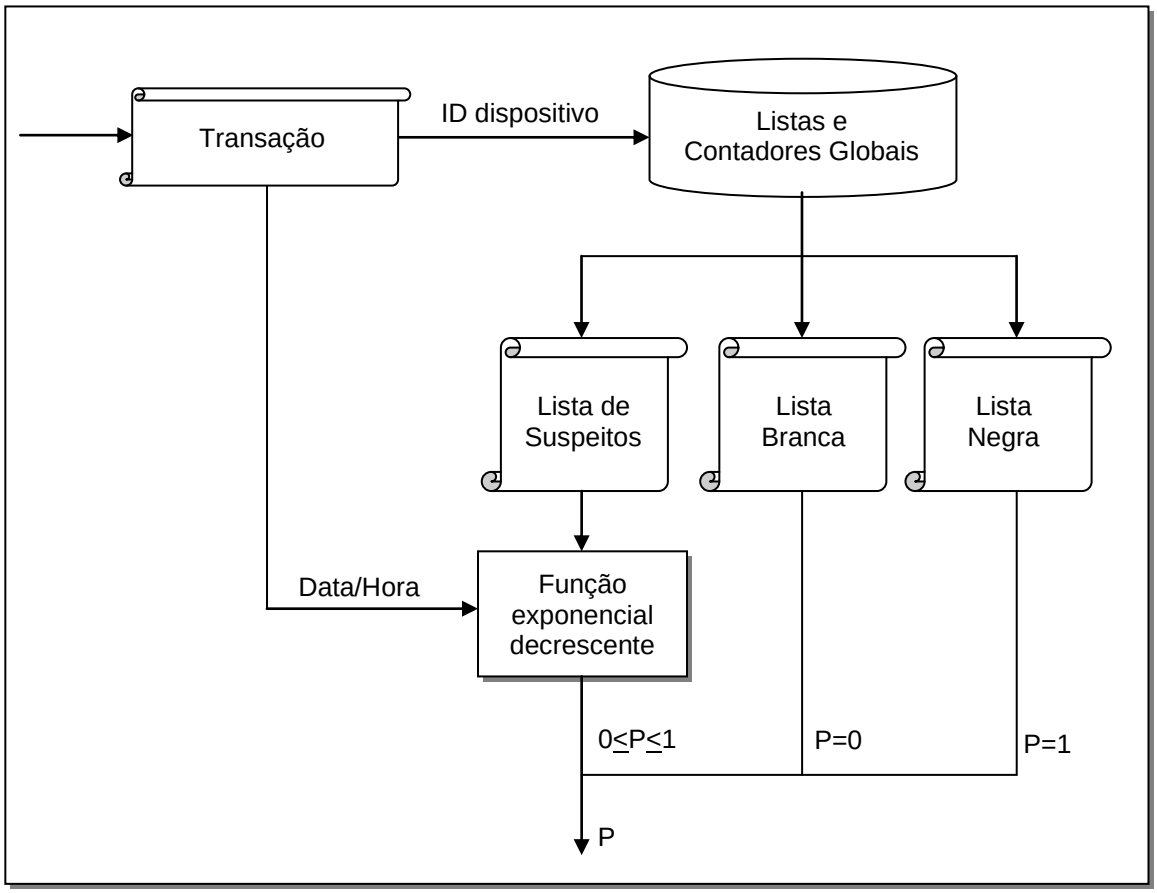

Figura 13 - Componentes básicos da análise global

4.12. COMBINAÇÃO DE EVIDÊNCIAS

A combinação dos resultados de vários detectores independentes fornece um desempenho melhor do que o resultado de um único detector. (SINGH R. et al., 2006; CHEN, Q.; AICKELIN U., 2006)

No sistema proposto, as evidências de fraude determinadas pelos módulos individuais de detecção de fraudes são combinadas utilizando a teoria matemática de Dempster-Shafer (DS) para dar o escore final sobre a suspeita de fraude da transação recém chegada.

O primeiro passo para utilizar a regra de combinação de DS é mapear as evidências de fraude geradas pelos módulos de detecção em um número probabilístico, m(*f*), conhecido também como bpa (*basic probability assignment*). Vide seção 2.5 para maiores explicações.

No sistema proposto, as evidências das fraudes determinadas pelos dois módulos já são fornecidos em valores probabilísticos, não existindo, portanto, a necessidade de fazer nenhum mapeamento.

A regra de combinação de Dempster fornece uma função para calcular o escore total de duas evidências. Dadas as massas de duas evidências de fraude $m_1(f)$ e $m_2(f)$, elas podem ser combinadas em uma terceira massa $m_3(f)$ pela seguinte expressão:

$$m_3(f) = m_1(f) \oplus m_2(f) = \frac{\sum_{x \cap y = f} m_1(x).m_2(y)}{1-K} \qquad (9)$$

onde, $K = \sum_{x \cap y = \phi} m_1(x).m_2(y)$

De acordo com a seção 2.5, o quadro de discernimento Θ no domínio de detecção de fraudes é constituído de dois valores mutuamente exclusivos, isto é,

$$\Theta = \{ f, -f \}$$

onde, *f = fraude; e -f = legítimo*.

O conjunto de todas as hipóteses possíveis de Θ corresponde a todos os subconjuntos de Θ incluindo ele mesmo. Este conjunto, denotado por 2^{Θ}é constituído de três possíveis hipóteses, $\{ f \}$, $\{ -f \}$ e $\Theta = \{ f, -f \}$ (representando a incerteza).

Supondo que o módulo tem uma evidência de fraude com probabilidade α, as massas a serem assinaladas são:

$m(f) = \alpha$
$m(-f) = 0$
$m(\Theta) = 1 - \alpha$

Com base nisso, resultado da combinação de Dempster, $m_3(f) = m_1(f) \oplus m_2(f)$, serão reduzidas a

$$m_3(f) = m_1(f) . m_2(f) + m_1(f) . m_2(\Theta) + m_1(\Theta) . m_2(f)$$

A título de exemplo, suponha que um detector tenha uma evidência de fraude com $m_1(f) = 0,8$ e um outro detector uma evidência de fraude com $m_2(f) = 0,6$.

Neste caso,

$m_1(-f) = 0$ e $m_1(\Theta) = 0,2$; e

$m_2(-f) = 0$ e $m_2(\Theta) = 0,4$; e

$$m_3(f) = 0,8 . 0,6 + 0,8 . 0,4 + 0,2 . 0,6 = 0,48 + 0,32 + 0,12 = 0,92$$

O resultado final é obtido através da aplicação de um limiar *t* ao $m_3(f)$,

$$\text{Resultado} = \begin{cases} Fraude, & se\ m_3(f) \geq t \\ Legítimo, & caso\ contrário \end{cases}$$

A figura 14 ilustra em blocos o módulo de combinação do sistema proposto utilizando a teoria matemática de Dempster-Shafer.

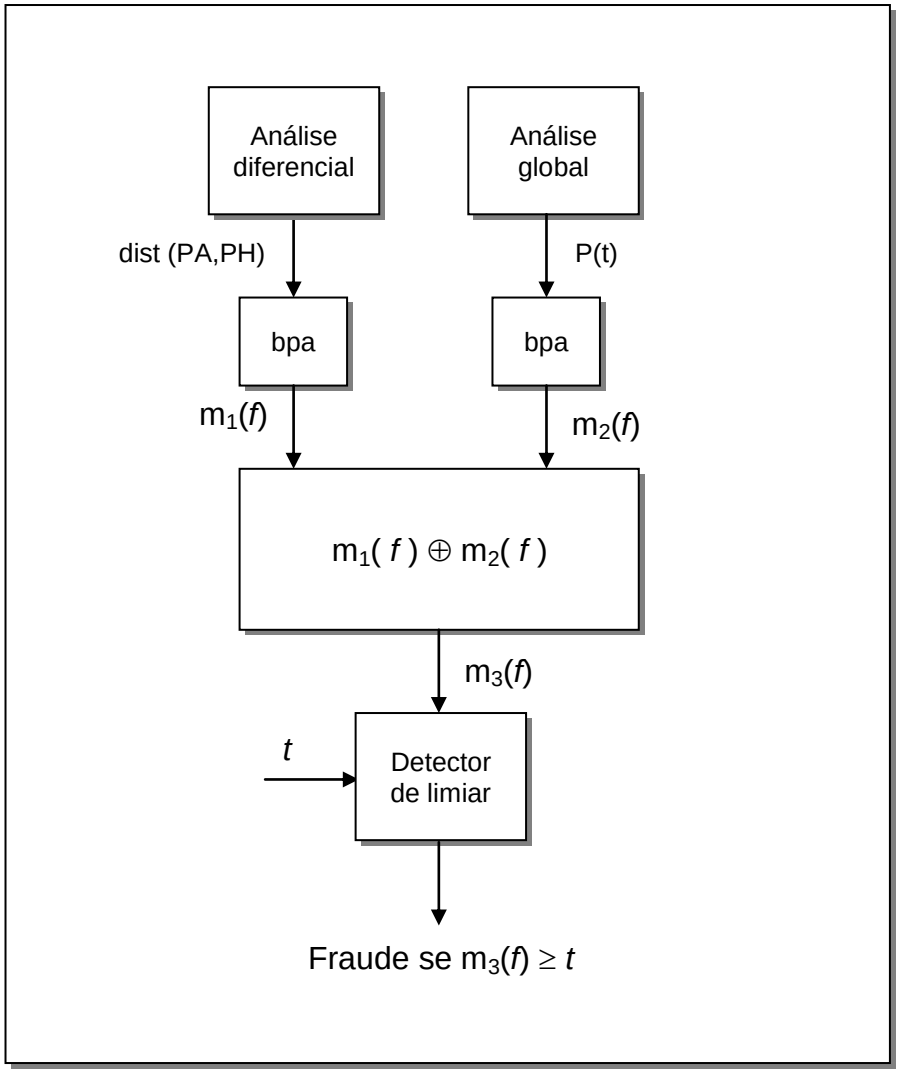

Figura 14 - Módulo de combinação de evidências

Duas observações devem ser feitas com relação ao módulo de combinação descrito nesta seção.

A primeira observação é a de que a função de combinação de Dempster é apresentada para duas massas de duas evidências apenas.

Caso tenha mais evidências para serem combinadas, sugere-se que a combinação seja feita de duas em duas evidências por vez. Isto é, combinam-se as duas primeiras evidências. O resultado desta combinação é combinado com a terceira evidencia. O resultado desta é combinado com a quarta evidência, e assim por diante.

A segunda observação vem do fato de que não houve nenhum mapeamento entre as evidências geradas pelos dois módulos de detecção em virtude deles fornecerem as evidências já em valores probabilísticos.

Caso um novo módulo de detecção seja adicionado ao sistema existente, a melhor situação seria que ele já fornecesse a evidência de fraude em valores probabilísticos.

Entretanto, caso isso não ocorra, haverá a necessidade de fazer um mapeamento em valores probabilísticos, isto é, em massas.

Existem várias abordagens utilizadas para efetuar mapeamento como em (CHEN, Q.; AICKELIN U., 2006) .

4.13. INTEGRAÇÃO COM OUTROS MODELOS DE DETECÇÃO

Esta seção aborda a integração de outros métodos de detecção no sistema proposto e a sua importância.

Em qualquer domínio de detecção de fraudes, a detecção baseada em apenas uma abordagem pode não ter um desempenho adequado, pois para cada tipo de fraude existe um método de detecção mais apropriado.

O desempenho aqui considerado se refere ao melhor equilíbrio entre a detecção correta e falsos alarmes, isto é, obter maior taxa de acertos com o menor número possível de falsos alarmes.

Dentre as abordagens adotadas para se atingir este objetivo, muitos utilizam a integração de vários métodos de classificação e depois combinando os resultados individuais.

Diversos exemplos descritos na literatura mostram que a combinação de métodos melhora o desempenho do sistema com relação a taxa de verdadeiros positivos versus taxa de falsos positivos.

A inclusão de outros métodos no sistema proposto nesta tese deve também melhorar o seu desempenho. Como exemplo de um método que pode ser agregado para complementar a análise diferencial e análise global é a abordagem baseada em regras. Nesta abordagem, regras baseadas em fraudes já ocorridas poderiam ser criadas automaticamente, mantendo assim uma base histórica de perfis de fraude.

4.14. CONSIDERAÇÕES FINAIS

Esta seção apresentou uma proposta de arquitetura para detecção de fraudes em transações financeiras via Internet em tempo real.

Quanto ao desempenho do sistema, que será avaliado no próximo capítulo, a abordagem utilizada para alcançá-lo foi a combinação de dois métodos diferentes e independentes de detecção, com a capacidade de adicionar outros métodos e ainda produzindo uma única métrica consolidada para detecção de fraudes.

Entretanto, apesar da combinação de vários métodos aumentarem o desempenho do sistema, o desafio está em como processar todos os métodos dentro do limite de tempo disponível para determinar o escore final antes de completar a transação, que é um dos requisitos do sistema proposto.

Embora este requisito possa ser alcançado através do aumento do poder de processamento, o que se procura é o máximo de desempenho em termos de detecção com relação a um determinado desempenho do sistema de processamento, o que limita as alternativas para serem usadas no sistema.

A arquitetura proposta na tese atinge este objetivo utilizando apenas dois módulos de detecção descritos neste capítulo.

5. VALIDAÇÃO DA PROPOSTA

5.1. CONSIDERAÇÕES INICIAIS

Neste capítulo é feita uma avaliação do sistema de detecção de fraudes proposto no capítulo quatro com intuito de validar a proposta desta tese.

O objetivo desta avaliação é demonstrar que a abordagem usando análise global como complemento aos métodos tradicionais para detecção de fraudes em transações financeiras via Internet, fornece um grande poder de discriminação de uma forma simples e eficiente, sem ter que recorrer a novos algoritmos ou métodos estatísticos complexos.

Esta avaliação foi efetuada através de simulações da arquitetura proposta junto com uma base de dados contendo registros de transações bancárias correspondentes a perfis reais de uso.

Essa base de dados consiste de uma amostra das transações *online* acorridas durante um período de quatro meses. Ela é composta de 44.147 transações de diversos tipos, como, de *login, fim de sessão, pagamentos,* entre outros, que foram executadas em 444 contas correntes diferentes.

Cada uma das transações foi identificada a priori como sendo *Legítimas* ou *Fraudulentas* servindo de base para a comparação dos resultados da simulação.

5.2. METODOLOGIA PARA A VALIDAÇÃO DA PROPOSTA

A metodologia utilizada para validar a proposta da arquitetura segue basicamente os principais itens que sintetizam o objetivo do trabalho:

- A determinação de atributos locais e globais mais adequados para definir o perfil de comportamento de um usuário e de seus dispositivos de acesso a partir dos registros das transações contidos na base de dados;
- A escolha dos modelos estatísticos mais adequados para caracterizar evidências de fraude com base nos atributos escolhidos;
- Aplicação dos modelos estatísticos nas transações registradas da base de dados e avaliar o desempenho da arquitetura proposta.

5.3. DADOS PARA AVALIAÇÃO DE DESEMPENHO

A figura 15, a seguir, ilustra os campos de uma transação na sua forma original, isto é, como registrada na base de dados com um exemplo do conteúdo de cada campo.

IDSessao	AGN	CTA	DATA	HORA	EVT	IDMQ
737956798465923	AA	AA	03/11/2010	09:01:03	LOGIN	X+tnQoiyram67grpBgKZNpoa/$auetbfTUhR/nPPh

ENDIP	iBrwsr	iSistOper	TPDISP	VALOR	FRAUDE
10.0.0.1	IE 6 (XP SP2 ou Server 2003)	Windows XP	Tabela	NULL	NAO

Figura 15 - Campos de uma transação

Os significados dos campos estão descritos a seguir:

- *IDSessao* - contém o identificador da sessão onde a transação em questão está sendo executada;
- *AGN e CTA* - indicam respectivamente, a agência e a conta corrente sendo acessada;
- *Data e Hora* - indicam a data e a hora, respectivamente, da execução da transação corrente;

- *EVT* - representa o tipo de transação sendo executada;
- *IDMQ* - indica a identidade do dispositivo de acesso que está executando a transação atual. Se o dispositivo não possuir identidade, este campo é deixado em branco;
- *ENDIP* - indica o endereço IP do dispositivo;
- *iBrwsr e iSistOper* - indicam respectivamente, o navegador e o sistema operacional utilizado no dispositivo de acesso;
- *TPDISP* - indica o tipo de autenticação (Tabela de códigos ou OTP) sendo feita;
- *Valor* - se a transação é uma transação que envolve pagamento ou transferência de dinheiro, este campo contém o valor da transação;
- *FRAUDE* - campo inserido pelo operador na base de dados para indicar que esta transação foi uma fraude ou não. Este campo só é sinalizado nas transações de pagamento

A primeira etapa da metodologia para a validação da proposta foi a identificação dos parâmetros necessários para caracterizar os atributos e os correspondentes perfis de atividades definidos para detectar fraudes. Os seguintes parâmetros foram identificados:

- *PseudoID:* Um “pseudo identificador” gerado através da concatenação dos conteúdos dos campos ENDIP, iBrwsr e iSistOp, para os dispositivos sem identificação. A necessidade deste identificador vem do fato de que muitas transações registradas na base de dados, em especial as fraudulentas, são provenientes de dispositivos sem identificação. Embora, não garanta uma identificação única em todas as situações como aquela gerada pelo componente instalado no dispositivo de acesso, o *PseudoID* mostrou ser uma boa alternativa para identificar um dispositivo de acesso. Para ocorrer uma colisão de identidade, seria necessário que o fraudador utilize a mesma versão do sistema operacional, a mesma versão do navegador, e ter o mesmo endereço IP, e a probabilidade desta ocorrência é bem baixa;

- *NumCC:* Número de contas correntes diferentes acessadas pelo dispositivo que está executando a transação. Este parâmetro corresponde a métrica utilizada na análise global; e

- *Pgm/Sessao:* Número total de pagamentos efetuados na sessão sendo executada. Esse parâmetro corresponde a métrica utilizada pela análise diferencial para caracterizar o perfil de atividade local de um usuário.

Foram criados três campos adicionais no registro original da transação para acomodar os parâmetros acima. Os conteúdos dos campos *PseudoID e Pgm/Sessão* são calculados durante a execução de cada transação, ao passo que o valor de NumCC é obtido através do acesso a base de dados.

Além da criação dos três campos mencionados acima, os registros da base de dados sofreram uma série de filtragens e simplificações com objetivo de facilitar a aplicação dos métodos estatísticos, assim como, para facilitar as eventuais análises visuais. Este processamento consistiu basicamente em:

- Substituir o conteúdo do campo *IDSessao* por uma sequência de números começando por um, para facilitar a análise visual;

- Substituir a identidade de máquina (*IDMQ*) e *PseudoID* por uma sequência de números inteiros (*IDMQsimpl*), também para facilitar a análise visual;

- Manter eventos apenas de *LOGIN*, pois os demais eventos são irrelevantes para a metodologia de avaliação utilizada;

- Eliminar contas correntes sem nenhuma transação de pagamento, pois elas não são necessárias para a metodologia de avaliação utilizada.

A figura 16 ilustra os campos de uma transação após filtragens e simplificações dos registros contidos na base de dados:

FRAUDE	AGN/CC	IDSessao	IDMQsimpl	NumCC	DATA_HORA	Pgm/Sessao
	AB	10	132	1	05/11/2010 02:59	1

Figura 16 - Registro simplificado de uma transação

A base de dados contendo os registros simplificados em ordem cronológica de chegada foi utilizada para avaliação de desempenho da abordagem proposta.

5.4. DETERMINAÇÃO DE ATRIBUTOS

Apenas um atributo local foi utilizado na análise diferencial para determinar o perfil de comportamento local dos usuários. Este atributo corresponde ao perfil de atividade caracterizado pelo número médio de pagamentos efetuados por sessão.

Outras opções de atributos locais para a análise diferencial foram consideradas. Estes atributos correspondem aos seguintes perfis de atividades: a frequência média de transações de pagamento dentro de uma sessão e a frequência média de sessões que envolvem pagamento. Entretanto, estes atributos não forneceram grande poder de discriminação no conjunto de registros contido na base de dados.

O que se verificou foi que para traçar o perfil histórico de um usuário, quaisquer que fossem os atributos escolhidos, seriam necessárias uma amostra maior de transações de cada usuário com uma abrangência maior do que quatro meses.

Dentro destas limitações, o atributo que apresentou melhor resultado na análise diferencial foi o atributo correspondente ao número médio de pagamentos por sessão.

Para o atributo global, foi escolhido o número de contas correntes diferentes acessadas por cada dispositivo em virtude das evidências apresentadas pela análise empírica de que os fraudadores acessam inúmeras contas diferentes.

5.5. AVALIAÇÃO DE DESEMPENHO DO DETECTOR DE FRAUDES

Para a avaliação de desempenho, as transações foram submetidas aos módulos de análise diferencial, global e combinação de Dempster-Shafer na ordem de sequência registrada na base de dados, simulando assim a situação real de chegada das transações em cada conta corrente. Os detalhes mais relevantes da implementação destes módulos são descritos a seguir.

5.5.1. Análise Diferencial

Para a análise diferencial, isto é, para determinar as variações significativas no comportamento local dos usuários, foram utilizados dois modelos estatísticos diferentes. O objetivo de usar dois modelos estatísticos foi para verificar se a escolha do modelo estatístico para a análise diferencial teria muita influência nos resultados junto com a análise global.

Considerando que o número de pagamentos por sessão corresponde ao perfil de atividade utilizado na construção do perfil de comportamento local, a análise diferencial determina se o número de pagamentos realizados na sessão corrente difere muito da média histórica de pagamentos por sessão deste usuário.

O resultado da análise diferencial é uma variável aleatória, denominada aqui de *distância probabilística (dist(pa,ph))* entre o perfil atual *(pa)* e o perfil histórico *(ph)*. A distância probabilística é determinada por um modelo estatístico cujo resultado é um valor entre zero (quando o perfil atual é igual ao perfil histórico) e um (quando o perfil atual é muito diferente do perfil histórico, denotando uma transação fraudulenta).

Um dos modelos estatísticos utilizados na análise diferencial foi baseado em *z-escore* e outro em média ponderada, ambos descritos no capítulo 4.

Para melhor se aproximar da situação real, a média e a variância do perfil histórico de um usuário foram calculados incrementalmente[3], durante a execução de cada transação.

Após calcular a distância probabilística de uma transação, ela é comparada com um limiar pré-estabelecido. Se estiver abaixo deste limiar, a transação é considerada legítima e neste caso, o perfil histórico é atualizado com o valor do perfil atual *(pa).* Caso contrário, a transação é considerada fraudulenta, e o perfil histórico não é atualizado.

No caso do método baseado em *z-score*, a distância probabilística é calculada como segue.

Sejam,

- i, o índice da instância atual;
- pa_i, o perfil atual do usuário, neste caso, o número de pagamentos realizados nesta sessão;
- ph_{i-1}, o perfil médio ou histórico do usuário a ser usado na instância i, neste caso, o número médio de pagamentos por sessão;
- dp_{i-1}, o desvio padrão de ph_{i-1};
- $z = (pa_i - ph_{i-1})/dp_{i-1}$, o *z-escore* correspondente a distribuição de pa_i, supondo uma distribuição normal. $P(Z = z)$ determina a distância probabilística entre pa_i e ph_{i-1}. Esta probabilidade *é* obtida a partir de z utilizando a tabela normal padrão, ou *tabela z*. $P(Z = z)$ se aproxima de 0,5, a medida que pa_i se aproxima de ph_{i-1}, e se aproxima de 1, a medida que se afasta de ph_{i-1}. Em virtude disso, a distância probabilística é normalizada, para que os resultados sejam mapeados entre zero e um.
- s_i, uma variável auxiliar acumulativa para cálculo de variância;
- var_i, a variância;

[3] FINCH, T., **Incremental Calculation of Weighted Mean and Variance**, University of Cambridge Computing Service, Fevereiro 2009.

- *dp_i, o desvio padrão.*

Os três passos a seguir determinam a distância probabilística entre pa_i *e* ph_{i-1}:

1) $z = ((pa_i - ph_{i-1}) / dp_{i-1} ;$

2) Obtém-se o valor de P(Z=z), usando a tabela z;

3) $dist(pa_i ,ph_i) = 2*(P(Z=z) - 0,5)$

O novo perfil histórico *(ph_i)* e o seu desvio padrão *(dp_i)* são atualizados como segue:

4) $ph_i = ph_{i-1} + (pa_i - ph_{i-1}) / i$

5) $s_i = s_{i-1} + (pa_i - ph_{i-1}) * (pa_i - ph_i)$

6) $var_i = s_i / (i-1)$

7) $dp_i = sqrt (var_i)$

No caso do método baseado em média ponderada, a distância probabilística é calculada como segue.

Sejam,

- $limiar_max = Max\{ ph_{i-1} + k*dp_{i-1} \}$, o maior valor de $ph + k* dp$ numa transação sem fraudes em uma conta corrente até o presente momento, desde o início da amostragem;

- $sqrt (var_i)$, a raiz quadrada de var_i

A distância probabilística entre pa_i *e* ph_{i-1} é obtida através de

1) $dist(pa,ph) = Abs (pa_i - ph_{i-1}) / limiar_max$

O perfil histórico *(ph_i)* e o seu desvio padrão *(dp_i)* são atualizados como segue:

2) $ph_i = (1-\alpha)\, ph_{i-1} + \alpha * pa_i$

3) $var_i = (1-\alpha) * (var_{i-1} + \alpha * (pa_i - ph_{i-1})^2)$

4) $dp_i = sqrt(var_i)$

Os valores de α e k são 0,2 e 2 respectivamente, escolhidos empiricamente através de testes;

5.5.2. Análise Global

O resultado da análise global é um valor probabilístico *(P(t))* que representa a distância entre o número de contas acessadas pelo dispositivo *(NumCC)* e um número máximo *(Nmax)* acima do qual o dispositivo é considerado fraudulento.

O valor de *P(t)* é dado pela seguinte expressão:

$$P(t) = \frac{NumCC}{Nmax}$$

Uma descrição mais detalhada de *P(t)* se encontra na sessão 4.10.

O cálculo de *P(t)* é feito para valores de *NumCC* iguais ou maiores do que 2. Para *NumCC* igual a um, o valor de P(t) é considerado zero, pois quaisquer que seja o dispositivo (fraudulento ou não), acessa pelo menos uma conta corrente.

Dois valores de *Nmax* foram utilizados nas simulações para a avaliação de desempenho: 5 e 7. A escolha destes valores baseou-se na observação feita na base de dados. Nesta observação constatou-se que 86,80% dos usuários legítimos acessam apenas uma conta e 99,15% acessam até três contas correntes diferentes. Por outro lado, 72,73% dos acessos realizados por dispositivos envolvidos em fraude acessaram mais de quatro contas correntes diferentes e 63,64% acessaram mais de sete contas correntes.

5.5.3. Combinação de Dempster-Shafer

Os resultados das análises diferencial e global são combinados para gerar um escore final da evidência de fraude utilizando a teoria matemática de Dempster-Shafer, conforme descrito na sessão 4.11:

$$m_3(f) = m_1(f) \oplus m_2(f) = \frac{\sum_{x \cap y = f} m_1(x).m_2(y)}{1-K}.$$

No caso da arquitetura proposta nesta tese, essa expressão pode ser reduzida a:

$$m3 = dist(pa,ph).P(t) + dist(pa,ph).(1 - P(t)) + P(t).(1 - dist(pa,ph))$$

onde, *dist(pa,ph)* é o resultado da análise diferencial, *P(t)* é o resultado da análise global e *m3*, o escore final da evidência de fraude.

A figura 17 ilustra uma sequência de transações que passaram pelas analises diferencial e global, junto com os resultados das combinações (*m3*). Na sessão 31 desta sequência, apesar da distância probabilística determinada pela análise diferencial ter indicado que esta transação é fraudulenta com 46% de probabilidade, a análise global determinou que esta transação é fraudulenta com 100% de probabilidade em virtude do número de contas correntes acessadas

FRAUDE	AGN	IDSessao	IDMQsimpl	NumCC	DATA_HORA	pa=Pgm/Sessao	dist(pa,ph)%	m3
	KW	30	1941	1	04/11/2010 16:15	0,00	5,00%	5,00%
SIM	KW	31	1945	9	04/11/2010 17:56	1,00	46,00%	100,00%
	KW	32	1941	1	11/11/2010 10:22	0,00	13,20%	13,20%
	KW	33	1941	1	11/11/2010 10:39	0,00	10,56%	10,56%
	KW	34	1944	1	12/11/2010 09:05	0,00	8,45%	8,45%
	KW	35	1941	1	23/11/2010 15:08	1,00	43,24%	43,24%
	KW	36	1941	1	24/11/2010 08:52	0,00	15,41%	15,41%

Figura 17 - Trecho de uma sequência de transações em uma conta corrente

Para a avaliação de desempenho do detector proposto, os valores de *m3* são comparados com uma série de limiares, e para cada valor de limiar, uma matriz de confusão é gerada para determinar taxa de verdadeiros positivos e taxa de falsos positivos. Estas taxas são usadas para a avaliação, conforme descrito nas próximas sessões.

5.5.4. Avaliação de desempenho e os seus resultados

Para avaliar o desempenho da abordagem proposta foram selecionados dois cenários baseando-se no conjunto de contas correntes incluídas para a avaliação. No primeiro cenário, foram incluídas apenas as contas correntes que foram fraudadas. No segundo cenário, foram incluídas todas as contas correntes da base de dados disponibilizada.

O objetivo de selecionar dois cenários foi para verificar se a proporção de contas correntes fraudadas no conjunto de teste tem algum efeito significativo nos resultados.

Para cada um dos cenários, foram aplicados os dois métodos estatísticos considerados para a análise diferencial, descritos no item 1.5.1, e a análise global usando dois valores de Nmax, descrito no item 1.5.2.

O resultado da combinação destas duas análises (*m3*) em cada situação foi comparado com valores de limiares, variando de 1,0 até 0,0 com decrementos de 0,01[4].

Para cada valor de *limiar* foi gerada a matriz de confusão, de onde foram calculados os valores das taxas de verdadeiros positivos (Tvp) e de falsos positivos (Tfp), correspondendo a um ponto da curva ROC.

[4] Esta ordem para os valores de *limiares* é em virtude do ponto 0,0 da curva ROC ocorrer com valores maiores de limiar.

Após determinar os pontos da curva ROC, calculou-se o valor de AUC correspondente.

Os gráficos a seguir, ilustram as curvas ROC dos cenários selecionados..

5.5.4.1. Cenário 1- Considerando apenas contas correntes que foram fraudadas

A figura 18 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *z-score* na análise diferencial sem a aplicação da análise global.

O melhor ponto de operação[5], marcado por um circulo na figura, ocorre quando o valor de limiar é igual a 0,980. Neste ponto, Tfp = 0,210 e Tvp = 0,644.

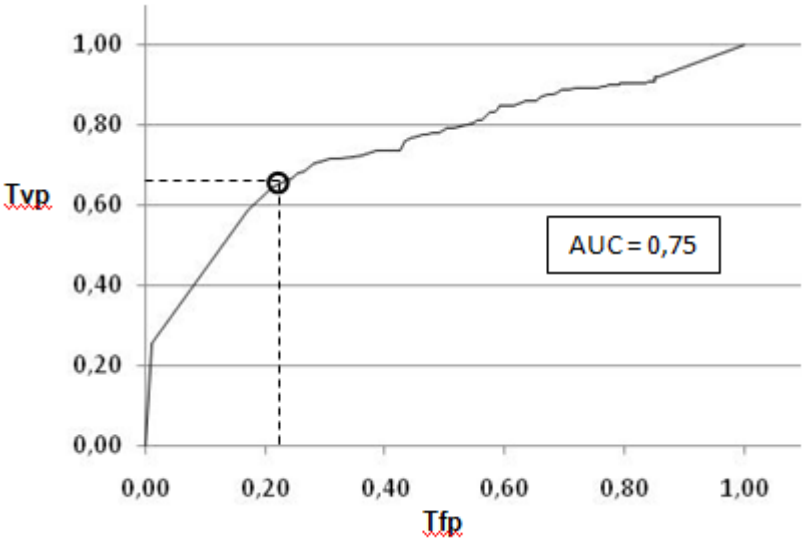

Figura 18 - Curva ROC e AUC aplicando z-score sem análise global, em apenas contas fraudadas

[5] O melhor ponto de operação é o ponto onde ocorre a melhor relação entre Tvp e Tfp, isto é, o ponto onde se tem a maior diferença entre Tvp e Tfp.

A figura 19 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *z-score* na análise diferencial e aplicando análise global com Nmax=5.

O melhor ponto de operação, marcado por um circulo na figura, ocorre quando o valor de limiar é igual a 0,999. Neste ponto, Tfp=0,010 e Tvp = 0,768.

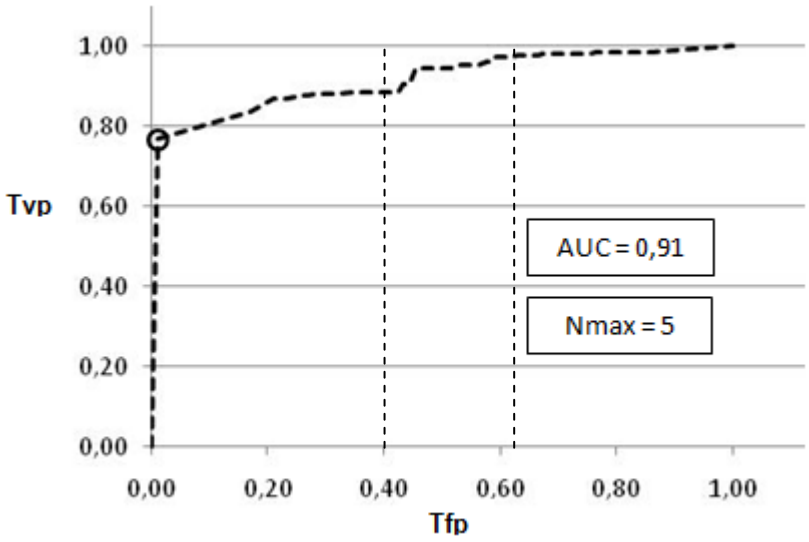

Figura 19 - Curva ROC e AUC aplicando z-score e análise global com Nmax=5, em apenas contas fraudadas

Comparando-se as figuras 18 e 19, observa-se um sensível aumento de desempenho do detector com a adição de análise global.

A figura 20 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *z-score* na análise diferencial e aplicando análise global com Nmax=7.

O melhor ponto de operação ocorre quando o valor de limiar é igual a 0,999. Neste ponto, Tfp=0,010 e Tvp = 0,768.

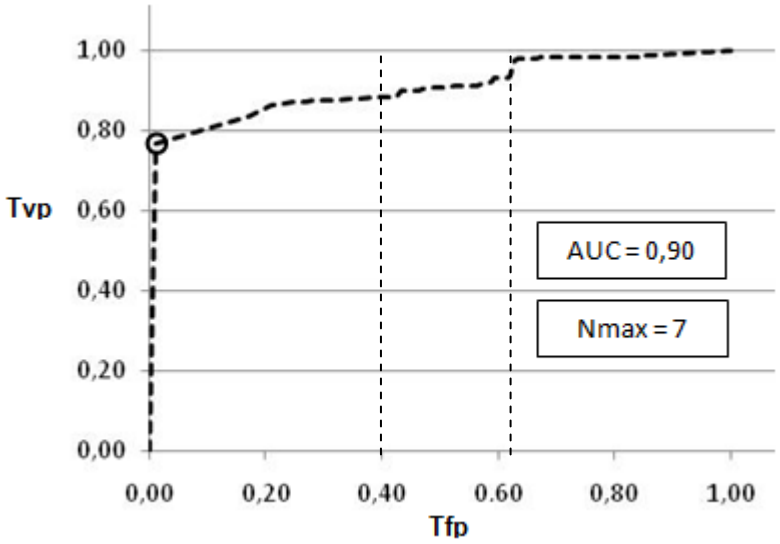

Figura 20 - Curva ROC e AUC aplicando z-score e análise global com Nmax=7, em apenas contas fraudadas

Embora pequena, nota-se pelo valor de AUC que houve uma queda de desempenho global quando Nmax passou de 5 para 7. Esse fato ocorre em virtude de algumas transações fraudulentas que acessam mais de 5 e menos de 7 contas correntes que contribuem bastante no escore quando Nmax é igual a 5, passam a contribuir menos com Nmax igual a 7. Esse efeito pode ser visto na região de Tfp entre 0,4 e 0,6 das figuras 19 e 20. Nesta região, o número de Tvp é maior quando Nmax é igual a 5, refletindo no respectivo valor de AUC.

A figura 21 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *média ponderada* na análise diferencial sem a aplicação da análise global.

O melhor ponto de operação ocorre quando o valor de limiar é igual a 0,520. Neste ponto, Tfp=0,170 e Tvp = 0,609.

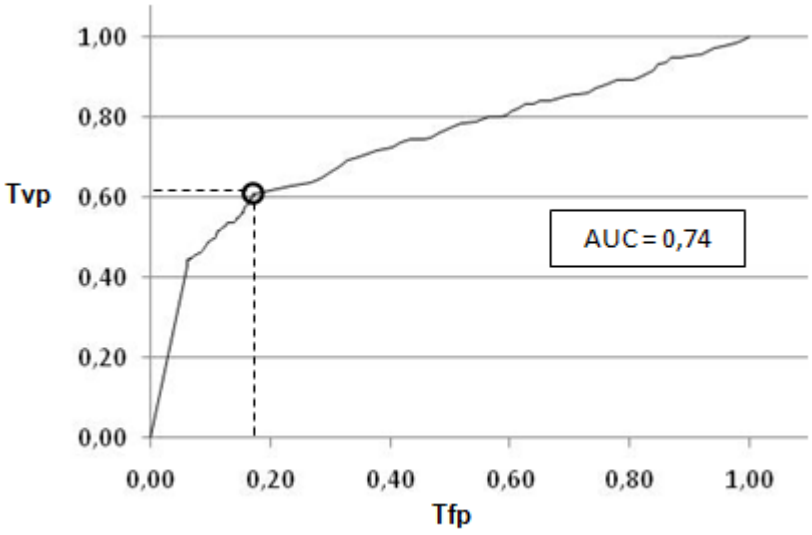

Figura 21 - Curva Curva ROC e AUC aplicando média ponderada sem análise global, em apenas contas fraudadas

A figura 22 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *média ponderada* na análise diferencial e aplicando análise global com Nmax=5.

O melhor ponto de operação ocorre quando o valor de limiar é igual a 0,990. Neste ponto, Tfp=0,097 e Tvp = 0,830.

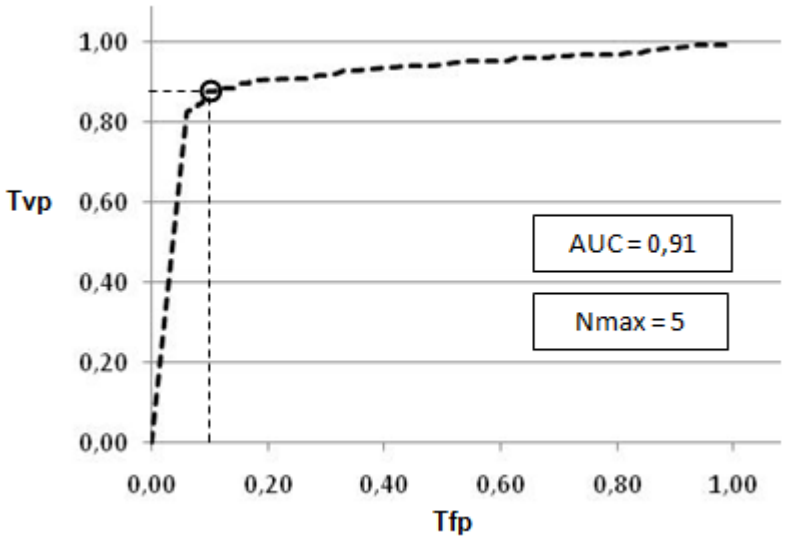

Figura 22 - Curva ROC e AUC aplicando média ponderada e análise global com Nmax=5, em apenas contas fraudadas

A figura 23 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *média ponderada* na análise diferencial e aplicando análise global com Nmax=7.

O melhor ponto de operação ocorre quando o valor de limiar é igual a 0,990. Neste ponto, Tfp=0,062 e Tvp = 0,830

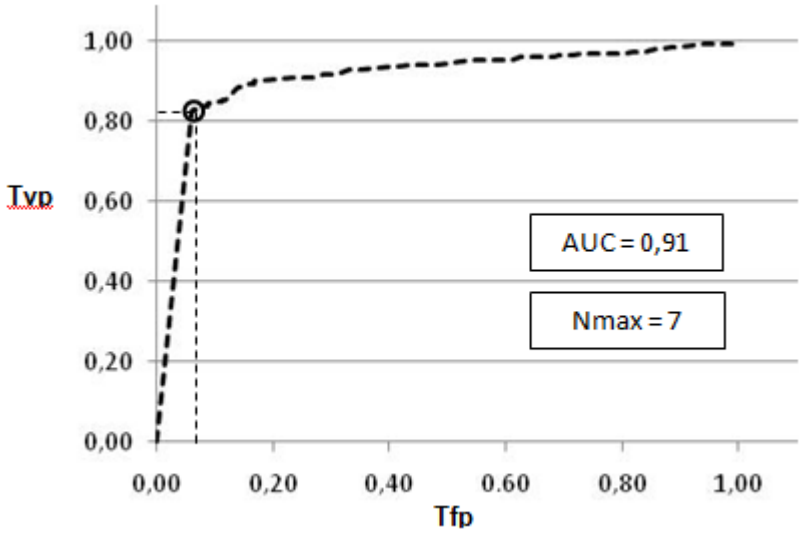

Figura 23 - Curva ROC e AUC aplicando média ponderada e análise global com Nmax=7, em apenas contas fraudadas

5.5.4.2. Cenário 2- Considerando todas as contas correntes da base de dados

A figura 24 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *z-score* na análise diferencial sem a aplicação da análise global.

O melhor ponto de operação ocorre quando o valor de limiar é igual a 0,980. Neste ponto, Tfp=0,196 e Tvp = 0,644.

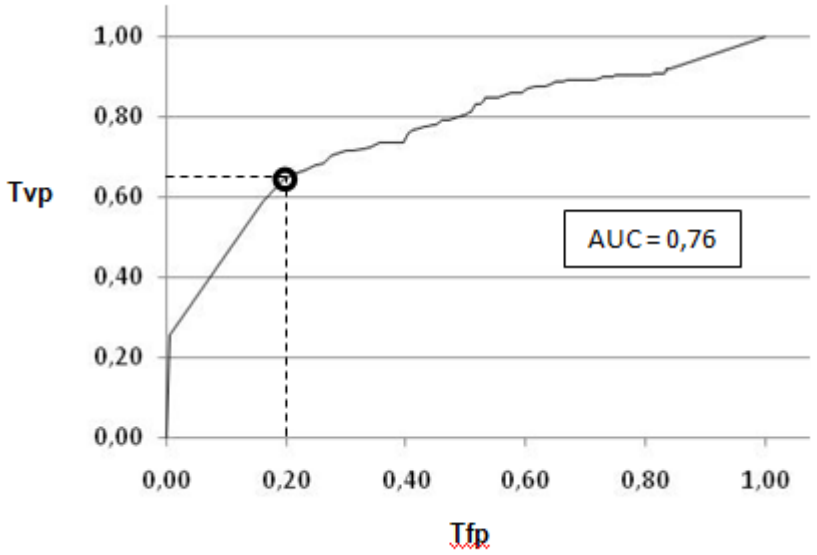

Figura 24 - Curva ROC e AUC aplicando z-score sem análise global, em todas as contas

A figura 25 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *z-score* na análise diferencial e aplicando análise global com Nmax=5.

O melhor ponto de operação ocorre quando o valor de limiar é igual a 0,999. Neste ponto, Tfp=0,007 e Tvp = 0,768.

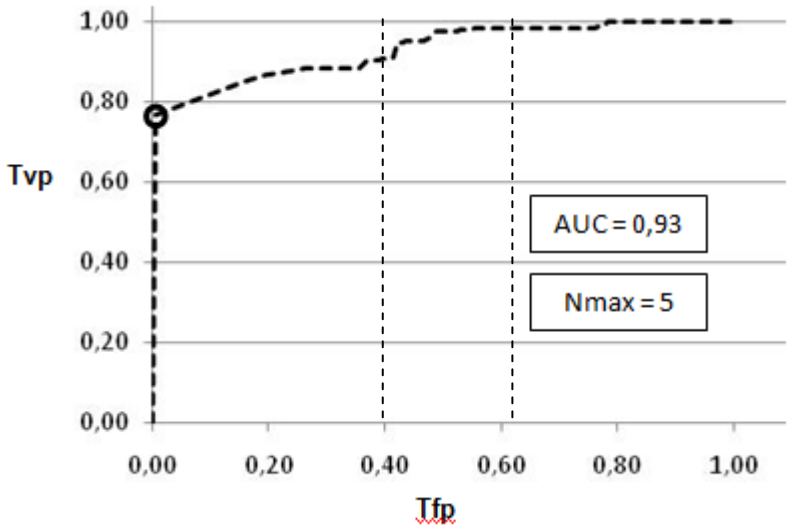

Figura 25 - Curva ROC e AUC aplicando z-score e análise global com Nmax=5, em todas as contas

A figura 26 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *z-score* na análise diferencial e aplicando análise global com Nmax=7.

O melhor ponto de operação ocorre quando o valor de limiar é igual a 0,999. Neste ponto, Tfp=0,007 e Tvp = 0,768.

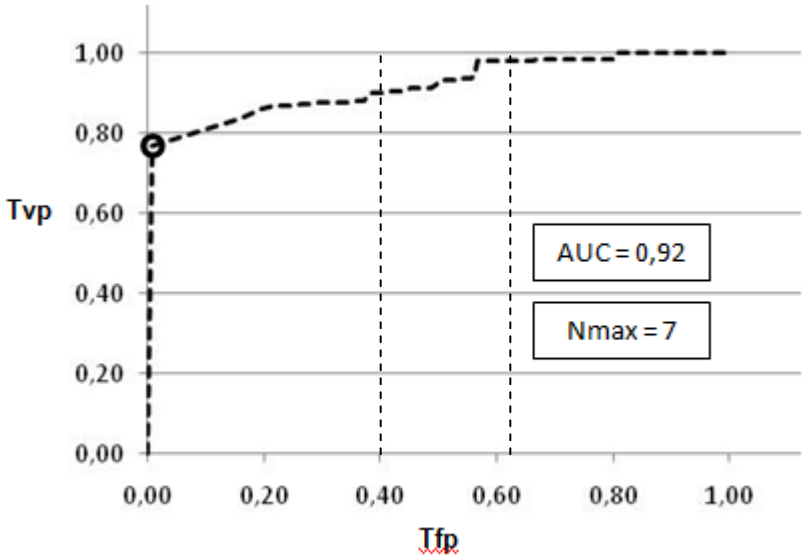

Figura 26 - Curva ROC e AUC aplicando z-score e análise global com Nmax=7, em todas as contas

Aqui também ocorre uma variação de desempenho quando Nmax passa de 5 para 7 pelos mesmos motivos que causou a variação de desempenho entre as situações ilustradas nas figuras 19 e 20.

A figura 27 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *média ponderada* na análise diferencial sem a aplicação da análise global.

O melhor ponto de operação ocorre quando o valor de limiar é igual a 0,520. Neste ponto, Tfp=0,162 e Tvp = 0,609.

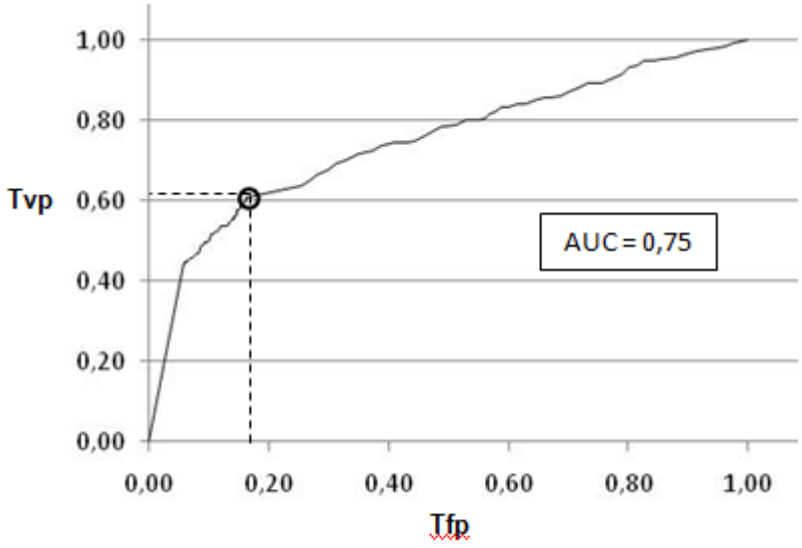

Figura 27 - Curva ROC e AUC aplicando média ponderada sem análise global, em todas as contas

A figura 28 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *média ponderada* na análise diferencial e aplicando análise global com Nmax=5.

O melhor ponto de operação ocorre quando o valor de limiar é igual a 0,860. Neste ponto, Tfp=0,087 e Tvp = 0,878.

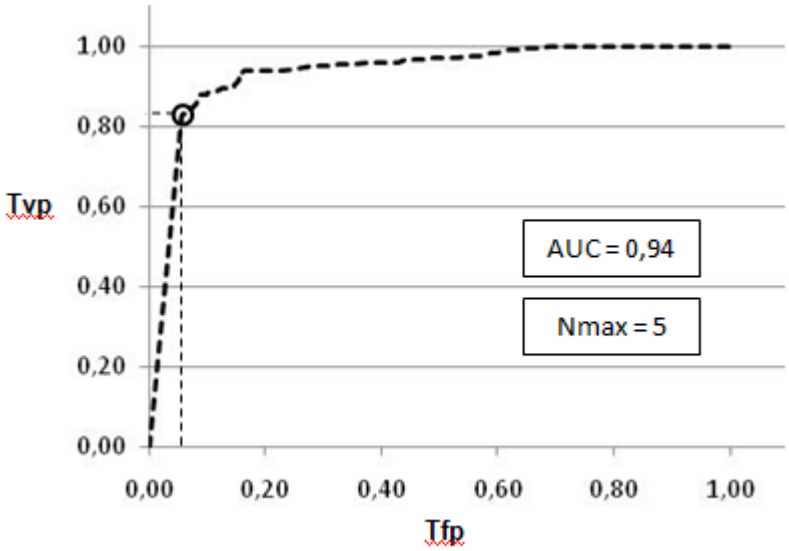

Figura 28 - Curva ROC e AUC aplicando média ponderada e análise global com Nmax=5, em todas as contas

A figura 29 ilustra a curva ROC e o valor de AUC resultante da aplicação do método estatístico baseado em *média ponderada* na análise diferencial e aplicando análise global com Nmax=7.

O melhor ponto de operação ocorre quando o valor de limiar é igual a 0,860. Neste ponto, Tfp=0,087 e Tvp = 0,878.

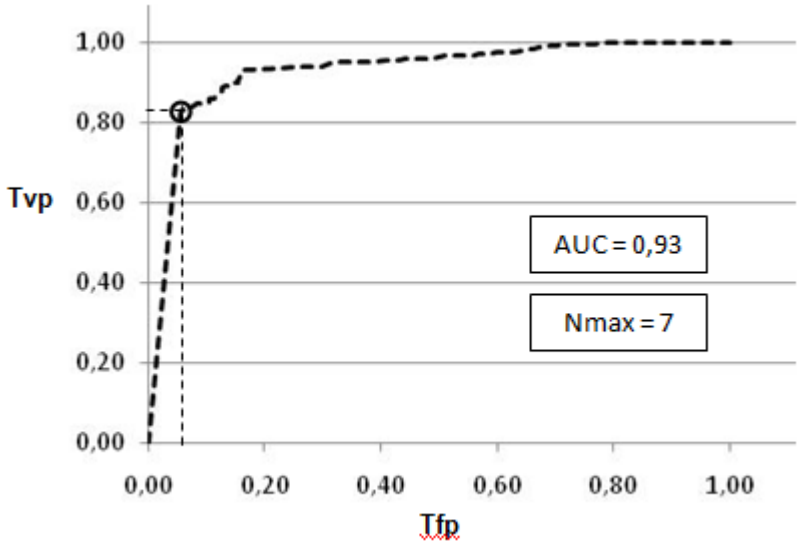

Figura 29 - Curva ROC e AUC aplicando média ponderada e análise global com Nmax=7, em todas as contas

5.5.4.3. Análise dos resultados

.A tabela 5 apresenta o resumo dos resultados obtidos em cada um dos cenários analisados.

Tabela 5 - Tabela comparativa dos resultados obtidos em cada um dos cenários

				Melhor ponto de operação		
Cenários	Análise Diferencial	Análise Global	AUC	Limiar	Tvp	Tfp
Cenário 1- Apenas contas correntes fraudadas	z-score	Sem análise global	0,75	0,980	0,644	0,210
		Nmax=5	0,91	0,999	0,768	0,010
		Nmax=7	0,90	0,999	0,768	0,010
	média-ponderada	Sem análise global	0,74	0,520	0,609	0,170
		Nmax=5	0,91	0.990	0,830	0,095
		Nmax=7	0,91	0,990	0,930	0,062
Cenário 2- Todas as contas correntes	z-score	Sem análise global	0,76	0,980	0,644	0,196
		Nmax=5	0,93	0,999	0,768	0,007
		Nmax=7	0,92	0,999	0,768	0,007
	média-ponderada	Sem análise global	0,75	0,520	0,609	0,162
		Nmax=5	0,94	0,860	0,878	0,087
		Nmax=7	0,93	0,860	0,878	0,087

Pode-se observar na tabela 5 que os valores de AUC das curvas ROC sem a análise global são muito próximos (~0,75). Isto significa que os dois algoritmos usados na análise diferencial têm praticamente as mesmas capacidades para discriminar fraudes dentro dos cenários considerados.

Ao se incluir a análise global, os valores de AUC das curvas ROC saltam para valores acima de 0,90, representando um aumento de 20% ou mais na capacidade de discriminação do detector dentro dos cenários considerados.

Estes resultados demonstram que o uso de atributo local junto com um atributo global melhora a capacidade de detecção, confirmando assim a hipótese inicial da tese.

A tabela 5 mostra também, o melhor ponto de operação de cada uma das curvas ROC, isto é, o ponto da curva onde se tem a maior diferença entre Tvp e Tfp.

Neste cálculo não foi levado em conta o custo associado a falso positivo nem a falso negativo.

5.6. CONSIDERAÇÕES A RESPEITO DE DETECÇÃO EM TEMPO REAL

A proposta desta tese é uma arquitetura para detecção de fraudes em tempo real.

A viabilidade desta proposição pode ser demonstrada somando-se os tempos envolvidos na execução de uma transação de pagamento e comparando o resultado da soma com o tempo disponível para a tomada de decisão.

A execução de uma transação de pagamento na arquitetura proposta envolve basicamente os seguintes tempos:

- *Tad:* Tempo para buscar os dados referentes a análise diferencial associada a conta corrente (média e variância histórica);
- *Tag:* Tempo para buscar os dados referentes a análise global associada ao dispositivo de acesso, entre os quais, a sua curva exponencial;
- *Tbn:* Tempo para buscar informações na Lista Branca e na Lista Negra.
- *Tex:* Tempos de execução dos algoritmos: da análise diferencial, da análise global, da combinação de Dempster-Shafer e da comparação com o valor de limiar pré-estabelecido.

Segundo o perfil de transações reais utilizadas nesta tese:

- 72,44% das transações foram executadas em menos de 25 ms, 17,03% entre 25 e 75 ms, 6,13% entre 75 ms e 150 ms, e o restante acima de 150 ms;
- A maior parcela do tempo de execução corresponde aos acessos as bases de dados.

- O tempo limite para a tomada de decisão é em torno de 500 ms.

Considerando as informações acima, se admitirmos uma média de 100 ms para cada acesso a base de dados, o tempo total gasto para buscar as informações no disco (*Tad* + *Tag* + *Tbn*) é de aproximadamente 300 ms.

Quanto ao tempo de execução dos algoritmos (*Tex*), os testes realizados indicaram um tempo total de execução da ordem de 50 ns e, portanto, desprezível com relação aos demais tempos.

Desta forma, o tempo de execução da abordagem proposta nesta tese depende basicamente dos acessos a base de dados que podem ser feitos junto com os acessos aos dados da própria conta corrente, portanto garantindo a detecção de fraudes em tempo real.

5.7. CONSIDERAÇÕES FINAIS

Os resultados das simulações mostraram que o uso de atributos globais podem melhorar bastante o poder de discriminação do detector de fraudes servindo de reforço aos métodos utilizados para a abordagem local.

. Os testes realizados nos dois cenários apresentaram curvas ROC e valores de AUCs bem próximos, significando que a curva ROC não depende muito da proporção entre transações fraudulentas e legítimas. Desta forma, as curvas ROC e AUCs podem ser consideradas como sendo métricas bem adequadas para a validação da abordagem proposta.

Entretanto, algumas observações devem ser feitas com relação a abordagem utilizada na avaliação de desempenho do sistema proposto.

Através de uma análise empírica feita por amostragem na base de dados verificou-se que não existe um padrão regular no comportamento dos acessos,

confundindo-se muitas vezes com ação dos próprios fraudadores. Portanto, para traçar o perfil de comportamento local de um usuário, seriam necessários volumes maiores de transações de pagamento abrangendo um período bem maior do que o disponibilizado.

Entretanto, não se pode garantir que, com um perfil de comportamento local mais preciso dos usuários, os resultados obtidos pela análise diferencial seriam melhores. Como a detecção de fraudes através de análise diferencial é baseada em desvio de comportamento, basta haver fraudadores que façam pagamentos na mesma taxa dos usuários legítimos que passará despercebido pelo detector.

O que se procurou demonstrar aqui foi que, quaisquer que sejam os métodos estatísticos usados para abordagem local, a sua combinação com uma abordagem global melhora significativamente o desempenho total do sistema de detecção de fraudes.

O único atributo global considerado nesta tese foi o número de contas correntes acessadas por um dispositivo. Essa escolha foi baseada no fato de que os fraudadores acessam muito mais contas correntes diferentes do que os usuários legítimos.

Isso não significa que os usuários legítimos nunca acessam diversas contas correntes. Existem casos, como por exemplo, de administradores de condomínios, em que centenas de contas são administradas através de uma única origem. Nestes casos, os dispositivos de acesso podem estar cadastrados previamente na Lista Branca, ou então, serem incluídos nesta lista após a confirmação da legitimidade da transação ao ocorrer a primeira detecção como fraudulenta.

Por outro lado, existem vários fraudadores que num dado instante são amostrados com números de contas correntes acessadas menores do que o limiar estabelecido (Nmax). Estes casos, apesar de serem reduzidos pela análise efetuada na base de dados, são fontes de falsos positivos.

A base de dados fornecida para a avaliação contém apenas uma amostra do total de transações que ocorreram num período de quatro meses. Entretanto, apesar de ser uma amostra, ela mantém a mesma realidade do perfil global das transações ocorridas neste período. Neste sentido, pode-se dizer que as conclusões obtidas pelas análises realizadas valem para o conjunto global de transações.

A abordagem proposta usando número de contas acessadas como atributo global mostrou que fornece um grande reforço para a análise local na detecção de fraudes. Entretanto, isso não significa que o uso deste atributo seja efetivo para sempre, pois os fraudadores normalmente mudam de estratégia quando a atual não funciona mais. Em outras palavras, pode-se dizer que os fraudadores se adaptam aos métodos de detecção através de novos mecanismos de ataque. Da mesma forma, outros atributos globais mais adequados poderão ser usados no lugar do atual para novos cenários de ataque. De qualquer forma, é importante notar que, independentemente dos atributos utilizados, o caráter global reforça o poder de detecção.

O simulador do sistema de detecção descrito neste capítulo foi desenvolvido utilizando Visual Basic do Microsoft Excel 2007 e executada em um notebook HP Pavillion dv2500 com processador Turion 64 x2 de 1,90 GHz da AMD.

O tempo total de execução dos algoritmos utilizados na análise diferencial foi calculado através de um programa desenvolvido em linguagem C. Como a precisão do temporizador do Sistema Operacional é de um milissegundo, as instruções que implementam o algoritmo foram executadas 10.000.000 de vezes e o tempo total de execução dado em milissegundos pelo sistema operacional foi dividido por este valor.

6. CONCLUSÕES E TRABALHOS FUTUROS

6.1. CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo propor uma arquitetura que permitisse detectar fraudes em transações financeiras via Internet e em tempo real, utilizando dois métodos: um baseado em observações locais, e outro baseado em observações globais.

O resultado apresentado foi uma arquitetura para a detecção de fraudes e os resultados da avaliação do seu desempenho em termos de detecção.

Existem poucas publicações sobre detecção de fraudes no domínio de aplicações de transações bancárias *online*. Desta forma a pesquisa bibliográfica sobre sistemas de detecção de fraudes se concentrou nos domínios de cartão de crédito, intrusão de computadores e comunicação móvel.

Esta pesquisa foi fundamental para entender as características de fraudes de cada uma destas áreas, assim como as alternativas específicas adotadas para detectar fraudes em cada domínio.

Baseando-se nos estudos bibliográficos realizados foi proposta uma arquitetura para detecção de fraudes em transações financeiras via Internet.

Nos ambientes pesquisados não foram encontrados trabalhos que utilizam conjuntamente os conceitos de atributos local e global que caracterizam um aspecto importante dentro da proposta deste trabalho.

Em primeiro lugar foram feitas algumas suposições sobre as características de transações bancárias baseando-se numa análise empírica efetuada num conjunto de transações do mundo real.

Estas características foram levadas em conta na determinação dos atributos local e global.

A arquitetura proposta foi dividida em três grandes módulos (Figura 8 – Capítulo 4):

- **Módulo de Análise Diferencial**, para detectar mudanças significativas nos padrões de transações em uma conta, baseando-se na observação dos atributos locais;
- **Módulo de Análise Global**, para detectar comportamento global anormal de usuários, baseando-se na observação dos atributos globais;
- **Módulo de Combinação de Dempester-Shafer**, para combinar as evidências de fraude determinadas pelas duas abordagens.

A técnica de detecção baseada em análise global proposta nesta tese teve como principal conceito, a noção de identidade do dispositivo de acesso.

Foram abordados os conceitos de perfil de atividade e das respectivas métricas utilizadas na construção do perfil comportamental de um usuário, assim como dos modelos estatísticos para detectar desvios de comportamento.

Dentre os modelos estatísticos para detectar desvios de comportamento no módulo de análise diferencial foi proposto um baseado em *z-score.* Entretanto, outro modelo, baseado em média ponderada, foi também utilizado durante a avaliação do sistema com objetivo de verificar se a escolha do modelo estatístico tem influência significativa no resultado final quando combinado com o resultado da análise global.

Pelos resultados da avaliação, observou-se que os dois modelos estatísticos utilizados para a análise diferencial tiveram desempenhos semelhantes dentro dos cenários considerados com pouca influência no resultado final quando combinado com o resultado da análise global.

Os conceitos relacionados com atributos globais e os efeitos da função exponencial decrescente para calcular a probabilidade de uma transação ser uma fraude baseada em um atributo global foram, então, apresentados e discutidos.

O capítulo 4 termina com algumas considerações sobre a integração com outros modelos de detecção e aspectos relacionados com os requisitos de tempo real.

O capítulo 5 detalha os passos para a validação da arquitetura do sistema proposto, cujos resultados confirmam a hipótese inicial desta tese, demonstrando que a integração de atributos globais com locais não só melhora a capacidade de detecção de fraudes, como também a execução desse procedimento pode ser viabilizado em tempo real.

6.2. TRABALHOS FUTUROS

No caso específico do sistema proposto, contadores foram utilizados como atributos globais para calcular o número de contas diferentes acessadas por cada dispositivo.

Estes números, junto com uma função exponencial decrescente determinam a probabilidade de uma transação ser uma fraude.

Duas observações podem ser feitas sobre esta abordagem para estender este trabalho.

A primeira observação é com relação a identificação do dispositivo.

Foi assumida na seção 4.4 de que a identidade do dispositivo é única para cada dispositivo. Entretanto, se um fraudador instalar um cavalo de tróia em um dispositivo, ele poderá usar a própria identidade da máquina para fazer o ataque na conta acessada por esta máquina sem ser detectado pelo módulo de análise global.

A sugestão neste caso seria sofisticar o componente gerador de identidade, incorporando um detector de vírus que avisaria o servidor de que o dispositivo está infectado e, portanto, não confiável.

A segunda observação é com relação ao atributo global.

Na análise global, utilizou-se apenas o contador de acessos como atributo global e a função exponencial decrescente para avaliar a probabilidade de fraude. Entretanto, outros atributos globais poderiam ser determinados, utilizando a mesma função.

Uma sugestão para evoluir este trabalho seria a definição de um atributo global para determinar a probabilidade de um dispositivo de acesso ser uma fraude baseando-se em sua localização.

O componente identificador de máquina teria uma função adicional de enviar a localização do dispositivo utilizando técnicas disponíveis, embora precária neste momento.

A integração com outros métodos de detecção é uma alternativa a ser explorada como uma forma de aumentar o desempenho do sistema desde que a condição de tempo real seja respeitada. A arquitetura proposta neste trabalho já está, funcionalmente, preparada para integrar novos métodos de detecção ou substituir os métodos atuais de detecção.

REFERÊNCIAS

ALESKEROV, E.; FREISLEBEN, B.; RAO, B., **CARDWATCH: A neural network based database mining system for credit card fraud detection.** *Computational Intelligence for Financial Engineering. Proceedings of the IEEE/IAFE,* p. 220-226, IEEE, Piscataway, NJ. 1997.

AHA, D. W.; KIBLER, D; ALBERT, M. K., **Instance-Based Learning Algorithms.** *Machine Learning Journal*, v.6, n.1, pp. 37-66, 1991.

BEBAR, H.; BECKER, M.: SIBOUNI, D., **A Neural Network Component for Intrusion Detection System,** *Proceedings of the 2002 IEEE World Congress on Computational Intelligence*, Honolulu, HI, pp. 1714-1719, 2002.

BOLTON, R. J.; HAND, D. J., **Unsupervised profiling methods for fraud detection**. *Conference on Credit Scoring and Credit Control 7*, Edinburgh, UK, 5-7 September., 2001.

_____., **Statistical Fraud Detection: A Review,** *Statistical Science,* vol. 17, no. 3, p. 235-255, 2002.

BREIMAN, L.; FRIEDMAN, J.; STONE, C. J.; OLSEN, R. A., **Classification and regression trees**, Chappman and Hall/CRC, 1984

BRAUSE, R.; LANGSDORF, T.; HEPP, M., **Neural data mining for credit card fraud detection.** *Proceedings of the 11th IEEE International Conference on Tools with Arificial Intelligence*. p. 103-106. IEEE Computer Society Press, Silver Spring, MD. 1999.

BRITOS, P.; GROSSER, H.; SIERRA, E.; GARCIA-MARTINEZ, R., **Unusual Changes of Consumption Detection in Mobile Phone Users.** *Special Issue in Neural Networks and Associative Memories Research in Computing Science*, pp. 195-204. 2006.

BURGE, P.; SHAWE-TAYLOR, J., **Detecting Cellular Fraud Using Adaptive Prototypes.** *Proceedings of the AAAI-97 Workshop and AI Approaches to Fault Detection and Risk Management.* Mento Park, CA: AAAI Press, pp. 9-13, 1997.

BURGE, P.; SHAWE-TAYLOR, J.; COOKE, C.; MOREAU, Y.; PRENEEL, B.; STOERMANN, C., **Fraud detection and Management in Mobile Telecommunication Networks.** *2nd European Conference and Detection, IEEE Conference.* Publication 437, pp. 91-96, London, 1997.

_____., **An Unsupervised Neural Network Approach to Profiling the Behavior of Mobile Phone Users for Use in Fraud Detection**. *Journal of Parallel and Distributed Computing*. pp. 915-925. 2001.

CAHILL, M. H.; LAMBERT, D.; PINHEIRO, J. C.; SUN, D. X., **Detecting fraud in real world**. *Handbook of Massive Datasets* (J. Abello, P. M. Pardalos and M. G. C. Resende, eds.). Kluwer, Dordrecht. 1999.

CHEN, T. M.; VENKATARAMANAN, V., **Dempster-Shafer Theory for Intrusion Detection in Ad Hoc Networks**. *In: Proceedings of the IEEE Internet Computing,* pp. 35–41. 2005.

CHEN, Q.; AICKELIN U., **Anomaly detection using the Dempster-Shafer method**, *in Proc. of the 2006 International Conference on Data Mining, DMIN 2006,* pp. 232–240, 2006.

CLARK, P.; NIBLETT, T., **The CN2 induction algorithm**. *Machine Learning* 3, pp. 261–285, 1989.

CLEARWATER, S.; PROVOST, F., **RL4: A tool for knowledge-based induction.** *In Proceedings of the Second International Conference on Tools for Artificial Intelligence*. IEEE CS Press, pp. 24-30, 1990.

COHEN, W., **Fast efective rule induction**. *In Proceedings of the 12th International Conference on Machine Learning,* pp. 115-123. Morgan Kaufmann, Palo Alto. CA, 1995.

CORTES, C.; PREGIBON, D., **Signature-based methods for data streams**. *Data Mining and Knowledge Discovery*, p. 167-83, 2001.

DENNING, D. E., **An intrusion detection model.** *IEEE Transactions on Software Engineering,* 13:222-232. 1987.

EDGE, M. E.; SAMPAIO, P. R., **Survey of signature based methods for financial fraud detection.** Manchester Business School, University of Manchester, Booth Street East, Manchester M16 6PB, United Kingdom, 2009.

ESMAILI, M., **Dempster-Shafer Theory and Network Intrusion Detection Systems**, Center for Computer Security Research, Scientia Iranica, Vol. 3, No. 4, Shrif Univerity of Technology, 1997.

FAWCETT, T., **An Introduction to ROC Analysis**, *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, 2006.

FAWCETT, T.; PROVOST, F., **Adaptive Fraud Detection.** *Data Mining and Knowledge Discovery,* Kluwe, 1, p. 291-316, 1997.

_____., **Combining data mining and machine learning for effective fraud detection.** *AAAI Workshop on AI Approaches to Fraud Detection and Risk Management*. p.14-19, AAAI Press, Menlo Park, CA, 1997.

_____., **Activity monitoring: Noticing interesting changes in behavior.** *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 53-62. ACM Press, New York, 1999.

FERREIRA, P.; ALVES, R.; BELO, O.; CORTESAO, L., **Establishing fraud detection patterns based on signatures**. *Industrial conference on data mining*, Leipzig, Germany, p. 526-538, 2006.

GIARRATANO J., RILEY G., **Expert Systems Principles and Programming**. Third Edition. PWS Publishing Company. 1998.

GOLDBERG, H; SENATOR, T. E., **Break detection systems.** *AAAI Workshop on AI Approaches to Fraud Detection and Risk Management*. p.22-28, AAAI Press, Menlo Park, CA, 1997.

GHOSH, S.; REILLY D. L., **Credit Card Fraud Detection with a Neural-Network**, *Proc. 27th Hawaii International Conference on System Sciences: Information Systems: Decision Support and Knowledge-Based Systems,* vol. 3, pp. 621-630, 1994.

GHOSH, A. K.; SCHWRTZBARD, A., **A study in using neural networks for anomaly and misuse detection**, *in Proceedings of the 8th USENIX Security Symposium*, 1999.

GUO T.; LI G. Y., **Neural Data Mining for Credit Card Fraud Detection**, *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming,* 12-15 July 2008.

HAMEL, L., **Model Assessment with ROC Curves**. *The Encyclopedia of Data Warehousing and Mining, 2nd edition, Idea Group Publishers*, http:// http://homepage.cs.uri.edu/faculty/hamel/pubs/hamel-roc.pdf. Acessado em 07 de Janeiro de 2011.

HENKIND S. J.; HARRISON M. C., **An analysis of four uncertainty calculi**. IEEE Transactions on Systems, MAN. and Cybernetics, Vol. 18, no. 5, p. 700-713, September/October 1988.

HILAS, C. S.; SAHALOS, J. N., **User profiling for fraud detection in telecommunications networks.** *Proceedings of the 5th International Conference Technology and Automation (ICTA'05*), Thessaloniki Greece, p. 382-387, 2005.

HOLLMÉN, J., **User Prifiling and Classification for Fraud Detection in Mobile Communication Network**s. Tese (Doutorado), Department of Computer Science and Engineering, Helsinki University of Technology, Finland. 2000.

KARSEN K. N. ; KILLINGBERG T. G., **Profile based intrusion detection for Internet banking systems**, Master Thesis, Norwegian University of Science and Technology, Norway, 2008

KOU, Y.; LU, C.; SIRWONGWATTANA, S.; HUANG, Y., **Survey of Fraud Detection Techniques,** *Proceedings of the 2004 IEEE International Conference on Networking, Sensing & Control*, Taipei, Taiwan, March 21-23, 2004.

KOVACH, S.; RUGGIERO, W. V., **Online Banking Fraud Detection Based on Local and Global Behavior**, *IPDS, The First International Workshop for Innovative Methods for Intrusion Prevention and Detection Systems, 2011*.

LANE, T.; BRODLEY, C. E., **An application of machine learning to anomaly detection.** In *Proceedings of the 20th National Conference on National Information Systems Security. Vol.1* (Baltimore, MD). National Institute of Standards and Technology, Gaithersburg, MD, 366–380, 1997.

LANE, T.; BRODLEY, C. E., **Temporal sequence learning and data reduction for anomaly detection.** *Proceedings of the 5th ACM Conference on Computer and Communications Security (CSS-98),* p. 150-158, ACM Press, New York, 1998.

LARSON, R.; FARBER, B., **Estatística Aplicada**. 4a edição, Pearson Prentice Hall, 2010.

LEE, W.; STOLFO, S., **Data mining approaches for intrusion detection.** *Proceedings of the 7th USENIX Security Symposium*, San Antonio, TX, p. 79-93, USENIX Association, Berkeley, CA, 1998.

LUNT, T. F.; TAMARU, A.; GILHAM, F.; JAGANNATHAN, R.; JALALI, C.; NEUMANN, P.; JAVITZ, H. S.; VALDES, A.; GARVEY, T. D., **A Real-Time Intrusion Detection Expert System (IDES), Final Technical Report.** Computer Science Laboratory, SRI International, Menlo Park, California, February 1992.

MITCHELL T. M., **Machine Learning.** McGraw Hill, 1997.

MOREAU, Y.; PRENNEL, B.; BURGE, P.; SHAWE-TAYLOR, J.; STOERMANN, C.; COOKE, C., **Novel Techniques for Fraud Detection in Mobile Telecommunication Networks**. *In: ACTS Mobile Summit*. 1997

MOREAU, Y.; LEROUGE, E.; VERRELST, H.; VANDEWALLE, J.; STÖRMANN, C.; BURGE, P., **A hybrid system for fraud detection in mobile communications**. *In: European Symposiumon Artificil Neural Networks, Bruges, Belgium, 1999.*

MOREAU, Y., VANDEWALLE, J., **Detection of Mobile Phone Fraud using Supervised Neural Networks: A First Prototype**. *In: Proceedings of the International Conference on Artificial Neural Networks.* 1997

MORANDI, M.; ZULKERNINE, M., **A Neural Network Based System for Intrusion Detection and Classification of Attacks,** *IEEE International Conference on Advances in Intelligent Systems - Theory and Applications,* Luxembourg-Kirchberg, Luxembourg, November 15-18, 2004

MURAD, U.; PINKAS, G., **Unsupervised Profiling for Identifying Superimposed Fraud.** *In : Proceedings of the 3rd European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 251-26, 1999.

PANIGRAHI, S.; KUNDU, A.; SURAL, S.; MAJUMBAR, A. K., **Use of Dempster-Shafer theory and Bayesian Inferencing for Fraud Detection in Communication Networks.** *Lecture Notes in Computer Science,* Spring Berlin/ Heidelberg, Vol. 4586, p.446-460, 2007.

PROVOST, F.; FAWCETT, T., **Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions**. *In: Proc. Third Internat. Conf. on Knowledge Discovery and Data Mining (KDD-97).* AAAI Press, Menlo Park, CA, pp. 43–48, 1997.

PROVOST, F., FAWCETT, T.,; KOHAVI, R., **The case against accuracy estimation for comparing induction algorithms.** *Proceedings of the Fifteenth International Conference on Machine Learning,* , 445–453, 1998.

QUINLAN, J. R., **Generating production rules from decision trees**. *In Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, pp. 304-307. 1987.

QUINLAN, J. R., **C4.5: Programs for Machine Learning.** Morgan Kaufmann, San Mateo, CA, 1993.

ROSSET S.; MURAD, U.; NEUMANN, E.; IDAN, Y.; PINKAS, G., **Discovery of fraud rules for telecommunications challenges and solutions.** *In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining,* pp. 409-413, New York, NY, USA, ACM Press, 1999.

RUGGIERO, W. V., **Medição e Distribuição de Confiança em redes Ad-Hoc** (versão em português). *International Conference on Wireless Security*, Las Vegas, Estados Unidos, 2002.

SHAFER, G., **A Mathematical Theory of Evidence.** Princeton University Press, 1976.

SINGH R.; VATSA, M.; NOORE, A.; SINGH, S. K., **Dempster Shafer Theory based Classifier Fusion for Improved Fingerprint Verification Performance**, *Indian Conference on Computer Vision, Graphics and Image Processing*, Springer, Vol. 4338, pp. 941-949, 2006.

TANIGUCHI, M.; HAFT, M.; HOLLMÉN, J.; TRESP, V., **Fraud Detection in Communication Networks Using Neural and Probabilistic Methods**. *In: Proceedings of The IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1241-1244, may 1996.

WALPOLE R. E.; MYERS, R. H.; MAYERS, S. L.; YE, K., **Probabilidade e Estatística para Engenharia e Ciências.** Oitava edição. Prentice Hall. 2009.

WITTEN I. H.; FRANK E., **Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations**. Morgan Kaufmann. 2000.

WANG, Y.; YANG, H.; WANG, X.; ZHANG, R., **Distributed Intrusion Detection System Based on Data Fusion Method.** *In: Proceedings of the 5th World Congress on Intelligent Control and Automation,* pp. 4331–4334. 2004.