

UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO

SANDRO DAROS DE LUCA

IDENTIFICAÇÃO DE PROBLEMAS NO ELUXO DE
FATURAMENTO DAS OPERADORAS DE
TELECOMUNICAÇÕES: UMA ABORDAGEM
EMPREGANDO LÓGICA FUZZY E REGRAS DE
PRODUÇÃO

Dissertação submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Mestre em Ciência da Computação

ORIENTADOR: PAULO JOSÉ DE FREITAS FILHO


Florianópolis, março de 2002

**IDENTIFICAÇÃO DE PROBLEMAS NO FLUXO DE
FATURAMENTO DAS OPERADORAS DE TELECOMUNICAÇÕES:
UMA ABORDAGEM EMPREGANDO LÓGICA FUZZY E REGRAS DE
PRODUÇÃO**

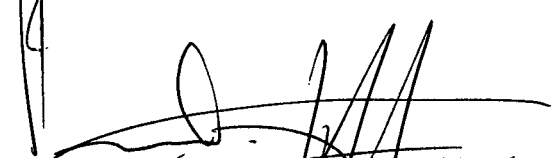
SANDRO DAROS DE LUCA

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, Área de Concentração Sistemas de Computação e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

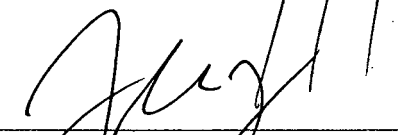
Banca Examinadora



Prof. Fernando A. Ostuni Gauthier, Dr.
Coordenador CPGCC, UFSC




Prof. Paulo José de Freitas Filho, Dr.
Orientador, INE, UFSC



Prof. José Marcos Silva Nogueira, Dr.
DCC, UFMG



Alexandre Moraes Ramos
UNIVALI



Elizabeth Sueli Specialski
INE, UFSC

Agradecimentos

Em primeiro lugar, a Deus e aos meus Pais, pois sem eles eu não conseguiria chegar até aqui.

À Thermus Comércio, Serviços e Representações pelo apoio e compreensão.

À instituição de ensino gratuito e de qualidade, Universidade Federal de Santa Catarina, pela oportunidade de realização do mestrado.

Aos membros da banca e em especial à Elizabeth Sueli Specialski e ao meu orientador Professor Paulo Freitas pelo tempo e auxílio despendidos.

À Débora Aparecida Ataíde Ampessan e Juliana Sell do Vale Pereira pela ajuda no desenvolvimento do trabalho.

Ao Professor Ilson Wilmar Rodrigues Filho, ao Professor Paulo José Ogliari e aos demais professores e funcionários do INE.

Aos meus amigos e colegas, ao meu irmão e à demais pessoas que colaboraram para a conclusão deste trabalho.

Em especial a Susi Meire Fátima Carvalho pelo apoio, compreensão e auxílio ao longo de todo o mestrado, principalmente na confecção do trabalho escrito.

Resumo

Este trabalho apresenta uma proposta de um mecanismo para analisar dados de chamadas telefônicas de teste. A idéia central é analisar problemas no fluxo de faturamento das operadoras de telecomunicações, a partir dos registros de chamadas gerados pela central telefônica (CDR - *Call Detail Recording*) e de registros gerados por um equipamento de teste. Com esta análise pretende-se indicar os pontos de falhas na geração dos CDRs. Para realizar esta análise, utilizar-se-á dos mecanismos computacionais de lógica fuzzy e regras de produção.

Uma das motivações deste trabalho é que as ferramentas de mercado normalmente enfocam mais a geração dos testes e menos a sua verificação ou análise dos resultados. Estas tarefas são na maioria das vezes repassadas ao usuário do sistema.

Abstract

This paper presents an automated tool to analyze test telephone call data. The main idea is to analyze the problems in the local exchange carrier billing process, based on the call record generated by the telephone switch (CDR – Call Detail Recording) and on the record generated by the test equipment. With this analysis, it intends to indicate the faults in the CDRs generation. To do this analysis, it will use the fuzzy logic computing technical and production rules.

One of the motivations for this work is that the tools on the market normally give more importance to the test generation and less to its verification or its result analysis. These tasks are generally re-passed to the system user.

Sumário

Capítulo 1 - Introdução.....	13
1.1 Apresentação.....	13
1.2 Justificativa.....	13
1.3 Abordagens da solução do problema.....	15
1.4 Objetivo do trabalho.....	16
1.5 Escopo do trabalho.....	16
1.6 Estrutura do Trabalho.....	17
Capítulo 2 - Bilhetagem e o processo de tarifação.....	19
2.1 Bilhetagem em centrais telefônicas de serviço de voz.....	19
2.2 Processo de faturamento das chamadas provenientes de CDRs.....	23
2.2.1 Fases do processo.....	24
2.2.2 Características do CDR.....	27
2.3 Considerações.....	30
Capítulo 3 - Lógica fuzzy e regras de produção.....	32
3.1 Introdução à lógica fuzzy.....	32
3.2 Conjuntos nebulosos.....	33
3.3 Inferência fuzzy.....	35
3.4 Intersecção, união e complemento.....	36
3.5 Regras de produção.....	38
3.6 Sistemas fuzzy.....	39
Capítulo 4 - Método proposto para identificação de problemas.....	43
4.1 Contexto do trabalho.....	43
4.1.1 Contexto do problema.....	43
4.1.2 Abordagens para o problema.....	48
4.1.3 Abordagem proposta para o problema.....	54
4.1.4 Restrições adotadas neste trabalho.....	58
4.2 Mecanismo de regras.....	60
4.2.1 Proposta de análises.....	63
4.2.2 Análises horizontais.....	64
4.2.3 Análises verticais.....	66
4.2.4 Enquadramento do mecanismo de regras nas operadoras.....	67
Capítulo 5 - Estudo de caso.....	70
5.1 Contexto do estudo de caso.....	70
5.2 Testes realizados.....	74
5.3 Resultados.....	79
Capítulo 6 - Conclusão e recomendações.....	82
6.1 Comentários e conclusões.....	82
6.2 Extensibilidade.....	85
6.3 Características para aplicações reais.....	86
6.4 Sugestões para trabalhos futuros.....	86
Referências bibliográficas.....	89
Anexo 1 – Proposta de uma linguagem de regras.....	93
Anexo 2 – Conjunto de regras utilizado no estudo de caso.....	95
Anexo 3 – Distribuição de probabilidade dos estudos de casos.....	98

Anexo 4 – Resultados das análises dos estudos de caso	100
Anexo 5 – Resultados da análise estatística	103

Lista de Figuras

Figura 1 – Demonstração dos conceitos de central bilhetadora, local e chamadas saintes e entrantes.....	23
Figura 2 – Processo genérico de tarifação.....	24
Figura 3 – Exemplos gráficos de união e intersecção de conjuntos fuzzy.....	37
Figura 4 – Exemplos de sistema fuzzy com duas entradas e uma saída.....	40
Figura 5 – Pontos de perda de receita.....	46
Figura 6 – Esquema hipotético de geração dos dados.....	55
Figura 7 – Esquema básico do trabalho.....	60
Figura 8 – Ordem de execução das regras e fluxo de informação.....	63
Figura 9 – Fluxo de dados nas análises horizontais e verticais.....	64
Figura 10 – Análise horizontal.....	65
Figura 11 – Análise vertical.....	66
Figura 12 – Operadora hipotética utilizando sistema de geração de testes controlados.....	68
Figura 13 – Operadora hipotética utilizando sistema de geração de testes controlados aliado a base teórica aqui apresentada.....	68

Lista de Tabelas

Tabela 1 - Índices de perdas típicos das operadoras da América Latina	44
Tabela 2 - Serviços não faturados	45
Tabela 3 - Irregularidades em faturas.....	45
Tabela 4 - Sumário dos resultados da análise estatística	80
Tabela 5 - Distribuição de probabilidade para o número de erros por CDR	98
Tabela 6 - Distribuição de probabilidade por tipo de erro	98
Tabela 7 - Distribuição de probabilidade dos campo.....	99
Tabela 8 - Resultados dos dados gerados nos estudos de casos.....	100

Lista de Quadros

Quadro 1 – Cálculo do grau de corretude	61
Quadro 2 – Aplicação das regras r_i , r e r_f	62
Quadro 3 – Esquema matemático da regra de 3.....	71
Quadro 4 – Regras utilizadas para a fuzificação.....	72
Quadro 5 – Exemplo de fuzificação utilizando as regras propostas	73
Quadro 6 – Listas dos erros utilizados nos testes.....	76
Quadro 7 – Regras utilizando o mecanismo proposto.....	96
Quadro 8 – Regras utilizando um algoritmo tradicional.....	97

Lista de Siglas e Abreviações

3G	Tecnologia celular de terceira geração, permite trabalhar com velocidades de transmissão máxima em torno de 2Mb/s.
2,5G	Tecnologia celular de transição entre a segunda e a terceira geração, permite trabalhar com velocidades de transmissão máxima em torno de 256Kb/s.
ANATEL	Agência Nacional de Telecomunicações, órgão regulador do setor no Brasil.
CDR	<i>Call Detail Recording</i> (registro detalhado de chamada) ou <i>Call Data Record</i> (registro de chamada).
CFMS	Compaq Fraud Management System, sistema antifraude da empresa Compaq.
CO	Status de uma chamada quando esta não pode ser completada por congestionamento.
CPF	Cadastro de Pessoa Física.
CPqD	Centro de Pesquisa de Desenvolvimento em Telecomunicações. Empresa que fornece soluções em telecomunicações.
DSC	Status de uma chamada quando esta não pode ser completada pois o assinante A desconecta antes que B atenda.
FTAM	<i>File Transfer, Access and Management</i> , protocolo de transferência de arquivos.
GAWK	GNU Aho Wenberg and Kernighan, linguagem script de programação.
GNU	Gnu Not Unix, organização internacional que desenvolve ou organiza softwares de livre distribuição.
GNU-Bash	GNU-Bourne Aghain Shell, interface de comandos comumente utilizada em ambiente UNIX.

GPL	General Public License, tipo de licença de software normalmente utilizado nas ferramentas da GNU (http://www.gnu.org/).
GVT	Global Village Telecom, operadora de telefonia fixa que atua na região centro sul do Brasil.
IA	Inteligência Artificial, ramo de informática.
IP	<i>Internet Protocol</i> , protocolo de comunicação comumente utilizado na Internet.
LO	Status de uma chamada quando esta resulta em linha ocupada.
NR	Status de uma chamada quando esta não pode ser completada porque o assinante B não respondeu.
OK	Status de uma chamada terminada com sucesso.
OM	<i>Operation Metrics</i> , medidas operacionais, também conhecidas como medidas de tráfego.
OU	Status da chamada quando outros problemas fazem com que uma chamada não seja completada ou efetivada com sucesso.
PAB	Status de uma chamada quando esta não pode ser completada por problemas entre os assinantes A e B.
Procon	Superintendência de Proteção e Defesa do Consumidor.
RI	Rede Inteligente, rede de telecomunicações normalmente utilizada para prover serviços como 0800, 0300 e 0900.
SAC	Sistema de Atendimento ao Cliente, setor de atendimento ao cliente que é realizado normalmente via telefone ou internet.
SE	Sistemas Especialistas
TCP/IP	<i>Transmission Control Protocol/Internet Protocol</i> , protocolo de comunicação comumente utilizado na Internet.
TI	Tecnologia de Informação, setor das empresas que trabalha com a área de informática.

Capítulo 1 - Introdução

O presente capítulo introduz o trabalho a ser apresentado. Após são descritas as justificativas, o objetivo e o contexto do trabalho. Ao final descreve-se como o trabalho está estruturado.

1.1 Apresentação

As empresas de telecomunicações passam por constantes mudanças e de forma muito rápida. O surgimento de novas tecnologias, a rotatividade de clientes, a forte concorrência e a necessidade de integração de equipamentos são alguns dos fatores que influenciam essas mudanças. No Brasil este quadro não é muito diferente. E adicionam-se ainda outros fatores, tais como recente privatização, abertura do mercado, fim do monopólio e entrada de investidores externos.

Em virtude dos fatores expostos, os processos, os sistemas computacionais e as pessoas envolvidas com telecomunicações passam também por constantes mudanças. Um processo adotado há quatro anos pode estar obsoleto, vindo a ser substituído ou abandonado em um futuro próximo.

Com tantas transformações em tão pouco tempo, fazer com que uma companhia de telecomunicações opere sem problemas é quase uma utopia. Estes problemas são causados principalmente por uma série de falhas espalhadas por toda a empresa. Um dos setores que possuem um dos maiores índices de erros é o de bilhetagem e cobrança das contas telefônicas.

O trabalho aqui apresentado expõe um mecanismo para tentar diminuir o alto volume de erros encontrado neste setor nas operadoras, empregando lógica fuzzy e regras de produção.

1.2 Justificativa

Os problemas apresentados afetam não só as operadoras, mas também os clientes, que são, muitas vezes, os principais prejudicados. Uma das principais reclamações dos clientes refere-se aos erros nas cobranças das contas telefônicas. Estes

erros de cobrança fizeram com que as operadoras assumissem os primeiros lugares em reclamações junto ao Procon (Superintendência de Proteção e Defesa do Consumidor). Um exemplo é a lista das empresas de São Paulo com maior número de reclamações junto ao Procon em 2001. Nesta lista [CARVALHO, 2002], os dois primeiros lugares são ocupados por operadoras de telecomunicações, sendo que, entre os vinte primeiros, aparecem as seis operadoras de telecomunicações do estado (duas de telefonia fixa, duas de celular e duas de longa distância). Este status de “líder em reclamações” degrada a imagem das operadoras. Com isso, os clientes buscam as operadoras concorrentes procurando serviços melhores e mais baratos.

Além de ter uma imagem lesada junto à sociedade, as operadoras perdem muito dinheiro. Segundo [SANTANA, 2001], estima-se que no mínimo R\$ 2,48 bilhões ao ano da receita são perdidos em todas as operadoras do país por falta de controle no processo de faturamento. Este valor é equivalente à receita operacional líquida de 2000 da Telesp Celular, a maior operadora móvel do país. Paulo Rogério Foina [FOINA, 1998] escreve que a fraude de assinantes corresponde a cerca de 1% da receita (variando de 0,5% em telefonia fixa, 1,5% em celular digital, e até 3% em celular analógico), a perda de receita provocada pelas falhas e erros de procedimentos pode chegar a 4% da receita de uma operadora. Já Stela Oliveira [OLIVEIRA, 2000] comenta que a maioria das chamadas “teleoperadoras” apresenta perdas da ordem de 2% a 5% de seus faturamentos anuais. Há outros estudos mais pessimistas [SANTANA, 2001] que mostram que o índice de evasão de receita pode chegar a 15% do faturamento total das operadoras.

Com esta perda de receita, aumenta o custo operacional das empresas, o que resulta em tarifas mais altas para os clientes, aumentando a evasão de clientes¹ ou diminuindo seu consumo. Assim sendo, estas cobranças errôneas não dão lucro nem para o cliente nem para a operadora.

¹ A taxa de *churn* (desconexão voluntária de assinantes) nas operadoras está atualmente em torno de 0,7% ao mês referente a uma base de cerca de 7 milhões de terminais instalados [MOURA, 2001].

Mas onde está exatamente o problema? Esta é a pergunta que as operadoras se fazem e para a qual não obtêm respostas conclusivas e satisfatórias. O que se sabe é que a quase totalidade destes erros está ligada ao processo de faturamento das empresas. Para tentar responder esta pergunta de uma forma mais eficaz pode-se utilizar uma gama de ferramentas de controle e verificação do processo de faturamento. Ao setor da empresa ou à categoria de sistemas que tratam deste problema dá-se o nome de garantia de receita ou *revenue assurance*. Este termo corresponde à tentativa de minimizar a perda de receita, ou seja, assegurar que os serviços oferecidos aos clientes sejam corretamente registrados, valorados, faturados e arrecadados [CPqD, 2001].

Portanto, sente-se a necessidade de ferramentas e/ou mecanismos que auxiliem no processo de garantia da receita, ou seja, ferramentas que auxiliem as operadoras a identificar falhas no processo de faturamento. As ferramentas comumente adotadas pelo mercado são descritas no próximo item.

1.3 Abordagens da solução do problema

Uma das formas de se identificar os erros no processo de tarifação é a geração de chamadas de teste controladas e seu posterior acompanhamento ao longo do processo de tarifação. Este acompanhamento pode ir desde o registro da chamada na central telefônica até o envio da conta telefônica ao cliente fictício. Com os dados deste acompanhamento pode-se fazer uma auditoria das chamadas realizadas e identificar possíveis problemas nas diversas fases do processo de faturamento.

Existem hoje no mercado algumas ferramentas que fazem a geração de chamadas de teste (The AbacusTM System², Ameritec's Crescendo^{®3}, OSPREY⁴ etc.). A maioria destas ferramentas tem por objetivo principal testar a disponibilidade de serviços. Poucas ferramentas procuram identificar se os serviços testados estão sendo cobrados de forma correta. Mesmo as ferramentas que executam esta identificação o fazem de forma bastante primária, necessitando de uma grande interação com o usuário durante as análises do problema.

² The AbacusTM System é de propriedade da empresa Zarak Systems Corporation, <http://www.zarak.com/>

³ Ameritec's Crescendo[®] é de propriedade da empresa Ameritec Corporation, <http://www.ameritec.com/>

⁴ OSPREY é de propriedade da empresa Rotadata Ltd., <http://www.rotadata.com/>

1.4 Objetivo do trabalho

O objetivo deste trabalho é desenvolver um mecanismo de regras para análise de chamadas de testes buscando verificar problemas no fluxo de faturamento das operadoras de telecomunicações. Este mecanismo permitirá ao usuário construir regras a fim de automatizar o trabalho de análise. A análise aqui proposta é efetuada no caso típico, a partir dos níveis de corretude (ou de erro) de cada chamada de teste. Para isto são utilizados artifícios computacionais como a lógica fuzzy e a lógica tradicional (binária), associadas a uma máquina de inferência e a regras de produção.

1.5 Escopo do trabalho

A base teórica desenvolvida neste trabalho tem um maior enfoque na identificação dos erros gerados pela central telefônica, sem preocupar-se com os múltiplos registros de uma mesma chamada que serão gerados nas várias centrais por onde esta passou. Também não é considerada a divisão de uma chamada em vários registros. Considera-se que há apenas um registro gerado pelas centrais envolvidas para cada chamada realizada. Além do registro de chamada gerado pelas centrais telefônicas, há um segundo registro gerado por um equipamento gerador de chamadas de teste. Este par de chamadas é a entrada de dados do sistema de regras aqui apresentado. O tamanho da massa de dados necessária para uma análise, ou como estes são obtidos, não faz parte do escopo deste trabalho.

A construção das regras será de responsabilidade dos futuros usuários. Cabe a estes usuários, analistas de sistemas, especialistas em telecomunicações, auditores etc., transmitir ao conjunto de regras a inteligência necessária para que estas identifiquem erros nas chamadas ou nos equipamentos.

Contextualizando o trabalho na arquitetura TMN⁵, pode-se enquadrá-lo na arquitetura TMN como uma aplicação de gerência de serviços. Apesar deste enfoque maior à gerência de serviços, o mecanismo aqui demonstrado pode ser utilizado em aplicações de gerência do negócio, da rede e dos elementos de rede. Este trabalho trata

⁵ TMN (*Telecommunications Management Network*) é uma arquitetura padronizada que serve de modelo genérico para uma rede de gerência de telecomunicações, possibilitando o gerenciamento completo dos diversos recursos da rede [AMPESAN, 1999].

principalmente da área funcional de gerência de falhas, mas, devido a sua flexibilidade, pode ser utilizado em outras áreas funcionais.

1.6 Estrutura do Trabalho

O Capítulo 2 aborda os conceitos básicos de telecomunicações relacionados à bilhetagem e ao fluxo de faturamento das empresas operadoras de telecomunicações. No início do capítulo é descrito o processo de bilhetagem das centrais da telefonia fixa. A segunda parte é dedicada ao fluxo de faturamento das operadoras, em que se destacam também algumas características importantes do registro de chamada gerado pela central telefônica (CDR). Ao final deste capítulo comentam-se algumas características importantes sobre o processo de faturamento.

No capítulo seguinte apresenta-se uma introdução à lógica fuzzy e a regras de produção. Inicialmente são descritos os princípios de lógica fuzzy, conjuntos nebulosos e as principais operações de conjuntos fuzzy. Após isso, são abordadas algumas características de regras de produção. Finalmente, apresenta-se um exemplo de uma estrutura básica utilizada nos sistemas que empregam lógica fuzzy e regras de produção.

O quarto capítulo é dedicado à descrição do trabalho desenvolvido. O item 4.1 evidencia o contexto em que o trabalho está inserido. São descritos também a problemática e os mecanismos utilizados para se tentar resolver o problema. O final do item expõe a solução adotada. Comentam-se também os motivos da utilização das técnicas adotadas na solução e as limitações do trabalho desenvolvido. O item 4.2 apresenta o mecanismo de regras desenvolvido neste item. São propostas também duas formas de análise utilizando a base teórica de regras aqui descrita. Por fim, tem-se a demonstração de como seria a interação das regras com o ambiente das operadoras de telecomunicações.

O Capítulo 5 descreve um estudo de caso do trabalho. Após a apresentação do contexto do estudo de caso e dos mecanismos utilizados para sua implementação, são apresentados os resultados deste, de forma estatística e teórica.

O sexto capítulo apresenta as conclusões sobre o trabalho. Este capítulo inicia com um breve resumo do trabalho desenvolvido e as principais conclusões. Seguem-se

alguns comentários e sugestões para trabalhos futuros. Ao final, são apresentadas algumas características relevantes para a implementação deste trabalho em um ambiente real.

O Anexo 1 descreve uma proposta de linguagem a ser utilizada pela base teórica de regras proposta. O Anexo 2 demonstra as regras utilizadas nos estudos de caso do Capítulo 5. Os anexos 3, 4 e 5 demonstram os dados e resultados do estudo de caso.

Capítulo 2 - Bilhetagem e o processo de tarifação

Este capítulo descreve como as centrais telefônicas registram, atualmente, as chamadas realizadas pelos usuários do serviço de voz das companhias telefônicas. Destaca-se ainda um modelo genérico do fluxo de faturamento das operadoras de telecomunicações nacionais. Pretende-se, através deste capítulo, facilitar o entendimento da problemática que o trabalho está abordando.

2.1 Bilhetagem em centrais telefônicas de serviço de voz

Uma central telefônica executa uma série de papéis dentro dos serviços das operadoras de telecomunicações. Algumas de suas principais funções, se não as principais, são comutar chamadas telefônicas e efetuar o posterior registro destas chamadas.

Quando um cliente, o originador da chamada ou assinante A, vai efetuar uma ligação para um outro cliente, o destino ou assinante B, esse - assinante A -, basicamente, retira o telefone do gancho, disca o número de destino, utiliza o serviço e encerra a ligação. A central registra esta chamada gerando um bilhete chamado CDR (*Call Detail Recording*).

O CDR contém informações das mais variadas, tais como duração da chamada, tipo de chamada, rota usada, hora de início da ligação, etc. Após sua geração, o CDR é gravado em um arquivo que conterá um conjunto de chamadas de um determinado período e/ou quantidade. A este arquivo dá-se o nome de arquivo de bilhetes ou arquivo de CDRs. Estes arquivos não contêm apenas CDRs válidos⁶, mas também CDRs de chamadas não completadas ou que apresentaram problemas.

Os CDRs provenientes de chamadas não válidas não são utilizados para cobrar do cliente, mas servem, entre outras coisas, para medir desempenho, encontrar erros nas

⁶ CDRs válidos: são CDRs cuja chamada foi completada com sucesso, isto é, a origem conseguiu falar com o destino com sucesso.

centrais ou entender o tráfego telefônico. É a partir dos CDRs que se sabe qual(is) chamada(s) cada assinante efetuou ao longo de um determinado período.

O sistema de bilhetagem que utiliza CDRs para registro da chamada é chamado bilhetador automático (*Toll Ticketing Equipment*).

O procedimento de chamada descrito no início deste item é o caso geral, a maioria das chamadas realizadas segue este padrão. Mas há casos particulares como o das ligações locais⁷. Neste caso, o processo é um pouco diferente, cada usuário tem um contador de pulsos ou “relógio”, semelhante ao de energia elétrica⁸.

Simplificadamente, quando a chamada é estabelecida soma-se 1 (gera-se um pulso) ao contador. Nos primeiros quatro minutos é gerado um pulso aleatório⁹ e, após isto, gera-se um pulso a cada quatro minutos. O valor destes contadores de pulsos é obtido visualmente ou gerando-se um arquivo com o valor destes contadores. No final do mês diminui-se o valor do contador do mês anterior do valor corrente. Ao valor resultante aplica-se uma fórmula para saber quanto o usuário irá pagar pelo serviço. A este mecanismo de tarifação dá-se o nome de tarifação por multimedição.

Porém, há casos em que se habilita a geração de CDRs de chamadas locais, mas isto, em geral, está desabilitado nas centrais. Esta funcionalidade é particularmente útil nos casos em que se quer conhecer melhor o perfil dos consumidores, coletar estatísticas, etc. Quando se fala em Brasil, existe uma tendência de se registrar chamadas locais utilizando registros de CDR e não contadores de pulsos. Isto se deve ao fato de os contadores de pulsos aleatórios serem mecanismos confusos e duvidosos para a maioria dos usuários.

⁷ Ligações locais referem-se a chamadas nas quais tanto o assinante chamador como o chamado estão no mesmo município.

⁸ Este “relógio” pode ser um dispositivo físico, como os medidores de água ou energia elétrica, ou pode ser um dispositivo lógico, só existindo virtualmente na memória da central.

⁹ O pulso aleatório pode ocorrer do segundo 0 até o segundo 240, de acordo com o sincronismo do relógio da central com o início da chamada. Portanto um usuário pode, por exemplo, falar 2 segundos e pagar 2 pulsos e um outro pode falar 238 segundos e pagar um pulso.

Há ainda casos de centrais que não possuem software/hardware para gerar CDRs¹⁰. Neste caso pode-se utilizar os contadores de pulsos para cobrança de chamadas interurbanas¹¹. Este método de cobrança é denominado de multimedição por tempo e distância e praticamente não existe mais no Brasil.

Outro método ou critério de tarifação utilizado é a tarifação por chamada completada (*message rate*). É um critério em que se contam as chamadas originadas por um assinante e que foram atendidas pelo assinante ou serviço tarifável chamado. Não leva em consideração o tempo de duração da chamada nem a distância. Basta prover cada assinante com um contador de chamadas [FERRARI, 1991]. Este método de tarifação é comumente usado no serviço de auxílio à lista (102).

O mecanismo mais simples de tarifação é a tarifa fixa (*flat rate*). Cobra-se do assinante uma quantia fixa por um período de assinatura independente de quantidade, duração ou distância das chamadas. Não requer equipamento de tarifação.[FERRARI, 1998] Este método de tarifação é usado em pequenas comunidades com baixo tráfego telefônico, sendo pouco utilizado no Brasil. Um dado que vale a pena ressaltar é que, com a abertura de mercado, este método de tarifação pode ser uma opção para cobrança de chamadas locais. Há empresas como GVT (Global Village Telecom¹²) que oferecem planos com este tipo de tarifação¹³.

Os métodos descritos anteriormente não são regra e não cobrem 100% das possibilidades de tarifação. Dependendo do tipo de serviço usado, da tecnologia da central e das características da operadora, os métodos podem variar bastante. Normalmente, são estes os encontrados nas operadoras nacionais e em boa parte do mundo. Ressalta-se que, de modo geral, isto não é aplicável para serviços especiais

¹⁰ Na planta telefônica, encontrar centrais que não têm a capacidade de bilhetar é bastante normal. A solução normalmente adotada é utilizar uma outra central para bilhetar as chamadas das centrais que não possuem este recurso. Porém isto nem sempre é viável. Neste caso são utilizados mecanismos alternativos como os descritos neste texto.

¹¹ Chamada interurbana é quando um assinante A liga para um assinante B, sendo que o assinante A está em uma cidade diferente do assinante B.

¹² Global Village Telecom, <http://www.gvt.net.br/>

¹³ Isto não significa necessariamente que a central, para este tipo de serviço, não irá gerar os dados (CDRs), mas que o sistema de tarifação irá ignorar estes dados.

como 0300, 0800, 0900, siga-me etc. Para estes casos pode haver variações dos métodos descritos.

Existem várias formas de se categorizar centrais. Uma delas é dividir as centrais em dois tipos, de acordo com sua função de bilhetagem. As centrais que possuem hardware/software para gerar registros de chamadas são denominadas bilhetadoras. As que não possuem este mecanismo são denominadas não-bilhetadoras. Normalmente as centrais que possuem função de bilhetagem estão nos ramos mais centrais da rede telefônica. As que não possuem estão localizadas mais na periferia da rede, ou mais perto dos assinantes. A maioria das centrais locais¹⁴ não possui função de bilhetagem, embora seja comum encontrar exceções. As centrais locais que possuem função de bilhetagem normalmente bilhetam chamadas de outras centrais locais. Vale a pena lembrar que existem outros tipos de centrais como trânsito e tandem. Para saber mais sobre rede telefônica e classificação de centrais examine [FERRARI, 1991] e [SORTICA, 1999].

Outros conceitos importantes são os de chamada sainte e de chamada entrante. Chamada sainte é a que sai da central bilhetadora do assinante A para algum outro destino (centrais intermediárias ou a central de B). Chamada entrante é a que entra na central bilhetadora de B, ou seja, são chamadas provenientes de outras centrais que não sejam as centrais locais ligadas à bilhetadora de B nem à própria bilhetadora de B.

Antes da privatização das operadoras só eram gerados CDRs para as chamadas saintes. Após a privatização, são gerados CDRs tanto para as chamadas entrantes como para as saintes. Porém isto não é regra. Pode-se, por exemplo, não gerar chamadas entrantes na bilhetadora de B, gerando esta chamada na central trânsito¹⁵.

A Figura 1 demonstra um esquema bastante simplificado da rede telefônica explanando os conceitos apresentados.

¹⁴ Central Local ou Pública é a central telefônica na qual os assinantes estão conectados. No caso típico, os assinantes estão a um raio de até 6km da central [SORTICA, 1999].

¹⁵ Central trânsito ou IU (interurbana) comuta chamadas originadas em centrais locais ou provenientes de centrais tandem [SORTICA, 1999].

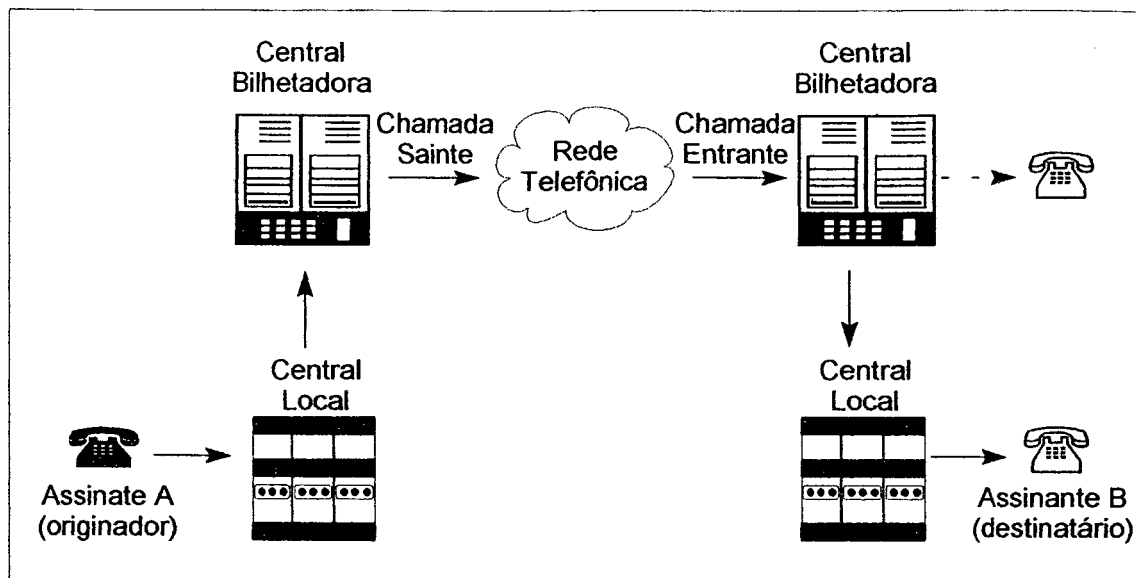


Figura 1 – Demonstração dos conceitos de central bilhetadora, local e chamadas saintes e entrantes

2.2 Processo de faturamento das chamadas provenientes de CDRs

Para transformar os arquivos de CDRs em contas telefônicas ou faturas, eles passam por uma série de sistemas desde sua geração na central até sua impressão e o posterior pagamento pelo usuário. Esta série de sistemas denomina-se processo ou fluxo de tarifação. O processo de tarifação pode ser visto como a produção de um bem qualquer. Entra matéria-prima, registro de chamadas telefônicas e demais serviços utilizados pelo cliente, e sai uma conta que deve ser paga pelo mesmo.

Os processos ou fases por que o CDR passa podem variar muito dependendo do tipo de serviço, das características da empresa e dos sistemas usados no processo. Não existe um modelo padrão deste processo, cada operadora o faz como melhor lhe convier. O que é demonstrado neste item é o que normalmente se utiliza nas empresas brasileiras e o que é considerado como modelo genérico.

No modelo genérico aqui apresentado, existem cinco fases básicas, são elas: geração dos bilhetes na central, mediação, valoração, encontro de contas e faturamento. A Figura 2 demonstra o fluxo do sistema de tarifação das operadoras bem como suas fases.

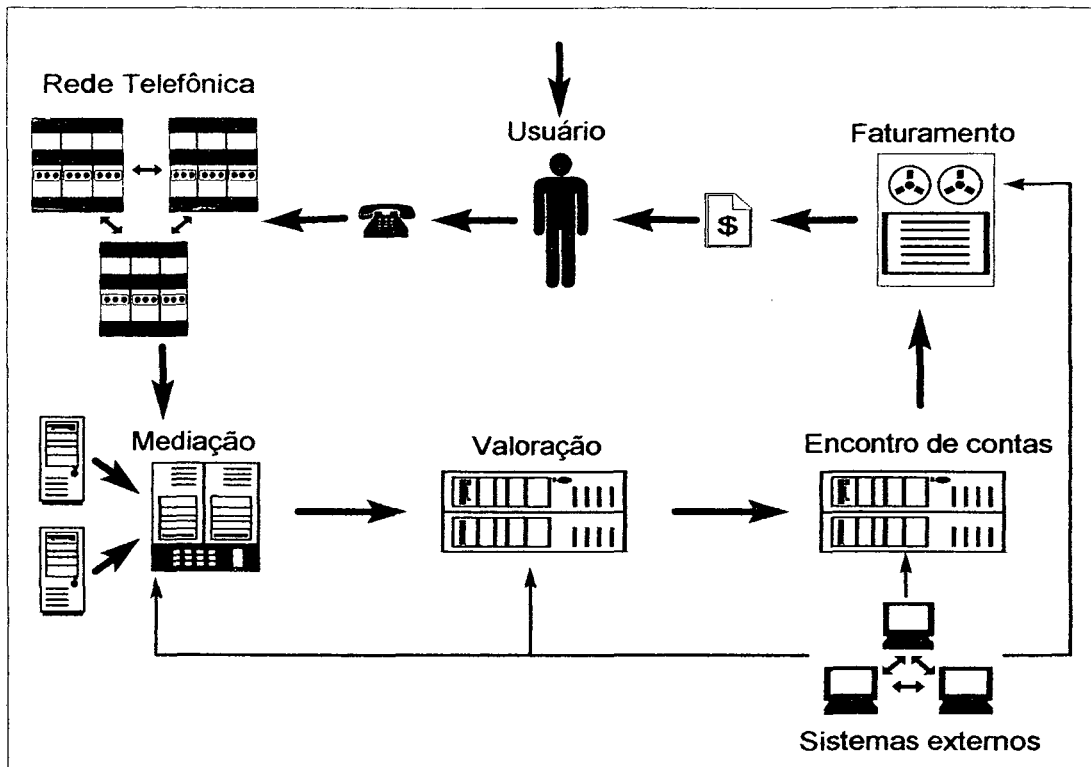


Figura 2 – Processo genérico de tarifação

2.2.1 Fases do processo

A primeira fase, ou geração dos bilhetes na central (ou na rede telefônica), consiste em criar um ou mais registros (CDRs) a cada chamada realizada. Este trabalho é feito normalmente após o encerramento da chamada e executado por um software proprietário de cada fabricante. O armazenamento do CDR gerado é feito em um meio magnético (normalmente fita ou disco rígido).

Na segunda fase ou mediação obtém-se o arquivo da central e faz-se a formatação/padronização destes CDRs para os vários sistemas que irão utilizá-los. A coleta dos arquivos de CDRs é feita de duas formas: (1) conectando-se à central via rede TCP/IP, X.25, FTAM (*File Transfer, Access and Management*) etc. e fazendo um *download* dos arquivos de bilhetes; (2) utilizando-se os dispositivos de fitas magnéticas/ópticas ou de rolo existentes na central.

Após a obtenção dos arquivos, inicia-se o processo de padronização dos mesmos. Isto é necessário, porque, em geral, cada central de fabricante diferente gera os CDRs com tamanhos diferentes, em formatos diferentes, com codificações diferentes. O arquivo padronizado fica então pronto para ser enviado para os sistemas que farão

uso de seus dados. Pode-se ainda anexar ao mediador funcionalidades como regra de validação dos CDRs¹⁶, filtro de bilhetes/campos a partir de regras, geração dos arquivos em formatos proprietários para cada sistema destino, uso de algoritmos de compactação para diminuir o volume de dados a ser transferido/armazenado, interfaces com banco de dados, etc.

Vale a pena ressaltar que o mediador ou *mediation device* tem a função de interface dos sistemas com a rede telefônica. O mediador não disponibiliza somente informações para o fluxo de faturamento. Muitas vezes é ele que disponibiliza informações para sistemas de análise de tráfego e sistemas antifraude. O mediador pode também trabalhar não somente com arquivos de CDRs, mas também com alarmes, medidas operacionais¹⁷, etc.

O mediador aqui descrito, assim como os outros sistemas das outras etapas, pode não ser um sistema único. Pode ser composto de uma série de sistemas executando em máquinas diferentes, de fabricantes diferentes, em locais diferentes e comunicando-se via rede ou através de qualquer outro meio.

Os arquivos de CDRs após passarem pelo mediador são enviados para os sistemas destino, em geral via rede TCP/IP. Estes sistemas destino, dentro do fluxo de faturamento, podem ser sistema de encontro de contas e/ou sistema de valoração de CDRs.

Depois de coletados e colocados em um formato padrão, os CDRs passam para a terceira fase, valoração ou *rating*. É nesta fase que se verifica se o bilhete coletado é válido para cobrança ao usuário e atribui-se um valor ao bilhete.

A primeira etapa desta fase, verificação dos bilhetes válidos, é basicamente um filtro. Nesta etapa retiram-se chamadas não completadas, chamadas de teste ou chamadas cuja duração é inferior à duração mínima para que seja cobrada. Há casos em que este filtro de bilhetes válidos é executado também no mediador, diminuindo o volume de dados que a fase de valoração terá de processar.

¹⁶ Por validação de CDRs entende-se o processo de verificação da geração do CDR de forma correta pela central. Existem muitos casos, principalmente em centrais mais antigas ou conectadas a centrais antigas, em que determinadas informações podem ser registradas inválidas ou nem serem geradas.

¹⁷ Medidas operacionais também conhecidas como medidas de tráfego ou OMs (*Operation Metrics*).

Após a definição de quais chamadas serão cobradas do usuário, é executada a segunda etapa deste processo, a valoração propriamente dita. Esta consiste em regras, fórmulas e cruzamento de tabelas, para saber quanto o usuário terá que pagar pela chamada executada.

Dependendo da tecnologia da central e do serviço utilizado, o custo da chamada pode estar incluso no CDR, evitando que isso tenha de ser calculado na fase de valoração. Porém isto é pouco comum. Nesta fase são consideradas iguais as chamadas de todos os usuários. Descontos, promoções e planos de cobrança especiais são aplicados em fases posteriores.

A fase seguinte, encontro de contas, não é necessariamente subsequente à valoração. Há casos em que esta fase une-se à de valoração, em outros casos é executada a partir de dados do mediador (em paralelo à valoração), e, em outros casos ainda, realizada por uma empresa externa à operadora, as chamadas “*cleaning houses*”. Nesta fase divide-se o valor dos CDRs entre as operadoras utilizadas para realizar a chamada. Verifica-se quem emitirá/cobrar a conta do usuário final, quanto caberá à(s) operadora(s) local(is) e quanto será pago à(s) operadora(s) interurbana(s) do valor total da chamada.

A tarifação, também chamada de faturamento ou *billing*, é a última fase do processo. Esta é a fase em que são somadas todas as chamadas que o usuário executou durante o mês, aplicando sobre estas os descontos. Além disso, incrementam-se tarifas mensais e/ou multas e somam-se outros serviços utilizados pelo cliente. Nesta fase também há a necessidade de um filtro para retirar chamadas de teste ou chamadas inválidas, caso as outras fases ainda não o tenham feito. Normalmente nesta fase há mais interação com sistemas externos, como cadastro de clientes, sistema contábil da empresa, SAC (Sistema de Atendimento ao Cliente), além dos sistemas das fases anteriores.

Outro fator diferenciador desta fase é que, nas fases anteriores, os CDRs são vistos como chamadas de um determinado número, enquanto nesta, são vistos como chamadas de um determinado cliente. Portanto, nesta fase pode haver a necessidade de agregar as várias contas, uma para cada número, em uma única conta para o cliente. Em

virtude disto, devem ser tratados casos como mudança de número de cliente, por exemplo. Se este usou o número N há cinco dias e o número X há 25 dias, estes valores devem ser somados na mesma conta. Encerrada esta fase, a conta é impressa e enviada ao cliente para que este efetue o seu pagamento.

O sistema de tarifação, em alguns casos, pode também fazer a valoração dos CDRs bem como fazer o encontro de contas. Pode ser ainda adicionada uma fase ao processo de tarifação, a fase de cobrança das contas. Nesta fase controla-se o envio das contas para os clientes, se os clientes pagaram ou não suas contas, etc. Em caso de não-pagamento ou de pagamento posterior à data de vencimento, o sistema de tarifação ou quaisquer outros sistemas devem tomar medidas cabíveis, alertar o não-pagamento, cobrar juros, cortar/religar o serviço, acionar mecanismo de cobrança judicial, etc. Além destas medidas, os sistemas de vendas devem evitar que o cliente tente reativar o serviço com outro nome, CPF (Cadastro de Pessoa Física), endereço etc.

2.2.2 Características do CDR

Um CDR é normalmente gerado quando se recebe ou quando se realiza uma chamada telefônica. Portanto, para cada chamada realizada podem ser gerados dois CDRs, um na central bilhetadora¹⁸ de origem e outro na bilhetadora de destino. Dependendo do número de pontos ou operadoras que a chamada percorre, podem ser gerados vários CDRs ao longo do trajeto. Por exemplo, se um assinante A ligar para um outro assinante B usando a operadora X (diferente da operadora de A e de B), será gerado, no mínimo, um CDR na central bilhetadora de A, um na de B e um na bilhetadora da operadora X.

Além de geração de CDRs duplicados, pode-se dividir o registro de uma chamada em *n* CDRs. Ao tipo de chamada que gera estes CDRs dá-se o nome de chamada de longa duração pois este caso normalmente só ocorre em chamadas com uma duração acima da média. Esta divisão de um CDR em vários CDRs é gerenciada, normalmente, por duas políticas. Na primeira, em determinados horários pré-configurados, gera-se um novo CDR. Por exemplo, se são seis horas e este é um dos

¹⁸ Central bilhetadora é uma central que tem a função de gerar os CDRs. Em geral, cada bilhetadora está associada a *N* centrais locais ou centrais que fazem somente comutação de chamadas.

horários pré-configurados, então para todas as chamadas com mais de trinta minutos será gerado um novo CDR. Na segunda política, se a chamada tiver duração superior a uma determinada duração pré-definida, gera-se um novo CDR. Por exemplo, se a chamada tem mais de uma hora de duração então um novo CDR é gerado ou se ocorrer um *handoff*¹⁹.

A adoção de uma ou outra política depende das capacidades dos modelos das centrais de trabalharem com um ou com outro tipo de política e das próprias políticas da empresa de telecomunicações. Outro motivo para geração de mais de um CDR para a mesma chamada é por comando do operador. Isto é normalmente executado em mudanças de horário de verão ou atualizações da central. Esta geração de CDR de forma não espontânea também depende da capacidade e da configuração da central. Um exemplo desta geração de CDRs é demonstrado a seguir:

Duração total da chamada: 1 hora e 12 minutos

Início da chamada: 12:01:00

Política: Se a chamada tem mais de 30 minutos gerar um novo CDR.

- CDR 1 gerado às 12:31:00 com duração de 30 minutos

- CDR 2 gerado às 13:01:00 com duração de 30 minutos

- CDR 3 gerado às 13:31:00 com duração de 12 minutos

O CDR gerado pela central possui uma grande variedade de informações. Dependendo do fabricante/modelo da central, da customização requerida pela operadora de telecomunicações ou do tipo de chamada ou serviço (chamada de telefone móvel para fixo, de móvel para móvel, de fixo para RI²⁰ etc.), estas informações podem variar muito. Em geral, encontram-se nos CDRs, no mínimo, as seguintes informações:

¹⁹ *Handoff* – transferência da ligação para uma outra célula [FERRARI, 1991], ou seja, mudança da ERB (Estação Radio Base) que servia o telefone móvel para uma outra que propicie uma maior qualidade da chamada telefônica. Esta operação é normalmente requerida quando há deslocamento do celular durante uma chamada. Para informações mais precisas e completas consultar: [FERRARI, 1991].

²⁰ RI (Rede Inteligente) – tecnologia de rede utilizada para prover serviços especiais ou diferenciados. Dois exemplos clássicos são os conhecidos serviços 0800 e 0900.

- Assinante A – número do telefone que realizou a chamada;
- Assinante B – número do telefone que recebeu a chamada;
- Duração – duração da chamada realizada;
- Data/Hora – data e hora de início da chamada;
- Usuário a tarifar – define que usuário irá pagar a chamada, particularmente útil em chamadas a cobrar;
- Status da chamada – define como a chamada acabou ou se não foi completada e, neste caso, por que não foi. Os *status* clássicos são: OK – chamada com sucesso, LO – linha ocupada, CO – congestionamento, PAB – perda do assinante A para o assinante B, DSC – assinante A desconecta antes que B atenda, NR – assinante B não responde e OU – outros problemas que fizeram com que a chamada não fosse completada ou efetivada com sucesso;
- Indicador de fonte – indica se esta chamada é entrante ou sainte, ou seja, indica se é uma chamada cuja origem é a central bilhetadora do assinante A ou B ou uma das centrais entre os dois assinantes;
- Custo da ligação – indica quanto o cliente terá de pagar pela chamada realizada;
- Prestadora²¹ – indica qual operadora de longa distância foi utilizada para realizar a chamada, caso isto se aplique.

As informações descritas anteriormente são as que comumente encontram-se CDRs, porém há casos em que estes campos não são obtidos de forma direta, por exemplo os campos Indicador de fonte e Custo da ligação podem aparecer de forma explícita ou serem calculados a partir de outros campos. Este cálculo de novos campos pode ser realizado em qualquer uma das fases do processo. Isto dependerá das características da operadora e/ou dos sistemas adotados. Há ainda informações que podem vir em formatos diferentes dependendo da configuração e/ou do modelo da central. Por exemplo, uma determinada central pode não possuir a capacidade de gerar a hora de início da chamada, mas somente a hora de fim. Ou em chamadas de longa

²¹ O campo “prestadora” só existe em chamadas provenientes de telefones fixos para fixos ou para móveis.

duração, pode colocar como hora de início da chamada a hora da geração do primeiro CDR gerado ou a hora de início do CDR corrente.

Outra característica interessante é que determinados campos do CDR podem não ser gerados pela central que registrou o CDR. Um caso típico é o campo Assinante A em uma chamada provinda de uma outra central. Neste caso, a central que originou a chamada passa o valor deste campo para a central destino. Se a central de origem está com um valor errôneo deste campo, esse será passado à frente, e será registrado de forma errada em todas as centrais bilhetadoras por onde esta chamada irá passar. Isto é mais um agravante para a complexidade do processo, pois uma central pode estar gerando informações erradas em virtude de um erro em outra central. Além disso, a central que está gerando o problema pode ser de outra empresa, o que dificultará e atrasará mais ainda a solução do problema.

2.3 Considerações

Um fator importante quando se fala em fluxo de faturamento é o tempo. A ANATEL²² só permite que chamadas nacionais sejam cobradas até 90 dias após sua realização e chamadas internacionais até 180 dias. Estes prazos, em um primeiro momento, podem parecer razoáveis. Mas se for considerada toda a complexidade do processo de faturamento descrita anteriormente, este tempo não é tão razoável assim. Suponha-se o caso onde duas operadoras nacionais registram um conjunto de chamadas de forma diferente. Até que se consiga identificar exatamente que chamadas estão com problema, onde estão os problemas, qual das operadoras cometeu o erro, e fazer com que estas chamadas passem novamente por parte do fluxo de faturamento e a nova conta chegue ao usuário, já decorreu um tempo considerável. Em muitos casos é mais fácil e barato para as operadoras ignorar estes erros, mesmo que isso signifique a perda de dinheiro.

²² ANATEL (Agência Nacional de Telecomunicações), <http://www.anatel.gov.br/>

Uma vez entendido melhor como funciona o ambiente de tarifação das operadoras de telecomunicações, o próximo capítulo pretende dar um apanhado geral de lógica fuzzy e regras de produção. Este embasamento é utilizado no Capítulo 4, onde é proposto o mecanismo de regras, tema deste trabalho.

Capítulo 3 - Lógica fuzzy e regras de produção

Neste capítulo são apresentados conceitos sobre lógica fuzzy e regras de produção. Este capítulo pretende fornecer o embasamento teórico necessário para o melhor entendimento dos conceitos utilizados nos capítulos posteriores.

3.1 Introdução à lógica fuzzy

Aristóteles, filósofo grego (384 - 322 a.C.), foi o fundador da ciência da lógica, e estabeleceu um conjunto de regras rígidas para que conclusões pudessem ser aceitas como logicamente válidas. O emprego da lógica de Aristóteles levava a uma linha de raciocínio lógico baseado em premissas e conclusões. Por exemplo: se for observado que “todo ser vivo é mortal” (premissa 1), a seguir é constatado que “Sarah é um ser vivo” (premissa 2), como conclusão temos que “Sarah é mortal”. Desde então, a lógica ocidental, assim chamada, tem sido binária, isto é, uma declaração é falsa ou verdadeira, não podendo ser ao mesmo tempo parcialmente verdadeira e parcialmente falsa. Esta suposição e a lei da não contradição, que coloca que “U e não U” cobrem todas as possibilidades, formam a base do pensamento lógico ocidental.[GSI, 2001]

Porém existem casos em que a lógica binária não é adequada. O conceito de “verdadeiro” e “falso”, principalmente, para variáveis lingüísticas não se mostra adequado. Por exemplo, uma pessoa medindo 1,70 m é alta ou baixa? E uma de 1,75m? E de 1,71? Se usarmos a regra de:

Se altura > 1,70 Então pessoa é alta Senão pessoa é baixa

Usando a regra descrita acima, uma pessoa com 1,71m é alta e uma com 1,69m é baixa. Porém estas duas pessoas possuem praticamente a mesma altura. Neste exemplo vê-se como a lógica binária pode não se adequar a determinadas situações.

Na lógica fuzzy o conceito de “verdadeiro” e “falso” é mais amplo, existe um grau de verdade associado à premissa. Por exemplo, uma pessoa com 1,71m possui um grau de “alto” de 0,7 e de “baixo” de 0,3, uma de 1,69m possui um grau de “alto” de

0,69 é de “baixo” de 0,31. Neste exemplo observa-se que o grau de pertinência da pessoa de 1,70m e da de 1,69 ao grupo dos altos são bastante próximos, coisa que não ocorre com a lógica binária. Além disso, a lógica fuzzy mostra-se eficiente quando utilizada para tratar de artifícios lingüísticos como “muito”, “pouco”, “não muito”. Estes artifícios associados a “alto” e “baixo” geram afirmações como “ele é pouco baixo”, “ela não é muito alta”, que podem ser mapeados para graus diferentes da lógica fuzzy.

De forma mais objetiva e preliminar, pode-se definir lógica fuzzy (difusa ou nebulosa) como sendo uma ferramenta capaz de capturar informações vagas, em geral descritas em uma linguagem natural, e convertê-las para um formato numérico, de fácil manipulação pelos computadores de hoje em dia. [GSI, 2001].

A lógica em questão foi desenvolvida por Lofti A. Zadeh da Universidade da Califórnia em Berkeley na década de 60, e combina lógica multivalorada, teoria probabilística, IA (Inteligência Artificial) e redes neurais para que possa representar o pensamento humano [GSI, 2001].

Devido às características da lógica fuzzy, esta é usada em várias áreas como:

- Processamento de linguagem natural;
- Sistemas especialistas;
- Robótica;
- Processos de tomada de decisão;
- Sistemas parcialmente abertos;
- Reconhecimento de padrões;
- Tratamento de imprecisões;
- Sistemas de inteligência artificial.

3.2 Conjuntos nebulosos

A lógica fuzzy está fundamentada na teoria de conjuntos nebulosos que foi criada por Lofti A. Zadeh em 1965 [ZADEH, 1965]. Na teoria clássica dos conjuntos,

os conjuntos são ditos “*crisp*”, de tal forma que um dado elemento do universo em discurso (domínio) pertence ou não pertence ao referido conjunto. Na teoria dos conjuntos nebulosos isso é um pouco diferente. Existe um grau de pertinência de cada elemento a um determinado conjunto. Zadeh em [ZADEH, 1973]. definiu os conjuntos difusos como uma classe com graus contínuos de pertinência, ou seja, dado um elemento a , do espaço A , pertencente a um conjunto difuso X (percebe-se aqui a relação com a teoria clássica dos conjuntos) existe uma função característica, ou de pertinência, dada por $f_A(a)$, que associa a esta pertinência um valor real no intervalo $[0,1]$. Uma descrição formal, descrita em [CROSS, 1994], utiliza a seguinte notação:

$$F_d = \{ \mu_{FI}(d,w)/(d,w) \mid d \in D \text{ e } w \in W \}$$

Onde F_d é o conjunto fuzzy de termos (w) que descrevem o documento d ; cada par (d,w) possui um grau de relação (entre zero e um) que é a relação entre o documento d e o termo w , e é expresso pela função $\mu_{FI}(d,w)$.

Mais simplificadamente, pode-se definir conjuntos fuzzy como:

$$f_A(x): U \rightarrow [0, 1]$$

Quando o conjunto de elementos é discreto, pode-se representar o conjunto fuzzy por um vetor:

$$A = (a_1, a_2, a_3, \dots, a_n) \quad \text{onde } a_i = f_A(x_i)$$

Pode-se também associar a cada elemento x_i o valor da pertinência a_i o símbolo “/”.

$$A = (a_1/x_1, a_2/x_2, a_3/x_3, \dots, a_n/x_n)$$

A notação padrão de conjuntos fuzzy representa a união das dimensões do vetor, utilizando a notação booleana de união, representada pelo símbolo “+” da seguinte forma [SPECIALSKI, 2000]:

$$A = (f_1/x_1, f_2/x_2, f_3/x_3, \dots, f_n/x_n) \quad \text{ou} \quad A = \sum f_i/x_i, i = 1, \dots, n$$

Se X é uma função contínua, então o conjunto A pode ser representado como:

$$A = f_x f_A(x_i)/x_i^{23}$$

3.3 Inferência fuzzy

A lógica fuzzy trata um conjunto fuzzy como uma proposição fuzzy. Uma proposição fuzzy é uma asserção que estabelece um valor para determinada variável lingüística, tal como 1,70 é alto. Normalmente uma proposição fuzzy é representada por [SPECIALSKI, 2000]:

U é A , onde A é o conjunto fuzzy sobre o universo de discurso U .

Uma regra fuzzy relaciona duas proposições da seguinte forma:

Se X é A Então Y é B por exemplo:

Se taxa de erros é alta Então qualidade do canal é ruim

A partir de uma regra ou conjunto de regras fuzzy, da mesma forma da lógica tradicional, pode-se tirar conclusões, como:

²³ Os Símbolos ‘+’, ‘ Σ ’ e ‘ f ’ não representam os símbolos ‘soma’, ‘somatória’ e ‘integral’ da matemática tradicional. Neste item representam elementos da lógica fuzzy.

Implicação: Se X é A Então Y é B

Premissa: X é A

Conclusão: Y é B

Implicação: Se taxa de erros é alta Então qualidade de canal é ruim

Premissa: taxa de erros é alta

Conclusão: qualidade do canal é ruim

3.4 Intersecção, união e complemento

Na matemática, a intersecção é formada pelos elementos pertencentes aos dois conjuntos que estão sendo analisados. Para representar a intersecção de conjuntos fuzzy é utilizado o menor dos valores dos conjuntos fuzzy, ou seja, dados os conjuntos **A** e **B**, sua intersecção será o menor valor de f_A e f_B para cada um dos valores de **A** e **B**. Na lógica fuzzy utiliza-se do símbolo ‘ \wedge ’ (**AND** ou **E**) para representar a função MIN (mínimo). A intersecção de dois conjuntos é expressa por:

$$f_A(x) \cap f_B(x) = f_A(x) \wedge f_B(x) = f_{A \wedge B}(x) = \text{MIN}(f_A(x), f_B(x))$$

Análoga à intersecção, a união é executada em lógica fuzzy pela função MAX. A função MAX é representada graficamente pelo símbolo ‘ \vee ’ (**OR** ou **OU**). A seguir é demonstrada a operação de união entre dois conjuntos A e B.

$$f_A(x) \cup f_B(x) = f_A(x) \vee f_B(x) = f_{A \vee B}(x) = \text{MAX}(f_A(x), f_B(x))$$

Outro conceito importante é o de complemento de conjuntos fuzzy. Para se representar o complemento utiliza-se o símbolo \sim (NOT ou negação). Esta operação é realizada utilizando a equação:

$$\bar{f}_A(x) = f_A(x) - 1$$

Segue um exemplo de união, intersecção e complemento em lógica fuzzy de forma gráfica (Figura 3) e outro utilizado conjuntos de valores:

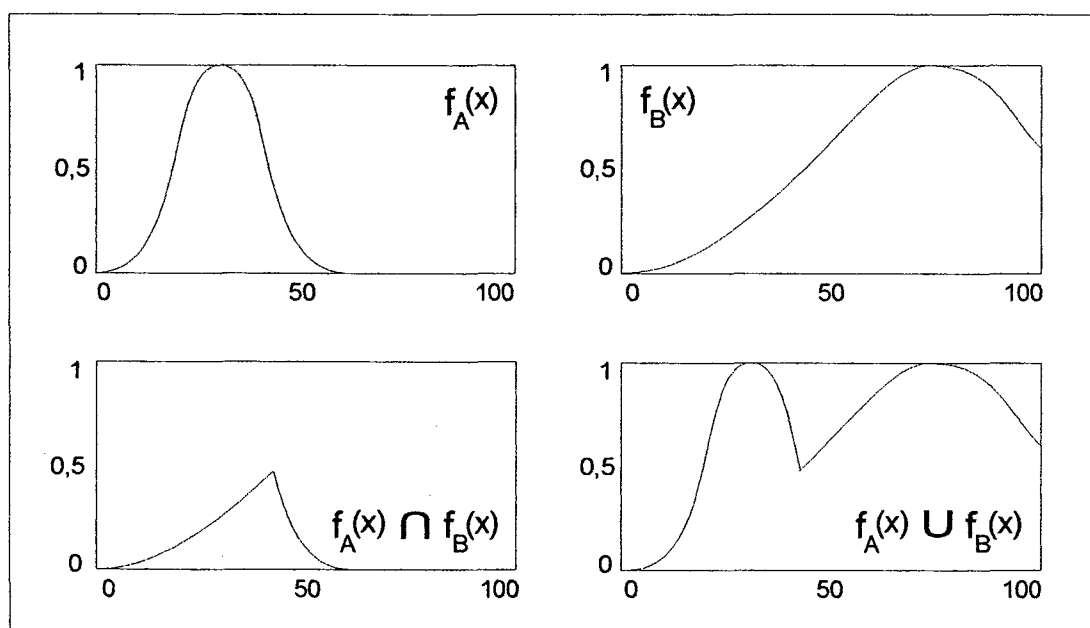


Figura 3 – Exemplos gráficos de união e intersecção de conjuntos fuzzy [SOUZA, 2001]

$$A = (0/5, 0.2/10, 0.4/15, 0.5/20, 0.3/25)$$

$$B = (0.7/5, 0.4/10, 0.3/15, 0.2/20, 1/25)$$

$$f_{A \wedge B}(x) = \text{MIN}(f_A(x), f_B(x)) = (0/5, 0.2/10, 0.3/15, 0.2/20, 0.3/25)$$

$$f_{A \vee B}(x) = \text{MAX}(f_A(x), f_B(x)) = (0.7/5, 0.4/10, 0.4/15, 0.5/20, 1/25)$$

$$\bar{f}_A = (1/5, 0.8/10, 0.6/15, 0.5/20, 0.7/25)$$

3.5 Regras de produção

Nas décadas de 1950 e 1960, muitos trabalhos foram desenvolvidos tendo como finalidade a busca por soluções de problemas gerais. Neste tema, dois pesquisadores deste período, Newell e Simon, se destacaram escrevendo uma série de artigos, depois reunidos no livro “Resolvendo Problemas Humanos” [NEWELL, 1972].

O resultado mais significativo do trabalho de Newell e Simon foi a demonstração de que muitas das soluções dos problemas humanos, frutos da cognição, poderiam ser expressas através de regras de produção na forma *Se-Então* (*If Then*). Segue um modelo básico de uma regra no formato *Se-Então*:

Se condição(ões)/premissa(s) *Então* Ação(ões)

A partir dos trabalhos de Newell e Simon, o uso de expressões no formato *Se-Então* se popularizou e é usado até hoje para resolver vários tipos de problemas.

O conceito de regras de produção²⁴ é genérico e define um conjunto de regras formais para se atingir um objetivo específico. De forma geral, as regras de produção podem ser utilizadas tanto com a lógica fuzzy como com a lógica binária (ou tradicional). Uma das primeiras uniões regras de produção a lógica fuzzy foi em 1974 nos trabalhos de Assilian e Mamdani [TANAKA, 1997]. Eles demonstraram que as regras nebulosas *Se-Então* podiam controlar o funcionamento de uma máquina a vapor [TANAKA, 1997] [TSOUKALAS, 1997]. Desde então, os controladores nebulosos têm despertado um crescente interesse.

Neste trabalho, o conceito de regras de produção está associado a um conjunto de regras dentro da estrutura *Se-Então*. A estrutura *Se-Então* é baseada no termo lógico da lógica proposicional²⁵ $P \rightarrow Q$ (P implica Q). Segue um exemplo de regra:

Se a rua está fechada *Então* não há carros trafegando

O termo **P** da sentença ou cláusula acima seria: a rua está fechada e o termo **Q** seria: não há carros trafegando.

²⁴ Neste trabalho quando for mencionado o termo “regras de produção” deve-se considerá-lo para regras utilizando tanto lógica binária como fuzzy

²⁵ Para maiores detalhes sobre lógica proposicional consultar [RABUSKE, 1995] e/ou [SUJIT, 2001].

Uma regra estabelece um relacionamento entre uma cláusula (asserções ou fatos) e, dependendo da situação, pode ser usada para gerar uma nova informação ou provar a verdade de uma asserção [CAROLINE, 2001].

A estrutura *Se-Então* é o método mais comum de representação do conhecimento em sistemas especialistas²⁶ [MAIA, 1998]. A um conjunto de N regras de produção se dá o nome de base de regras ou base de conhecimento. Esta base de regras tem, como objetivo, representar, de forma sistemática, o controlador (especialista) que gerenciará o sistema sob sua supervisão [DRIANKOV, 1996].

Um das principais vantagens deste tipo de estrutura é que é fácil de ser criada e entendida por pessoas [CAROLINE, 2001]. Outra vantagem é que a interpretação/tradução destas regras para o entendimento do computador é uma tarefa, na maioria dos casos, simples e rápida. A velocidade de execução de um sistema utilizando regras *Se-Então* é bastante satisfatória em termos computacionais, principalmente se as regras são traduzidas e compiladas em uma linguagem nativa do computador.

Tanto o termo **P** como o **Q** podem ser termos compostos. No termo **P** faz-se uso de conectores lógicos (conjunções): \neg (negação), \vee (Ou), \wedge (E). No termo **Q** não se utiliza nenhum conectivo específico, todas as conclusões da regra serão tidas como verdadeiras. Um exemplo de uso de termos compostos é mostrado a seguir:

Se José é casado com Maria E Maria é mulher Então José é homem, Maria é casada com José.

3.6 Sistemas fuzzy

Sistemas fuzzy são sistemas baseados em regras que usam lógica fuzzy, preferivelmente a lógica booleana, para tomar decisões [CAROLINE, 2001].

Mais formalmente, um sistema fuzzy é formado pelo conjunto de regras de inferência que mapeia regiões no espaço das entradas $U_1 \times U_2 \times \dots \times U_r$ para regiões

²⁶ Segundo Edward Feigenbaum [FEIGENBAUM, 1982], define-se um Sistema Especialista como sendo um “programa computacional que usa o conhecimento e procedimentos de inferências para resolver problemas complexos que exigem especialistas humanos em suas soluções”, ou seja, um sistemas especialista é um sistema computacional que procura comportar-se como um especialista humano na tomada de decisão.

no espaço das saídas $V_1 \times V_2 \times \dots \times V_m$, sendo que os conjuntos U_i e V_i são conjuntos fuzzy. A Figura 4 mostra um sistema nebuloso de duas entradas e uma saída. Assim, este sistema mapeia regiões do espaço R_2 para o espaço R_1 . Ele é formado por "n" regras que codificam o conhecimento de um especialista descrito como "Se X_1 é A_i e X_2 é B_i Então Y é C_i ". Como exemplo de tal conhecimento, poder-se-ia ter: "Se a renda do cliente é alta e sua dívida é pequena, então seu crédito é muito alto". Desta forma, a variável lingüística X_1 estaria associada à renda do cliente (Cliente.renda); a variável lingüística X_2 estaria associada à dívida do cliente (Cliente.dívida); a variável lingüística Y estaria associada ao crédito do cliente (Cliente.crédito); o conceito de renda alta estaria associado a um conjunto fuzzy A_i , digamos A_3 (alta.renda); o conceito de dívida pequena estaria associado a um conjunto nebuloso B_i (pequena.dívida) e, finalmente, o conceito de crédito muito alto estaria associado a um conjunto nebuloso C_i (muito_alto.crédito). Do ponto de vista da lógica nebulosa e neste exemplo, as variáveis lingüísticas e os conjuntos nebulosos X_1 e A_i , X_2 e B_i , e Y e C_i pertencem aos universos de discurso U , V e W , respectivamente [SOUZA, 2001]. A Figura 4, a seguir, demonstra um exemplo de sistema fuzzy.

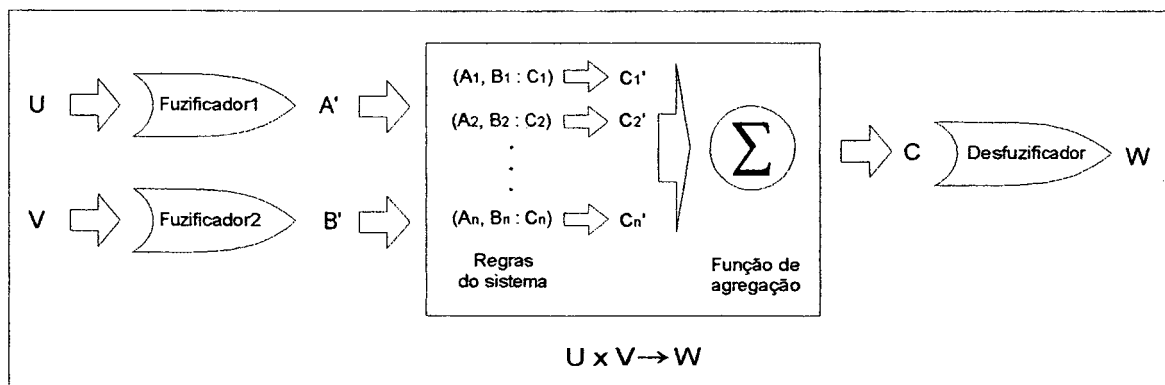


Figura 4 – Exemplos de sistema fuzzy com duas entradas e uma saída²⁷

Para se obter o valor de resposta, interagem as entradas A' e B' com as n regras aqui representadas de forma compactada $(A_i, B_i : C_i)$. As saídas destas regras são C_i' . Os conjuntos C_i' são somados logicamente para produzir a saída C . Além disto, os valores brutos de entrada (u e v) utilizam funções fuzificadoras para mapear os valores

²⁷ Figura adaptada do trabalho [SOUZA, 2001] Figura 6 de Carlos Pimentel.

dos conjuntos U e V nos conjuntos A e B . Depois de obtida a resposta C utiliza-se uma função desfuzificadora para mapear C em W . A idéia básica do fuzificador e do desfuzificador é mapear dados coletados (dados não fuzzy, entre 0 e 1) para valores fuzzy ou o contrário. Estes elementos são colocados em geral na entrada e na saída do sistema fuzzy permitindo que internamente se trabalhe somente com valores fuzzy.

O exemplo aqui apresentado representa somente um caso de sistema nebuloso com duas entradas e uma saída. Normalmente o mapeamento é feito de n entradas para uma saída. Porém pode se mapear as n entradas para m saídas. No exemplo demonstrado aqui foram utilizados mecanismos fuzificadores e desfuzificadores, no entanto obtêm-se entradas entre 0 e 1 e/ou se a saída do sistema foram valores entre 0 e 1 podemos dispensar estes mecanismos.

Klir e Yuan (1995) apud [VIEIRA, 1999] destacam também que alguns controladores difusos não fuzificam as variáveis de entrada, sendo estas colocadas diretamente na máquina de inferência, como fatos. Com isso pode-se ter máquinas de inferência que são híbridas ou em dois estágios (um híbrido e outro só fuzzy), trabalhando tanto com valores fuzzy como valores não fuzzy.

Uma outra característica importante das máquinas de inferência é a capacidade de guardar informações para as várias interações do sistema, ou seja, ter determinadas variáveis que são residentes do sistema, por exemplo: um sistema fuzzy X gera informações da qualidade da rede a cada trinta segundos. O sistema fuzzy poderia ter uma variável “qualidade da rede”, com isto poder-se-ia saber qual é o nível de qualidade da rede não somente nos últimos trinta segundos, mas nas últimas duas horas, por exemplo. As máquinas de inferência podem ter várias outras características particulares dependendo das necessidades da aplicação. As características aqui apresentadas não limitam nem encerram o assunto.

Uma diferença entre regras de produção tradicionais e regras de produção fuzzy é o fato de que os termos de uma asserção fuzzy retornam valores entre 0 e 1 e são utilizados os conectores lógicos da lógica fuzzy.

Segue um exemplo de aplicação de uma regra fuzzy:

Se a taxa de erros é alta E a informação trafegada é importante *Então* status da rede é ruim

Fuzificadores:

$$\text{Taxa de erros alta}(N) = \begin{cases} 0 & \text{se } N = 0 \\ N/30 & \text{se } 1 \leq N \leq 30 \\ 1 & \text{se } N > 30 \end{cases}$$

$$\text{Informação importante}(I) = \begin{cases} 0,3 & \text{se } I \text{ é informação operacional} \\ 0,7 & \text{se é informação financeira} \\ 1 & \text{se é informação gerencial} \end{cases}$$

Supondo os valores de $N = 15$ e Tipo da informação = financeira

a taxa de erros é alta E a informação é importante

$$\text{Taxa de erros alta}(N) \wedge \text{Informação importante}(I)$$

$$0,5 \wedge 0,7 = 0,5$$

$$\text{Então: rede é ruim } (x) = 0,5$$

Assim, o resultado da regra seria, segundo os dados de entrada, que o grau de pertinência ao conjunto das redes ruins é de 0,5.

Uma vez mais bem compreendidos o ambiente de telecomunicações e a lógica fuzzy, o próximo capítulo descreve uma solução para os problemas levantados no Capítulo 1, utilizando estes conceitos.

Capítulo 4 - Método proposto para identificação de problemas

O presente capítulo expõe a solução proposta para o problema. No início deste, descreve-se o problema, de forma aprofundada. São descritas também algumas soluções utilizadas no mercado para tentar resolver o problema. Em seguida a solução proposta é apresentada, bem como suas limitações e o motivo para o emprego das técnicas descritas. O mecanismo de regras é definido na seqüência. Demonstram-se também dois tipos de análises que podem ser feitas com a base teórica apresentada. No final do capítulo apresenta-se um exemplo de como os mecanismos descritos poderiam ser inseridos no ambiente de telecomunicações.

4.1 Contexto do trabalho

Este item aborda o problema a ser tratado. Inicialmente, são apresentados alguns números que refletem a gravidade do problema, além dos principais pontos de erro ao longo do processo de tarifação. Em seguida, são descritas as principais soluções adotadas pelo mercado para solucionar o problema. E, por fim, são apresentados os motivos que levaram à escolha das técnicas utilizadas na solução, bem como as limitações e restrições adotadas neste trabalho.

4.1.1 Contexto do problema

O faturamento é um dos setores mais importantes de uma empresa de telecomunicações. A partir deste setor são geradas as contas telefônicas para que os clientes façam o pagamento. Erros neste setor normalmente geram muitos prejuízos. Considerando o alto faturamento das operadoras, estes erros podem representar somas consideráveis. Em termos numéricos, as operadoras deixaram de faturar no ano passado o equivalente a 5% de sua receita bruta, com fraudes e perdas no processo de cobrança de contas dos clientes. Isto decorre da falta de uma política adequada de garantia de receita (*revenue assurance*). O cálculo foi feito com base no faturamento estimado pelo

Yankee Group²⁸ para o ano 2000, envolvendo as operadoras de telefonia fixa, de telefonia móvel e de dados. De acordo com estudo da KPMG Consulting²⁹, realizado no final do ano 2000, esses números podem ser ainda bem maiores, considerando-se que os índices de perdas no Brasil, assim como em outros mercados latino-americanos, podem chegar a 15% em casos mais críticos [MOURA, 2001].

Todavia, uma parcela substancial das suas perdas de receita decorre de falhas de procedimentos internos ou erros em todo o fluxo de faturamento. Enquanto a fraude de assinantes corresponde a cerca de 1% da receita, a perda provocada pelas falhas e erros de procedimentos pode chegar a 4% da receita de uma operadora [FOINA, 1998].

Este problema não ocorre apenas no Brasil. Em uma pesquisa realizada nos Estados Unidos pela PriceWaterhouseCoopers³⁰ junto a 250 operadoras indicou que 41% das empresas pesquisadas perdem 6% ou mais de suas receitas por ano [MOURA, 2001].

Seguem algumas tabelas que exemplificam melhor estas perdas. A Tabela 1 descreve os índices encontrados normalmente nas operadoras da América Latina. A Tabela 2 descreve os serviços que possuem um maior índice de perdas nas operadoras. Na Tabela 3 demonstram-se alguns problemas encontrados nas faturas telefônicas.

Tabela 1 - Índices de perdas típicos das operadoras da América Latina

Categoria	Percentual
Médias dos percentuais de perdas totais de receita das operadoras	3% a 15%
Erros de reciclagem de dados de chamada	0,3% a 1%
Perdas de CDRs desconhecidos	1%
Terminais ativos, porém não bilhetados	0,1% a 0,2%
Terminais bloqueados, porém ativos	14%

Fonte: KPMG Consulting [MOURA, 2001]

²⁸ The Yankee Group, <http://www.yankeegroup.com.br/>

²⁹ KPMG Consulting, Inc. <http://www.kpmg.com/>

³⁰ PriceWaterhouseCoopers Consulting <http://www.pwcconsulting.com/>

Tabela 2 - Serviços não faturados

Categoria	Percentual
Identificador de chamadas	12%
Chamada em espera	5%
Reversão de créditos em disputas	até 80%

Fonte: KPMG Consulting [MOURA, 2001]

Tabela 3 - Irregularidades em faturas

Categoria	Percentual
Erros nos valores	12,9%
Erros de classificação	3,2%
Problemas de formatação	até 40%
Variações em interconexão entre redes	até 15%

Fonte: KPMG Consulting [MOURA, 2001]

Segundo Paulo Rogério Foina [FOINA, 1998], os principais pontos potenciais de perda de receita podem ser:

- 1) Captura dos CDRs;
- 2) Tarifação dos CDRs;
- 3) Configuração das tabelas internas do sistema de faturamento;
- 4) Inconsistência entre cadastro de assinantes do sistema e tabelas de assinantes das centrais;
- 5) Ajuste ou desconto concedido sem autorização;
- 6) Faturas não processadas por inconsistência de dados;
- 7) Faturas não impressas;
- 8) Faturas não recebidas pelos assinantes;
- 9) Cobrança efetuada (pela agência de cobrança) e não repassada.

Além destes destacam-se:

- 1a) Erros na geração dos CDRs pela central;

2a) Diferença entre os registros das chamadas (CDRs) gerados por diferentes operadoras/centrais;

3a) Erros provocados por provedores de serviços externos (terceirização);

4a) Fraudes geradas por usuários mal intencionados;

5a) Descartes errôneos de CDRs tanto na mediação como na tarifação;

6a) Erros de impressão.

A Figura 5 demonstra os principais pontos de perda de receita.

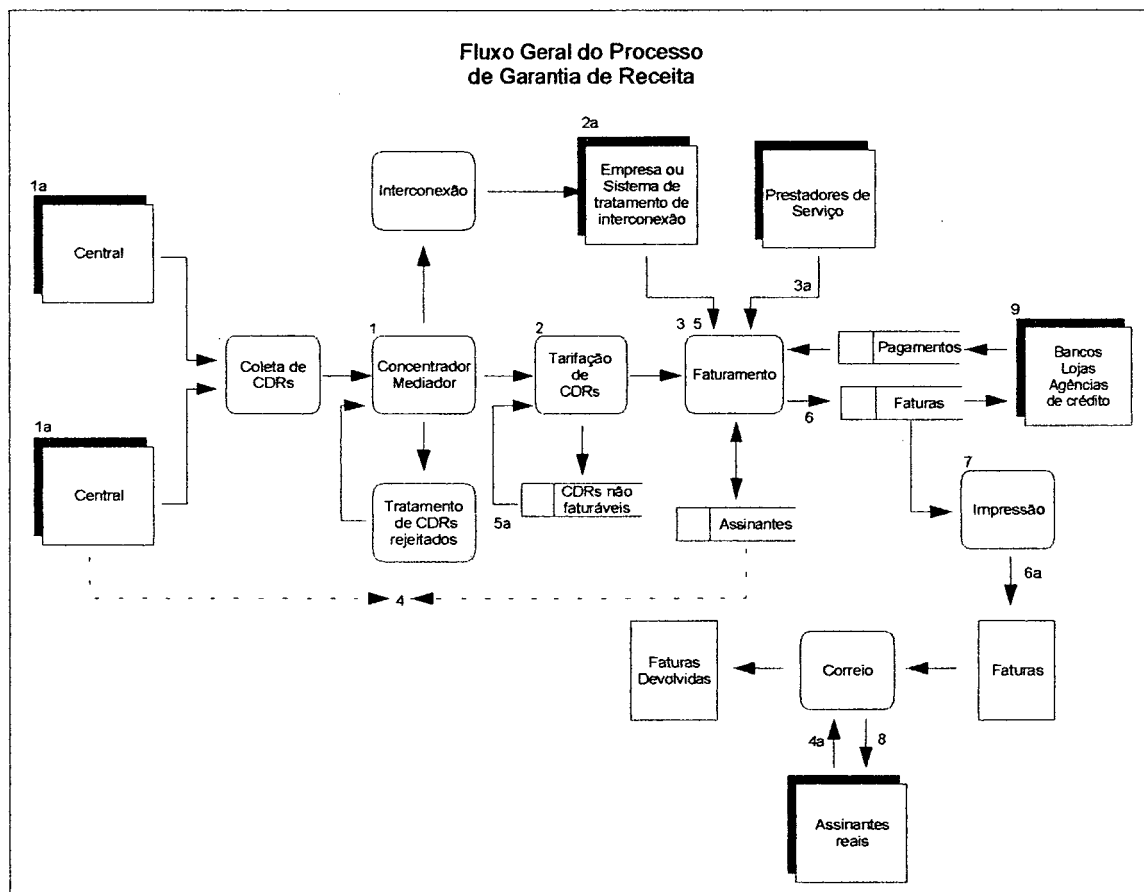


Figura 5 – Pontos de perda de receita³¹

Podem existir outros pontos de perda de receita que dependerão da operadora de telecomunicações e do fluxo do processo de faturamento adotados. Segundo Vicente

³¹ Figura adaptada do artigo [FOINA, 1998] de Paulo Rogério Foina

Paiva Lima [MOURA, 2001], há cerca de 35 itens do ciclo produtivo das operadoras em que se pode atuar para evitar evasão de receita.

Analisando mais a fundo o ambiente das companhias de telecomunicações brasileiras, pode-se classificar os erros no processo de tarifação³², descritos, em três grupos. São eles:

- Erros de geração do bilhete pela central;
- Erros ao longo do processo de conversão de CDRs em contas telefônicas;
- Erros de impressão/envio das contas aos assinantes.

Os principais motivos para a geração de erros no processo são falta de sincronismo ou desatualização de tabelas, erros operacionais humanos, problemas com equipamentos e alta complexidade do processo (muitas variáveis em um grande volume de dados)³³.

Existem muitos fatores que contribuem para que o sistema de cobranças das operadoras não funcione adequadamente. Alguns desses principais fatores estão listados a seguir³⁴:

- União/desmembramento³⁵ das operadoras nacionais;
- Saída de funcionários por corte de pessoal ou saída voluntária;
- Necessidade de integrar as várias soluções “caseiras” a sistemas corporativos ou de transformar soluções “caseiras” em soluções corporativas;
- Necessidade de adequar sistemas, equipamentos e funcionários a novas realidades como internet, intranet, seleção de prestadora pelo cliente, tarifação de IP (*Internet Protocol*), 2,5G, 3G etc.;
- Necessidade de interoperar equipamentos ou sistemas muito antigos a equipamentos extremamente modernos e/ou de fabricantes diferentes;
- Necessidade de adequar regras de negócio, sistemas e funcionários para oferecer aos clientes planos diferenciados de serviços e cobranças;
- Mudança de enfoque de prestadora de serviço público para operadora privada;

³² Excluem-se deste item problemas gerados após o envio da conta ao usuário.

³³ No Capítulo 1, descrevem-se com mais detalhes os prováveis motivos que levam as operadoras de telecomunicações a ter estes problemas.

³⁴ Alguns dos fatores mencionados podem não ter contribuído diretamente com os problemas de tarifação mas foram citados por terem mudado as características das empresas ou do mercado de telecomunicações.

³⁵ União/desmembramento se refere à separação das antigas operadoras estaduais em operadoras de telefonia fixa e de telefonia móvel e às fusões das mesmas com companhias de outras regiões ou estados.

- Implantação de novos serviços;
- Substituição dos antigos funcionários “políticos” (nomeados pelo governo) por funcionários provindos de outras empresas e de outros países;
- Surgimento da concorrência;
- Problemas de hardware/software diversos;
- Uso de protocolos nem sempre seguros ou adequados à função;
- Falta de integração ou duplicidade das bases de dados;
- Falta de sincronismo entre o que há na planta e o que há nas bases de dados;
- Fraudes geradas por usuários mal intencionados.

4.1.2 Abordagens para o problema

Para o cenário descrito fica clara a necessidade da utilização de mecanismos para tentar diminuir este problema. Genericamente, os processos que geram as perdas de receita podem ser divididos em três grandes grupos: *Service Fulfillment*, *Service Assurance* e *Service Billing*. O *Service Fulfillment* implementa o provisionamento oportuno e correto das solicitações feitas pelo cliente. Começa com atividades de pré-vendas e estende-se até a correta instalação dos serviços requisitados, ou seja, o cliente satisfeito. O *Service Assurance* são ações que asseguram o perfeito funcionamento da rede, possibilitando a entrega ao cliente final do serviço com a qualidade esperada. O grupo que atualmente merece maior esforço no combate as perdas é o *Service Billing* [CPqD, 2001]. Este consiste em controlar o fluxo de faturamento assegurando que o cliente irá pagar exatamente os serviços por ele utilizados.

Observando as soluções/necessidades existentes no mercado de telecomunicações no processo de *Service Billing*, identifica-se a existência de seis grupos de ferramentas para controlar/identificar falhas no processo [CPqD, 2001]: auditorias especializadas, ferramentas de controle do processo, identificação de fraudes, monitoramento baseado em perfil histórico, identificação de erros em base de dados e execução de testes controlados.

O primeiro grupo, auditores externos, consiste na contratação de especialistas com alto grau de conhecimento e experiência. Este tipo de ferramenta de identificação normalmente é a primeira etapa de um processo de garantia de receita. Com base nestas informações decide-se como será atacado o problema.

As ferramentas de controle do processo, segundo grupo, procuram identificar problemas no fluxo dos arquivos do processo de tarifação³⁶. Sumarizando, estas ferramentas fazem a verificação de quantos arquivos/bilhetes/registros entraram e quantos arquivos/bilhetes saíram de cada fase do fluxo de faturamento. Além disso, é feita uma espécie de fluxo de caixa com os arquivos e suas informações tentando encontrar sistemas ou pessoas que possam estar executando tarefas de forma errônea ou inadequada. Um exemplo de sistema de controle de processo é o Sub-sistema de Gestão Flexível de Tarifários³⁷.

As ferramentas do terceiro grupo, identificação de fraudes, são os chamados sistemas antifraudes. As soluções desta categoria são mais comumente encontradas em companhias de telefonia celular ou grandes companhias de telefonia fixa. Estes sistemas procuram identificar usuários, ou não-usuários, mal intencionados tentando burlar as companhias de telecomunicação, levando vantagem com isso. Neste grupo de ferramentas trabalha-se em geral com dados amostrais. De forma geral estas ferramentas utilizam históricos e perfis dos clientes para identificar possíveis desvios de comportamento e inferir possíveis fraudes. Estas ferramentas podem ir um pouco além disto, verificando integridade de arquivos, comparando resultados de meses anteriores, volumes de tráfegos em centrais, etc. Este subconjunto de ferramentas trabalha, normalmente, com toda a massa de dados do faturamento. Alguns exemplos de sistemas antifraude são: CFMS³⁸ e AntiFraude³⁹.

O quarto grupo, monitoramento baseado em perfil histórico, é bastante parecido com o anterior. Abstraindo particularidades, a diferença básica é que um tenta identificar fraudes enquanto o outro tenta identificar erros no fluxo de faturamento. Estes sistemas procuram, através do uso de sistemas do tipo *tending*⁴⁰, montar perfis de clientes, equipamentos ou sistemas a fim de identificar problemas. Um caso clássico

³⁶ Os arquivos analisados são normalmente arquivos com o registro das chamadas, mas esta classe de ferramentas pode trabalhar com arquivos com registro de pulsos (ver Capítulo 3), configurações da planta, tabelas de entroncamento entre as centrais telefônicas, prefixos das centrais etc.

³⁷ Sub-sistema de Gestão Flexível de Tarifários é de propriedade da empresa Integer - Informática de Telecomunicações e Gestão SA, <http://www.integer.pt/>

³⁸ CFMS (Compaq Fraud Management System) é de propriedade da empresa Compaq Computer Corporation, <http://www.compaq.com/>

³⁹ AntiFraude é de propriedade da empresa Innovatec, <http://www.innovatec.com.co/>

⁴⁰ Sistemas do tipo *tending* são sistemas de análise de tendências

deste tipo de ferramenta são os sistemas de “gap”, que procuram ao longo do tempo desvios no número de CDRs gerados.

As ferramentas do quinto grupo, identificações de erros em base de dados, procuram encontrar erros de sincronismo, incompatibilidade ou falta de integridade de base de dados. Estas ferramentas checam se as bases de todos os sistemas do fluxo de faturamento ou de toda a operadora estão íntegras. Estas checagens podem ir desde a base de dados da central telefônica até a base de dados dos clientes, passando por sistemas de contabilidade, SAC, planta da rede, etc. Um exemplo de sistema de verificação de base de dados é o Sagent's Suite⁴¹.

O último grupo, execução de testes controlados, basicamente, faz a execução de testes controlados e parte da idéia de se gerar um conjunto de ligações controladas⁴² e verificar se estas foram executadas e registradas de forma correta. Estas ferramentas trabalham em geral com dados estatísticos em ambientes controlados tentando simular clientes ou tráfego telefônico. Em sua maioria necessitam de hardwares específicos para gerar o tráfego telefônico⁴³ a ser utilizado na análise. Estes sistemas normalmente testam as primeiras fases do processo de faturamento. Em sua maioria, não têm enfoque no faturamento das empresas, mas sim em medir a disponibilidade e a qualidade do serviço prestado do ponto de vista do cliente e da empresa prestadora do serviço. Estas ferramentas possuem enfoque principal no teste por elas executado e normalmente possuem mecanismos pouco eficientes de análise e comparação de dados, quando possuem estes mecanismos. Na existência de mecanismos de análise, estes requerem grande interação com o usuário, bem como usuários com grande conhecimento em várias áreas da empresa para trabalhar com tais mecanismos. Desta forma, o trabalho de identificar problemas no processo de tarifação, por ser manual, torna-se árduo e bastante complicado para o usuário.

⁴¹ Sagent's Suite of Business Intelligence Applications é de propriedade da empresa Sagent Technology, <http://www.sagent.com/>

⁴² Definem-se aqui como ligações controladas as ligações em que se determina de antemão sua duração, data e hora de início e demais parâmetros da ligação. Estas ligações podem também resultar em falhas, de forma pré-determinada.

⁴³ Entende-se por hardwares para gerar tráfego telefônico hardwares/softwarewares com a capacidade de gerar e receber ligações utilizando a rede telefônica convencional.

Para um melhor entendimento do atual status dos sistemas de geração de testes, foram analisadas algumas ferramentas de mercado. Nos parágrafos posteriores são listadas as principais características destas.

O primeiro equipamento analisado foi o Abacus™ System, da Zarak Systems Corporation (<http://www.zarak.com/>). Dentre os equipamentos estudados, este é o que possui os maiores recursos, no que se refere à variedade de testes que podem ser executados. Em termos de análise, este é o mais simples dos equipamentos. Tem-se a seguir uma lista das principais características do Abacus™ System:

- Enfatiza a execução dos testes;
- Enfoca em testes de disponibilidade de serviços e em testes de qualidade do serviço;
- Permite que o equipamento simule tráfego de usuário ou uma central. Ou seja o equipamento pode simular uma central ou um usuário ou conjunto de usuários para a central que se está executando o teste;
- Possui uma API (*Application Programming Interface*) permitindo que o usuário desenvolva seu próprio gerador de chamadas de teste ou sistema auditor;
- Permite configurações precisas (volume, tom etc.) dos testes através de scripts;
- Permite implementação de novos protocolos usando interface gráfica;
- Permite, através de equipamentos de hardware, trabalhar com vários dispositivos diferentes (E1, T1, ISDN, etc) e com crescimento adaptável às necessidades;
- Possui software para configuração e obtenção de resultados simples (sem grandes recursos gráficos e facilidades ao usuário);
- Não compara seus dados com os de outras fontes. Isto só é possível com o auxílio da API e com implementação do usuário;
- Permite obter resultados e estatísticas em tempo real durante a execução dos testes;
- Permite implementar protocolos proprietários via *scripts*;

- Seus dados de saída podem ser em arquivos texto, tabelas ou gráficos;
- Permite integração com geradores de testes de outros fabricantes.

A segunda análise foi na família de produtos da Ameritec Corporation, <http://www.ameritec.com/>, o Ameritec's Crescendo® Niagara®, Squirte® e FeatureCall®. Os dois primeiros produtos, Crescendo® Niagara® e Squirte® são os geradores de testes. Estes produtos são compostos pelo hardware/software para geração dos testes. O último produto FeatureCall® é o gerente de execução de testes. Ele permite gerenciar os geradores de teste de forma centralizada. Com isso pode-se ter vários geradores de teste espalhados geograficamente e somente um gerente, possibilitando ao usuário executar vários testes em vários locais sem a necessidade de se deslocar até o gerador de teste. Seguem algumas características do produto:

- É um mecanismo de análise simples;
- Enfoca em testes de disponibilidade de serviços e em testes de qualidade do serviço;
- Permite gerar alguns relatórios ou dados em arquivos de *log*;
- Não compara seus dados com os de outras fontes;
- Permite implementar protocolos proprietários via *scripts*;
- Seus dados de saída podem ser em arquivos texto, tabelas, gráficos ou relatórios configuráveis;
- Permite obter resultados e estatísticas em tempo real;
- Permite que determinados parâmetros do teste possam ser modificados durante a execução do mesmo;
- Permite utilizar equações baseadas em campos estatísticos, tempo e operadores matemáticos (+, -, *, /) para geração de relatórios de análise;
- Permite geração de gráficos customizados;
- Permite a geração dos testes via *scripts*;

- Faz parte de uma família de produtos com arquitetura modular e flexível, permitindo que o sistema cresça conforme a necessidade da empresa.

O último produto analisado é o OSPREY da Rotadata Ltd., <http://www.rotadata.com/>. Dos produtos analisados é o único que foi elaborado com o objetivo principal de comparar os dados do sistema de teste com os dados do faturamento. Segue algumas características:

- Tem comparação e análises de pontos do faturamento;
- Requer programação manual ou integração via banco de dados para unificar os dados do faturamento com os dados gerados pelo sistema;
- Seu gerador de teste é o mais simples, se comparado com os outros produtos especializados em teste;
- Permite customizações via software (não se comenta nada sobre análise);
- Divide os resultados dos testes em arquivos de resultado do tipo: com chamadas OK, com erro, com informações de tempo e tráfego de cada teste;
- Permite adicionar aos dados obtidos do gerador de teste dados do custo da chamada. Os dados do custo da chamada são calculados pelo sistema e editados pelo usuário via interface gráfica;
- Suas saídas são pré-formatadas para disponibilização dos resultados em interface web;
- Necessita de antecedência de no mínimo 24h para execução dos testes;
- Não permite a configuração de análises automáticas sem necessidade de programação ou customizações;
- É uma arquitetura cliente/servidor modular e adaptável às necessidades dos usuários.

No que se refere, mais especificamente, à análise de dados para identificação de erros no processo de faturamento, nota-se que as ferramentas analisadas possuem mecanismos primários de análise. Ou seja, boa parte da tarefa mais complexa de análise fica a cargo do analista (auditor e/ou operador do sistema). Com este alto grau de

dependência e exigência do analista, sendo este um ser humano, é pouco provável que, utilizando-se estas ferramentas, seja possível identificar um alto índice de erros ou trabalhar com grandes volumes de dados. São estes problemas, entre outros, que este trabalho pretende resolver ou minimizar.

Nos parágrafos anteriores foram apresentados seis grupos de ferramentas para controlar/identificar falhas no fluxo. Porém existem várias outras soluções que podem ser utilizadas para atingir o objetivo de maximizar o lucro das operadoras. Alguns exemplos de outras ferramentas são: sistemas de faturamento paralelo ao oficial e conferência manual de contas. Os grupos de ferramentas ou soluções mencionadas ao longo deste capítulo não são soluções excludentes, podendo-se adotar todas ou apenas uma delas. São formas complementares de se atingir o objetivo. O mais importante é tentar cercar o problema por todos os lados.

Salienta-se aqui que sistemas como os descritos são aplicáveis não somente no controle de um processo existente de tarifação, mas também na implantação de novos serviços ou planos de tarifação. Quando novos serviços ou planos de tarifação são implantados, normalmente, isto requer mudanças manuais em equipamentos, bem como adição ou atualização dos existentes. Esta mudança no ambiente pode provocar erro não somente no que se está implantando, mas também em serviços ou planos de tarifação existentes. Este controle de novos serviços é extremamente importante, pois existe uma tendência de que as operadoras de telecomunicações, cada vez mais, gerem novos planos ou serviços, chegando até o ponto de atendimento individualizado das necessidades dos clientes.

4.1.3 Abordagem proposta para o problema

Das seis categorias de soluções descritas, são utilizadas neste trabalho as ferramentas de execução de testes controlados. Este trabalho apresenta um mecanismo de análise das chamadas de teste, buscando identificar se estas possuem problemas e quais são os problemas.

Mais especificamente, este mecanismo serve como um mecanismo de auditoria no processo de geração dos CDRs, com a função de identificação de desvios de determinados valores do bilhete. Com isso, pretende-se oferecer uma orientação ao

auditor (sendo este ser humano ou máquina) de quais erros estão acontecendo, sem que este tenha de fazer a verificação chamada a chamada (como fazem alguns sistemas que possuem mecanismos de análise).

A identificação de problemas ou análise a ser realizada terá como base dois registros da chamada, o CDR gerado pela central e um registro gerado de uma outra forma (hardware, manualmente etc.). Este dois registros serão da mesma chamada. O primeiro será gerado pelos meios tradicionais, ou seja, gerado pela central bilhetadora. O segundo utilizar-se-á de um sistema paralelo à tarifação, um gerador de chamadas de teste. Como estes dois registros são gerados em locais diferentes, existe a necessidade de algum mecanismo de associação dos dois registros da chamada (para saber que CDR corresponde exatamente ao registro gerado de forma alternativa). A Figura 6 demonstra um caso hipotético de geração dos dados de entrada do mecanismo aqui proposto.

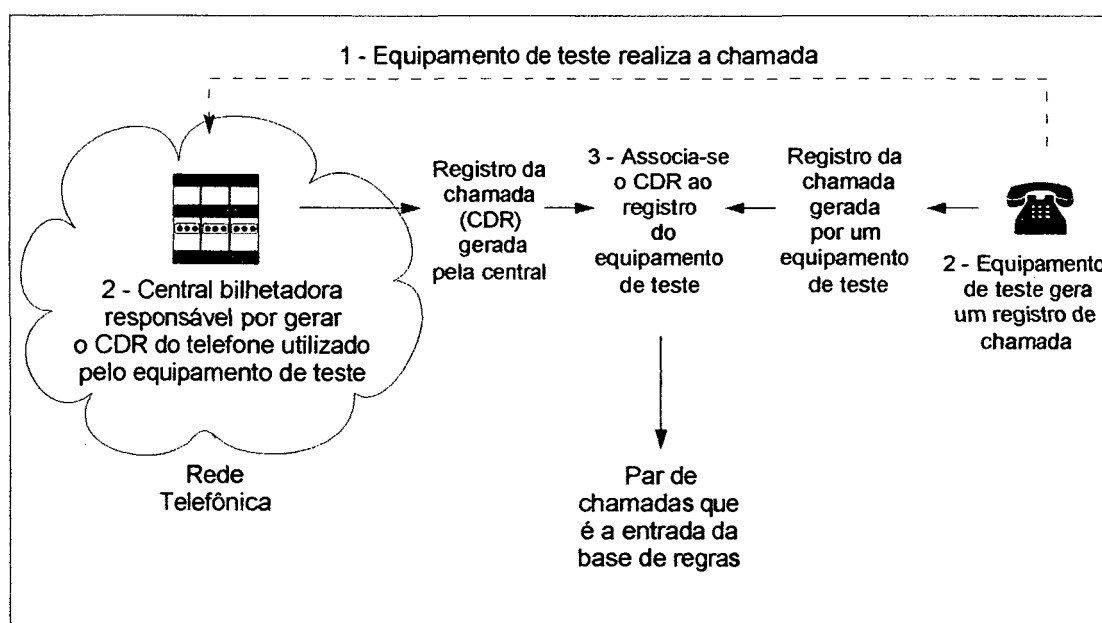


Figura 6 – Esquema hipotético de geração dos dados

Uma vez obtido o par registro/CDR e existindo vários destes pares, obtém-se uma tabela (base de dados) com as chamadas de teste realizadas. De posse desta base de dados de chamadas de teste, surge a questão e uma das motivações deste trabalho: de que forma identificar que fase(s) possui(em) problema(s)? A forma mais primária de fazer isto é através de um operador humano. Este verificaria, chamada a chamada, se

esta saiu da central de forma correta e se, durante as fases da tarifação, não ocorreu nenhum erro, para, a partir disso, proceder à análise. Porém, esta forma de identificação de problemas é bastante limitada, principalmente quando se considera o alto volume de chamadas de teste e a complexidade para identificar os erros, bem como as próprias falhas do operador humano. Nestes casos, esta forma primária tornar-se-ia praticamente inviável. Uma forma alternativa seria utilizar processos computacionais para identificar falhas no processo de tarifação.

Na literatura pesquisada até esta data, não foi encontrado nenhum algoritmo, ou ferramenta, que possa atender a todas as necessidades das operadoras de telecomunicações. As pesquisas que fizeram com que se chegasse a esta conclusão se deram através da internet e de conversas informais com técnicos, supervisores e consultores na área de telecomunicações. O que se encontra no mercado de telecomunicações são sistemas não automáticos de identificação de falhas e que, em geral, possuem apenas enfoque nos testes. Alguns exemplos destes sistemas foram mencionados no item 4.1.2.

O mecanismo aqui proposto para a análise é composto de duas fases. A primeira fase consiste em identificar o quanto cada um dos campos do par de CDR/registo está correto. Uma vez obtido o nível de correção de cada um dos campos, inicia-se a segunda fase ou processo de análise. Esta fase dependerá das necessidades do usuário. Vale a pena ressaltar que estas duas fases são propostas, o mecanismo resultante deste trabalho permite outros tipos de análise, de acordo com as necessidades do usuário.

Para se calcular o grau de corretude (ou de erro⁴⁴) de uma chamada utilizou-se técnicas de lógica fuzzy. Esta etapa, o cálculo de grau de corretude, é equivalente à etapa de fuzificação em sistemas fuzzy tradicionais. Em casos particulares, esta fuzificação pode não ser necessária como, por exemplo, em aplicações que não necessitam trabalhar com lógica fuzzy. Ressalta-se que o mecanismo de regras a ser apresentado permite trabalhar simultaneamente com a lógica fuzzy e com a lógica tradicional.

⁴⁴ O grau de corretude indica o quanto uma chamada ou campo está correto. Neste trabalho, este termo está associado ao intervalo de $[0,1]$ e pode ser substituído por grau de erro, ressaltando que um é o complemento-um do outro (grau de erro = $1 - \text{grau de corretude}$).

Escolheu-se utilizar estas técnicas, pois, para o problema em questão, não é adequado o uso de lógica tradicional. Para este problema, respostas binárias (0 ou 1) são insuficientes. Existe a necessidade de uma maior flexibilidade de resposta: valores fuzzy, variáveis lingüísticas, contadores etc. A necessidade de uso de fuzzy pode ser exemplificada como segue: supondo que um usuário realize uma chamada de 10 minutos e a central registre 9 minutos e 30 segundos. A central registrou esta chamada de forma correta? E se fossem registrados 9 min e 58 seg? Nota-se aqui que esta diferença é melhor representada através da lógica fuzzy que por valores 0 e 1 da lógica tradicional.

Como alternativa à lógica fuzzy e à lógica tradicional poderiam ser utilizados mecanismos de faixa de valores para cada nível de erro, porém a lógica fuzzy é mais rica que este mecanismo. Com o mecanismo de faixas ocorreria o mesmo problema de valores de fronteira da lógica binária. Por exemplo, define-se duas faixas, uma de 0 a 5 e outra de 6 a 10. O valor 5 estaria na primeira faixa, porém muito próximo da segunda faixa. Neste trabalho é interessante que o grau varie conforme variem os valores, ou seja, não utilizar faixas fixas de valores. Poderia-se utilizar também algum mecanismo de média aritmética para evitar o problema de faixas fixas. No entanto, este mecanismo de médias não possuiria o embasamento matemático que a lógica fuzzy possui, como os operadores E e OU, etc.

Além disso, em aplicações reais deste trabalho, muito provavelmente, será necessário o auxílio de um especialista que, não necessita ser grande conhecedor de programação ou informática. O especialista utilizará termos como: “campos com muita diferença” ou “Se isto ocorrer, então o problema é grave”. Estes são exemplos típicos de situações onde a lógica fuzzy é bastante utilizada.

Outra característica importante que não é abordada a fundo no trabalho é a desfuzificação dos valores resultantes. Optou-se por não tratar este assunto, pois se acredita que os possíveis usuários do sistema, por estarem acostumados a trabalhar com valores numéricos ou por estes sistemas serem computacionais, preferirão receber valores fuzzy a valores desfuzificados. Mas, se houver a necessidade da etapa de

desfuzificação, a base teórica aqui apresentada contemplará este requisito completamente, ou necessitará de uma pequena adaptação.

Optou-se também pelo uso de regras de produção no formato *SE-ENTÃO*, pois estas se adaptam perfeitamente à lógica fuzzy. São também as mais bem sucedidas formas de representação do conhecimento. Elas são fáceis de criar e são entendidas por pessoas em geral [CAROLINE, 2001]. Tanto para um especialista leigo em informática quanto para o desenvolvedor fica mais fácil trabalhar com regras deste tipo, pois estas facilitam bastante o mapeamento do conhecimento do especialista para o sistema. Uma outra vantagem do mecanismo de regras proposto é que, geralmente, sistemas que fazem uso de regras são facilmente modificáveis com simples edição de texto. Esta facilidade é visível quando se compara o mecanismo proposto aos sistemas tradicionais. O que é bastante interessante, principalmente, na implantação da solução, e, também, na adaptação desta solução às constantes mudanças no setor de telecomunicações.

A máquina de inferência que é tomada como base neste trabalho é do tipo híbrida (não trabalha só com a lógica fuzzy). Optou-se por este tipo de máquina de inferência, pois para máquinas de inferência totalmente fuzzy trabalhar com variáveis de status ou operações matemáticas simples não seria viável. Outra vantagem é que usando uma máquina de inferência híbrida não é necessário mapear todas as variáveis para fuzzy. Algumas variáveis, como tipo da chamada, são mais facilmente trabalhadas de forma direta (sem a fuzificação).

Portanto, é utilizada uma associação de regras de produção *SE-ENTÃO*, lógica fuzzy e uma máquina de inferência híbrida para se identificar os CDRs com problema e, a partir destes dados, realizar análises.

4.1.4 Restrições adotadas neste trabalho

Admite-se que este trabalho manipula somente as chamadas que geram CDRs. As chamadas que utilizam pulsos (contadores) para fazer a cobrança do usuário bem como outros serviços que não sejam o de telefonia fixa não são tratados por este trabalho, embora os conceitos aqui apresentados não se limitem somente à telefonia fixa, a uma fase do processo de tarifação ou a determinado serviço. É importante salientar que a base teórica aqui apresentada também se adapta a boa parte dos serviços

existentes hoje no mundo de telecomunicações. Apesar do enfoque do trabalho estar em um tema específico, este tem aplicações em outros setores de telecomunicações e em outras áreas, que não a de telecomunicações. As demais aplicações deste trabalho serão comentadas no Capítulo 6.

O mecanismo de associação do CDR ao registro não é tratado neste trabalho. Este mecanismo consiste em encontrar os pares de chamadas a partir de um dado conjunto de chamadas registradas pela central e do mesmo conjunto registrado pelo equipamento de teste. Considera-se que o par de chamadas já tenha sido associado por algum mecanismo externo antes de passar pelo conjunto de regras.

O trabalho enfatiza a identificação de erros no nível de central telefônica. Esta ênfase deve-se ao fato de que, segundo [CPqD, 2001], a maioria dos erros estão concentrados nas primeiras fases do processo de faturamento.

Além disso, não existe a preocupação, neste trabalho, com a identificação das causas do erro ou com o ponto exato do erro. Esta identificação do ponto da rede telefônica que gerou o erro e sua respectiva causa poderão ser temas de outros trabalhos. O trabalho aqui apresentado tem a função de identificar os campos do CDR que estão com erro e calcular um grau de corretude de uma chamada ou conjunto de chamadas. São apresentadas duas formas de como realizar análises para tentar identificar problemas ou extrair informações, a partir dos dados de corretude dos campos, análise horizontal e análise vertical (ver item 4.2).

Não faz parte do escopo deste trabalho o destino dos resultados das saídas das regras. São aqui indicadas algumas possíveis saídas. Optou-se por retirar este item do trabalho, pois a forma como estes dados serão utilizados dependerá da situação e do objetivo das regras implementadas. Por exemplo, se o resultado das regras é apresentado diretamente ao usuário, ou através da geração de alarmes, pode haver a necessidade de uma desfuzificação. Enquanto que, se os dados forem enviados para um sistema de correlação de eventos, os dados de saída poderão ser enviados em valores fuzzy.

Pressupõe-se neste trabalho que cada chamada gerará somente um CDR, ou seja, não haverá registro de CDRs em mais de um ponto na mesma empresa nem CDRs divididos em c CDRs (comumente gerados em chamadas de longa duração).

4.2 Mecanismo de regras

Como apresentado, o trabalho pretende, a partir de um par de registros de uma chamada (uma registrada pelo fluxo normal de faturamento e outra registrada de outra forma), identificar, em um primeiro momento, o grau de corretude de cada campo do CDR. Uma vez obtido este grau, executa-se a segunda etapa, a análise. Esta análise pode ser realizada de várias formas. As duas formas aqui propostas são: (1) horizontal, que dá enfoque ao cálculo de um grau de corretude geral para o CDR, e (2) vertical, que dá enfoque ao cálculo de um grau de corretude para alguns campos ou conjuntos de campos.

Estes dois tipos de análise não são excludentes e não impedem a realização de outros tipos de análise. A base teórica aqui apresentada não se restringe somente aos dois tipos de análise apresentados. No decorrer deste capítulo estes dois tipos de análise são mais bem explicados (itens 4.2.1, 4.2.2 e 4.2.3). A Figura 7 apresenta um esquema primário do fluxo de informações do mecanismo de regras aqui proposto.

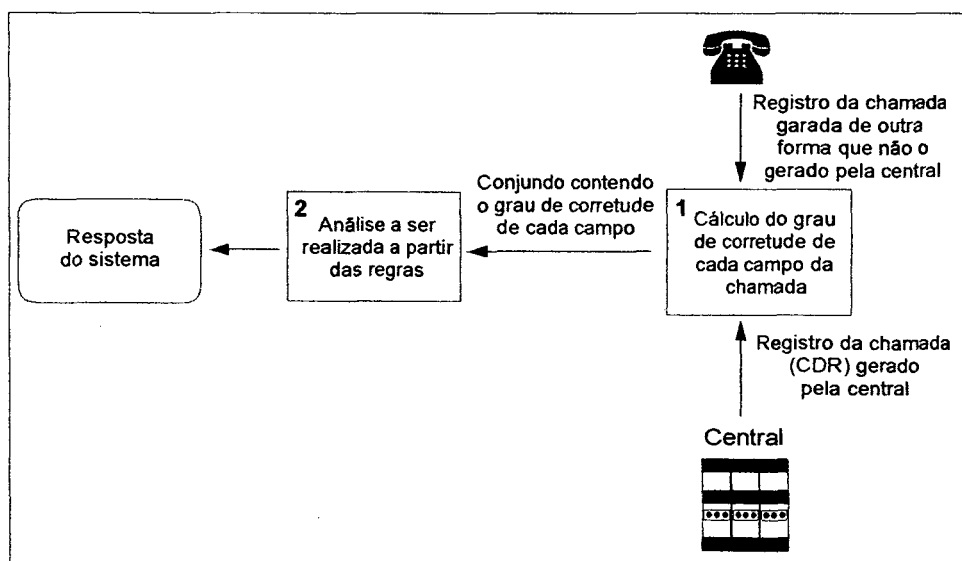


Figura 7 – Esquema básico do trabalho

A primeira etapa é o cálculo do grau de corretude (quadro 1 da Figura 7) de cada campo através da comparação do registro base (CDR gerado pela central) com o registro modelo (gerado por um sistema paralelo). Esta comparação é realizada campo a campo, dos dois registros, e mostra o grau de corretude de cada campo da chamada. A geração do grau de corretude de cada campo é a etapa de fuzificação. O Quadro 1 apresenta uma descrição matemática de como é obtido o grau de corretude da chamada.

Obtendo o grau de corretude de um dos n campos da chamada:

$$\epsilon_i = f_i (Re_i, Co_i)$$

onde:

- $f_i(x)$ - função fuzificadora para o campo i
- Re_i - valor do campo do registro obtido no teste
- Co_i - valor do campo do CDR gerado pela central
- ϵ_i - grau de corretude do campo i
- i - i -ésimo campo do CDR, varia de 1 até n

Quadro 1 – Cálculo do grau de corretude

Uma vez obtido o grau de corretude de cada campo (ϵ_i), é executada a segunda etapa, a análise. Esta etapa consiste na aplicação dos valores ϵ_i a um conjunto de regras definidas com o auxílio de um especialista (quadro 2 da Figura 7). O conjunto de regras aqui proposto é um conjunto de regras híbridas, ou seja, regras que trabalham tanto com lógica fuzzy como com lógica tradicional. Existem três níveis de execução das regras: inicialização, processamento e finalização. As regras de inicialização são o princípio da execução das regras, ou seja, inicializam seu contexto. Estas regras atribuem valores iniciais às variáveis ou fazem a leitura destes quando as variáveis tiverem sido salvas em execuções anteriores. Esta primeira etapa não possui valores de entrada (CDRs) e é executada somente uma vez para cada conjunto de dados.

O segundo conjunto, regras de produção, tem a função de tratar os valores de entrada (CDRs e registros) de forma individualizada. Estas regras servem tanto para obtenção de informações das entradas quanto para a geração de resultados individuais ou parciais. Este conjunto de regras é executado c vezes, dependendo do número de

entradas. Simplificando, é executado um conjunto de regras r a cada novo par CDR/registro.

O último grupo, regras de finalização, tem a função de agregar valores, salvar o contexto das regras ou variáveis e gerar resultados de análises. O conjunto de regras rf é executado apenas uma vez para cada conjunto de dados de entrada. Esta regra é executada sempre depois que todos os dados de entrada (CDRs) tiverem sido processados (terminada a execução das regras r).

No Anexo 1 encontra-se a proposta de uma linguagem a ser empregada por aplicações que utilizarão a base teórica aqui apresentada. A linguagem proposta contempla todas as características descritas anteriormente e que são utilizadas por este trabalho. Segue no Quadro 2, um esquema matemático da aplicação dos valores de entrada ao conjunto de regras.

Aplicando os valores de corretude dos campos às regras:

$$ri_j$$

$$\varepsilon_{c,1}, \varepsilon_{c,2} \dots \varepsilon_{c,n} \rightarrow r_j = \lambda_j$$

$$rf_j = \lambda f_j$$

onde:

- $\varepsilon_{c,1}, \varepsilon_{c,2} \dots \varepsilon_{c,n}$ - valores de entrada das regras r ^{45 46}.
- r_j - regra de produção número j , varia de 1 até m
- ri_j - regra de inicialização número j , varia de 1 até a
- rf_j - regra de finalização número j , varia de 1 até b
- λ_j - consequência ou resultado da regras r_j
- λf_j - consequência ou resultado da regras rf_j
- n - número de campos do CDR/registro
- c - número de CDR/registro de entrada
- m, a e b - número de regras respectivamente de produção, inicialização e finalização.

Quadro 2 – Aplicação das regras ri , r e rf

⁴⁵ Em aplicações reais os valores de entrada $\varepsilon_{x,1}, \varepsilon_{x,2} \dots \varepsilon_{x,n}$ serão os graus de corretude dos campos do CDR (valores já fuzificados) ou os valores brutos dos campos do CDR e do registro gerado pelo equipamento de teste. Mas, dependendo da aplicação, isto pode variar.

⁴⁶ A notação $\varepsilon_{c,1}, \varepsilon_{c,2} \dots \varepsilon_{c,n}$ diz respeito aos c CDR/registro de entrada das regras r . Para denotar apenas um par CDR/registro utilizar-se-á a notação $\varepsilon_1, \varepsilon_2 \dots \varepsilon_n$.

inferência e a estrutura das regras são as mesmas para os dois tipos. A Figura 9 demonstra o fluxo de dados nos dois tipos de análises.

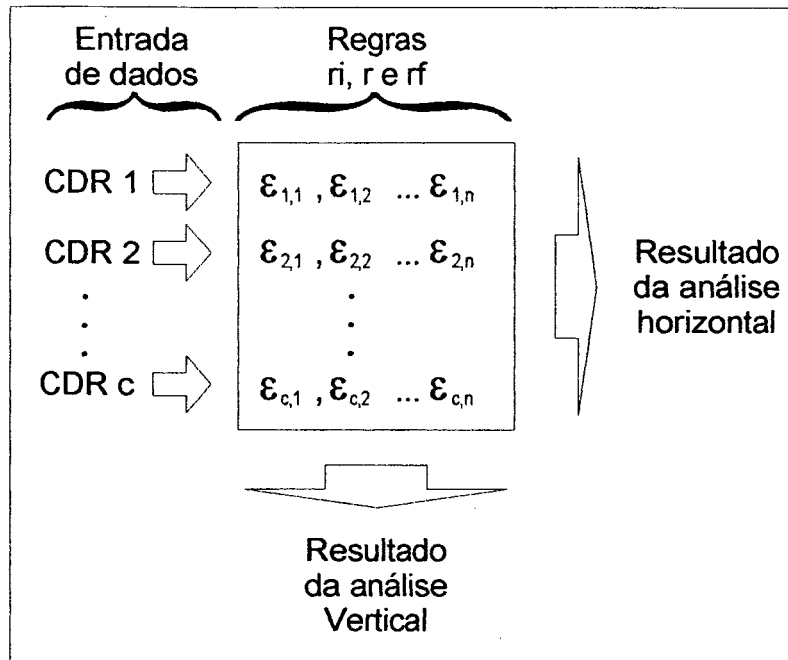


Figura 9 – Fluxo de dados nas análises horizontais e verticais

4.2.2 Análises horizontais

O primeiro tipo de análise, a horizontal, consiste em utilizar o conjunto de regras para se tirar conclusões dos bilhetes de forma geral. Ou seja, olham-se os valores de entrada $\epsilon_1, \epsilon_2 \dots \epsilon_n$ como campos de um CDR individual e o resultado das regras r_j são normalmente valores que dizem respeito ao CDR por completo. Neste tipo de análise há normalmente maior enfoque nas regras r_j do que nas regras r_{fj} . Este tipo de análise é habitualmente útil quando se deseja retirar informações sobre qualidade de equipamentos, qualidade individual de CDRs ou de conjunto de CDRs. A Figura 10 apresenta um modelo de uma análise horizontal.

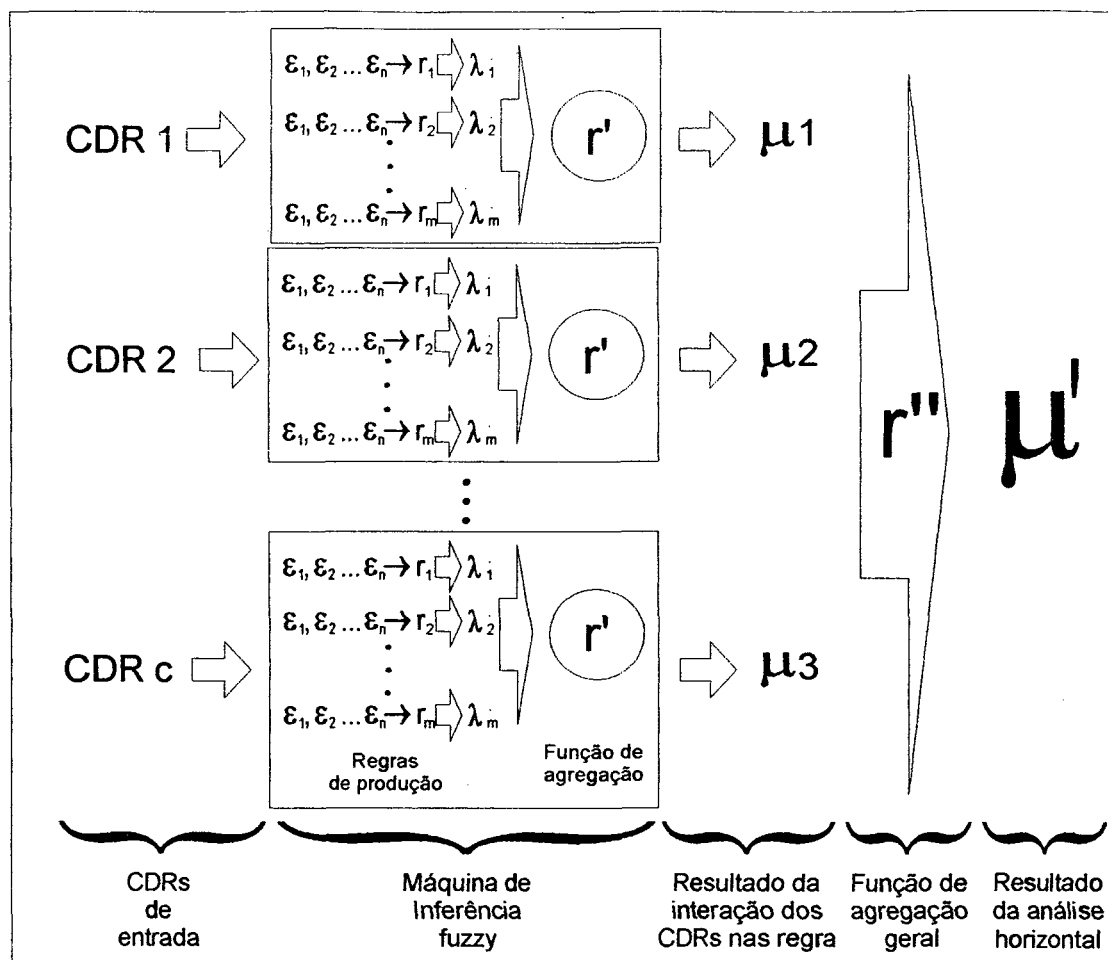


Figura 10 – Análise horizontal

Para contextualizar, as regras r e r' , por serem regras de análise de bilhetes de forma individual, são implementadas no conjunto das regras de produção (regras r_j). A regra r'' , por ser uma regra de agregação de valores de c CDRs, é implementada nas regras de finalização (regras rf_j). O resultado μ para cada CDR é o grau de corretude da chamada. Em um sistema que necessite apenas obter o grau de corretude de cada chamada para análise posterior não se aplicaria a regra r'' e utilizar-se-ia os vários valores de μ como resultado. Em uma análise de qualidade dos CDRs de uma central ou de um horário, por exemplo, utilizar-se-iam as regras r'' para a geração destes dados (geração do valor μ').

4.2.3 Análises verticais

O segundo tipo de análise aqui apresentado é a vertical. Neste conjunto, as regras procuram trabalhar com os campos do CDR de forma separada, pois seu objetivo é analisar cada campo ou conjunto de campos do CDR. Neste tipo de análise utilizam-se, por exemplo, as m ocorrências de um campo específico para tentar identificar se este campo está ou não sendo gerado de forma correta. Com esta informação pode-se identificar onde está o problema e quem o está gerando⁴⁷.

Neste tipo de análise há normalmente enfoque maior nas regras r_{f_j} do que nas regras r_j . Isto se deve ao fato de que todos os valores dos campos só estarão prontos com o término das execuções das regras r_j . Este tipo de análise é normalmente útil quando se necessita retirar informações sobre determinados campos ou problemas específicos de geração, principalmente quando relacionados com outros equipamentos. A Figura 11 demonstra o uso deste tipo de análise.

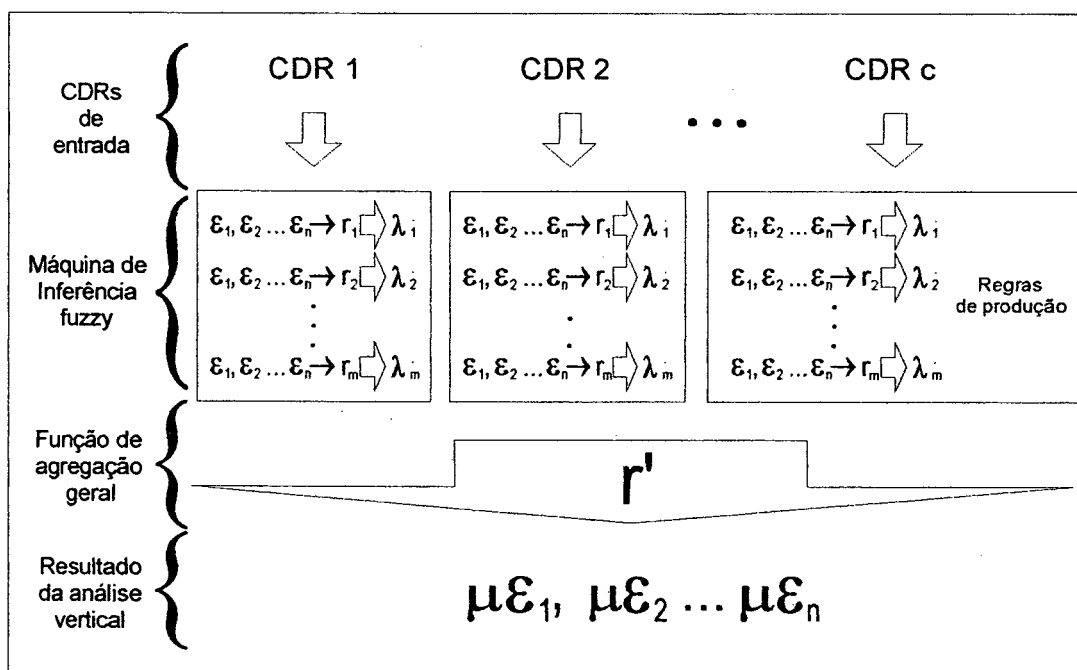


Figura 11 – Análise vertical

⁴⁷ Como já foi comentado no Capítulo 2, item 2.2, algumas informações contidas nos CDRs podem ter sido geradas por uma outra central/equipamento. Sabendo que campo tem problema podemos identificar qual equipamento está com problema e com isso tentar resolvê-lo.

Fazendo um paralelo ao que já foi explicado, as regras r são regras do tipo produção (regras r_j). Em análises verticais, estas regras têm função de análise reduzida, ficando a cargo das regras de finalização a maior parte do trabalho. As regras de finalização são demonstradas na regra r' . Estas regras têm maior importância, pois na maioria dos casos há interesse em analisar um conjunto de CDRs, e isto é mais interessante nas regras de finalização. Outra característica interessante na Figura 11 é que os resultados gerados: $\mu_{\epsilon_1}, \mu_{\epsilon_2} \dots \mu_{\epsilon_n}$ não correspondem a um valor único, mas sim a um valor independente para cada campo do CDR.

Vale ressaltar que o resultado de uma análise pode ser um valor único ou não. Nos exemplos aqui apresentados demonstrou-se a geração de um valor μ para cada CDR e um único valor μ' para o conjunto de regras, no caso das regras horizontais. Nas análises verticais, o resultado é um vetor de dados $\mu_{\epsilon_1}, \mu_{\epsilon_2} \dots \mu_{\epsilon_n}$. Esta diferença no número de valores resultantes não necessariamente irá qualificar o tipo de análise. O que se espera em aplicações reais é uma tendência de que análises horizontais gerem valores individuais e análises verticais gerem mais valores. Porém isto pode variar com o objetivo da análise a ser realizada.

4.2.4 Enquadramento do mecanismo de regras nas operadoras

O mecanismo de regras apresentado irá se encaixar em algum ponto no processo de tarifação das operadoras de telecomunicações. Este mecanismo será a interface entre os usuários e os equipamentos de execução de testes controlados. Para traçar um paralelo entre as soluções atuais de sistemas de testes controlados e um sistema de teste que utilizaria o embasamento teórico aqui apresentado, seguem as Figuras 12 e 13. A Figura 12 demonstra o fluxo de dados dentro de uma hipotética operadora de telecomunicações utilizando um sistema de geração de testes controlados com técnicas tradicionais. Já a Figura 13 demonstra como um sistema utilizando a base teórica aqui apresentada unir-se-ia ao exemplo demonstrado na Figura 12.

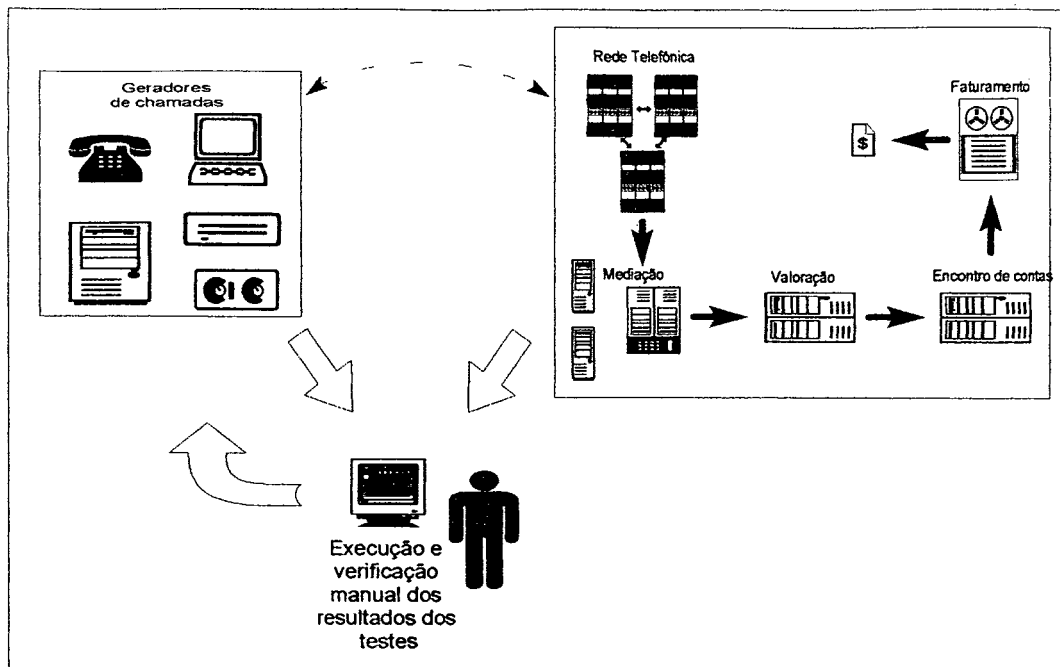


Figura 12 – Operadora hipotética utilizando sistema de geração de testes controlados

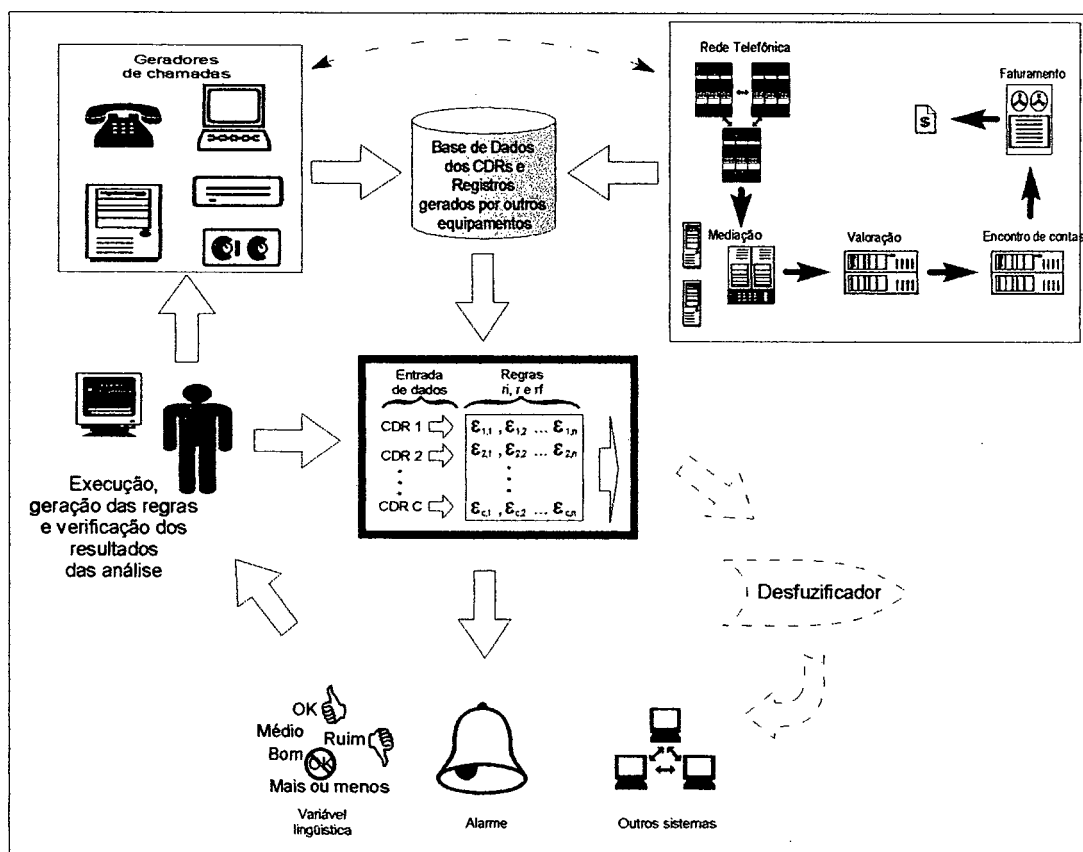


Figura 13 – Operadora hipotética utilizando sistema de geração de testes controlados aliado a base teórica aqui apresentada

Nota-se que na primeira figura os usuários do sistema de testes controlados têm de fazer a análise de forma manual. Na segunda há todo um aparato computacional, não só para gerar análises ou relatórios para os usuários, mas também para que haja uma interação com todo o ambiente das operadoras. Não se exclui, neste caso, a possibilidade de que os sistemas tradicionais tenham interação com outros sistemas. Porém, se fossem consideradas as reduzidas capacidades de análise e de extração de informação destes sistemas, a interação seria prejudicada. Por este motivo, não foi demonstrada na Figura 12 esta interação.

Na Figura 13, nota-se que a desfuzificação aparece em tracejado, para demonstrar que a desfuzificação não é um item obrigatório deste trabalho. Isto se deve ao fato de que os possíveis usuários de um sistema com esta base teórica são usuários avançados ou outros sistemas computacionais. Para estes tipos de usuário ficaria mais fácil e claro trabalhar com valores fuzzy do que com variáveis lingüísticas, principalmente se este usuário for um outro sistema. Mas isto dependerá muito do ambiente no qual será implantado o sistema.

Compreendido o mecanismo de regras, o próximo capítulo descreve um estudo de caso. Este estudo de caso tem o intuito de ser um exemplo prático do mecanismo de regras apresentado. Além disso, é feita uma comparação de um mecanismo de análise empregado por algumas soluções de mercado e um conjunto de regras criadas como exemplo.

Capítulo 5 - Estudo de caso

O presente capítulo tem por objetivo apresentar os conceitos demonstrados no Capítulo 4 em uma aplicação e fazer uma comparação de um exemplo de regras utilizando o mecanismo apresentado neste documento com um mecanismo adotado por algumas empresas.

5.1 Contexto do estudo de caso

O protótipo implementado neste estudo de caso é um mecanismo de análise. Para este mecanismo utilizou-se como base a linguagem descrita no Anexo 1 e a base teórica descrita no Capítulo 4. Todos os exemplos a serem apresentados neste estudo de caso utilizam como base a linguagem descrita no Anexo 1. Este mecanismo de análise tem como entrada um conjunto de valores fuzzificados. Estes valores correspondem ao grau de erro de cada campo dos CDRs a serem utilizados na análise. A saída é também um conjunto de valores fuzzy um para cada campo do CDR de entrada. Este valor fuzzy é um indicativo de problema do campo, ou seja, quanto maior for este valor, maior será o grau de problema do campo.

Os campos do CDR a serem utilizados no protótipo são: assinante A, assinante B, duração, data inicial, hora e status da chamada. Escolheram-se estes campos por serem estes os principais campos de um CDR.

O algoritmo implementado como estudo de caso é bastante simples. Este algoritmo gera como saída o valor médio de erro de cada campo, utilizando média aritmética simples. Suas entradas são os graus de erro dos campos descritos acima. Antes do cálculo da média, as regras verificam se, para cada conjunto de valores de entrada, os campos Assinante A e Assinante B possuem erro. Se os dois campos possuírem erro então se define que para esta entrada o grau de erros destes dois campos será igual ao maior valor entre os dois. Após o cálculo do grau médio de erro de cada campo, realiza-se uma projeção dos valores obtidos, de forma que o maior grau de erro seja considerado 1. Esta projeção é basicamente uma regra de três simples entre o maior

grau de erro e cada um dos graus dos campos. O Quadro 3 descreve um esquema matemático deste cálculo.

MÁXIMO VALOR DE (Campo1, Campo2... CampoN)	-	1
VALOR DO CAMPO (j)	-	x

Ou seja:

$$x = (\text{VALOR DO CAMPO (j)} * 1) / \text{MÁXIMO VALOR DE (Campo1, Campo2... CampoN)}$$

Quadro 3 – Esquema matemático da regra de 3

A projeção demonstrada no quadro anterior tem como objetivo principal facilitar a análise dos possíveis usuários de um sistema com estas regras. Tendo em vista que o grau de erro médio para cada campo será um valor bastante pequeno (nos testes realizados sempre menor que 0,001), esta projeção torna os valores mais visíveis, principalmente se o usuário do resultado das regras for uma pessoa. No Quadro 7 do Anexo 2 apresenta-se uma listagem do conjunto de regras definido para este estudo de caso.

Admite-se, neste protótipo, que os valores de entrada são valores já fuzificados correspondentes ao grau de erro de cada campo. Esta fuzificação está levando em conta o seguinte conjunto de regras apresentado no Quadro 4:

Para o campo Assinante A:
1 - Se o campo possuir 1 dígito errado Então grau de erro é 0,7
2 - Se o campo possuir 2 ou mais dígitos errados Então grau de erro é 1
Se não grau de erro é 0
Para o campo Assinante B:
3 - Se o campo possuir 1 dígito errado Então grau de erro é 0,3
4 - Se o campo possuir 2 dígitos errados Então grau de erro é 0,5
5 - Se o campo possuir mais de 2 dígitos errados Então grau de erro é 1
6 - Se o campo possuir valor nulo Então grau de erro é 1
Se não grau de erro é 0
Para o campo Duração da Chamada:
7 - Se o campo possuir valor nulo Então grau de erro é 1
8 - Se o campo possuir erro de precisão maior que 30 segundos

Então grau de erro é 0,7

9 - Se o campo possuir erro de precisão menor ou igual a 30 segundos Então grau de erro é 0,1

Se não grau de erro é 0

Para o campo Hora da Chamada:

10 - Se o campo possuir valor nulo Então grau de erro é 1

11 - Se o campo possuir erro de precisão maior que 1 hora Então grau de erro é 0,6

12 - Se o campo possuir erro de precisão menor que 1 hora e maior que 2 minutos Então grau de erro é 0,4

13 - Se o campo possuir erro de precisão menor que 2 minutos Então grau de erro é 0,2

14 - Se o campo possuir erro no subcampo "hora" Então grau de erro é 0,75

15 - Se o campo possuir erro no subcampo "minuto" Então grau de erro é 0,4

Se não grau de erro é 0

Para o campo Data da Chamada:

16 - Se o campo possuir valor nulo Então grau de erro é 0,5

17 - Se o campo possuir erro de precisão maior que 1 dia Então grau de erro é 0,75

18 - Se o campo possuir erro no subcampo "dia" Então grau de erro é 0,6

19 - Se o campo possuir erro no subcampo "mês" Então grau de erro é 0,4

Se não grau de erro é 0

Para o campo Status da chamada:

20 - Se o campo possuir valor nulo Então grau de erro é 1

21 - Se o campo possuir valor errado Então grau de erro é 1

Se não grau de erro é 0

Quadro 4 – Regras utilizadas para a fuzificação

Os fuzificadores descritos no quadro anterior foram construídos partindo-se do princípio que um campo com grau de erro maior que 0,5 é considerado como erro grave, ou erro dificilmente recuperável. Já um com grau menor que 0,5 é um erro menos grave, ou seja, provavelmente recuperável⁴⁸.

As regras numeradas de 1 a 21 são regras que representam as categorias de erro que este estudo de caso irá trabalhar. Por exemplo, a regra:

16 - Se o campo possuir valor nulo Então grau de erro é 0,5

⁴⁸ Erro recuperável é um erro, em um campo, que pode ser corrigido. Esta correção pode ser realizada com análise dos possíveis valores para este campo ou dos outros campos do CDRs ou com dados a partir de análise histórica.

Esta regra representa o erro referente ao campo Data da Chamada do CDR ter vindo com o valor nulo. Estas 21 categorias de erros serão listadas no item 5.2.

A base teórica apresentada no Capítulo 4, assim como a regra definida anteriormente, parte da prerrogativa de que o conjunto de regras receberá o valor fuzzificado para o grau de erro dos campos. Entretanto podem existir casos em que a obtenção destes valores fuzzificados seja inviável. Nestes casos a fuzzificação deverá ocorrer dentro das regras. Ou seja, a etapa 1, fuzzificação, e a etapa 2, análise, serão executadas dentro do conjunto de regras. Vale a pena lembrar que, em virtude da máquina de inferência ser híbrida (trabalhar com elementos fuzzy e da lógica tradicional), pode-se trabalhar com valores fuzzy ou não. Para exemplificar a fuzzificação utilizando a linguagem aqui proposta, o Quadro 5 demonstra um trecho de código de como seria a fuzzificação do campo Status da Chamada utilizando as regras definidas anteriormente (regras 20 e 21).

```
// Está se considerando que o status gerado pela central está na
  entrada 1 e o gerado por outro equipamento está na entrada 2.
SE ENTRADA1 = ENTRADA2 ENTÃO
{
  GrauDeErroStatus = 0f
} SENÃO {
  GrauDeErroStatus = 1f
}
SE (ENTRADA1 = "") ENTÃO
{
  GrauDeErroStatus = 1f
}
```

Quadro 5 – Exemplo de fuzzificação utilizando as regras propostas

Para fazer a comparação com as regras definidas anteriormente, utilizou-se um mecanismo de análise empregado por algumas soluções de geração de testes controlados. Este mecanismo consiste em verificar se os campos do CDR e do registro de teste são iguais ou não. Se forem iguais, então é incrementado 1 ao contador de erros por campo. Ao final gera-se uma saída com estes contadores. Para efeitos de comparação desta regra com a descrita anteriormente, houve a necessidade de realizar a mesma projeção dos valores dos contadores para valores entre [1,0]. Lembrando que esta projeção já é uma melhoria nos resultados desta análise.

Para a implementação deste teste utilizou-se a mesma linguagem definida no Anexo 1. Implementou-se este mecanismo de análise nesta linguagem para mostrar que a base teórica definida aqui é um superconjunto dos mecanismos de análise existentes. Ou seja, as mesmas regras de análise utilizadas atualmente podem ser implementadas empregando os conceitos aqui apresentados. Para simplificar a implementação da projeção de valor para [1,0], utilizou-se uma função fictícia MAXIMO_VALOR que retornará o maior valor de uma lista de valores. Na implementação real deste conjunto de regras utilizou-se uma série de estruturas *SE/ENTÃO* aninhada. No Quadro 8 do Anexo 2 segue esta implementação.

Destaca-se que alguns geradores de teste sequer possuem o mecanismo descrito, simplesmente geram as chamadas e colocam o resultado em tabelas, ficando a cargo do usuário fazer toda a análise de forma manual. Outro ponto importante é que os mecanismos de análise existentes no mercado são pouco flexíveis. São poucas as soluções que permitem que o usuário possa realizar análises com um grau razoável de customização. Ou seja, está se comparando um mecanismo normalmente implementado de forma estática (sem que o usuário possa redefinir regras sem alteração direta no programa) com outro que permite customizações, no qual o usuário pode realizar uma série de análises sem a necessidade de mudança no código do programa (apenas nas configurações e na base de regras).

Com relação aos algoritmos expostos anteriormente, está se admitindo que os CDRs que servirão de entrada para o mecanismo de regra dos dois métodos serão somente CDRs com problema. Foi retirada esta verificação das regras para torná-las mais simples e claras. Na implementação prática destas regras há um conjunto de estruturas *SE/ENTÃO* para realizar este filtro de CDRs sem erros.

5.2 Testes realizados

Para a realização dos testes desenvolveu-se um simulador. Este tem a função de simular as ligações telefônicas de teste e a função de bilhetagem da central telefônica.

Ou seja, o simulador gera, para cada chamada, dois registros: um CDR, simulando o gerado pela central, e um registro de teste, simulando o equipamento gerador de chamadas.

Utilizou-se este simulador na geração dos dados por dificuldades na obtenção de dados reais. Esta dificuldade na obtenção de informações junto às operadoras deve-se, em parte, à privatização no setor. Ainda assim tentou-se obter dados reais junto a uma operadora e a uma empresa de consultoria. Nenhuma das duas empresas contatada respondeu satisfatoriamente. Vale lembrar que a não utilização de dados reais não invalida o trabalho, pois seu objetivo central é o mecanismo de regras.

Outro motivo é que os dados gerados na simulação representam da forma mais fiel possível os dados reais, com base nos conhecimentos que se tem dos mesmos (mais de quatro anos trabalhando com CDRs) e em conversas informais com alguns funcionários de uma operadora de telecomunicações.

Para tentar suprir esta deficiência os casos foram definidos com características distintas de modo que o conjunto de regras fosse testado em diversas situações.

Nos testes realizados utilizou-se uma massa de dados de 5.000 chamadas e uma probabilidade de 10% de ocorrência de erro em uma chamada. Destes 10% que podem apresentar erro, definiu-se que o número de campos errados pode variar de 1 a 4. A distribuição de probabilidade de geração de 'e' erros no mesmo CDR é demonstrada no Anexo 3. Foram realizadas quatro repetições de cada um dos cenários a serem estudados.

Nos testes, foram definidas 21 categorias de erros, cada uma representando uma regra de fuzificação descrita no item 5.1. O Quadro 6 apresenta a lista dos erros que poderão ser gerados nos testes. Os códigos que estão na frente de cada erro correspondem aos mesmos códigos utilizados nas regras de fuzificação.

- 1 - Diferença de 1 dígito entre o campo Assinante A do CDR e o do registro de teste (gerado pelo gerador de chamadas de teste)
- 2 - Diferença de mais de 2 dígitos entre o campo Assinante A do CDR e o do registro de teste
- 3 - Diferença de 1 dígito entre o campo Assinante B do CDR e o do registro de teste
- 4 - Diferença de 2 dígitos entre o campo Assinante B do CDR e o do registro de teste
- 5 - Diferença de mais de 2 dígitos entre o campo Assinante B do CDR e o do registro de teste
- 6 - O campo Assinante B do CDR veio com valor nulo
- 7 - O campo Duração da Chamada do CDR veio com valor nulo
- 8 - Diferença entre o campo Duração da Chamada do CDR e o do registro de teste maior que 30 segundos
- 9 - Diferença entre o campo Duração da Chamada do CDR e o do registro de teste menor que 30 segundos
- 10 - O campo Hora da Chamada do CDR veio com valor nulo
- 11 - Diferença entre o campo Hora da Chamada do CDR e o do registro de teste é maior que 1 hora
- 12 - Diferença entre o campo Hora da Chamada do CDR e o do registro de teste é menor que 1 hora e maior que 2 minutos
- 13 - Diferença entre o campo Hora da Chamada do CDR e o do registro de teste é menor que 2 minutos
- 14 - Não foi possível identificar o valor do subcampo "hora" no CDR
- 15 - Não foi possível identificar o valor do subcampo "data" no CDR
- 16 - O campo Data da Chamada do CDR veio com valor nulo
- 17 - Diferença entre o campo Data da Chamada no CDR e no registro de teste maior que 1 dia
- 18 - Não foi possível identificar o valor do subcampo "dia" no CDR
- 19 - Não foi possível identificar o valor do subcampo "mês" no CDR
- 20 - O campo Status da chamada do CDR veio com valor nulo
- 21 - O campo Status da chamada veio com valores diferentes no CDR e no registro de teste

Quadro 6 – Listas dos erros utilizados nos testes

Foram definidos doze cenários comparando os dois mecanismos de análise apresentados. Os doze cenários foram divididos em dois grupos. Cada grupo possui seis cenários simulando uma central diferente. Os cenários de 1 a 6 são situações diferentes para a central 1. Já os cenários 7 a 12 representam as mesmas situações, 1 a 6, para a central 2. Ou seja, os pares de cenários 1,7; 2,8; 3,9; 4,10; 4,11 e 6,12 representam situações semelhantes aplicadas a centrais diferentes.

Segue uma descrição dos cenários utilizados nos testes:

Cenários 1 e 7 – São os cenários-padrão das centrais 1 e 2 respectivamente. Estes cenários são utilizados como base para os cenários de 2 a 6 e de 8 a 12. Os erros do cenário 1, da central 1, foram distribuídos em maior quantidade para o campo Duração da Chamada e em menor quantidade para o campo Status da Chamada. Os

demais campos tiveram distribuições de erros parecidas quanto ao número de erros. A distribuição de erros do cenário 7, da central 2, concede um maior número de erros ao campo duração e menor número de erros no campo status. Os demais campos tiveram distribuições de erros parecidas quanto ao número de erros.

Cenários 2 e 8 – Partem como base dos cenários 1 e 7, porém simulam um aumento no número de erros do tipo 3. Com estes cenários procura-se simular um aumento no número de erros com baixa gravidade para verificar o comportamento dos dois métodos.

Cenários 3 e 9 – Partem como base dos cenários 1 e 7, porém simulam um aumento no número de erros do tipo 12. Com estes cenários procura-se simular um aumento no número de erros com gravidade média para verificar o comportamento dos dois métodos.

Cenários 4 e 10 – Partem como base dos cenários 1 e 7, porém simulam um aumento no número de erros do tipo 21. Com estes cenários procura-se simular um aumento no número de erros com alta gravidade para verificar o comportamento dos dois métodos.

Cenários 5 e 11 – Partem como base dos cenários 1 e 7, porém simulam um aumento no número de erros do tipo 3 e 21. Com estes cenários procura-se simular um aumento no número de dois tipos de erros. Os tipos de erros escolhidos são um com baixa gravidade (tipo 3) e um com alta gravidade (tipo 21). Procura-se, através destes dois cenários, verificar o comportamento dos dois métodos.

Cenários 6 e 12 – Partem como base dos cenários 1 e 7, porém simulam um aumento no número de erros do tipo 12 e 21. Com estes cenários procura-se simular um aumento no número de dois tipos de erros. Os tipos de erros escolhidos são um com gravidade média (tipo 12) e um com alta gravidade (tipo 21). Procura-se, através destes cenários, verificar o comportamento dos dois métodos.

Para cada um dos cenários descritos definiu-se uma probabilidade de ocorrer um dos 21 erros definidos. Estas distribuições tiveram como base as descrições dos doze cenários mencionadas anteriormente. A distribuição de probabilidade utilizada nos doze cenários é apresentada no Anexo 3.

Ressalta-se que os cenários e as regras aqui demonstrados são somente situações hipotéticas, Dependendo do contexto das aplicações reais, os resultados apresentados a posteriori podem variar.

Para implementação da simulação utilizaram-se scripts⁴⁹ implementados no programa GNU⁵⁰-Bash⁵¹. Escolheu-se este programa, Poer ser uma linguagem bastante poderosa em termos de recursos e que facilita bastante a implementação de protótipos. O script implementado nesta linguagem gera os dados para os testes. Também é ele quem executa as regras, obtendo as respostas e gerando as saídas. Para a implementação das regras utilizou-se o programa GAWK⁵². Optou-se por utilizar este programa, pois não existe nenhum interpretador ou compilador para a linguagem proposta e por este ter um sistema de entrada de dados semelhante ao utilizado pelas regras definidas neste trabalho. Um outro motivo que levou à utilização dos programas GAWK e GNU-Bash é o fato de estes programas serem GPL⁵³ e possuírem versões para praticamente todas as plataformas computacionais existentes atualmente. Uma das desvantagens da utilização destas ferramentas é que não trabalham com lógica fuzzy de forma nativa. Para suprir esta restrição fez-se necessária a implementação das funções fuzzy necessárias para as regras, além de outras adaptações.

Os resultados gerados dos doze cenários, utilizando os programas descritos anteriormente, são apresentados no Anexo 4.

Para comparação dos dois tipos de análise utilizou-se o software SAS⁵⁴. Os resultados destes testes estão no Anexo 5. Os testes realizados foram baseados em um teste de análise de variância de um fatorial 2 x 12, sendo que foram abordados dois métodos (fatores) comparados em doze cenários (experimentos). Para estes testes está se considerando um nível de significância de 5% ou 0,05.

⁴⁹ Script é um tipo de código de computador, neste caso utilizado na implementação aplicações

⁵⁰ GNU (Gnu Not Unix), <http://www.gnu.org/>

⁵¹ GNU-Bash (GNU- Bourne Aghain SHell), <http://www.gnu.org/>

⁵² GAWK (GNU Aho Wenberg and Kernighan), <http://www.gnu.org/>

⁵³ GPL - General Public License, tipo de licença de software normalmente utilizado nas ferramentas da GNU, <http://www.gnu.org/>

⁵⁴ The SAS System é de propriedade da empresa SAS Institute, <http://www.sas.com/>

Para verificar se as médias são estatisticamente iguais utilizaram-se dois testes. No primeiro teste verifica-se a hipótese de que os dois métodos têm efeitos diferentes. A partir deste teste, é executado o segundo e mais importante teste. Este mostra se há interação entre os métodos para cada um dos doze cenários. Este segundo teste verifica se os valores médios dos resultados dos dois métodos são iguais ou diferentes estatisticamente.

5.3 Resultados

A partir dos dados gerados pelas análises descritas, verifica-se claramente uma diferença entre os dois métodos em nível estatístico. Isto pode ser verificado observando-se o campo **p-valor** (campo “Pr > F” no relatório do SAS, Anexo 5) nos itens nomeados “metodo*cenarios Effect Sliced by cenarios for <nome do campo>”. Para saber se estatisticamente os valores médios das repetições são estatisticamente diferentes deve-se comparar o **p-valor** com o nível de significância (0,05). Se **p-valor** < **nível de significância**, rejeita-se a hipótese de que as médias são iguais [VIEIRA S., 1999].

De posse destas informações, verifica-se claramente que nos cenários 3, 4, 9, 10 e 12 os resultados dos dois métodos são iguais para os campos em que se procurou dar maior ênfase. Já nos cenários 2, 5, 6, 8 e 11 observa-se que os dois métodos geraram resultados diferentes. Os cenários 3, 4, 9, 10 e 12 são cenários onde foram gerados erros com gravidade alta ou baixa em maior intensidade. Nota-se, então, que, quando se trabalhou com erros de maior gravidade, tanto o método tradicional quanto o exemplo aqui apresentado tiveram os mesmos resultados. Este comportamento semelhante foi observado nos campos em que se procurou dar maior enfoque. Porém, para a maioria dos outros campos, os resultados foram diferentes, enquanto que para os cenários 2, 5, 8 e 11 observou-se o contrário. Ou seja, nos campos que se pretendeu dar um maior enfoque, para estes cenários, não se observou igualdade nos resultados. Com isto, nota-se que para campos com gravidade baixa há diferença nos resultados dos dois tipos de análise.

Uma exceção para a regra de que campos com gravidade alta e média têm resultados iguais é o cenário 6. Neste caso observou-se diferença nos resultados.

Nos outros campos, além dos já citados, observa-se claramente que há diferença nos métodos. Sumarizando, os dois métodos apresentam para a maioria dos cenários diferenças consideráveis. As poucas exceções para isto aparecem quando se trabalha com grande quantidade de erros com alta gravidade. Neste caso, observa-se uma igualdade nos dois métodos.

Para melhor observar os resultados comentados neste item, veja a Tabela 1. Esta apresenta uma síntese dos **p-valores** obtidos na análise estatística do SAS (apresentado no anexo 5). Vale a pena lembrar a regra mencionada anteriormente: se **p-valor < 0,05** então as médias são estatisticamente diferentes, se não, as médias são estatisticamente iguais, com nível de significância de 0,05 (ou 5%). Para melhor visualizar estes resultados na Tabela 1, as colunas em vermelho apresentam casos onde os dois métodos são iguais. As colunas em preto referem-se a casos onde os dois métodos são diferentes.

Tabela 4 - Sumário dos resultados da análise estatística

Campo Método	Assinante A	Assinante B	Data	Hora	Duracao	Status
1	0.0035	<.0001	1.0000	<.0001	<.0001	0.0110
2	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	<.0001	0.0048	<.0001	1.0000	0.1014	<.0001
4	<.0001	<.0001	<.0001	<.0001	<.0001	1.0000
5	0.0319	<.0001	0.0110	0.0408	0.0041	0.0541
6	0.0807	0.0009	0.0369	<.0001	0.0107	0.0005
7	1.0000	0.0012	0.0571	0.0055	0.0052	0.0013
8	<.0001	<.0001	<.0001	0.0007	0.0850	<.0001
9	<.0001	0.0161	<.0001	1.0000	0.1014	<.0001
10	<.0001	<.0001	0.0009	<.0001	<.0001	1.0000
11	<.0001	<.0001	0.1263	0.0334	0.0135	0.0226
12	<.0001	0.0025	0.1514	<.0001	0.0085	0.0541

Depois de observados os dados gerados pelos cenários e a análise estatística feita pelo SAS, constata-se que, quando ocorrem erros de menor gravidade, o método utilizado pelas ferramentas de mercado acaba desviando a atenção do analista para erros sem muita importância, deixando erros mais graves escondidos.

Ressalta-se que os resultados aqui apresentados correspondem à comparação de um mecanismo simples e pouco flexível com um exemplo de regra definido em um mecanismo que permite uma maior customização por parte do usuário. Se os resultados aqui apresentados fossem gerados por regras mais bem elaboradas, os dados resultantes

seriam ainda mais favoráveis para o mecanismo de regras aqui proposto. Isso se deve ao fato de que os mecanismos encontrados no mercado possuem pouca capacidade de modificação se comparados a este mecanismo de regras. Ele permite que sejam gerados resultados diferentes para diferentes objetivos. Por exemplo, pode-se gerar um relatório para o setor de garantia da receita enfatizando o nível de erro das chamadas. Já para o setor de engenharia seria gerado um outro relatório dando destaque aos prováveis pontos de falhas. Com isso nota-se sua alta capacidade de adaptação às necessidades das operadoras.

Um problema observado nos métodos de análise das ferramentas encontradas no mercado, demonstradas no Capítulo 4, é que estas não permitem trabalhar com diferentes níveis de erro, grau de importância de campos, casos particulares ou exceções. Vale a pena lembrar que boa parte dos sistemas de geração de teste sequer possui este mecanismo pouco eficiente de análise.

Este capítulo demonstrou a potencialidade do mecanismo de regras apresentado, tanto em forma de exemplo como na comparação com outro método. No próximo capítulo é feito o fechamento do trabalho, comentando seus resultados e levantando sugestões para trabalhos futuros.

Capítulo 6 - Conclusão e recomendações

O presente capítulo tem por objetivo finalizar o trabalho apresentando as conclusões do mesmo e fazendo os comentários finais. Inicia-se com os comentários gerais sobre o trabalho e sobre outros usos para a base teórica aqui apresentada. Em seguida destacam-se propostas de continuações ou trabalhos futuros. O último item é composto por uma série de recomendações para implementações reais deste trabalho.

6.1 Comentários e conclusões

As operadoras nacionais de telecomunicações sofreram nos últimos anos muitos prejuízos por problemas operacionais no seu fluxo de faturamento. Estes problemas estão espalhados por vários pontos e são difíceis de serem encontrados sem uma ferramenta adequada.

Em termos de valores, os problemas operacionais das operadoras geraram um prejuízo de aproximadamente R\$ 2,5 bilhões ao ano nas operadoras nacionais de telecomunicações, ou seja, 5% da sua receita bruta [CPqD, 2001]. Para se tentar diminuir estas perdas, o trabalho aqui apresentado propôs um mecanismo de análise de chamadas telefônicas de teste. Este mecanismo foi desenvolvido a partir de técnicas de lógica fuzzy e de mecanismos de regras de produção e permite trabalhar tanto com a lógica tradicional como com a lógica fuzzy.

Como resultado, obteve-se um mecanismo de análise flexível a fim de atender algumas necessidades atuais e futuras das operadoras. Com esta flexibilidade, pode-se gerar regras e saídas diferentes para vários usuários dentro da empresa de telecomunicações. Setores como garantia de receita, TI (Tecnologia de Informação), engenharia e SAC têm necessidades de saídas diferentes dos sistemas de auditoria. A área de TI, por exemplo, procura identificar erros nos sistemas computacionais ou no fluxo de dados da empresa. Já a área de engenharia procura identificar problemas mais específicos em equipamentos. Os sistemas existentes atualmente não possuem nível de customização para atender todas a estas áreas. O mecanismo de regras aqui proposto conseguiria, em grande parte, atender a estas necessidades.

O mecanismo de regras proposto deve trabalhar com as saídas geradas pelos equipamentos de teste e com as saídas geradas pelas centrais telefônicas. Com estas informações, pode-se gerar análises de várias formas e para vários fins. Propuseram-se duas formas de análise, uma olhando os dados de forma horizontal e outra de forma vertical. A horizontal permite identificar um grau único de erro referente a um registro ou conjunto de registros. Esta análise permite que sejam formuladas asserções como:

Se +20% dos registros possuírem grau de erro $> 0,4$ Então

Gere Alarme “Número excessivo de registros errados”

Na segunda forma de análise, vertical, é possível identificar o campo que está com problema e com isto o possível ponto de falha. Neste caso pode-se formular questões como:

Se grau de erro do campo data/hora $> 0,6$ Então

Mande mensagem para operador da central: “Verifique problema na data/hora da central. Registros sendo gerados de forma incorreta”.

Estas formas de análise não são únicas ou excludentes, são simplesmente duas formas de se realizar uma análise utilizando o mecanismo de regras apresentado.

Com as regras aqui descritas pode-se medir a eficiência das centrais e/ou tecnologias, identificar problemas, auxiliar na correção de problemas, etc. Porém, deve-se ter cuidado com a massa de dados de entrada. Esta deve ser coerente quanto a seu tamanho e valores para que se possa retirar informações corretas das análises. Ou seja, se for feito um teste de uma determinada central, deve-se verificar se os dados de entrada refletem somente esta central, se os dados dos campos foram gerados por esta central ou por outra e se a massa de dados tem tamanho suficiente para que se possa fazer inferências estatisticamente comprovadas.

Esta confiança nos dados de entrada deve ser estendida para todo o processo de garantia de receita. Estas ferramentas devem ter um alto grau de confiabilidade para que os usuários possam confiar nos seus resultados. Para aumentar-se o grau de confiança dos sistemas de auditoria deve-se evitar ao máximo o uso de dados dos sistemas do fluxo de tarifação. Isso se deve ao fato de que, se forem utilizados dados previamente errados do fluxo de faturamento, estes dados gerarão resultados errados também nos sistemas de auditoria.

Uma das limitações do trabalho apresentado é a necessidade de um especialista ou pessoa capacitada para desenvolver as regras. Porém o conjunto de regras aqui apresentado não impede que a geração das regras seja feita utilizando mecanismos como redes neurais, algoritmos genéticos, etc.

Com a utilização de boas ferramentas de monitoramento dos processos internos, as empresas de telecomunicações podem recuperar cerca de 4% da sua receita. Para uma empresa com faturamento anual de US\$ 500 milhões, por exemplo, com uma perda de receita estimada em US\$ 20 milhões, um projeto de *revenue assurance* ao custo inicial de US\$ 500 mil e mais US\$ 50 mil mensais pode gerar um retorno de investimento em menos de um mês [FOINA, 1998]. Em outros estudos [CPqD, 2001] é previsto que para cada R\$ 1,00 aplicado em garantia de receita obtêm-se cerca de R\$ 25,00 em receita. Estes dados são mais que suficientes para justificar a implantação de ferramentas deste tipo nas operadoras de telecomunicações.

Vale a pena ressaltar a importância da utilização de mais de um tipo de ferramenta de garantia de receita. Se for utilizado apenas um mecanismo corre-se o risco de estar ignorando alguns erros gravíssimos. Um bom planejamento e um estudo prévio das necessidades e dos objetivos é fundamental para o sucesso das ferramentas de garantia de receita nas operadoras.

Encontrar boas referências bibliográficas do setor de telecomunicações foi uma das dificuldades deste trabalho. Principalmente após a privatização, as operadoras ficaram mais fechadas dificultando bastante a obtenção de informações sobre seus dados e processos. Estas informações, mesmo dentro das operadoras, ficam restritas a poucas pessoas e, muitas vezes, não são documentadas. Para tentar sanar esta

dificuldade, utilizou-se de um pouco de conhecimento adquirido nesta área bem como o auxílio informal de algumas pessoas mais experientes do setor de telecomunicações. Outra fonte relevante de dados foram revistas do setor. Apesar de algumas delas enfocarem determinados produtos ou situações específicas, não deixa de referenciar pessoas diretamente ligadas à área, com conhecimento no assunto. Em virtude desta dificuldade, pode-se considerar este trabalho como uma contribuição relevante para a academia e para o mercado de trabalho. Esta contribuição se dá principalmente na geração de referências relacionadas a telecomunicações (principalmente no Capítulo 2).

6.2 Extensibilidade

Apesar de focar a identificação de problemas utilizando mecanismos de testes controlados, o trabalho aqui apresentado pode ter várias outras aplicações. Pode-se utilizar não somente os dados gerados pela central, mas simular os dados gerados pelas outras fases do processo de tarifação e fazer a comparação. É possível, por exemplo, desenvolver um sistema paralelo de faturamento. Com os dados deste sistema e do sistema real e utilizando a base teórica aqui definida, pode-se fazer análises e identificar problemas no sistema de faturamento da empresa.

Além de executar as análises, o mecanismo de regras aqui apresentado pode ser utilizado em outras áreas das operadoras. Um exemplo de sua utilização seria na realização de análises de logs/alarmes da central ou validação dos CDRs. Na validação de CDRs, particularmente, tem-se um caso bastante claro de aplicação para este trabalho. A entrada de dados para as regras seria composta apenas pelo CDR, em vez do CDR e do registro gerado por um equipamento de teste. Na análise, o resultado seria a informação de o CDR ter sido gerado de forma correta ou não.

Abstraindo o ambiente de telecomunicações, pode-se utilizar o mecanismo de regras apresentado para, por exemplo, identificar o ponto de geração de erro em uma fábrica de camisetas. Neste caso, tem-se como entrada os graus de qualidade dos itens que compõem a produção de uma camiseta: tecido, corte, costura, acabamento etc. De posse destes dados, as regras identificariam qual, ou quais, dos itens gera maior prejuízo à fábrica. Pode-se ainda identificar quais dos turnos ou setores da fábrica apresentam

maior qualidade. Além deste exemplo, pode-se aplicar o trabalho em empresas de TV a cabo, montadoras de automóveis, setor agrícola etc.

6.3 Características para aplicações reais

Quando se implementa um protótipo em um ambiente real ou faz-se agregação a um sistema pronto, normalmente há a necessidade de algumas adaptações. Este item foi escrito com o intuito de levantar algumas destas adaptações necessárias. As características não invalidam o trabalho apresentado; são apenas características a serem levantadas e necessidades de adaptação do sistema ao se utilizar as regras aqui propostas.

O primeiro ponto a ser considerado é a obtenção dos CDRs junto à central. Estes CDRs, muito provavelmente, serão obtidos do mediador. Isto se deve às dificuldades de se obter os CDRs diretamente da central telefônica. Com esta restrição não se terá garantias totais de que os erros encontrados nestes CDRs são erros do mediador e não erros da central. Para tentar diminuir os erros gerados pelo mediador e enfocá-los na central, deve-se tentar usar os arquivos brutos (sem manipulação pelo mediador). O mediador para estes arquivos deverá ter apenas a função de coletor de CDRs.

Outro ponto importante é como se obter os CDRs dentro dos arquivos originais, dada a complexidade de leitura e de execução de filtro dos CDRs de teste, dentro deste arquivo. Uma alternativa seria obter estes dados do mediador, caso este tenha capacidade de realização de filtros. Esta mesma característica de realização de filtros é também desejável nas outras fases do processo de faturamento.

Com relação aos CDRs, muito provavelmente haverá a necessidade de utilização de um mecanismo de consolidação de CDRs, ou seja, juntar os vários CDRs de uma chamada em um único CDR. Esta característica será normalmente útil em CDRs provenientes de chamadas de longa duração. Além disso, deve-se desenvolver mecanismos para tratar a geração do mesmo CDR em vários pontos.

6.4 Sugestões para trabalhos futuros

Uma primeira sugestão de trabalho futuro seria a geração de um estudo mais aprofundado da extensibilidade comentada no item anterior, bem como a aplicabilidade

para outros serviços de telecomunicações como 0800, 0900, celular, tarifação IP (*Internet Protocol*), etc. Estes estudos poderiam ter início em um aprofundamento das necessidades específicas do tema abordado e um estudo de caso, demonstrando a eficiência ou não do mecanismo de regras aqui apresentado.

Outra sugestão bastante interessante é adicionar às regras o poder de se modificarem. Uma idéia inicial seria adicionar dois novos comandos “adicione regra” e “remova regra”. Com estes dois comandos pode-se modificar a base de regras conforme o contexto da aplicação. Por exemplo, se ocorrer determinado evento o conjunto de regras poderia adicionar uma nova regra para tratar este caso específico. No Anexo 1 este mecanismo é colocado na gramática como um item opcional para ser utilizado em futuros trabalhos.

A definição de uma forma de integração do mecanismo aqui apresentado com outros tipos de mecanismos de garantia da receita também é uma proposta de trabalho. Desta forma, este mecanismo não trabalharia somente com dados de chamadas de teste, mas sim com estes dados e com os de outros sistemas de garantia da receita. Além disto, esta integração é extremamente importante, pois diminuirá o número de interfaces com o usuário, melhorando a qualidade dos dados de saída.

Uma extensão também interessante seria utilizar algumas técnicas de IA para construção ou manutenção das regras. Poderiam ser utilizados algoritmos genéticos, redes neurais, lógica proposicional, etc.

Para se ter um maior grau de confiança nos dados e nas análises geradas seria interessante o uso de técnicas de estatísticas para geração de amostras, definição do grau de confiança dos resultados, geração de perfis de consumo, etc.

Outra proposta seria a geração de mecanismos de identificação das chamadas pares. Ou seja, dado um conjunto de c CDRs e c' (ou um c' diferente de c) registros gerados por um equipamento de teste, identificar que registro corresponde a que CDR. Este é um trabalho bastante interessante, se considerada a sua complexidade.

Esta complexidade se deve ao fato de que nem sempre os dados registrados no equipamento de teste são os mesmos obtidos na central. Isso ocorre por vários motivos, dentre os quais podem-se citar como principais: falta sincronismo entre o relógio da

central e do equipamento de teste, erro na geração de determinados campos pela central, falta de índice ou chave unida que permita associação direta e suposições erradas ou imprecisas do equipamento de teste. Além disso, pode haver casos que o mecanismo de identificação tem que procurar dentro dos arquivos de CDRs quais CDRs correspondem ao gerados pelo equipamento de teste. Para tentar resolver esta complexidade podem-se utilizar algoritmos com heurística, mecanismo de faixa de valores, lógica fuzzy, etc.

Uma limitação do mecanismo de regras aqui apresentado e da linguagem proposta é a incapacidade de trabalhar com várias fontes externas de dados diferentes. Por exemplo, utilizar dados de outras ferramentas de garantia de receita, juntamente com tabelas de base de dados e com as informações dos CDRs/registros, não seria viável hoje. Além destas várias fontes externas de informação, deve-se pensar em mecanismos para tratar a geração do mesmo CDR em locais (centrais) diferentes. Um bom trabalho seria uma ferramenta, ou extensão das regras aqui apresentadas, capaz de trabalhar com várias fontes de dados.

Referências bibliográficas

- [AMPESAN, 1999] AMPESAN, Débora Aparecida Ataíde, **Especificação de um Modelo de Informação de um Agente para Gerência de Desempenho do Sistema de Telefonia Móvel Celular**. 1999. v.1. Dissertação (Mestre em Ciência da Computação) - Universidade Federal de Santa Catarina, PPGCC, Florianópolis, SC.
- [CAROLINE, 2001] CAROLINE, Anne; LIN, Maurício e CARIBAS, Ville, **Sistema de Raciocínio**. Departamento de Ciências da Computação, Universidade do Amazonas, Mar-2000. Disponível em: http://www.dcc.fua.br/~dcc_ia/reason/Reasoning.html Acesso em fevereiro de 2001.
- [CARVALHO, 2002] CARVALHO, Leonardo. Telefônica e Embratel lideram reclamações no Procon em 2001. Disponível em: <http://ultimosegundo.ig.com.br/>. Acesso em: março de 2002.
- [CPqD, 2001] CPqD (Centro de Pesquisa de Desenvolvimento em Telecomunicações). Como as empresas de telecomunicações podem garantir a fidelização e as receitas e diminuir as perdas. **rnt – Revista Nacional de Telecomunicações**. São Paulo/SP, v. 23, n. 266, out 2001.
- [CROSS, 1994] CROSS, Valerie. **Fuzzy Information Retrieval**. **Journal Of Intelligent Information Systems**, Boston, v.3, n.1, p. 23-56, Fevereiro. 1994.

- [DRIANKOV, 1996] DRIANKOV, Dimiter; HELLENDORRN, Hans, REINFRANK, Michael. **An Intoduction to Fuzzy Control**. 2. ed. Berlin: Springer Verlag, 1996. 316 p.
- [FEIGENBAUM, 1982] FEIGENBAUM, E. A. **Knowlegde Engineering in the 1980's**, Dept. of Computer Science, Stanford University, Stanford CA, 1982.
- [FERRARI, 1991] FERRARI, Antonio Martins. **Telecomunicações Evolução & Revolução**. São Paulo: Ed. Érica, 1991. 297 p.
- [FOINA, 1998] FOINA, Paulo Rogério, Revenue Assurance: Como Identificar Perdas de Receita, **Revista Teletime**, São Paulo/SP v. 2, n. 13, out, 1999. Disponível em: <<http://www.teletime.com.br/>>. Acesso em: dezembro 2001.
- [GSI, 2001] GSI - Grupo de Sistemas Inteligentes. **Lógica Difusa**. Universidade Estadual de Maringá. Disponível em: <http://www.din.uem.br/ia/control/fuz_prin.htm>. Acesso em: março de 2001.
- [MAIA, 1998] MAIA, Wagner Ubiratan Lanzieri De Azevedo, **Sistema Integrado De Operação e Diagnóstico De Falhas Para Sistemas De Energia Elétrica – SODF**. 1998. Dissertação (Mestre em Ciência da Computação) - Universidade Federal de Santa Catarina, PPGEP, Florianópolis, SC.
- [MOURA, 2001] MOURA, Luiz, Vazamentos bilionários. **Revista Teletime**, São Paulo/SP v. 4, n. 29, fev, 2001. Disponível em: <<http://www.teletime.com.br/>>. Acesso em: dezembro 2001.
- [NEWELL, 1972] NEWELL, A.; SIMON H. A. **Human Problem Solving**. Englewood Cliffs, N.J.: Prentice-Hall., 1972.

- [OLIVEIRA, 2000] OLIVEIRA, Stela. A dor-de-cabeça das operadoras. **Revista Tele.com**. São Paulo, v. 1, n. 7, p. 25, set. 2000.
- [RABUSKE, 1995] RABUSKE, Renato Antônio. **Inteligência Artificial**. Florianópolis: EDUFSC, 1995. 240 p.
- [SANTANA, 2001] SANTANA, Ivone, Só tecnologia não garante eficiência. **Revista Teletime**, São Paulo/SP v. 4, n. 29, fev, 2001. Disponível em: <<http://www.teletime.com.br/>>. Acesso em: dezembro 2001.
- [SORTICA, 1999] SORTICA, Eduardo Almansa. **Redes de Telecomunicações TMN e gerência de redes e Serviços**. Salvador: Edição Independente, 1999. 265 p.
- [SOUZA, 2001] SOUZA, Carlos Pimentel de. **Lógica Fuzzy Aplicada**. Disponível: em: <<http://www.dee.ufc.br/~pimentel/fuzzy/>>. Acesso em: fevereiro de 2001.
- [SPECIALSKI, 2000] SPECIALSKI, Elizabeth Sueli. **Modelo de Informação Baseado em Relacionamentos entre Objetos Gerenciados para Gerência Integrada de Ambientes de Telecomunicações**. 2000. 140 p.1. Tese (Doutorado em Engenharia Elétrica) - Universidade Federal de Santa Catarina, PPGEP, Florianópolis, SC.
- [SUJIT, 2001] SUJIT Nath Pant e KEITH E. Holbert. **Fuzzy Logic in Decision Making and Signal Processing**. Última atualização: 21 outubro 1999 Disponível: em: <<http://ceaspub.eas.asu.edu/PowerZone/FuzzyLogic/index.htm>>. Acesso em: novembro de 2001.
- [TANAKA, 1997] TANAKA, K. **An introduction to fuzzy logic for practical applications**. New York: Spring-Verlag, 1997.

- [TSOUKALAS, 1997] TSOUKALAS, L. H.; UHRIG, R. E. **Fuzzy and neural approaches in engineering**. New York: John Wiley & Sons, Inc, 1997.
- [VIEIRA S., 1999] VIEIRA, Sônia. **Estatística Experimental**. 2. ed. São Paulo: Editora Atlas S.A., 1999. Segunda edição, 185 p.
- [VIEIRA, 1999] VIEIRA, Rodrigo De Sousa. **Protótipo de um Sistema de Monitoramento Remoto Inteligente**.1999. v.1. Dissertação (Mestre em Engenharia de Produção) - Universidade Federal de Santa Catarina, PPGEP, Florianópolis, SC.
- [ZADEH, 1965] ZADEH, Lotfi A. **Fuzzy Sets. Information and control**, n.8, p.338-353, 1965.
- [ZADEH, 1973] ZADEH, Cof. A. **Outline of a New Approach to the Analysis of Complex Systems and Decision Processes**. IEEE Transactions on Systems, Man and Cybernetics, V. SMC-3, n.1. Janeiro. 1973.

Anexo 1 – Proposta de uma linguagem de regras

Como forma de facilitar o entendimento dos exemplos apresentados neste trabalho e auxiliar futuras implementações sugere-se a linguagem proposta neste anexo, utilizada para definição das regras.

A linguagem de regras aqui proposta trabalha tanto com lógica tradicional como com lógica fuzzy. Outra característica interessante é a capacidade de ler e escrever dados de forma que estes sejam persistentes. Ou seja, permitir salvar ou ler valores de um *buffer* em disco.

Para a especificação sintática desta linguagem, este anexo apresenta uma GLC (gramática livre de contexto) especificada em BNF (*Backus Naur Form*, notação utilizada na especificação formal da sintaxe de linguagens)⁵⁵.

Visando facilitar o entendimento da linguagem para uma futura implementação, não se preocupou com a fatoração da gramática, bem como com a eliminação de recursões à esquerda. Isto deverá ser feito caso o analisador sintático implementado seja descendente preditivo, ou qualquer outra implementação determinística.

Os itens em vermelho da gramática apresentada só serão necessários se for implementada uma linguagem tipada, caso contrário devem ser ignorados.

Os itens em verde serão necessários apenas caso se queira dar ao conjunto de regras a capacidade de adicionar ou remover regras da base de conhecimento de forma dinâmica.

Ressalta-se que a gramática apresentada a seguir é apenas uma proposta. Para uma real implementação deve-se melhor especificar a linguagem.

```
<conjunto_de_regras_gerais> ::= <regras_inicializacao> <regras_producao>
    <regras_finalizacao>
<regras_inicializacao> ::= 'INICIO' '{' <conjunto_de_regras> '}' | ε
<conjunto_de_regras> ::= <regra> <conjunto_de_regras> | <regra>
<regra> ::= <if_logico> | <fuzzy> | <comando> | <comentario>
<if_logico> ::= 'SE' <expressao_logica> 'ENTAO' '{' <conjunto_de_regras> '}'
    <resto_se>
```

⁵⁵ AHO, Alfred V., SETHI, Ravi, ULLMAN, Jeffrey D. **Compiladores, Princípios, Técnicas e Ferramentas**. LTC - Livros Técnicos e Científicos Editora S.A. Rio de Janeiro, 1995. Tradução de Daniel de Ariosto Pinto. (Compilers, Principles, Techniques and Tools – 1986 Estados Unidos Addison - Wesley Massachusetts)

```

<resto_se> ::= 'SENAO' '{' <conjunto_de_regras> '}' | ε
<expressao_logica> ::= <expressao_logica_simples> | '(' <expressao_logica>
    <operador> <expressao_logica> ')' | '(' 'NAO' <expressao_logica> ')'
<expressao_logica_simples> ::= <atributo> <operador_comparacao> <atributo>
<atributo> ::= <constante> | <variavel> | <valor_de_entrada>
<operador> ::= 'E' | 'OU' | 'XOU'
<operador_comparacao> ::= '>' | '<' | '>=' | '<=' | '!=' | '='
<constante> ::= <constante_string> | <constante_numerica> |
    <constante_ponto_flutuante> | <constante_data> | <constante_hora> |
    <constante_data_hora> | <constante_fuzzy> | <constante_booléana>
<constante_string> ::= <sequencia_de_caracteres>
<constante_numerica> ::= <sequencia_de_numeros>
<constante_ponto_flutuante> ::= <sequencia_de_numeros>'x' |
    <sequencia_de_numeros>'.'<sequencia_de_numeros>'x'
<constante_fuzzy> ::= '0.'<sequencia_de_numeros>'f' | 'lf' | '0f'
<constante_data> ::=
    <dígito><digito>/'<digito><digito>/'<digito><digito><digito><digito>
<constante_hora> ::= <digito><digito>':'<digito><digito>':'<digito><digito>
<constante_data_hora> ::= <constante_data>'-'<constante_hora>
<constante_booléana> ::= 'VERDADEIRO' | 'FALSO'
<sequencia_de_caracteres> ::= <caracter> | <caracter>
    <sequencia_de_caracteres>
<caracter> ::= <digito> | <letra> | '!' | '@' | '#' | '$' | '%' | '^' | '&' |
    '*' | '(' | ')' | '|' | '-' | '+' | '=' | '[' | ']' | '{' | '^' |
    '~' | ';' | ':' | '-' | '>' | ',' | '<' | '/' | '|'
<letra> ::= 'a..z' | 'A..Z' | 'á' | 'é' | 'í' | 'ó' | 'ú' | 'Á' | 'É' | 'Í'
    | 'Ó' | 'Ú' | 'ç' | 'Ç' | 'à' | 'À' | 'ù' | 'Û' | 'ä' | 'Ä' | 'ö'
    | 'Ö' | 'i' | 'á' | 'é' | 'ó' | 'Á' | 'É' | 'Ö'
<digito> ::= '0..9'
<sequencia_de_numeros> ::= <digito> | <digito> <sequencia_de_numeros>
<variável> ::= <identificador>
<valor_de_entrada> ::= 'ENTRADA' <sequencia_de_numeros>
<identificador> ::= <caracter_identificador> <resto_caracter_identificador>
<resto_caracter_identificador> ::= <identificador> | ε
<caracter_identificador> ::= <digito> | <letra> | ε
<comando> ::= <atribuicao> | <leia> | <escreva> | <gere_resultado>
    | <declaracao_variavel> | <mudanca_na_base_regra>
<atribuicao> ::= <variável> '<->' <valor_atribuido>
<valor_atribuido> ::= <resto_atribuicao> | <resto_atribuicao>
    <operador_matematico> <resto_atribuicao>
<resto_atribuicao> ::= <atributo> | '(' <resto_atribuicao>
    <operador_matematico> <resto_atribuicao> ')'
<operador_matematico> ::= '+' | '-' | '*' | '/'
<leia> ::= 'LEIA' <variavel>
<escreva> ::= 'ESCREVA' <variavel>
<gere_resultado> ::= 'GERE_RESULTADO' <variavel> <lista_variaveis>
<lista_variaveis> ::= ','<variavel> | ε
<declaracao_variavel> ::= <tipo> <variavel>
<tipo> ::= 'String' | 'Inteiro' | 'PontoFlutuante' | 'Data' | 'Hora' |
    'DataHora' | 'Fuzzy' | 'Booleano'
<mudanca_na_base_regra> ::= <operador_de_mudanca_de_regra> |
    <resto_da_mudanca_de_regra>
<operador_de_mudanca_de_regra> ::= 'ADICIONE' | 'REMOVA'
<resto_da_mudanca_de_regra> ::= <if_logico> | <fuzzy>
<fuzzy> ::= <expressao_logica_fuzzy> '->' <variavel>
<expressao_logica_fuzzy> ::= <resto_expressao_logica_fuzzy> <operador_fuzzy>
    <resto_expressao_logica_fuzzy> | 'NAO' <resto_expressao_logica_fuzzy>
<resto_expressao_logica_fuzzy> ::= '(' <expressao_logica_fuzzy> ')' |
    <atributo>
<operador_fuzzy> ::= 'E' | 'OU'
<comentario> ::= '///' <sequencia_de_caracteres>
<regras_producao> ::= '{' <conjunto_de_regras> '}'
<regras_finalizacao> ::= 'FIM' '{' <conjunto_de_regras> '}' | ε

```

Anexo 2 – Conjunto de regras utilizado no estudo de caso

A seguir apresentam-se os dois conjuntos de regras implementados para os estudos de caso. O Quadro 7 apresenta um conjunto de regras utilizando algumas das características da linguagem apresentada no Anexo 1. O Quadro 8 apresenta um conjunto de regras implementadas para a comparação com o mecanismo aqui apresentado. Este conjunto de regras representa a implementação de um mecanismo comumente utilizado em algumas ferramentas de execução de testes controlados.

```

INICIO
{
  ContadorGrauDeErroAssinanteA <- 0f
  ContadorGrauDeErroAssinanteB <- 0f
  ContadorGrauDeErroDuração <- 0f
  ContadorGrauDeErroDataInício <- 0f
  ContadorGrauDeErroHoraInício <- 0f
  ContadorGrauDeErroStatus <- 0f
  ContadorDeCDRs <- 0
}
{
  // Foi definido que os campos entrarão na ordem: Número do
  // Assinante A (chamador), Número do Assinante B (chamado),
  // Duração da chamada, Data que a chamada iniciou, Hora que a
  // chamada iniciou, Status de fim de chamada
  // Verifica se o Assinante A e B possuem erro
  ENTRADA1 E ENTRADA2 -> GrauErroAssinantesAB
  SE (GrauErroAssinanteAB != 0f)ENTÃO
  {
    // Se Sim faz um OU entre os valores
    ENTRADA1 OU ENTRADA2 -> GrauErroAssinanteA
    ENTRADA1 OU ENTRADA2 -> GrauErroAssinanteB
  }
  SENÃO
  {
    GrauErroAssinanteA <- ENTRADA1
    GrauErroAssinanteB <- ENTRADA2
  }
  ContadorGrauDeErroAssinanteA <- ContadorGrauDeErroAssinanteA +
  GrauErroAssinanteA
  ContadorGrauDeErroAssinanteB <- ContadorGrauDeErroAssinanteB +
  GrauErroAssinanteB
  ContadorGrauDeErroDuração <- ContadorGrauDeErroDuração + ENTRADA3
  ContadorGrauDeErroDataInício <- ContadorGrauDeErroDataInício +
  ENTRADA4
  ContadorGrauDeErroHoraInício <- ContadorGrauDeErroHoraInício +
  ENTRADA5
  ContadorGrauDeErroStatus <- ContadorGrauDeErroStatus + ENTRADA6
  ContadorDeCDRs <- ContadorDeCDRs + 1
}
FIM
{
  ResultadoDoGrauDeErroAssinanteA <- ContadorGrauDeErroAssinanteA /

```

```

ContadorDeCDRs
ResultadoDoGrauDeErroAssinanteB <- ContadorGrauDeErroAssinanteB /
ContadorDeCDRs
ResultadoDoGrauDeErroDuração <- ContadorGrauDeErroDuração /
ContadorDeCDRs
ResultadoDoGrauDeErroDataInício <- ContadorGrauDeErroDataInício /
ContadorDeCDRs
ResultadoDoGrauDeErroHoraInício <- ContadorGrauDeErroHoraInício /
ContadorDeCDRs
ResultadoDoGrauDeErroStatus <- ContadorGrauDeErroStatus /
ContadorDeCDRs
ResultadoDoGrauDeErroAssinanteA E ResultadoDoGrauDeErroAssinanteB
E ResultadoDoGrauDeErroDuração E
ResultadoDoGrauDeErroDataInício E
ResultadoDoGrauDeErroHoraInício E
ResultadoDoGrauDeErroStatus -> ValorMaximo
ResultadoDoGrauDeErroAssinanteA <-
ResultadoDoGrauDeErroAssinanteA / ValorMaximo
ResultadoDoGrauDeErroAssinanteB <-
ResultadoDoGrauDeErroAssinanteB / ValorMaximo
ResultadoDoGrauDeErroDuração <- ResultadoDoGrauDeErroDuração /
ValorMaximo
ResultadoDoGrauDeErroDataInício <-
ResultadoDoGrauDeErroDataInício / ValorMaximo
ResultadoDoGrauDeErroHoraInício <-
ResultadoDoGrauDeErroHoraInício / ValorMaximo
ResultadoDoGrauDeErroStatus <- ResultadoDoGrauDeErroStatus /
ValorMaximo
GERE_RESULTADO ResultadoDoGrauDeErroAssinanteA
ResultadoDoGrauDeErroAssinanteB ResultadoDoGrauDeErroDuração
ResultadoDoGrauDeErroDataInício
ResultadoDoGrauDeErroHoraInício ResultadoDoGrauDeErroStatus
}

```

Quadro 7 – Regras utilizando o mecanismo proposto

```

INICIO
{
ContadorAssinanteA <- 0
ContadorAssinanteB <- 0
ContadorDuração <- 0
ContadorDataInício <- 0
ContadorHoraInício <- 0
ContadorStatus <- 0
ContadorDeCDRs <- 0
}
{
// Foi definido que os campos entrarão na ordem: Número do
Assinante A proveniente da central, Número do Assinante A
proveniente de um equipamento de teste, Número do Assinante
B proveniente da central, Número do Assinante B proveniente
de um equipamento de teste, Duração da chamada proveniente
da central, Duração da chamada proveniente de um equipamento
de teste, Data que a chamada iniciou proveniente da central,
Data que a chamada iniciou proveniente de um equipamento de
teste, Hora que a chamada iniciou proveniente da central,
Hora que a chamada iniciou proveniente de um equipamento de
teste, Status de fim de chamada proveniente da central
Status de fim de chamada proveniente de um equipamento de

```



```

teste
SE ENTRADA1 != ENTRADA2 ENTÃO
{
  ContadorAssinanteA <- ContadorAssinanteA + 1
}
SE ENTRADA3 != ENTRADA4 ENTÃO
{
  ContadorAssinanteB <- ContadorAssinanteB + 1
}
SE ENTRADA5 != ENTRADA6 ENTÃO
{
  ContadorDuração <- ContadorDuração + 1
}
SE ENTRADA7 != ENTRADA8 ENTÃO
{
  ContadorDataInício <- ContadorDataInício + 1
}
SE ENTRADA9 != ENTRADA10 ENTÃO
{
  ContadorHoraInício <- ContadorHoraInício + 1
}
SE ENTRADA11 != ENTRADA12 ENTÃO
{
  ContadorStatus <- ContadorStatus + 1
}
ContadorDeCDRs <- ContadorDeCDRs + 1
}
FIM
{
  valorMaximo <- MAXIMO_VALOR(ContadorAssinanteA,
    ContadorAssinanteB, ContadorDuração, ContadorDataInício,
    ContadorHoraInício, ContadorStatus)
  ResultadoAssinanteA <- ContadorAssinanteA / valorMaximo
  ResultadoAssinanteB <- ContadorAssinanteB / valorMaximo
  ResultadoDuração <- ContadorDuração / valorMaximo
  ResultadoDataInício <- ContadorDataInício / valorMaximo
  ResultadoHoraInício <- ContadorHoraInício / valorMaximo
  ResultadoStatus <- ContadorStatus / valorMaximo
  GERE_RESULTADO ResultadoAssinanteA ResultadoAssinanteB
    ResultadoDuração ResultadoDataInício ResultadoHoraInício
    ResultadoStatus
}

```

Quadro 8 – Regras utilizando um algoritmo tradicional

Anexo 3 – Distribuição de probabilidade dos estudos de casos

O presente anexo descreve as distribuições de probabilidade utilizadas nos estudos de caso apresentados no Capítulo 5. Os dados presentes neste item foram baseados na experiência do autor do trabalho bem como em conversas informais com alguns funcionários de uma operadora de telecomunicações. Estes dados procuram descrever casos hipotéticos que podem ser encontrados em operadoras de telecomunicações em geral. Para maiores informações sobre os casos consulte o Capítulo 5, item 5.2.

Tabela 5 - Distribuição de probabilidade para o número de erros por CDR

Quantidade de erros	Probabilidade
1	82%
2	9%
3	5%
4	4%

Tabela 6 - Distribuição de probabilidade por tipo de erro

Cenário	1	2	3	4	5	6	7	8	9	10	11	12
Tipo erro												
1	10,22%	7,50%	7,50%	7,50%	4,15%	4,15%	23,51%	17,60%	17,60%	17,60%	10,04%	10,04%
2	4,37%	3,20%	3,20%	3,20%	1,80%	1,80%	10,10%	7,55%	7,55%	7,55%	4,30%	4,30%
3	5,84%	31,02%	4,27%	4,27%	32,04%	2,37%	4,69%	28,66%	3,51%	3,51%	30,65%	2,00%
4	5,84%	4,27%	4,27%	4,27%	2,37%	2,37%	4,69%	3,51%	3,51%	3,51%	2,00%	2,00%
5	5,11%	3,74%	3,74%	3,74%	2,07%	2,07%	3,35%	2,51%	2,51%	2,51%	1,43%	1,43%
6	1,46%	1,06%	1,06%	1,06%	0,60%	0,60%	2,01%	1,50%	1,50%	1,50%	0,85%	0,85%
7	10,94%	8,02%	8,02%	8,02%	4,45%	4,45%	6,71%	5,02%	5,02%	5,02%	2,86%	2,86%
8	14,60%	10,70%	10,70%	10,70%	5,93%	5,93%	6,71%	5,02%	5,02%	5,02%	2,86%	2,86%
9	3,65%	2,70%	2,70%	2,70%	1,48%	1,48%	2,01%	1,50%	1,50%	1,50%	0,85%	0,85%
10	5,11%	3,74%	3,74%	3,74%	2,07%	2,07%	4,69%	3,51%	3,51%	3,51%	2,00%	2,00%
11	2,20%	1,60%	1,60%	1,60%	0,89%	0,89%	2,68%	2,01%	2,01%	2,01%	1,14%	1,14%
12	1,46%	1,06%	27,85%	1,06%	0,59%	30,26%	1,34%	1,00%	26,15%	1,00%	0,57%	29,22%
13	4,37%	3,20%	3,20%	3,20%	1,80%	1,80%	4,69%	3,51%	3,51%	3,51%	2,00%	2,00%
14	1,46%	1,06%	1,06%	1,06%	0,59%	0,59%	1,34%	1,00%	1,00%	1,00%	0,57%	0,57%
15	1,47%	1,06%	1,06%	1,06%	0,59%	0,59%	1,34%	1,00%	1,00%	1,00%	0,57%	0,57%
16	7,30%	5,34%	5,34%	5,34%	2,96%	2,96%	6,71%	5,02%	5,02%	5,02%	2,86%	2,86%
17	1,46%	1,10%	1,06%	1,06%	0,59%	0,59%	1,34%	1,00%	1,00%	1,00%	0,57%	0,57%
18	3,65%	2,67%	2,67%	2,67%	1,48%	1,48%	3,35%	2,51%	2,51%	2,51%	1,45%	1,45%
19	3,65%	2,67%	2,67%	2,67%	1,48%	1,48%	3,35%	2,51%	2,51%	2,51%	1,43%	1,43%
20	5,11%	3,74%	3,74%	3,74%	2,07%	2,07%	4,69%	3,51%	3,51%	3,51%	2,00%	2,00%
21	0,73%	0,55%	0,55%	27,34%	30,00%	30,00%	0,70%	0,55%	0,55%	25,70%	29,00%	29,00%

Tabela 7 - Distribuição de probabilidade dos campo

Cenário \ Campo	1	2	3	4	5	6	7	8	9	10	11	12
Ass.A	14,59%	10,70%	10,70%	10,70%	5,95%	5,95%	33,61%	25,15%	25,15%	25,15%	14,34%	14,34%
Ass.B	18,25%	40,09%	13,34%	13,34%	37,08%	7,41%	14,74%	36,18%	11,03%	11,03%	34,93%	6,28%
Dur.	29,19%	21,42%	21,42%	21,42%	11,86%	11,86%	15,43%	11,54%	11,54%	11,54%	6,57%	6,57%
Hora	14,60%	10,66%	37,45%	10,66%	5,94%	35,61%	14,74%	11,03%	36,18%	11,03%	6,28%	34,93%
Data	17,53%	12,84%	12,80%	12,80%	7,10%	7,10%	16,09%	12,04%	12,04%	12,04%	6,88%	6,88%
Status	5,84%	4,29%	4,29%	31,08%	32,07%	32,07%	5,39%	4,06%	4,06%	29,21%	31,00%	31,00%

Anexo 4 – Resultados das análises dos estudos de caso

Este anexo descreve os resultados das regras geradas nos estudos de caso. Os valores das tabelas nas colunas Assinante A, Assinante B, Data, Hora, Duração e Status são valores fuzificados. Os valores da tabela no campo Método referem-se aos dois métodos analisados, sendo que 1 corresponde ao método alternativo e 2 ao método usado por alguns sistemas de teste.

Tabela 8 - Resultados dos dados gerados nos estudos de casos

Método	Cenário	Repetição	Assinante A	Assinante B	Data	Hora	Duração	Status
1	1	1	0.47	0.38	1	0.42	0.37	0.23
1	1	2	0.55	0.53	1	0.47	0.44	0.3
1	1	3	0.5	0.47	1	0.43	0.38	0.3
1	1	4	0.44	0.45	1	0.39	0.4	0.2
1	2	1	0.5	0.7	1	0.51	0.35	0.24
1	2	2	0.52	0.65	1	0.42	0.32	0.2
1	2	3	0.43	0.68	1	0.47	0.36	0.2
1	2	4	0.61	0.68	1	0.51	0.5	0.19
1	3	1	0.39	0.34	0.72	1	0.32	0.26
1	3	2	0.37	0.35	0.94	1	0.34	0.17
1	3	3	0.39	0.4	0.91	1	0.35	0.2
1	3	4	0.36	0.34	0.81	1	0.29	0.19
1	4	1	0.21	0.2	0.46	0.23	0.2	1
1	4	2	0.25	0.25	0.6	0.27	0.2	1
1	4	3	0.21	0.22	0.49	0.21	0.18	1
1	4	4	0.23	0.21	0.5	0.19	0.17	1
1	5	1	0.09	0.19	0.21	0.12	0.1	1
1	5	2	0.13	0.19	0.25	0.09	0.1	1
1	5	3	0.12	0.22	0.25	0.16	0.09	1
1	5	4	0.08	0.19	0.2	0.12	0.09	1
1	6	1	0.13	0.1	0.25	0.5	0.07	1
1	6	2	0.09	0.1	0.25	0.39	0.1	1
1	6	3	0.12	0.1	0.23	0.46	0.11	1
1	6	4	0.13	0.1	0.26	0.5	0.07	1
1	7	1	1	0.39	0.49	0.45	0.36	0.22
1	7	2	1	0.35	0.53	0.37	0.28	0.24
1	7	3	1	0.38	0.5	0.49	0.42	0.31
1	7	4	1	0.41	0.51	0.4	0.37	0.22
1	8	1	1	0.48	0.58	0.58	0.36	0.3
1	8	2	1	0.55	0.55	0.43	0.33	0.26
1	8	3	1	0.5	0.48	0.47	0.32	0.2
1	8	4	1	0.51	0.61	0.54	0.34	0.31
1	9	1	0.91	0.34	0.46	1	0.36	0.23
1	9	2	0.84	0.3	0.42	1	0.24	0.19

1	9	3	0.92	0.36	0.52	1	0.29	0.26
1	9	4	0.9	0.31	0.49	1	0.37	0.27
1	10	1	0.54	0.21	0.31	0.25	0.13	1
1	10	2	0.49	0.15	0.24	0.18	0.2	1
1	10	3	0.48	0.21	0.27	0.23	0.19	1
1	10	4	0.55	0.2	0.31	0.24	0.14	1
1	11	1	0.28	0.19	0.14	0.12	0.09	1
1	11	2	0.27	0.2	0.14	0.18	0.08	1
1	11	3	0.24	0.17	0.13	0.12	0.1	1
1	11	4	0.3	0.19	0.13	0.13	0.08	1
1	12	1	0.25	0.09	0.12	0.4	0.08	1
1	12	2	0.29	0.1	0.14	0.48	0.09	1
1	12	3	0.24	0.14	0.15	0.5	0.08	1
1	12	4	0.27	0.1	0.12	0.42	0.09	1
2	1	1	0.53	0.53	1	0.58	0.53	0.17
2	1	2	0.63	0.75	1	0.58	0.6	0.21
2	1	3	0.58	0.63	1	0.56	0.55	0.22
2	1	4	0.51	0.65	1	0.48	0.58	0.15
2	2	1	0.26	1	0.46	0.28	0.22	0.08
2	2	2	0.3	1	0.53	0.25	0.23	0.07
2	2	3	0.24	1	0.48	0.27	0.24	0.07
2	2	4	0.34	1	0.51	0.29	0.35	0.07
2	3	1	0.27	0.29	0.46	1	0.28	0.11
2	3	2	0.26	0.29	0.58	1	0.29	0.07
2	3	3	0.28	0.31	0.54	1	0.3	0.09
2	3	4	0.26	0.28	0.5	1	0.24	0.09
2	4	1	0.34	0.38	0.68	0.4	0.4	1
2	4	2	0.42	0.45	0.78	0.42	0.38	1
2	4	3	0.32	0.39	0.65	0.35	0.36	1
2	4	4	0.37	0.38	0.66	0.36	0.34	1
2	5	1	0.15	1	0.28	0.18	0.19	0.98
2	5	2	0.18	1	0.32	0.15	0.2	0.95
2	5	3	0.16	1	0.3	0.23	0.15	0.86
2	5	4	0.14	0.96	0.28	0.17	0.18	1
2	6	1	0.16	0.18	0.29	1	0.12	0.83
2	6	2	0.15	0.18	0.33	0.9	0.2	1
2	6	3	0.17	0.18	0.29	1	0.22	0.95
2	6	4	0.16	0.17	0.3	1	0.11	0.83
2	7	1	1	0.48	0.44	0.54	0.44	0.14
2	7	2	1	0.44	0.46	0.46	0.35	0.15
2	7	3	1	0.43	0.47	0.55	0.51	0.2
2	7	4	1	0.48	0.46	0.49	0.46	0.14
2	8	1	0.72	1	0.37	0.46	0.32	0.14
2	8	2	0.65	1	0.33	0.34	0.27	0.11
2	8	3	0.69	1	0.3	0.39	0.28	0.09
2	8	4	0.69	1	0.36	0.42	0.28	0.13
2	9	1	0.61	0.28	0.27	1	0.3	0.1
2	9	2	0.6	0.26	0.25	1	0.21	0.08
2	9	3	0.64	0.28	0.3	1	0.25	0.11
2	9	4	0.6	0.27	0.28	1	0.31	0.12

2	10	1	0.83	0.36	0.4	0.49	0.27	1
2	10	2	0.75	0.3	0.33	0.31	0.38	1
2	10	3	0.74	0.37	0.36	0.37	0.38	1
2	10	4	0.85	0.37	0.4	0.48	0.27	1
2	11	1	0.43	1	0.19	0.19	0.18	0.97
2	11	2	0.36	1	0.17	0.21	0.13	0.82
2	11	3	0.36	0.91	0.16	0.2	0.19	1
2	11	4	0.45	1	0.18	0.2	0.14	0.96
2	12	1	0.38	0.17	0.17	0.92	0.17	1
2	12	2	0.41	0.15	0.16	1	0.15	0.9
2	12	3	0.34	0.21	0.17	1	0.15	0.89
2	12	4	0.41	0.18	0.18	0.98	0.18	1

Anexo 5 – Resultados da análise estatística

Segue um relatório da análise fatorial 2 x 12 gerada pelo software The SAS System (<http://www.sas.com/>) versão 8.0 de 1999. Os dados deste relatório foram gerados a partir dos dados do Anexo 4.

Análise de variância de um fatorial 2 x 12

The GLM Procedure

Class Level Information

Class	Levels	Values
metodo	2	1 2
cenarios	12	1 2 3 4 5 6 7 8 9 10 11 12

Number of observations 96

Análise de variância de um fatorial 2 x 12

The GLM Procedure

Dependent Variable: Assinante A

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	7.62196250	0.33138967	287.82	<.0001
Error	72	0.08290000	0.00115139		
Corrected Total	95	7.70486250			

R-Square	Coeff Var	Root MSE	Assinante A Mean
0.989241	7.115517	0.033932	0.476875

Source	DF	Type I SS	Mean Square	F Value	Pr > F
metodo	1	0.00166667	0.00166667	1.45	0.2329
cenarios	11	6.86586250	0.62416932	542.10	<.0001
metodo*cenarios	11	0.75443333	0.06858485	59.57	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
metodo	1	0.00166667	0.00166667	1.45	0.2329
cenarios	11	6.86586250	0.62416932	542.10	<.0001
metodo*cenarios	11	0.75443333	0.06858485	59.57	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
central A vs B	1	2.93300417	2.93300417	2547.36	<.0001

Análise de variância de um fatorial 2 x 12

The GLM Procedure

Dependent Variable: Assinante B

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	7.99404896	0.34756735	348.90	<.0001

Error	72	0.07172500	0.00099618
Corrected Total	95	8.06577396	

R-Square	Coeff Var	Root MSE	Assinante B Mean
0.991107	7.247031	0.031562	0.435521

Source	DF	Type I SS	Mean Square	F Value	Pr > F
metodo	1	1.51755104	1.51755104	1523.37	<.0001
cenarios	11	4.57293646	0.41572150	417.32	<.0001
metodo*cenarios	11	1.90356146	0.17305104	173.71	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
metodo	1	1.51755104	1.51755104	1523.37	<.0001
cenarios	11	4.57293646	0.41572150	417.32	<.0001
metodo*cenarios	11	1.90356146	0.17305104	173.71	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
central A vs B	1	0.05367604	0.05367604	53.88	<.0001

Analise de variancia de um fatorial 2 x 12

The GLM Procedure

Dependent Variable: Data

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	6.85324896	0.29796735	222.72	<.0001
Error	72	0.09632500	0.00133785		
Corrected Total	95	6.94957396			

R-Square	Coeff Var	Root MSE	Data Mean
0.986139	7.971289	0.036577	0.458854

Source	DF	Type I SS	Mean Square	F Value	Pr > F
metodo	1	0.11275104	0.11275104	84.28	<.0001
cenarios	11	5.85431146	0.53221013	397.81	<.0001
metodo*cenarios	11	0.88618646	0.08056241	60.22	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
metodo	1	0.11275104	0.11275104	84.28	<.0001
cenarios	11	5.85431146	0.53221013	397.81	<.0001
metodo*cenarios	11	0.88618646	0.08056241	60.22	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
central A vs B	1	1.77398438	1.77398438	1326.00	<.0001

Analise de variancia de um fatorial 2 x 12

The GLM Procedure

Dependent Variable: Hora

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	8.65264063	0.37620177	226.62	<.0001
Error	72	0.11952500	0.00166007		

Corrected Total	95	8.77216562			
	R-Square	Coeff Var	Root MSE	Hora Mean	
	0.986375	7.935523	0.040744	0.513438	

Source	DF	Type I SS	Mean Square	F Value	Pr > F
metodo	1	0.32783438	0.32783438	197.48	<.0001
cenarios	11	7.29245313	0.66295028	399.35	<.0001
metodo*cenarios	11	1.03235312	0.09385028	56.53	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
metodo	1	0.32783438	0.32783438	197.48	<.0001
cenarios	11	7.29245313	0.66295028	399.35	<.0001
metodo*cenarios	11	1.03235312	0.09385028	56.53	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
central A vs B	1	0.00467604	0.00467604	2.82	0.0976

Analise de variancia de um fatorial 2 x 12

The GLM Procedure

Dependent Variable: Duracao

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	1.46613333	0.06374493	38.90	<.0001
Error	72	0.11800000	0.00163889		
Corrected Total	95	1.58413333			

	R-Square	Coeff Var	Root MSE	Duracao Mean
	0.925511	15.47128	0.040483	0.261667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
metodo	1	0.06720417	0.06720417	41.01	<.0001
cenarios	11	1.18633333	0.10784848	65.81	<.0001
metodo*cenarios	11	0.21259583	0.01932689	11.79	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
metodo	1	0.06720417	0.06720417	41.01	<.0001
cenarios	11	1.18633333	0.10784848	65.81	<.0001
metodo*cenarios	11	0.21259583	0.01932689	11.79	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
central A vs B	1	0.01500000	0.01500000	9.15	0.0034

Analise de variancia de um fatorial 2 x 12

The GLM Procedure

Dependent Variable: Status

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	15.57644583	0.67723678	470.89	<.0001
Error	72	0.10355000	0.00143819		

Corrected Total 95 15.67999583

R-Square Coeff Var Root MSE Status Mean
0.993396 6.555022 0.037924 0.578542

Source	DF	Type I SS	Mean Square	F Value	Pr > F
metodo	1	0.15360000	0.15360000	106.80	<.0001
cenarios	11	15.36824583	1.39711326	971.44	<.0001
metodo*cenarios	11	0.05460000	0.00496364	3.45	0.0007

Source	DF	Type III SS	Mean Square	F Value	Pr > F
metodo	1	0.15360000	0.15360000	106.80	<.0001
cenarios	11	15.36824583	1.39711326	971.44	<.0001
metodo*cenarios	11	0.05460000	0.00496364	3.45	0.0007

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
central A vs B	1	0.00350417	0.00350417	2.44	0.1229

Analise de variancia de um fatorial 2 x 12

The GLM Procedure
Least Squares Means

metodo	cenarios	Assinante A LSMEAN	Assinante B LSMEAN	Data LSMEAN	Hora LSMEAN	Duracao LSMEAN
1	1	0.49000000	0.45750000	1.00000000	0.42750000	0.39750000
1	2	0.51500000	0.67750000	1.00000000	0.47750000	0.38250000
1	3	0.37750000	0.35750000	0.84500000	1.00000000	0.32500000
1	4	0.22500000	0.22000000	0.51250000	0.22500000	0.18750000
1	5	0.10500000	0.19750000	0.22750000	0.12250000	0.09500000
1	6	0.11750000	0.10000000	0.24750000	0.46250000	0.08750000
1	7	1.00000000	0.38250000	0.50750000	0.42750000	0.35750000
1	8	1.00000000	0.51000000	0.55500000	0.50500000	0.33750000
1	9	0.89250000	0.32750000	0.47250000	1.00000000	0.31500000
1	10	0.51500000	0.19250000	0.28250000	0.22500000	0.16500000
1	11	0.27250000	0.18750000	0.13500000	0.13750000	0.08750000
1	12	0.26250000	0.10750000	0.13250000	0.45000000	0.08500000
2	1	0.56250000	0.64000000	1.00000000	0.55000000	0.56500000
2	2	0.28500000	1.00000000	0.49500000	0.27250000	0.26000000
2	3	0.26750000	0.29250000	0.52000000	1.00000000	0.27750000
2	4	0.36250000	0.40000000	0.69250000	0.38250000	0.37000000
2	5	0.15750000	0.99000000	0.29500000	0.18250000	0.18000000
2	6	0.16000000	0.17750000	0.30250000	0.97500000	0.16250000
2	7	1.00000000	0.45750000	0.45750000	0.51000000	0.44000000
2	8	0.68750000	1.00000000	0.34000000	0.40250000	0.28750000
2	9	0.61250000	0.27250000	0.27500000	1.00000000	0.26750000
2	10	0.79250000	0.35000000	0.37250000	0.41250000	0.32500000
2	11	0.40000000	0.97750000	0.17500000	0.20000000	0.16000000
2	12	0.38500000	0.17750000	0.17000000	0.97500000	0.16250000

metodo	cenarios	Status LSMEAN
1	1	0.25750000
1	2	0.20750000
1	3	0.20500000
1	4	1.00000000
1	5	1.00000000
1	6	1.00000000
1	7	0.24750000
1	8	0.26750000
1	9	0.23750000
1	10	1.00000000
1	11	1.00000000
1	12	1.00000000
2	1	0.18750000
2	2	0.07250000
2	3	0.09000000
2	4	1.00000000
2	5	0.94750000
2	6	0.90250000

2	7	0.15750000
2	8	0.11750000
2	9	0.10250000
2	10	1.00000000
2	11	0.93750000
2	12	0.94750000

Analise de variancia de um fatorial 2 x 12

The GLM Procedure
Least Squares Means

metodo*cenarios Effect Sliced by cenarios for Assinante A

cenarios	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	1	0.010513	0.010513	9.13	0.0035
2	1	0.105800	0.105800	91.89	<.0001
3	1	0.024200	0.024200	21.02	<.0001
4	1	0.037813	0.037813	32.84	<.0001
5	1	0.005512	0.005512	4.79	0.0319
6	1	0.003613	0.003613	3.14	0.0807
7	1	1.390521E-31	1.390521E-31	0.00	1.0000
8	1	0.195312	0.195312	169.63	<.0001
9	1	0.156800	0.156800	136.18	<.0001
10	1	0.154013	0.154013	133.76	<.0001
11	1	0.032512	0.032512	28.24	<.0001
12	1	0.030012	0.030012	26.07	<.0001

Analise de variancia de um fatorial 2 x 12

The GLM Procedure
Least Squares Means

metodo*cenarios Effect Sliced by cenarios for Assinante B

cenarios	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	1	0.066613	0.066613	66.87	<.0001
2	1	0.208013	0.208013	208.81	<.0001
3	1	0.008450	0.008450	8.48	0.0048
4	1	0.064800	0.064800	65.05	<.0001
5	1	1.256112	1.256112	1260.93	<.0001
6	1	0.012013	0.012013	12.06	0.0009
7	1	0.011250	0.011250	11.29	0.0012
8	1	0.480200	0.480200	482.04	<.0001
9	1	0.006050	0.006050	6.07	0.0161
10	1	0.049613	0.049613	49.80	<.0001
11	1	1.248200	1.248200	1252.99	<.0001
12	1	0.009800	0.009800	9.84	0.0025

Analise de variancia de um fatorial 2 x 12

The GLM Procedure
Least Squares Means

metodo*cenarios Effect Sliced by cenarios for Data

cenarios	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	1	7.800016E-33	7.800016E-33	0.00	1.0000
2	1	0.510050	0.510050	381.25	<.0001
3	1	0.211250	0.211250	157.90	<.0001
4	1	0.064800	0.064800	48.44	<.0001
5	1	0.009113	0.009113	6.81	0.0110
6	1	0.006050	0.006050	4.52	0.0369
7	1	0.005000	0.005000	3.74	0.0571
8	1	0.092450	0.092450	69.10	<.0001
9	1	0.078012	0.078012	58.31	<.0001
10	1	0.016200	0.016200	12.11	0.0009
11	1	0.003200	0.003200	2.39	0.1263
12	1	0.002812	0.002812	2.10	0.1514

Analise de variancia de um fatorial 2 x 12

The GLM Procedure
Least Squares Means

metodo*cenarios Effect Sliced by cenarios for Hora

cenarios	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	1	0.030012	0.030012	18.08	<.0001
2	1	0.084050	0.084050	50.63	<.0001
3	1	0	0	0.00	1.0000
4	1	0.049613	0.049613	29.89	<.0001
5	1	0.007200	0.007200	4.34	0.0408
6	1	0.525313	0.525313	316.44	<.0001
7	1	0.013613	0.013613	8.20	0.0055
8	1	0.021012	0.021012	12.66	0.0007
9	1	2.46519E-32	2.46519E-32	0.00	1.0000
10	1	0.070313	0.070313	42.36	<.0001
11	1	0.007813	0.007813	4.71	0.0334
12	1	0.551250	0.551250	332.06	<.0001

Analise de variancia de um fatorial 2 x 12

The GLM Procedure
Least Squares Means

metodo*cenarios Effect Sliced by cenarios for Duracao

cenarios	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	1	0.056113	0.056113	34.24	<.0001
2	1	0.030013	0.030013	18.31	<.0001
3	1	0.004512	0.004512	2.75	0.1014
4	1	0.066613	0.066613	40.64	<.0001
5	1	0.014450	0.014450	8.82	0.0041
6	1	0.011250	0.011250	6.86	0.0107
7	1	0.013613	0.013613	8.31	0.0052
8	1	0.005000	0.005000	3.05	0.0850
9	1	0.004512	0.004512	2.75	0.1014
10	1	0.051200	0.051200	31.24	<.0001
11	1	0.010513	0.010513	6.41	0.0135
12	1	0.012012	0.012012	7.33	0.0085

Analise de variancia de um fatorial 2 x 12

The GLM Procedure
Least Squares Means

metodo*cenarios Effect Sliced by cenarios for Status

cenarios	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	1	0.009800	0.009800	6.81	0.0110
2	1	0.036450	0.036450	25.34	<.0001
3	1	0.026450	0.026450	18.39	<.0001
4	1	1.165188E-32	1.165188E-32	0.00	1.0000
5	1	0.005513	0.005513	3.83	0.0541
6	1	0.019013	0.019013	13.22	0.0005
7	1	0.016200	0.016200	11.26	0.0013
8	1	0.045000	0.045000	31.29	<.0001
9	1	0.036450	0.036450	25.34	<.0001
10	1	1.627411E-32	1.627411E-32	0.00	1.0000
11	1	0.007813	0.007813	5.43	0.0226
12	1	0.005513	0.005513	3.83	0.0541