Utilização de sistemas inteligentes para o processamento de alarmes – Mineração de dados usando OLAP

L. Scheunemann, Ernani A. da Silva Neto, Alexandre C. Maciel, Eletrosul e
J. M. de Souza, M. Sanches, FITec

***Resumo* -O presente artigo mostra os métodos utilizados na etapa inicial do projeto de pesquisa entre a FITec e a Eletrosul que visa a construção de uma ferramenta de Gerência de Alarmes visando a identificação da causa-raiz para a rede de telecomunicações da Eletrosul. A etapa inicial consiste na análise da base de dados de alarmes e seus atributos, da topologia da rede e sua organização em sub-redes, da identificação de cenários significativos de alarmes que representem o dia-a-dia do centro de operação. Discutimos a técnica OLAP (On-Line Analytical Processing) e técnicas estatísticas como Pareto e Gráfico de Controle, na identificação e construção dos cenários de alarmes.**

Palavras-chave: **Gerência de alarmes, OLAP, Redes SDH,.Gráfico de Controle, Pareto.**

I. INTRODUÇÃO

As redes de comunicação estão cada vez mais complexas abrangendo um grande número de diferentes equipamentos que funcionam juntos para oferecimento do serviço fim-a-fim aos usuários. A esse aumento de complexidade corresponde um aumento no número de falhas permanentes/transientes e um aumento substancial no número de alarmes no centro de gerência da rede devido à interdependência entre os elementos de rede (ER), a ocorrência primária gerando múltiplas ocorrências secundárias. Isso implica que um alarme gerado por um ER se propague rapidamente causando um efeito avalanche de alarmes que chegam até o sistema de gerência.

No contexto de Gerência de Redes, a gerência de alarmes é uma facilidade importante no sentido de permitir ao operador uma análise e ação rápidas sobre as causas da ocorrência evitando um possível colapso da rede.

O presente trabalho de pesquisa entre a FITec e a Eletrosul visa a construção de uma ferramenta de Gerência de Alarmes para a rede de telecomunicações da Eletrosul visando a rápida identificação da causa-raiz na ocorrência de alarmes que sinalizam indisponibilidade parcial ou total de enlaces do sistema.

A rede de telecomunições da Eletrosul é composta das sub-redes:

- o anel SDH (Synchronous Digital Hierarchy), cerca de 35 nós,
- as rotas rádio necessárias à comunicação de algumas subestações (SE), são cinco rotas rádio, 1 e 3 a 6. A rota 2 ainda não está disponível.
- a rede de supervisão com quatro servidores.

Os alarmes gerados pelas sub-redes tem características diferentes e devem ser analisados separadamente.

A rede SDH é o conjunto de equipamentos e meios físicos de transmissão que compõem um sistema digital síncrono de transporte de informações. Um elemento de rede SDH quando percebe um problema na recepção sinal ou se não há recepção gera os alarmes LOF (Loss Of Frame) e LOS (Loss Of Signal). Para localização do enlace com problema, gera mais dois alarmes: o alarme AIS (Alarm Indication Signal) é enviado na direção do terminal de destino, que chamaremos para frente, e o sinal RDI (Remote Defect Indication) é enviado na direção contrária, na direção do terminal de origem, que chamaremos para trás. Esses são repetidos pelos nós intermediários entre origem e destino e permitem rastrear, baseado na topologia, em que ponto do enlace entre os terminais houve o problema. A forma de geração dos alarmes AIS e RDI pode no entanto variar entre fornecedores o que requer a análise de cenários reais para entendimento desse procedimento. Existem outros alarmes associados à degradação do sinal e perda de sincronismo.

A rotas rádio não tem um padrão comum de alarmes e dependem das definições dos equipamentos rádio de cada fornecedor. Nesse cenário devemos levantar os alarmes rádio presentes na rede, as respectivas funções e a que tipo de equipamento rádio se referem.

Finalmente rede de supervisão é composta de servidores que monitoram o anel óptico e as rotas rádio. A falta de comunicação com um nó gera o alarme CLF (Communication Link Failure with NE, Network Equipment) para cada nó isolado.

Os alarmes gerados pelo sistema vão para a tela de operação permanecendo na tela enquanto ativados. Na desativação saem da tela e são gravados em uma base de histórico. São gravados na base histórica de alarmes (BHA) diversos atributos do alarme que serão mostrados posteriormente.

Como explicado anteriormente, é necessário o levantamento dos alarmes nas diversas sub-redes e como se correlacionam. Nos casos de avalanche de alarmes. precisamos ter regras de identificação dos alarmes primários que servirão para a identificação da causa-raiz.

Para esse levantamento e o estabelecimento de regras precisamos minerar a base de dados de ocorrência à procura de cenários significativos e nesses cenários levantar a cronologia de ocorrências dos alarmes. A base contém de 5000 a 10000 ocorrências por mês. Para efetuar a procura de cenários recorremos a uma ferramenta, OLAP (On Line Analytical Processing).

Nesse artigo tratamos da etapa de procura e criação de cenários desse projeto. Queremos mostrar que o uso de OLAP facilita:

- a análise de cenários diretamente na base de produção, sem bases intermediárias,
- a interação entre as equipes usando cenários que estão ocorrendo e portanto ainda bem presente na memória dos operadores,
- a avaliação da eficiência de regras e métodos estatísticos na precisa identificação e localização do problema.

Na seção seguinte discutimos brevemente algumas linhas de pesquisa na área de gerência de alarmes. Na seção III apresentamos o conceito de mineração de dados e a ferramenta para análise e criação de cenários que estamos utilizando, OLAP. A base de dados de alarmes e seus atributos é apresentada na seção IV. Os métodos que estão sendo usados para análise e criação de cenários são apresentados na seção V.

II. PESQUISAS EM GERÊNCIA DE ALARMES

Podemos identificar as seguintes vertentes de pesquisa na gerência de falhas:

1. monitoração estatístico dos alarmes e identificação dos pontos críticos [3],
2. redução do número de alarmes filtrando os alarmes secundários [4],
3. uso de sistemas inteligentes [4, 5],

A. Monitoração estatística de Alarmes (vertente 1)

Redes e equipamentos mal ajustados ou perto do limiar de falha tendem a gerar mais alarmes, seja em relação a um limite esperado ou comparando-se com seus equivalentes. A deteção desses pontos fora da curva contribui para a confiabilidade da rede e para o aumento de confiabilidade de equipamentos quando submetidos a ajustes automáticos desnecessários.

Medidas como a freqüência de alarmes, tempo de solução de ocorrência, etc podem ser derivadas dos alarmes e analisadas quantitativamente. Técnicas de Controle Estatístico de Processos [1, 2] podem ser usadas para detectar as medidas fora de controle e que portanto devem ser explicadas pela operação o que muitas vezes leva a um ajuste ou reparação.

A vantagem desse tipo de monitoração é que ele seleciona os pontos a serem analisados agindo como um filtro de ocorrências permitindo à operação focar nas áreas críticas.

B. Análise da correlação entre alarmes (vertentes 2 e 3)

Nas redes de telecomunicações existe geralmente uma grande interdependência entre os Elementos de Rede (ER) que cooperam para a prestação de um determinado serviço.

O principal objetivo dessa vertente é determinar a(s) causa(s) com precisão, filtrando alarmes secundários (que são causados pelos alarmes primários) para em seguida iniciar a análise de causa-raiz, resultando nas prováveis causas com um certo nível de confiança (probabilidade de acerto).

Basicamente dois elementos funcionais são necessários: Analisador de Correlação de Eventos, para filtragem, e Analisador de Causa-Raiz, para identificação das causas.

Podemos considerar que essa é ainda uma área de pesquisa, onde as soluções atuais apresentam complexidade crescente em função da complexidade da rede, causando um esforço de implantação importante e constante, devido às rápidas mudanças na rede em função da introdução de novos serviços [4].

Nossa abordagem combina as vertentes 1 e 3, ou seja, usamos métodos estatísticos para análise de mudanças de tendência e métodos de mineração de dados para estabelecimento de regras para filtro e localização dos problemas.

III. MINERAÇÃO DE DADOS

A mineração de dados é uma estratégia que procura em uma base de dados, criando modelos para identificação de padrões escondidos. O modelo criado pelo algoritmo de mineração de dados é uma generalização conceitual dos dados. A generalização pode ser na forma de uma árvore, uma rede, uma equação ou um conjunto de regras.

Um processo de pesquisa (“data query”) pode ajudar a encontrar respostas às questões sobre informações armazenadas nos dados, mas a mineração de dados difere de um processo de pesquisa nos dados ao dar a habilidade de encontrar respostas às questões que não haviam sido pensadas. Os sistemas especialistas, por sua vez, usam o conhecimento humano ao invés de construir modelos para tomada de decisão. Quando não existem dados suficientes, a abordagem de sistemas especialistas pode ser uma alternativa viável.

A mineração de dados é um processo que compreende acessar e preparar os dados para o algoritmo de mineração de dados, minerar os dados, analisar os resultados, e agir de forma apropriada. Os dados a serem acessados podem estar armazenados em uma ou mais bases de dados operacionais, “data warehouse”, ou arquivo.

No desenvolvimento da. ferramenta de Gerência de Alarmes vamos usar técnicas de mineração para geração de regras visando a rápida identificação dos alarmes primários. Para chegar nesse ponto precisamos identificar cenários significativos de alarmes do dia-a-dia da gerência, identificando para cada cenário os atributos importantes e separando os alarmes primários do secundários.

Usamos uma técnica que permite a mineração “manual” dos dados, “On-Line Analytical Processing”, OLAP. Chamamos de manual [6] porque é o operador que dirige a procura indicando os atributos que quer analisar e a apresentação dos resultados (quantidade, média, etc.).

A. OLAP (On-line Analytical Processing)

O OLAP [6] é uma metodologia baseada em pesquisa (“query-based”) que suporta análise de dados em um ambiente multidimensional. OLAP é uma ferramenta

muito útil para verificar ou rebater alguma hipótese e para executar, conduzida pelo operador, uma mineração de dados.

Um processamento OLAP estrutura logicamente dados multidimensionais na forma de um cubo. O cubo pode apresentar várias dimensões que são subconjuntos de atributos.

Como um cubo é projetado para um propósito específico, não é usual ter vários cubos estruturados a partir de um único "warehouse". O projeto dos dados do cubo inclui a decisão sobre quais atributos serão incluídos no cubo, bem como a granularidade de cada atributo. Um cubo bem projetado é configurado de forma a conter somente informação útil.

Cada atributo em um cubo OLAP pode ter uma ou mais hierarquias conceituais associadas. Uma hierarquia conceitual define um mapeamento que permite que o atributo possa ser visualizado em diversos níveis de detalhe. Por exemplo: atributo de localidade, poderia ter a região como o nível mais alto, depois as cidades que compõe aquela região, e depois os endereços contidos naquela cidade.

Operações suportadas no OLAP:

1. Operação fatiar ("slice") - seleciona dados de uma única dimensão de um cubo OLAP.
2. Operação cortar um subcubo ("dice") - extrai um subcubo do cubo original executando um operação de seleção em duas ou mais dimensões.
3. Operação de agregação ("roll-up") - é a combinação de células de uma ou mais dimensões definidas num cubo. Uma forma de agregação usa o conceito de associação hierárquica com uma dimensão para atingir um nível maior de generalização.
4. Operação de "drill-down" - é o reverso da agregação ("roll-up"), implica em examinar dados com algum nível maior de detalhe.
5. Operação de rotação ("rotation") - permite visualizar dados de uma nova perspectiva.

As ferramentas de OLAP possuem uma interface ao usuário amigável e são capazes de mostrar os dados de diversas perspectivas, executar análises estatísticas e fazer pesquisas sucessivas para menor e/ou maior nível de detalhe. Para pequena quantidade de dados, o "MS Excel pivot table" pode ser utilizado oferecendo algumas das funcionalidades disponíveis em ferramentas de OLAP complexas. As funcionalidades do "Pivot Table" incluem a habilidade de sumarizar e agrupar dados e mostrar dados em diversos formatos.

IV. A BASE HISTÓRICA DE ALARMES (BHA)

A base histórica de alarmes (BHA) foi usada para o levantamento de alarmes nas três sub-redes, a criação de cenários e a identificação de regras.

A BHA possui os seguintes atributos:

- (Time) Data/hora da ativação do alarme,
- (NE) Elemento de Rede: identificação do elemento,
- (Object Type) Objeto (enlace, tributário, camada SDH, etc.),
- (Location) Localização no equipamento,
- (Cause) Causa: identificação do alarme,
- (Severity) Severidade do alarme,
- (Alarm Class) Classe do alarme,
- (Cleared Time) Data/hora de desativação.

Time (Local)	NE	Object Type	Location
23/05/2007 10:49:09	DUALQ-GRA,,Dual Q-Adapter GFL	NE	
23/05/2007 10:45:05	DUALQ-GRA,,Dual Q-Adapter GFL	NE	
23/05/2007 10:43:27	PFU-01,B25/S4,SLD16B 2.5	MS4-TTP	405
23/05/2007 10:44:59	PFU-02,B25/S4,SMA1/4C 2.3M	VC4-TTP	411.01
23/05/2007 10:44:57	PFU-02,B25/S4,SMA1/4C 2.3M	VC4-TTP	407.01
23/05/2007 10:43:19	PFU-01,B25/S4,SLD16B 2.5	MS4-TTP	405
23/05/2007 10:44:55	PFU-02,B25/S4,SMA1/4C 2.3M	VC12-TTP	403.17
23/05/2007 10:44:55	PFU-02,B25/S4,SMA1/4C 2.3M	VC12-TTP	403.16

(a)

Cause	Severity	Alarm Class	Cleared Time (Local)
Communication Link Failure with NE	Critical	Communication	23/05/2007 10:55:54
Equipment Problem	Indetermin	Equipment	23/05/2007 10:55:54
Threshold Crossed: UAT Far End (UAT far)	Minor	Quality	23/05/2007 10:46:04
Remote Defect Indication (RDI)	Warning	Communication	23/05/2007 10:47:33
Remote Defect Indication (RDI)	Warning	Communication	23/05/2007 10:47:32
Remote Defect Indication (RDI)	Warning	Communication	23/05/2007 10:46:04
Remote Defect Indication (RDI)	Warning	Communication	23/05/2007 10:47:32
Remote Defect Indication (RDI)	Warning	Communication	23/05/2007 10:47:32

(b)

Figura 1. Instâncias da BHA

A figura 1 (a)/(b) (divisão necessária para formatá-las em uma coluna) mostra 8 instâncias da BHA. Com exceção da data/hora de desativação, os demais atributos são gerados em tempo real no momento da ativação.

Um primeiro procedimento referente à topologia é a organização dos ERs por sub-rede e identificação da rota rádio a que pertence. Também é calculada a duração do alarme conhecendo-se na BHA a data/hora de ativação/desativação.

V. ANÁLISE DE CENÁRIOS

Para a construção e análise de cenários por meio da colaboração entre a FITec e a Eletrosul, foi instalada na Eletrosul a ferramenta OLAP com tela de rosto ("front-end") Excel. A BHA para geração do cubo OLAP coleta em tempo real todos os alarmes gerados atualizando o cubo de 10 em 10 minutos ou no momento que o Excel é ativado. Isso permite a colaboração usando a mesma base e o mesmo processo de análise.

A. Alarmes mais freqüentes: Rotas rádio

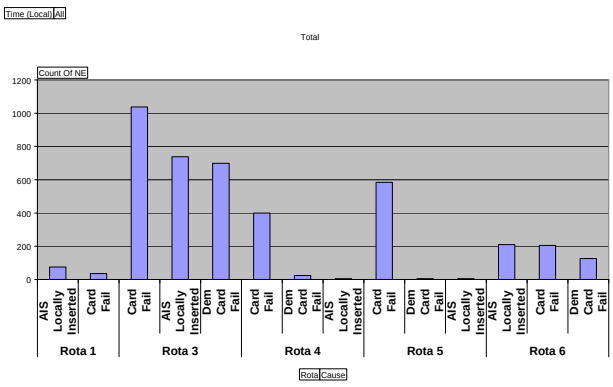

Figura 2. Alarmes mais freqüentes nas rotas rádio

A BHA usado no artigo foi coletada entre 29/04/2007 e 23/05/2007. A primeira análise feita foi sobre os alarmes mais freqüentes nas rotas rádio e no anel óptico.

A figura 2 exemplifica o uso do Excel como tela de rosto para análise do cubo. Na análise usamos apenas a operação "fatiar" onde selecionamos os dados em uma única dimensão. No caso fatiamos "Rota/Causa" pela quantidade de ocorrências. Usamos o diagrama de Pareto mostrando as três causas mais freqüentes por rota. Vemos que em todas as rotas aparece a causa "card fail", "AIS locally inserted" e "Dem (Demodulator) Card Fail". Observe a coluna do lado esquerdo. Ela contém todos os atributos considerados importantes para essa etapa inicial de entendimento e criação de cenários. Para "fatiar" usando outro atributo, basta arrastá-lo para o eixo X, que indica o atributo. O eixo Y contém a quantidade de ocorrências do(s) atributo(s) escolhido(s). Os demais gráficos são gerados pelo Excel.

Para analisar o comportamento desses três alarmes no período estudado arrastamos para o eixo X os atributos data/hora de ativação e rotas, selecionando apenas as rotas rádio. No atributo "cause" selecionamos apenas os alarmes em questão. O resultado é mostrado na figura 3.

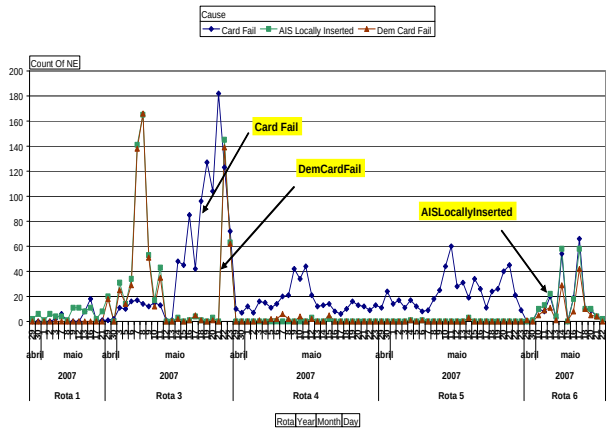

Figura 3. Rotas rádio, análise dos três alarmes com maior número de ocorrências.

A figura 3 sugere forte correlação entre os alarmes "DemCardFail" e "AISLocally Inserted" nas rotas 3 a 6 que pode ser verificado usando o teste de tendência linear, como mostrado na figura 4 para as rotas 3 e 6 que têm maior número de ocorrências.

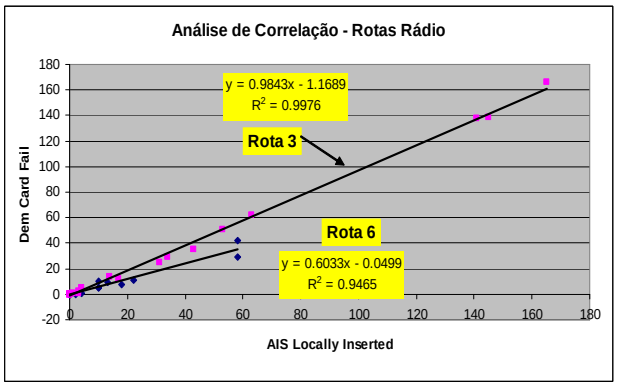

Figura 4. Análise de correlação, Rotas rádio 3 e 6

A análise anterior mostra a flexibilidade da ferramenta OLAP na ajuda à compreensão do problema.

B. Alarmes mais freqüentes: Anel Óptico

A figura 5 mostra os alarmes com maior número de ocorrências. A rede SDH é organizada em camadas de transmissão ("physical" e "section") e via ("path") e os alarmes indicam a camada em que foi gerada. A figura 5 não mostra essa divisão em camadas. Essa divisão é possível pois o atributo "object type" indica a camada. Não vamos entrar nesse detalhe pois nosso objetivo nesse artigo é mostrar o uso de OLAP para análise e criação de cenários.

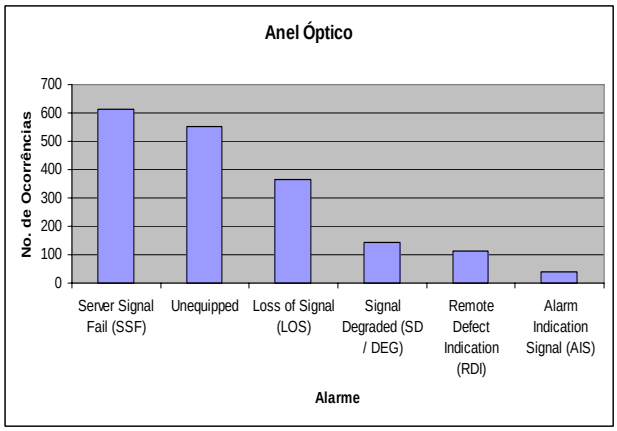

Figura 5. Anel Óptico, alarmes com maior freqüência

C. Avalanche de alarmes

O efeito avalanche, como explicado na introdução, é devido à interdependência entre os elementos de rede (NE), a ocorrência de alarmes primários gerando múltiplos alarmes secundários. A tela do centro de gerência é sobrecarregada de alarmes dificultando ao operador a análise da correlação entre os alarmes para chegar à causa-raiz. Essa análise é geralmente manual, sendo o objeto final desse trabalho a automatização dessa análise. A figura 6 mostra que no dia 8 de maio ocorreram cerca de 1400 alarmes. O que aconteceu nesse dia?

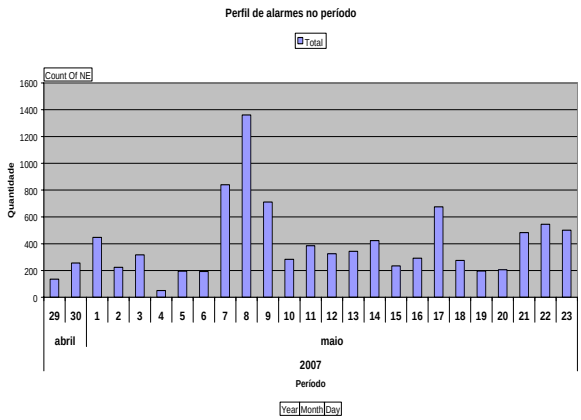

Figura 6. Perfil de alarmes no período estudado

Vamos usar OLAP para fazer uma pesquisa de profundidade. Filtramos então o dia 8 de maio e

detalhamos as ocorrências por hora, como mostrado na figura 7.

Vemos que a maioria das ocorrências ocorre às 13h. Continuando o aprofundamento, chegamos que 249 alarmes ocorreram às 13:47:59 e permaneceram cerca de 10min.

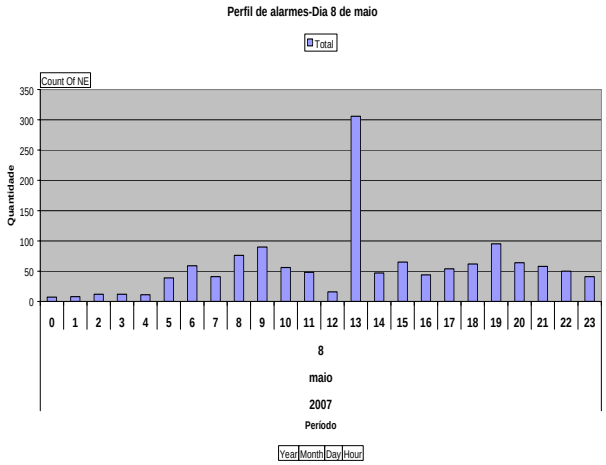

Figura 7. Detalhamento do dia 8 de maio

Novo aprofundamento nos leva às causas e local das ocorrências, que é mostrado na figura 8.

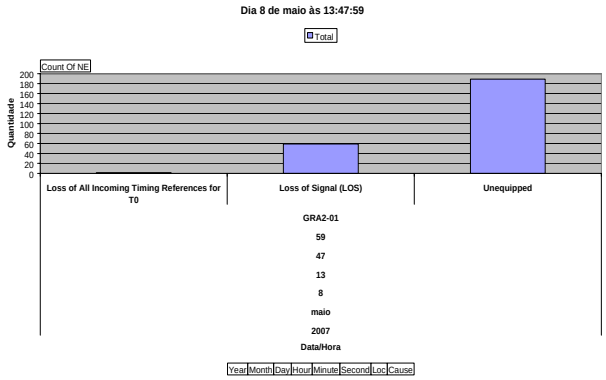

Figura 8. Localidade: Gravataí e Causas, dia 8/5 às 13:47:59

Nesse cenário consideramos:

Alarmes primários: "Loss of Signal (LOS)" e "Unequipped"

Alarme secundário: "Loss of All Incoming Timing References for T0",

O mesmo procedimento aplicado ao dia 7 de maio, figura 9, mostra o mesmo local porém com um alarme primário adicional: "Loss of Timing Source T3".

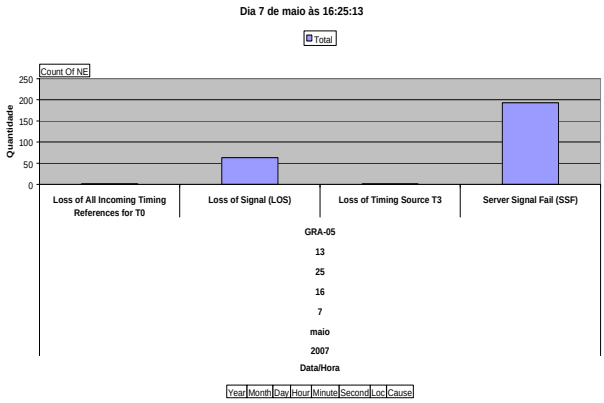

Figura 9. Localidade: Gravataí e Causas, dia 7/5 às 16:25:13

Esse tipo de procedimento está sendo aplicado em dados históricos coletados desde Dezembro de 2006. A seguir mostramos alguns dos cenários encontrados.

D.Exemplos de cenários de alarme

Cenário 1

Data: dia 18 de Dezembro de 2007 (11, 12, 14 e 16 horas)
Enlace: BLU-02 e ITA-02
NE2: SMA
Evento "Cause": LOS, SSF, RDI e Unequipped
Seqüência de eventos: Unequipped-> SSF->LOS->RDI; SSF->LOS; SSF->Unequipped->RDI
Duração: até 3 horas.
Causa provável: Causa: ativação de feixe. Nesse caso os alarmes "RDI Remote Defect Indication", "SSF Server Signal Fail" e "Unequipped" são alarmes temporários. O LOS fica presente até a conexão do equipamento no feixe.

Cenário 2

Data: 25 de Jan de 2007
Enlace: ITJ-01 e FLO-01 (link Óptico)
NE1: ITJ-01 e FLO-01
NE2: SLD
Evento "Cause": LOS (Óptico)
Seqüência de eventos: LOS as 11 e 20 Horas.
Duração: até 3 minutos
Causa provável: Falha do Link Óptico entre ITJ e FLO. É a causa do problema de SMA nos outros sites do link PAL,FLO,CBA,BLU. Na verdade o problema foi entre Sede e Florianópolis, e como conseqüência houve perda de relógio em ITJ (Loss of Time Source)

Cenário 3

Data: dia 2 de Jan 2007
Enlace: VNI-02_RVE-01
Evento "Cause": CARD FAIL
Seqüência de eventos: CARD Fail ODU, seguido de TX ODU Failure.
Duração: ~ 6000 s = 100 min
Causa provável: Instabilidade ou falha intermitente na TX ODU de VNI-02. O lado RVE-01 não detecta problema.

E. Mineração usando a duração do alarme

Essa mineração parte da observação que, devido à interdependência entre os elementos de rede, um alarme primário gera múltiplos alarmes secundários. Notamos que os alarmes secundários podem ser divididos em função da sua correlação com o primário em:

- alarmes com baixa correlação: são geralmente transitórios e sua duração (tempo entre a ativação e desativação) é menor que a duração do primário,
- alarmes com alta correlação: nesse caso a extinção do alarme secundário está correlacionada com a extinção do primário.

Aplicando essa mineração ao anel óptico no período estudado, detectamos um conjunto de alarmes com duração aproximada de 30 min no dia 23 de maio. A tabela I mostra o resultado da mineração.

TABELA I
Mineração usando a duração – Anel Óptico

Time (Local)	Cause	Loc
23/05/2007 10:16:56	Server Signal Fail (SSF)	MCL-02
23/05/2007 10:16:56	Alarm Indication Signal (AIS)	MCL-02
23/05/2007 10:16:56	Server Signal Fail (SSF)	GRA-03
23/05/2007 10:16:56	Server Signal Fail (SSF)	IVP-02
23/05/2007 10:16:57	Alarm Indication Signal (AIS)	SED-04
23/05/2007 10:16:57	Alarm Indication Signal (AIS)	JLB-02
23/05/2007 10:16:57	Server Signal Fail (SSF)	SED-02
23/05/2007 10:16:57	Server Signal Fail (SSF)	SED-02
23/05/2007 10:16:57	Server Signal Fail (SSF)	SED-02
23/05/2007 10:16:57	Server Signal Fail (SSF)	SED-02
23/05/2007 10:16:57	Server Signal Fail (SSF)	SOS-02
23/05/2007 10:16:58	Loss of Timing Source T1 #1	GPE-01
23/05/2007 10:16:58	Server Signal Fail (SSF)	GPE-01
23/05/2007 10:16:58	Loss of Signal (LOS)	GPE-01
23/05/2007 10:17:05	Threshold Crossed: UAT	GPE-01
23/05/2007 10:17:23	Server Signal Fail (SSF)	NSR-01

Identificamos uma falha de sincronismo “Loss of Timing Source T1 #1” na regeneradora de Guaporé, causada pela perda de sinal. Nesse minuto foram gerados 60 alarmes mas apenas os 16 da Tabela I têm duração próxima do alarme primário.

Outro exemplo analisando a Rota 6 de rádio permite minerar o cenário da Tabela II. que durou cerca de 40 min. O problema é na transmissão em CHA causando alarmes AIS em MPO.

TABELA II
Mineração usando a duração – Rota 6

Time (Local)	Cause	Loc
21/05/2007 10:29:22	Loss of Protection Signal	CHA-01
21/05/2007 10:29:23	AIS Prot Side	MPO-01
21/05/2007 10:29:23	AIS	MPO-01
21/05/2007 10:29:23	Tx On Line Channel Failure	CHA-01
21/05/2007 10:29:24	Loss of Signal	CHA-01

F. Mineração usando Gráfico de Controle

Para entender o conceito do Gráfico de Controle (“Control Chart”) de uma maneira bem simples, primeiro precisamos entender o conceito de variação.

Considere a ação de ir para o trabalho um processo, que leva em média 30 minutos. Um intervalo entre 25 a 35 minutos é um intervalo esperado para a execução desse processo. Medindo esse processo se o tempo esperado está dentro desse intervalo dizemos que está dentro da variação esperada. Caso o pneu fure, esse tempo médio será elevado para cerca de 50 minutos, será um caso excepcional, um ponto “fora da curva”. Se colocarmos essas informações num gráfico, esse gráfico será um “control chart”, onde o limite inferior (25 minutos) e superior (35 minutos) definem o intervalo normal e os pontos fora desses intervalos são os pontos “fora da curva” ou “fora de controle”. No caso o “pneu furado” é um caso de exceção que justificou o ponto fora de controle.

Num processo, os pontos “fora da curva”, ou tipos de variação, precisam ser investigados e diagnosticados como causas especiais ou causas comuns. Se existirem causas comuns, o processo precisa ser reavaliado para que seja reduzido o número de causas comuns. Isto é, ações podem ser tomadas tais como: redefinir os limites, justificar o acontecido ou identificar algo que pode ser melhorado no processo.

Portanto a variação pode ser entendida como algo que acontece dentro de um intervalo esperado. Utilizando esse conceito, o Gráfico de Controle poderá ser aplicado na mineração de alarmes com ocorrência “fora de controle” podendo identificar situações de anormalidade.

O método do gráfico de controle supõe que um processo (fabricação, medição, alarmes, etc.) exibe variações ao longo do tempo ou entre itens de um conjunto com as mesmas características. O gráfico de controle usa três parâmetros para análise da variabilidade:

- CL - “Central Line” - Linha Central de Controle
- UCL - “Upper Control Limit” - Limite Superior de Controle
- LCL - “Lower Control Limit” - Limite Inferior de Controle.

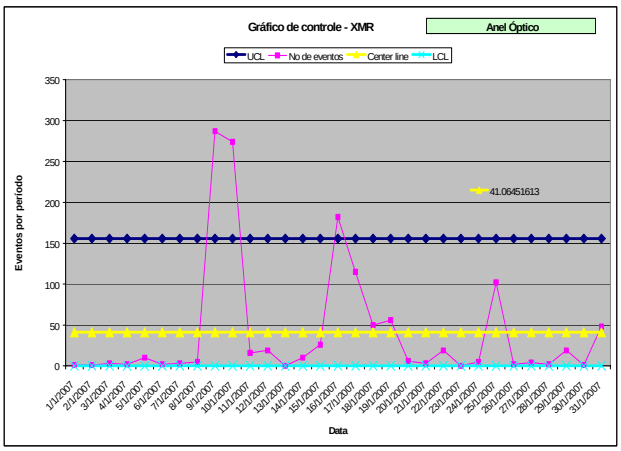

Figura 10. Exemplo de “Control Chart” - XmR

Nessa análise vamos usar um tipo de gráfico de controle que permitam uma análise diária como é o caso dos gráficos XmR e “Moving Range (mR)”. Para uma explicação detalhada desse tipo de gráfico consulte [1].

Os gráficos das figuras 10 e 11, referem-se a uma análise do mês de janeiro de 2007 nos alarmes do “Anel Óptico.

Em [1] é ressaltado que a variação móvel, mR, “não consegue indicar precisamente mudanças na variação do processo”. Ou seja, se o processo muda de patamar mas em cada patamar mantém uma variação controlada, a variação móvel apenas indicará um pico no momento da mudança. Isso pode ser útil para a indicação dos momentos das variações mas não no estudo das variações.

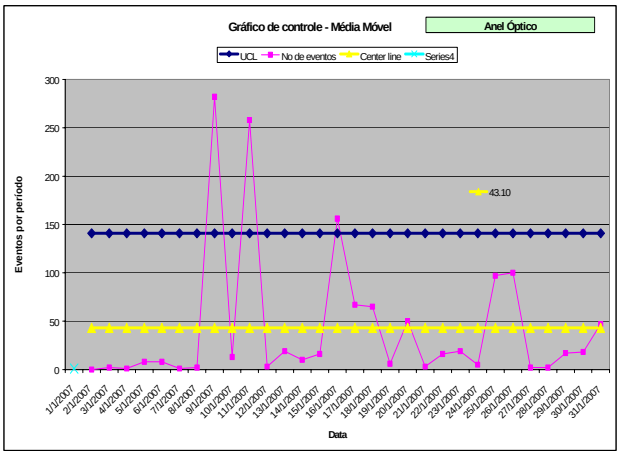

Figura 11. Exemplo de “Control Chart” - “Moving Range”

A figura 10 mostra no mês de janeiro, três pontos “fora de controle” nos dias 9,10 e 16 que causam a variação da figura 11.

O gráfico mR da figura 11 mostra três variações fora de controle que correspondem aos saltos para cima no dia 9, o salto para baixo no dia 11 e o salto para cima no dia 16.

Para prosseguir na análise, explicado os eventos do dia 9, 10 e 16 devemos retirar da análise os dias com eventos fora de controle mostrados na figura 10 que provocam os saltos na variação. Fazendo esse procedimento teríamos os gráficos da figura 12 e 13.

TABELA III
Mineração usando Gráfico de Controle

Data	Dia da semana	No eventos
09/01/07	3a	287
10/01/07	4a	274
16/01/07	3a	182
17/01/07	4a	115
18/01/07	5a	50
19/01/07	6a	56
25/01/07	5a	102
31/01/07	4a	48

Como existem variações fora de controle no gráfico mR, figura 13, o processo de análise e retirada dos pontos prosseguiria. Se continuássemos esse processo deveríamos analisar os dias mostrados na Tabela III.

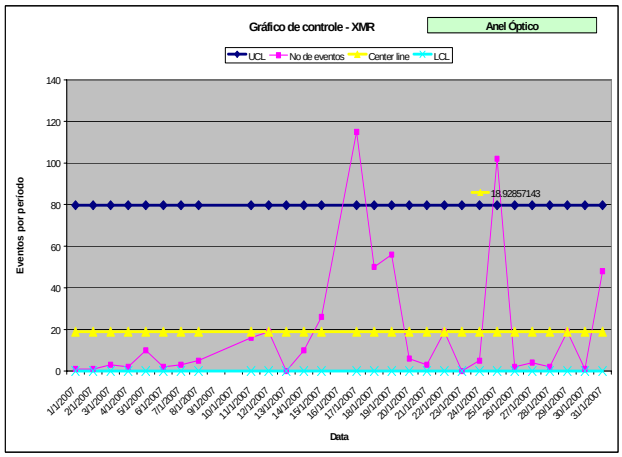

Figura 12. Gráfico XmR retirados os dias 9, 10 e 16

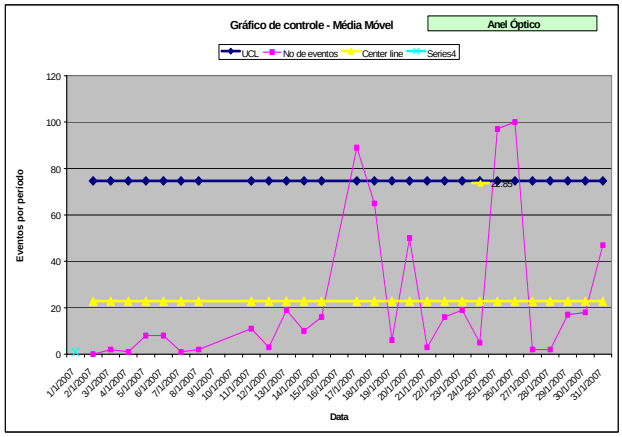

Figura 13 - Gráfico mR retirados os dias 9, 10 e 16

Note que à medida que prossegue o processo de retirada de dias fora de controle, a taxa média de eventos diminui (variou de 41 para 19 por dia). Outro critério de parada pode ser a taxa média abaixo de um determinado valor. O gráfico indica pontos a serem analisados.

O processo neste caso requer a análise ponto a ponto dos itens fora de controle, remoção dos pontos analisados e explicados tecnicamente, com novos ciclos de análise dos dados remanescentes até o fim do processo ou estabilização dos dados.

Essa mineração mostrou-se menos eficiente que as anteriores para construção de cenários. A sua maior contribuição, a nosso ver, é na análise preventiva visando a otimização de sistemas, ou seja, ajudar a evidenciar potenciais problemas ocorridos que podem ser controlados, facilmente reconhecidos ou evitados no futuro.

VI. CONCLUSÃO

O presente artigo mostra os métodos utilizados na etapa inicial do projeto de pesquisa entre a FITec e a

Eletrosul que visa a construção de uma ferramenta de Gerência de Alarmes visando a identificação da causa-raiz para a rede de telecomunicações da Eletrosul.

As etapas do trabalho estão representadas de forma resumida na figura 11.

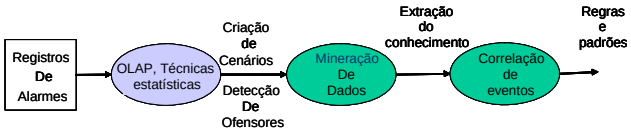

Figura 11. Etapas do projeto

A etapa inicial consiste na análise da base de dados de alarmes e seus atributos, da topologia da rede e sua organização em sub-redes, da identificação de cenários significativos de alarmes que representem o dia-a-dia do centro de operação.

Dado o alto volume de alarmes por mês, cerca de 10000/mês, precisávamos automatizar o processo de procura e análise. Nesse sentido escolhemos o uso de OLAP e técnicas estatísticas como Pareto e Gráfico de Controle.

No artigo mostramos, por meio de casos reais coletados da base histórica de alarmes, o uso das ferramentas citadas acima.

Principais conclusões:

- A ferramenta OLAP mostrou-se de fácil implantação, os cubos sendo gerados a partir dos alarmes coletados,
- A combinação Excel/OLAP permite a análise rápida de cenários. A geração do cubo pode ser no Excel ou no servidor da base de dados. No primeiro caso a pesquisa fica lenta para mais de 20000 registros. Para maiores volumes de dados aconselha-se que o cubo fique no servidor da Base de Dados,
- A geração de cenários com dados reais e recentes facilita a interação com a equipe de operação na identificação de cenários significativos, seus ofensores e eventos causadores.

VII. AGRADECIMENTOS

Os autores agradecem a colaboração de Arthur José Pierozzi e Alessandro Povero da Silva no entendimento e nas implantações da ferramenta OLAP, peça fundamental nessa etapa do projeto.

VIII. REFERÊNCIAS

[1] D. C. Montgomery, *Introduction to Statistical Quality Control*, J. Wiley, 1997.

[2] W. Florac, A Carleton, *Statistical Process Control for Software Process Improvement*, Addison-Wesley, 1999.

[3] D. Levy, R. Chillarege, "Early Warning of Failures through Alarm Analysis - A Case Study in Telecom Voice Mail System", *IEEE Int. Symp. On SW Reliability Engineering (ISSRE)* Nov 2003.

[4] N. Garofalakis, R. Rastogi, Data Mining Meets Network Management: The Nemesis Project, Bell Labs, 2002.

[5] A. Arnaud; R. Cunha; G. Vasconcelos; P. Adeodato; J. Genu e B. Regueira, "Abordagem Inteligente para Tratamento de Alarmes e Diagnóstico de Falhas em Sistemas Elétricos**",** *III Citenel*, 5/6 de dezembro 2005, Florianópolis.

[6] R. J. Roiger, M. W. Geatz, *Data Mining – A Tutorial Primer*, Addison Wesley, 2003.