Uma Abordagem de Monitoração de Tráfego de Rede utilizando Lógica Difusa

Enio Rovere Silveira[1], M.A.R. Dantas[2]

Universidade Federal de Santa Catarina (UFSC),
Centro Tecnológico (CTC), Departamento de Informática e Estatística (INE),
Campus Universitário, Trindade 88040-900 – Florianópolis – SC – Brasil
[1]enios@correios.com.br, [2]mario@inf.ufsc.br

Resumo: Neste artigo, apresentamos um trabalho de pesquisa de monitoração de tráfego de rede utilizando uma abordagem de lógica difusa. Nossa contribuição está focada na criação de um protótipo que pudesse ser avaliado em uma configuração real de produção de uma grande organização, utilizando-se de hardware e software existentes na empresa. Resultados experimentais obtidos através de estudos de casos reais no ambiente, indicam o sucesso da abordagem na melhoria da monitoração de tráfego da rede.
Palavras-chaves: Difuso, Tráfego de Rede.

Abstract: In this article we present a research work in traffic monitoring employing a fuzzy logic approach. The main contribution is based in building a prototype which could be verified in real production environment in a large organization, using available hardware and software. Experimental results, from several real case studies, indicate that the proposal enhanced successfully the monitoring of the network.
Keywords: Fuzzy, Network Traffic.

(Received March 22, 2005 / Accepted May 13, 2005)

1. Introdução

A complexidade das configurações físicas das arquiteturas de redes de computadores, em adição aos diversos componentes de hardware e software existentes nesses ambientes, sinaliza para a importância de uma monitoração eficaz. As organizações tem se tornado altamente dependentes das tecnologias de rede, sentindo, imediatamente, o impacto quando seus recursos não estão disponíveis. Torna-se evidente a necessidade de estabelecer monitoramento e controle sobre o comportamento do tráfego de rede, como forma de garantir a identificação de problemas. Um aspecto relevante na monitoração de rede diz respeito à maneira pela qual a configuração de rede será efetivamente monitorada. Em outras palavras, uma monitoração do tipo intrusiva pode levar a criação de um ambiente artificial que não espelha as condições originais do ambiente. Por outro lado, empregando-se um método passivo de monitoração, pode-se incorrer em perda de informações relevantes. Uma abordagem interessante, que pode representar uma melhoria no processo de monitoração de redes, caracteriza-se pela utilização de heurística de conhecimento empregada em conjunto com técnicas de amostragem.

A lógica difusa pode ser considerada uma resposta para o processamento de uma base de conhecimento relativa a uma configuração de rede, construída através de regras condicionais relativas às tarefas inexatas, que usualmente ocorrem nos ambientes de rede. Por outro lado, uma técnica de amostragem parece adequada para o estudo de comportamento do universo do tráfego de redes, utilizando-se, para tanto, um subconjunto representativo do conjunto global, no sentido de evitar o problema de monitoração intrusiva de tráfego.

Como [3] menciona, a pesquisa por medição nas infra-estruturas de rede é o atual objetivo de muitos grupos de pesquisa, e a proposta de trabalho de grupos de gerência de redes. Uma arquitetura para a medição de pacotes pode ser encontrada na RFC 2722, onde algumas métricas comuns são empregadas para medir o fluxo de tráfego.

Neste artigo, é apresentada uma pesquisa sobre como aplicar e avaliar a utilização da lógica difusa para a monitoração de tráfego de uma rede de computadores. Nosso objetivo não inclui medidas de desempenho, ou projeto de nova ferramenta de monitoração. Nossa meta de investigação é o uso de conceitos relativos à monitoração de tráfego de rede, efetuando-se o

mapeamento de alguns parâmetros característicos de uma configuração real e suas relações, através da lógica difusa.

O artigo está estruturado da seguinte forma. Apresenta-se na seção 2 alguns trabalhos de pesquisa correlatos. A seção 3 trata dos aspectos importantes relativos à lógica difusa, tráfego de rede e *clusterização*. Características do protótipo desenvolvido são apresentadas na seção 4. Na seção 5 é apresentado os resultados experimentais do ambiente real de teste. Finaliza-se o artigo com conclusões e propostas para trabalhos futuros.

2. Trabalhos Correlatos

Na literatura sobre redes existem pesquisas empregando a lógica difusa para solucionar problemas referentes à monitoração e segurança. Em nossa investigação por trabalhos correlatos foram encontradas algumas propostas antigas e outras mais recentes quanto ao aspecto de monitoração utilizando a abordagem difusa. Inúmeros ambientes (com características dispares) são propostos para atacar diferentes problemas relativos à monitoração de rede.

Um sistema baseado na técnica difusa, utilizando-se das experiências de administradores de rede e de relatórios de notificações de problemas, foi projetado por [6]. O ambiente proposto indicava os possíveis níveis de falha dos dispositivos de uma rede.

Em [15] existe uma proposta de construção de um sistema distribuído e integrado de agentes difusos, focado na monitoração de aspectos de rede, através da observação do comportamento de alguns parâmetros de componentes de rede.

Na área de congestionamento de tráfego de rede, o trabalho de [7] propõem atualizações no RED (*Random Early Discard*) por intermédio de um controlador difuso que é responsável por executar o descarte automático de pacotes, baseado no *status* da rede. Esse mecanismo é importante, uma vez que com a crescente tendência de diferentes tipos de pacotes nas redes, uma política de descarte mais eficiente pode representar um diferencial de desempenho.

Um estudo sobre o controle de fluxo de rede, empregando a técnica difusa, pode ser encontrado em [5]. Nesse trabalho, os autores propõem a definição do tamanho de janela do protocolo de transporte TCP, baseado na característica da rede, inferida atráves de um conjunto de variáveis difusas.

O crescimento da quantidade de ataques às redes, com diferentes formas, também é alvo de propostas difusas. Entendendo os padrões de alguns tipos de ataques, a ferramenta Fire (*Fuzzy Recognition Engine*) [8] pode auxiliar no envio de alertas para os administradores, quando uma intrusão estiver em curso. Uma outra contribuição relativa a IDS (*Intrusion Detection System*) é o ambiente *snort* [1]. Esse pacote de software foi projetado com um conjunto de regras, conhecidas como regras *snort*, por um processo de difusão automática, pode prever novos ataques.

O trabalho de pesquisa apresentado nesse artigo abrange alguns aspectos relativos aos trabalhos encontrados na literatura. Nossa contribuição está focada na coleta de informações de uma rede real, em uma grande organização brasileira, e na construção de regras de inferência utilizando-se de hardware e pacotes de software existentes na empresa. Em outras palavras, nossa investigação tem como objetivo a melhoria de desempenho na monitoração de tráfego de uma rede, empregando a abordagem difusa. Como resultado final indica-se, para o grupo responsável pela gerência de rede, alguns aspectos importantes da monitoração, através de uma linguagem mais amigável e natural. Desta forma, nossa solução é semelhante à encontrada em [6], no que diz respeito a uma maior preocupação no gerenciamento de falhas da rede. Por outro lado nossa abordagem permite, de maneira análoga a [15], a construção de um sistema distribuído de agentes. Finalmente, o nosso propósito se identifica com os trabalhos encontrados em [7] e [5], pois está relacionado ao uso de uma base de regras para o estabelecimento de um conhecimento especializado.

3. Lógica Difusa, Tráfego de Rede e Clusterização

Nesta seção aborda-se alguns conceitos importantes sobre lógica difusa, tráfego de rede e *clusterização*, posto que estes são os pilares de nossa contribuição.

3 .1. Lógica Difusa

A lógica difusa é caracterizada como um eficiente método para tratar informações inexatas ou incompletas, utilizando uma abordagem sistêmica e mais rigorosa. A descrição matemática relacionada à lógica difusa foi apresentada por Zadeh [21]. Como comentado em [20],

o principal aspecto da lógica difusa é a captura clara e concisa de vários conceitos utilizados por humanos em um raciocínio convencional. Linguagens naturais possuem um conjunto de expressões com significados imprecisos, ou seja, palavras idênticas podem representar idéias diferentes. Utilizando a técnica de lógica difusa é possível a manipulação simultânea de parâmetros numéricos e de informações de linguagem.

O conceito principal da teoria de conjuntos lógicos difusos é a função de pertinência. Esta função representa, em números, o nível de certeza de que um determinado elemento pertença a um específico conjunto. A função de pertinência mapeia cada elemento do universo de discurso em um valor entre 0 e 1, representando, assim, o grau de pertinência do elemento ao conjunto. Nos conjuntos ordinários, essa função assume o valor 1 para os elementos pertencentes ao conjunto, e 0 para os elementos não-pertencentes, possibilitando considerar esses conjuntos como casos particulares dos conjuntos difusos.

Se um elemento denominado genericamente por *x*, do universo de discurso *U*, pertence a um conjunto difuso *A*, então este conjunto difuso pode ser definido através da seguinte relação:

$A = \{(\mu_A(x), x) \mid \mu_A(x) \in [0,1]\}$, onde $\mu_A(x)$ é a função de pertinência.

Conforme demonstrado, os elementos pertencentes a um conjunto difuso são especificados através de um par, constituído do elemento propriamente dito e de seu grau de pertinência ao conjunto.

Considerando-se dois conjuntos difusos *A* e *B*, definidos em *X*, com o uso das funções de pertinência μ_A e μ_B, a função de pertinência da união ou disjunção de *A* com *B*, é definida ponto a ponto para todos os elementos $x \in X$, tal que [17]:

$\mu_{AuB}(x) = \max \{\mu_A(x), \mu_B(x)\}$

De modo análogo, a função de pertinência da intersecção ou conjunção de *A* com *B*, é definida ponto a ponto para todos os elementos $x \in X$, tal que [17]:

$\mu_{AuB}(x) = \min \{\mu_A(x), \mu_B(x)\}$

A função de pertinência do complemento de *A*, é definida ponto a ponto para todos os elementos $x \in X$, tal que [17]: $\mu(x) = 1 - \mu_A(x)$

Os resultados dessas operações entre os conjuntos difusos *A* e *B* são demonstrados graficamente através da figura 3.1 [17].

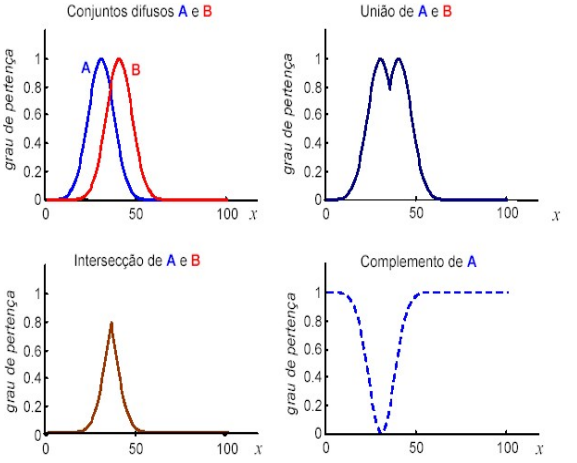

Figura 3.1: Operações difusas realizadas entre conjuntos *A* e *B* genéricos.

3.2. Tráfego de Rede

A medição de tráfego é uma prática usual na caracterização de uma rede. Para [2], atualmente se verifica um esforço no sentido de padronizar ou estabelecer uma linguagem comum para medição de tráfego de redes.

O uso do conceito de fluxos, como base para medições e para distribuição de tráfego em uma rede, é proposto em [12]. Neste trabalho o autor sugere a confecção dos fluxos a partir das informações sobre os endereços *IP*, número das portas fonte e destino, protocolos, tipos de serviço, marcação de tempo de início e fim do fluxo, contadores de pacotes e octetos.

A necessidade de uma garantia mínima para que as aplicações possam ser executadas com um desempenho satisfatório, em um ambiente de rede, está desencadeando pesquisas em várias áreas de conhecimento.

O *CAIDA (Cooperative Association for Internet Data Analysis)* [4] tem como proposta a unificação dos recursos de coleta, classificação e visualização dos dados, em um só pacote. Através dos vários resultados experimentais obtidos e de intensas pesquisas, o *CAIDA* alerta para o fato de que nenhuma análise sobre tráfego de rede pode ser realizada com o uso de apenas um parâmetro, pois, é importante uma avaliação do conjunto de variáveis e dos seus efeitos.

O grupo de trabalho *IPPM (IP Performance Metrics)* do *IETF* (*Internet Task Force*) tem como objetivo

desenvolver um conjunto padrão de métricas que possam ser empregadas para medir o desempenho e a qualidade dos serviços de dados.

Na *RFC 2722* está definida uma arquitetura para mensuração de pacotes. Na proposta procura-se métricas comuns para mensurar fluxos de tráfego de rede com objetivo de disponibilizar informações da rede sobre: comportamento dimensionamento, expansão, e desempenho.

Nas abordagens de medições passivas, considera-se o monitoramento por fluxo de tráfego. Essa técnica é proposta visando à redução do volume de dados coletados através da utilização do conceito de fluxos de tráfego, em substituição a utilização dos pacotes de rede. Essa proposta é passível de consideração pois quando um fluxo é criado, esse se torna apenas uma entidade caracterizada e individualizada através de parâmetros retirados, na maioria das vezes, dos cabeçalhos dos pacotes e, a partir daí, um contador tem seu valor incrementado no aparecimento de fluxos com as mesmas características.

3.3. Clusterização

Segundo [10], o objetivo dos métodos de classificação é dividir em subconjuntos (classes), os mais semelhantes possíveis, um conjunto de elementos (indicadores), a partir de distâncias dois a dois.

Para [18], os métodos de clusterização (*clustering*) podem ser caracterizados como qualquer procedimento estatístico que, utilizando um conjunto finito e multi-dimensional de informações, classifica seus elementos em grupos restritos e homogêneos internamente, permitindo gerar estruturas agregadas significativas.

Por outro lado, para [11] se o conjunto de informações, seja pelas peculiaridades do objeto a que representam, seja pela ambigüidade da própria estrutura de dados, possui uma fonte de imprecisão, que não há aleatoriedade derivada de processos estocásticos, e sim derivada da ausência de fronteiras abruptamente definidas entre as classes, deve-se voltar à atenção para a utilização da teoria dos conjuntos difusos.

Um dos algoritmos de *clusterização* difusa mais utilizados é o *Fuzzy C-Mean*s (FCM). O objetivo do FCM, para [18], é minimizar uma função do tipo abaixo, onde *m* é um fator que controla a difusividade dos *clusters*. *m>=1*.:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{ij}^{m} d^2(z_j, v_i)$$

Quanto maior seu valor, mais difusas ficam as regiões de transição entre os *clusters*. Um valor típico é *m=2*. $d^2(z_j,v_i)=(z_j-v_i)^T(z_j-v_i)$ é a norma euclidiana que representa a distância entre o ponto z_j e o centro v_i i-ésimo *cluster*. v_i é a variável de livre escolha no algoritmo.

4. Características do Protótipo

O protótipo proposto tem como meta aplicar e avaliar o uso das técnicas da lógica difusa na procura de relações, entre os dados presentes no tráfego de rede, que possibilitem apurar diferentes estados de comportamento, reportando esses eventos à administração da rede, no momento em que ocorram. Aliado a aplicação das técnicas da lógica difusa, procurou-se utilizar os conceitos de monitoração por fluxos de tráfego como forma de diminuir a massa de dados necessária para geração de contadores e somadores utilizados como variáveis de entrada do protótipo.

A idéia básica é conseguir um perfil de comportamento de determinado segmento de rede e gerar informações sobre possíveis desvios. A crescente complexidade das redes e sua heterogeneidade em termos de hardware e software, dificultam muitas vezes uma avaliação mais precisa dessas configurações. A impossibilidade de se estabelecer limites (ou escopos fixos) de avaliação de desempenho das configurações de rede pode ser entendida através da possibilidade de acréscimo a cada momento de um novo software ou hardware. Em outras palavras, a possibilidade de inclusão dinâmica de recursos leva, muitas vezes, a uma incerteza e uma imprecisão no gerenciamento do tráfego.

Por outro lado, quando se trabalha com mensuração de tráfego de rede, deve-se ter em mente que qualquer estratégia não poderá interferir (ou contribuir) para modificar o comportamento natural do tráfego. As medições passivas são caracterizadas por não interferir no tráfego da rede, utilizando-se de um dispositivo que auxilia na observação de todo o tráfego gerado na rede, e armazena informações em arquivos. Nas medições

passivas, o monitoramento por fluxo de tráfego propõe a redução do volume de dados coletados através da utilização do conceito de fluxos.

O tempo de monitoramento é, também, um fator importante quando se trabalha com medições passivas, pois existirá um custo de processamento e armazenamento, que exigirá uma capacidade computacional adequada para um pós-processamento e geração de informação em tempo hábil à tomada de decisão. O tempo de monitoramento, adequado para o protótipo, foi estimado através de diversos ensaios experimentais na rede real utilizada como laboratório de testes.

Para alcançar o objetivo de caracterizar o comportamento de um segmento de rede, empregou-se algumas técnicas apresentadas na tabela 4.1, bem como as suas respectivas motivações.

Técnica	Motivo
Lógica difusa	Possibilitar a modelagem de um raciocínio, por inferência lógica, nas situações de incertezas encontradas na subjetividade de se modelar o comportamento de uma rede.
Medições passivas de tráfego	Evitar ruídos no tráfego de rede durante o monitoramento
Cabeçalhos dos protocolos de rede e o conceito de fluxo de tráfego.	É a fonte de informação para geração dos contadores/somadores que compõem as variáveis de entrada do modelo difuso.
Consultoria de profissional de administração de redes	Necessidade de apoio na construção da base de regras difusa e na avaliação dos resultados.
Utilização do método *fuzzy c-means*	Possibilitar o ajuste de parâmetros das funções de pertinência.
Conceitos de administração e modelagem de dados	Possibilitar a modelagem de uma base de dados para armazenamento de informações de tráfego de rede, servidores e de serviços de rede.

Tabela 4.1: Técnicas utilizadas no modelo.

O funcionamento do modelo difuso proposto para monitorar o comportamento de um segmento de rede real, fez uso de um processo do tipo *sniffer*. Esse observa todo o tráfego de rede e armazena as informações dos cabeçalhos dos protocolos em *logs* durante um horário pré-estabelecido.

Em determinados intervalos de tempo é desencadeada uma tarefa que efetua a execução, em seqüência, de procedimentos, da seguinte maneira:

1. O primeiro procedimento da série executa a importação de todo tráfego de rede armazenado no *log* do *sniffer* e atualiza tabelas de uma base de dados;
2. De posse das informações do *log*, e em seqüência, o processo de geração de contadores e somadores entra em execução e gera, em forma tabular, as variáveis de entrada para o controlador difuso. Para geração dos contadores e somadores são selecionados os pacotes dos protocolos *TCP, UDP* e *ICMP*, identificados os servidores existentes, as portas disponíveis por servidor, máscara da rede, e números de IPs origem e destino;
3. O controlador difuso baseado no modelo [Mamdani, 1975], executa o processo de inferência que terá como resultado a avaliação do comportamento do segmento de rede monitorado. Uma notificação é gerada, para a administração de rede, nos casos de mudança desse comportamento;
4. A notificação é realizada através de um *e-mail* endereçado à administração da rede com uma planilha anexa contendo os valores das variáveis de entrada e o resultado do processo difuso.

Não é objetivo do presente trabalho de pesquisa estruturar o modelo difuso para que gere sugestões de ações corretivas. Nossa meta é avaliar a possibilidade de uso das técnicas difusas na busca de relações entre os dados presentes no tráfego de rede, que possam caracterizar o comportamento das redes. As ações corretivas surgirão em conseqüência da evolução do uso das técnicas difusas em ferramentas desenvolvidas com objetivo de monitoramento e gestão de redes.

A utilização da lógica difusa e dos conceitos de medição por fluxo de tráfego, nesse trabalho, tem como objetivo direcionar os procedimentos a serem adotados pela administração da rede, auxiliando, dessa forma, na

utilização de softwares específicos para atender situações específicas.

Os contadores e somadores, utilizados como variáveis de entrada do modelo difuso, são gerados a partir dos campos que compõem o cabeçalho dos protocolos. Os campos escolhidos foram: o IP origem, o IP destino, porta origem, porta destino e tamanho do pacote.

Os contadores e somador considerados no projeto foram estruturados com o objetivo de expressar o perfil da rede, de uma maneira mais inteligente, fornecendo meios para avaliar e informar mudanças de comportamento. É importante mencionar que os contadores e o somador foram especificados juntamente com a administração da rede da organização. O protótipo deverá auxiliar também na eventual atualização, modificação ou extinção de contadores e somador. O somador caracteriza-se pela função de totalizar os bytes do tráfego, por unidade de tempo, e a relação de contadores de tráfego do protótipo está caracterizada na tabela 4.2.

Os termos lingüísticos *lower* (pequeno), *normal* (normal), *acceptable* (aceitável) e *abnormal* (anormal), foram escolhidos para descrever o comportamento de cada variável.

De posse das considerações bibliográficas, optou-se pelas formas triangulares para referenciar os termos lingüísticos de normalidade (*normal*) e aceitabilidade (*acceptable)*, e funções de formato *Z* e *S* para descrever, respectivamente, os termos com valores baixos (*lower*) e anormais (*abnormal)*, representando, assim, o mapeamento dos números reais dos contadores e somadores em números difusos. Esses termos foram igualmente espaçados em cada universo de discurso, de cada variável de entrada, nos primeiros testes do modelo. Posteriormente, utilizou-se o algoritmo *Fuzzy C-Means* para o ajuste das funções de pertinência de cada termo lingüístico, e a ferramenta de *Clustering* (*clusterização*) para ilustrar, graficamente, os resultados obtidos. O algoritmo *Fuzzy C-Means* e a ferramenta de *Clustering* estão disponíveis no pacote de software Matlab [14]. Para execução dos ajustes, considerou-se a existência de um *log* de tráfego de rede, com dezenas de megabytes, gerado e armazenado durante três meses de coleta, bem como os quatro *cluster*s (*lower, normal, acceptable e abnormal)* que foram sugeridos pelos adminstradores da rede utilizada nos testes.

TCP por unidade de tempo
UDP por unidade de tempo
ICMP por unidade de tempo
Contador de tráfego por unidade de tempo
Com destinos externos a rede monitorada por unidade de tempo
Com origens externas a rede monitorada por unidade de tempo
Com mesmo endereço origem e destino
Porta destino de servidores
Destinado às estações de trabalho por unidade de tempo
Destinado a servidores de trabalho por unidade de tempo
Com mesmo endereço IP origem, porta origem, endereço IP destino e porta destino por unidade de tempo

Tabela 4.2: Contadores de tráfego empregados no protótipo.

Esses 4 *cluster*s foram ratificados, posteriormente, pela execução do algoritmo *Fuzzy C-Means*. O modelo apresentado em [13] foi escolhido para o desenvolvimento do protótipo, pois, segundo [16], matematicamente não há diferenças entre a abordagem de [13] e [19]. A diferença entre as duas abordagens está caracterizada nas relações difusas existentes nos conseqüentes das regras. A saída de cada regra é representada, no protótipo, por funções de pertinência triangulares com três termos lingüísticos, assim caracterizados: *normal*, *acceptable* e *abnormal*.

5. Resultados Experimentais

As ferramentas de monitoramento de tráfego de rede, proprietárias ou não, distribuídas comercialmente ou de acesso livre, apresentam as mais variadas carcterísticas. Apesar das facilidades de monitoramento de variáveis e visualizações gráficas de tráfego, o processo de diagnóstico ainda é um tarefa difícil. O protótipo implementado, sob paradigma da lógica difusa, objetiva identificar a **possibilidade** de diminuição no esforço de monitoramento do tráfego de rede, através do uso de um racicínio difuso. Cabe mencionar que todas as técnicas utilizadas nessa pesquisa são passíveis de serem implementadas em ferramentas de gerenciamento de tráfego, que terão como diferencial a utilização do racicínio difuso.

Para este trabalho experimental foram especificadas 12 variáveis, entre contadores e somador, que servem

como entrada de dados para o protótipo. Com objetivo de provar processo de ajuste dos parâmetros das funções de pertinência, escolheu-se uma das variáveis para demonstração. Utilizou-se o pacote de software Matlab para ilustrar o processo implementado no protótipo. Empregou-se o utilitário de lógica difusa *Fuzzy Logic Toolbox*, do Matlab, que demonstra a distribuição dos termos lingüísticos e funções de pertinência para a variável chamada de *somador de bytes por segundo.*

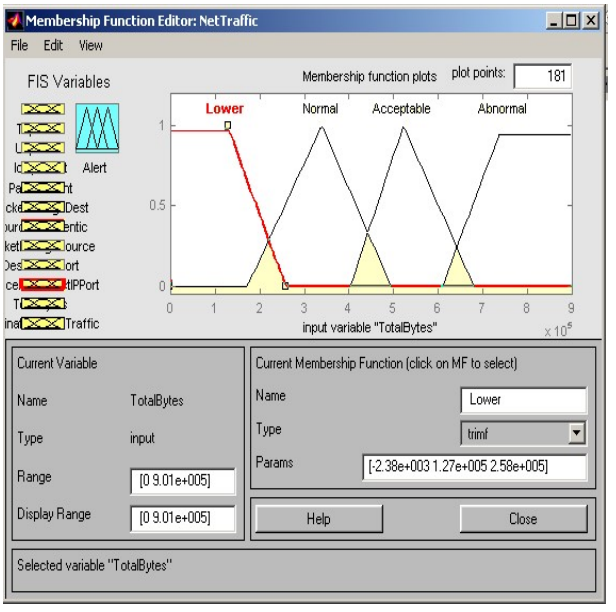

Figura 5.1: Somador de bytes por segundo

A figura 5.1 retrata a distribuição simétrica das funções de pertinência que compõem o universo de discurso da variável escolhida para ilustração.

Por desconhecimento sobre o perfil do tráfego de rede dos segmentos existentes, utilizou-se funções matemáticas para apuração de valores máximos e mínimos de cada um dos contadores e somador, com base em uma amostra de tráfego coletada durante uma semana. Esses valores foram utilizados como universo de discurso de cada variável e, sobre eles, distribuídas, simetricamente, as 4 funções de pertinência associadas aos quatro termos lingüísticos que objetivam representar o comportamento do segmento de rede monitorado.

Para o ajuste dos parâmetros das funções de pertinência que melhor retratasse o comportamento de cada termo lingüístico, recorreu-se ao algoritmo *Fuzzy C-Means* implementado no Matlab. Esse algoritmo implementa um método de agrupamento difuso, muito utilizado como técnica de reconhecimento de padrões, onde um dado pode ser classificado em várias categorias (*clusters*) com diferentes graus de associação.

Para o ajuste preciso dos parâmetros das funções de pertinência é imprescindível que a amostra utilizada no processo seja representativa, retratando com a maior fidelidade possível o comportamento do segmento em estudo. Com esse propósito, foram coletadas amostras de tráfego durante um período de três meses. Durante o processo de coleta foram constatados eventos de irregularidades que proporcionaram uma maior representatividade para a amostra.

Em [Bezdek e Pal, 1992] verifica-se que a qualidade dos resultados obtidos com o algoritmo *Fuzzy C-Means* é influenciada pela escolha dos parâmetros do algoritmo, que devem gerar, a princípio, conjuntos de *clusters* compactos e bem separados. Todavia, entende-se que uma avaliação criteriosa deva ser realizada posteriormente, através da realização de exaustivos testes com variações nos parâmentos do algoritmo.

De posse dos valores dos *clusters*, que estabelecem um agrupamento padrão de cada termo lingüístico, em cada variável de entrada, e da amostra utilizada para o ajuste dos parâmetros das funções de pertinência, empregou-se o utilitário de plotagem do software Matlab, utilizando o eixo *x* para retratar a massa de dados da amostra, e o eixo *y* para retratar o grau de pertinência de cada dado amostral a cada um dos quatro *clusters.* Através desse processo foi possível uma representação dos valores a serem ajustados para cada função de pertinência, conforme demonstra a figura 5.2.

Utilizou-se o mesmo procedimento para o ajuste de todas as funções de pertinência que compõem o protótipo. Uma vez determinado os ajustes, procedeu-se a atualização das funções implementadas no protótipo. Pode-se verificar o resultado prático da adoção do algoritmo *FCM*, fazendo-se uma comparação entre as figuras 5.1 e 5.3, representadas graficamente através do utilitário de lógica difusa *Fuzzy Logic Toolbox* .

A base de regras, as funções de pertinência e os termos lingüísticos, no presente trabalho, são estruturados com objetivo de encontrar relações que possam evidenciar e caracterizar o comportamento de um segmento, através dos dados presentes no tráfego da rede. O sistema de regras estará completo quando puder responder satisfatoriamente a todas as ocorrências que caracterizem as mudanças de comportamento.

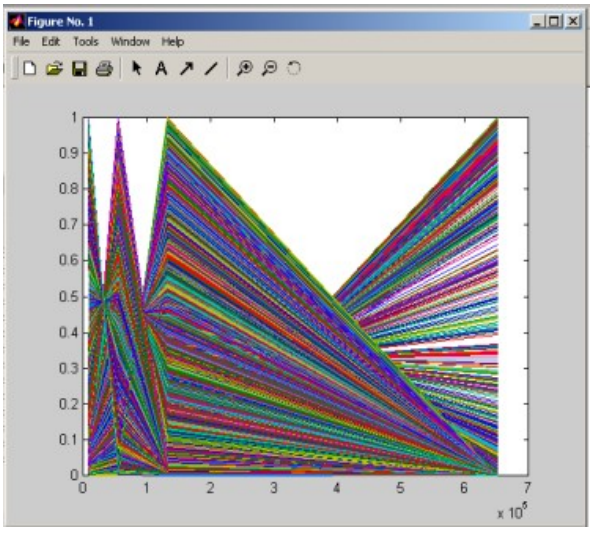

Figura 5.2: Distribuição amostral do somador de bytes por segundo

O modelo difuso está definido da seguinte forma:

- Para o conectivo *OR* utilizou-se a *normas-t min.*
- Para o conectivo *AND* utilizou-se a *co-normas max.*
- No processo de implicação usou-se a *normas-t min.*
- No processo de agregação usou-se *co-normas max*.
- No processo de defuzzificação utilizou-se o método do *centróide*.

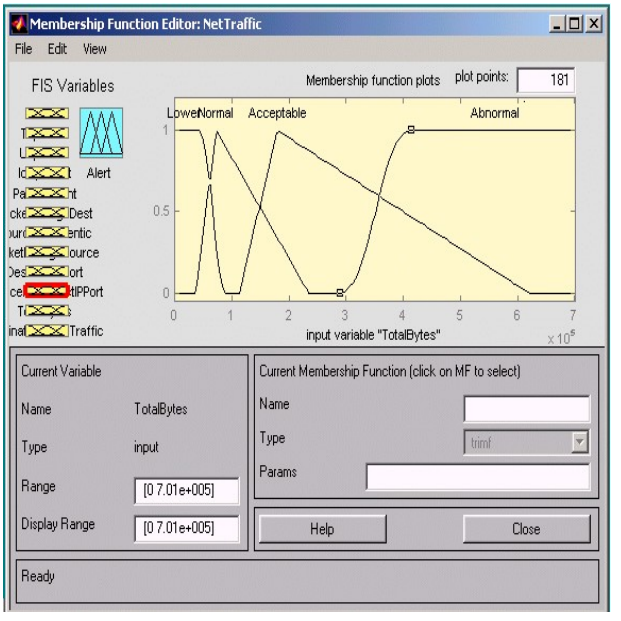

Figura 5.3: Ajuste das funções de pertinência do somador de bytes por segundo.

A figura 5.4 descreve regras usadas no modelo.

```
If TcpCount is Normal
and UdpCount is Normal
and IcmpCount is Normal
and PacketCount is Normal
and PacketForeignDest is Normal
and SourceDestIdentic is Normal
and PacketForeignSource is Normal
and DestinationPort is Normal
and SourceIPPortDestIPPort is Normal
and TotalBytes is Normal
and DestinationClientTraffic is Normal
and DestinationHostTraffic is Normal
Then  Alert is Normal

If  TcpCount  is Lower
and UdpCount is Lower
and  IcmpCount is Lower
and PacketCount is Lower
and PacketForeignDest is Lower
and SourceDestIdentic is Lower
and PacketForeignSource is Lower
and DestinationPort is Lower
and  SourceIPPortDestIPPort is Lower
and TotalBytes is Lower
and DestinationClientTraffic is Lower
and DestinationHostTraffic is Lower
Then Alert is Abnormal

If TcpCount  is Abnormal
or UdpCount is Abnormal
or  IcmpCount is Abnormal
or PacketCount is Abnormal
or PacketForeignDest is Abnormal
or SourceDestIdentic is Abnormal
or PacketForeignSource is Abnormal
or DestinationPort is Abnormal
or SourceIPPortDestIPPort is Abnormal
or TotalBytes is Abnormal
or DestinationClientTraffic is Abnormal
or DestinationHostTraffic is Abnormal
Then  Alert is Abnormal

If DestinationClientTraffic is Acceptable
and TotalBytes is Acceptable
Then  Alert is Acceptable

If DestinationClientTraffic is Acceptable
and DestinationHostTraffic is Lower
Then Alert is Abnormal

If DestinationClientTraffic is Lower
and DestinationHostTraffic is Acceptable
Then Alert is Acceptable

If PacketForeignSource is Acceptable
and DestinationHostTraffic is Acceptable
and SourceIPPortDestIPPort is Acceptable
Then Alert is Abnormal
```

Figura 5.4: Trecho da base de regras do protótipo.

5.1. Estudo de Caso A

Um dos eventos rastreados pelo protótipo notifica uma situação anormal resultante da execução de um backup de banco de dados que utilizou o segmento de rede para gravação do arquivo de backup. Este tipo procedimento não é indicado por motivos de segurança e pela conseqüente degradação do tempo de resposta da rede. Este procedimento foi executado, conscientemente, pela equipe de administração de dados, que trabalhava em uma migração de versão de software. O protótipo gerou a notificação sobre uma mudança de comportamento do segmento, informando altos valores escalares para descrever a criticidade do evento.

5.2. Estudo de Caso B

Outro evento interessante, capturado pelo protótipo, ocorreu por ocasião de um problema técnico em um dos *switches* do nível 3 que atendem os diversos andares do prédio que abriga a rede local escolhida para o monitoramento. O referido *switch* desencadeou um processo de embaralhamento de tráfego que provocou a geração de um grande tráfego *TCP*, *UDP*, e um aumento significativo no tráfego *ICMP*. Através da notificação do protótipo foi possível constatar, rapidamente, a origem do problema de degradação de rede.

5.3. Estudo de Caso C

Com a utilização do protótipo constataram-se situações que já eram de conhecimento da equipe técnica, como por exemplo, a degradação do tempo de resposta do segmento de rede nos intervalos de horários entre 12:00 horas e 14:00 horas, e horários próximos às 18:00 horas. Nessas ocasiões nota-se um aumento significativo, apesar de aceitável, na quantidade de bytes transmitidos e na quantidade de tráfego com destino externo ao segmento.

5.4. Estudo de Caso D

O comportamento do protótipo é alterado, exigindo um ajuste nos parâmetros das funções de pertinência, a cada inclusão de um novo segmento de rede. O mesmo não se verifica com a inclusão de novos dispositivos no segmento já monitorado, caracterizando, assim, um diferencial em relação a outros modelos, suscetíveis a inclusão de dispositivos. Este último fato se deve a representatividade da amostra utilizada no ajuste inicial dos parâmetros do protótipo, que já retrata a movimentação de dispositivos.

Deve-se considerar que o segmento de rede monitorado possui um inventário significativo de dispositivos. Quando nos referimos a dispositivos, deve-se desconsiderar a inclusão de servidores, pois para os mesmos, obrigatoriamente, deve-se proceder as atualizações nos parâmetros do protótipo, que possui variáveis dependentes da informação de números de *IPs*, portas e protocolos.

Durante os primeiros ensaios experimentais, devido aos ajustes iniciais das funções de pertinência e a construção gradativa da base de regras, constatou-se a geração de uma grande quantidade de falsos alarmes. Uma vez concluída a etapa de ajuste de parâmetros, observou-se a inexistência de notificações por parte do protótipo, caracterizando, assim, o padrão de utilização do segmento e o sucesso na caracterização do comportamento desse segmento.

É importante mencionar que apesar da utilização do algoritmo *FCM*, para o ajuste dos parâmetros das funções de pertinência, foram necessários mínimos ajustes em algumas variáveis, cujo particionamento dos termos lingüísticos geraram sobreposição de funções.

6. Conclusões e Trabalhos Futuros

Este trabalho de pesquisa teve como objetivo principal avaliar o uso das técnicas difusas na apuração de diferentes estados de comportamento de um segmento de rede, utilizando-se das possíveis relações existentes entre os dados que compõem o tráfego de rede. Com base na implementação de um protótipo e sua efetiva utilização em um ambiente de produção, pode-se constatar a eficácia do uso da técnica.

Esse trabalho proporcionou a utilização de abordagens interessantes, tais como, técnicas para o tratamento de um volume gigantesco de informações coletadas no tráfego de rede, e mesmo, a utilização de técnicas de reconhecimento de padrões que possibilitaram ajustar os parâmetros do modelo difuso proposto. Entre os resultados obtidos pode-se citar a possibilidade de: minimizar as atividades de monitoramento das equipes de administração de rede, que passam a atuar no momento em que o comportamento usual da rede tenha sido afetado; de conhecer, melhor, o perfil dos usuários através do tráfego gerado; de descrever o conhecimento

especializado através de uma base de regras; de conhecer o tráfego restrito a segmentos de redes específicos. No entanto, se percebe dificuldades de assimilação no uso prático da técnica por parte dos usuários. Esse fato é mencionado como uma das limitações dos sistemas inteligentes em [9].

A adoção de diferentes tipos e formas de funções de pertinência, ou ainda, uma análise na base de regras do protótipo, poderia ser objeto de estudo para investigações futuras. A utilização de simuladores para geração de vários tipos de tráfego também poderia ser objeto de interesse, onde o objetivo seria avaliar o comportamento do protótipo sob diferentes cenários.

7. Referências

[1] AICKELIN, U, HESKETH, T.**,** *Fuzzy Rule Learning in Intrusion Detection Systems***,** Submitted & Under Review Paper, Computer Science - ASAP group, 2003. www.cs.matt.ac.uk.

[2] ANGELIS, A., *Um Modelo de Tráfego de Rede para Aplicação de Técnicas de Controle Estatístico de Processos*, Tese de Doutorado. São Paulo: Instituto de Física de São Carlos - USP, 2003.

[3] BROWNLEE, N., MILLS, C., RUTH, G., *Traffic Flow Measurement: Architecture*, RFC 2722, IETF, October, 1999.

[4] CAIDA, *Cooperative Association for Internet Data Analysis*, http://www.caida.org, 2003.

[5] CARBONELL, P; JIANG, Z. P.; PANWAR, S. S., *Fuzzy TCP: A Preliminary Study*, Proceedings Of the 15th IFAC World Congress (IFAC 2002), Barcelona, Spain, July, pp. 21-26, 2002.

[6] CHEN, J-L. HUANG, P-H., *A fuzzy expert system for network fault management*, IEEE International Conference on Systems, Man and Cybernetics, Information Intelligence and Systems, Vol. 1, pp. 328-331, 1996.

[7] CHRYSOSTOMOU, C; PITSILLIDES, A; ROSSIDES, L. *Fuzzy Logic Controlled RED: Congestion Control in TCP/IP Differentiated Services Networks*, Special Issue on The Management of Uncertainty in Computing Applications in Soft Computing Journal - A Fusion of Foundations, Methodologies and Applications, Vol 8, Number 2, pp. 79 - 92, December 2003.

[8] DICKERSON, J. E. ; DICKERSON, J. A., *Fuzzy Intrusion Detection*, IFSA World Congress and 20th North American Fuzzy Information Processing Society (NAFIPS) International Conference, Vancouver, British Columbia, Volume 3, pp. 1506-1510, July, 2001.

[9] GÜRER, D., KHAN, I., OGIER, R**.,** *An Artificial Intelligence Approach to Network Fault Management*, SRI International, Menlo Park, California, USA, 1999.

[10] KAGEYAMA, A., LEONE, E. T., *Uma Tipologia de Municípios Paulistas com Base em Indicadores*, Notas de Aula. Campinas: Unicamp, 1999.

[11] KAUFMAN, L.; ROUSSEEUW, P.J., *Finding Groups in Data: An Introduction to Clusters Analysis*, New York: Wiley, 1990.

[12] LAI, W. S., *A Framework for Internet Traffic Engineering Measurement*, IETF, Internet Draft. Informational, Work in Progress, November, 2001.

[13] MAMDANI, E. *Application of Fuzzy Algorithm for Control of Simple Dynamic Plant.* Proceedng of IEE Control and Science 121(12), pp. 1585-1588. 1975.

[14] Matlab, *Fuzzy Logic Toolbox*, Mathworks, 1998.

[15] NDOUSSE, D. T. *Distributed Fuzzy Agents: A Framework for Intelligent Network Monitoring*, IEEE International Conference on Communications, ICC '97, Towards the Knowledge Millennium, Montreal, Québec, Canada, Conference Record IEEE, 867-871. 8-12 June 1997.

[16] SHAW, I., SIMÕES, M. G. *Controle e Modelagem Fuzzy*. São Paulo: Edgard Blücher Ltd, 1a ed. 1999.

[17] SILVA, G., *Controle Não Linear*, Escola Superior de Tecnologia Setúbal. Artigo. Portugal. 2001.

[18] Simões, R.F, *Uma Análise de Fuzzy Cluster*, Notas de Discussão No. 26. Belo Horizonte: UFMG, 2003.

[19] TAKAGI, T., SUGENO, M., *Fuzzy identification of systems and its applications to modeling and control*, IEEE Trans. on Systems, Man and Cybernetic SMC-15, pp. 116-132, 1985.

[20] WEBER, L., KLEIN T. A. P., *Aplicações da Lógica Fuzzy em Software e Hardware***,** Canoas (RS): Ed. Ulbra, 2003.

[21] ZADEH, C. J., *Fuzzy sets*, Information and Control, pp 338-353, 1965.

[22] ZADEH, C. J. , PAL, S.K, *Fuzzy Models for Pattern Recognition*. ISBN 0780304225. IEEE Press, New York, 1992).