Universidade Federal de Pernambuco
Departamento de Sistemas de Computação
Graduação em Ciência da Computação

ADS-Fusion: Fusão de dados para detecção de anomalias baseada na teoria de evidência de Dempster-Shafer

Proposta de Trabalho de Graduação

Aluno: Bruno Filipe de Oliveira Lins
Orientador: Prof. Dr. Djamel Sadok
Co-orientador: Prof. MSc. Eduardo Feitosa

Recife, 19 de agosto de 2008

ADS-Fusion

Fusão de dados para detecção de anomalias baseada na teoria da evidência de Dempster-Shafer

Contexto

Detecção de Anomalia

Uma anomalia é um desvio do funcionamento normal ou corriqueiro. No tráfego de redes essas anomalias podem ser causadas por uma série de fatores como: vírus, *worms*, ataques a protocolos e serviços ou falhas de roteamento. Essas falhas podem causar desperdício de recursos e falta de confidencialidade e autenticidade de dados, muitas vezes importantes. A detecção de anomalia tem como princípio que qualquer comportamento que fuja do padrão é classificado como uma anomalia, logo um ataque[1]. Baseado em [2][3][4] podemos classificar as principais anomalias como:

a) **Comportamento do usuário**: usuários não autorizados acessando serviços restritos ou usuários inexperientes responsáveis por serviços importantes;

b) **Exploração de falhas em sistemas e softwares**: elementos de rede, computadores de usuários finais, protocolos ou serviços sendo atacados;

c) **Resposta a anomalias**: alvos de um ataque bem sucedido podem ter a função de gerar tráfego para a rede ou enviar dados ao atacante;

d) **Evasão**: atacantes deliberadamente manipulam os protocolos de rede para ocultar seus ataques dos IDSs.

Diante das variedades de anomalias no trafego de rede e dos prejuízos causados por elas percebemos a importância de um sistema ou arquitetura eficiente na detecção de anomalias de tráfego.

Fusão de dados

Fusão de dados é comumente definida como o uso de técnicas que trabalham sobre os dados de várias fontes e agrupam essas informações a fim de conseguir inferências, que serão, muitas vezes, mais eficientes e mais precisas do que se fossem geradas por meio de uma única fonte.

Muitas vezes métodos meramente estatísticos são usados para fusão de dados, no entanto muitos desses não são capazes de modelar conhecimentos incompletos, imprecisos ou ate mesmo incertos, daí a necessidade de se trabalhos com modelos mais adequados a este ambiente.

Teoria da evidência de Dempster-Shafer

Todo sistema que trabalha com conhecimento tem de ter a capacidade e manipular, transportar e modelar dados para sua representação. Um dos grandes problemas que esses sistemas baseados em conhecimento têm que

tratar é a capacidade de lidar com dados imprecisos, incompletos ou inconsistentes. Encontramos na literatura vários modelos tradicionais para a modelagem da incerteza [5][6], entre eles esta a teoria da evidência de Dempster-Shafer.

A teoria de Dempster-Shafer, originada dos trabalhos de Dempster e estendida pelas atividades de Shafer, teve como finalidade inicial, tratar falhas na modelagem da incerteza quando se trabalhava com métodos meramente probabilísticos. Tem por objetivo estabelecer as crenças de cada um dos elementos estudados para que assim tenha-se plena capacidade de demonstrar mais claramente o acúmulo de evidências e gerar assim funções de crenças para definir uma inferência mais clara.

ADS-Fusion

O ADS-Fusion representa um módulo capaz de baseado nas saídas geradas pelas ferramentas de detecção, fundir os dados e produzir uma inferência com um grau de certeza maior que as certezas geradas pelos módulos de detecção individualmente.

A idéia central deste trabalho é implementar um modelo estatístico que seja capaz de lidar com as incertezas das detecções de anomalia minimizando desta forma o número de falsos positivos, aumentando a eficiência da detecção.

Logo o desenvolvimento deste modulo tem como objetivo principal estudar o método de fusão de dados baseado na teoria de Dempster-Shafer e encontrar propostas de variações deste modelo, para que possamos sugerir uma inovadora e eficiente forma de utilização da técnica de fusão de dados visando segurança da informação.

ADS-Fusion

Fusão de dados para detecção de anomalias baseada na teoria da evidência de Dempster-Shafer

Objetivos

Objetivos Gerais

O objetivo desse trabalho de graduação é estudar técnicas de detecção de anomalia e desenvolver um sistema capaz de aumentar a eficiência da detecção através da fusão de dados. Para isso iremos adequar os sensores de tráfego (anomalias) para que possam gerar entradas ao módulo de fusão em um padrão adequado.

Depois de modificar os sensores de detecção precisaremos aplicar os conhecimentos adquiridos sobre a teoria da evidência de Dempster-Shafer em um módulo de fusão de dados que seja capaz de minimizar a incerteza ou incompletude das informações geradas pela detecção de anomalias aumentando assim a segurança da rede.

Objetivos Específicos

Espera-se no fim deste trabalho de graduação termos desenvolvido as seguintes atividades:

1. Modificação dos sensores de detecção de anomalia para que possam gerar as entradas do modulo de fusão de dados;

2. Implementação, em software, de um módulo capaz de fundir os dados tomados como entrada e gerar uma inferência com um grau de certeza maior que os da entrada;

3. Estudar as saídas geradas pelo modelo de fusão para tomar a ação adequada quando detectada uma anomalia.

ADS-Fusion

Fusão de dados para detecção de anomalias baseada na teoria da evidência de Dempster-Shafer

Cronograma

Tabela

Lista de Atividades	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Pesquisa bibliográfica						
Definição de escopo e cronograma						
Confecção da proposta do TG						
Elaboração e comparação de técnicas de fusão de dados						
Implementação e teste do protótipo						
Escrita do documento						
Revisão do documento						
Confecção da Apresentação						

Tabela 1. Cronograma do Trabalho de Graduação

Descrição das Atividades

Segue abaixo uma breve descrição das atividades listadas na tabela acima:

Pesquisa Bibliográfica: Este trabalho consiste na busca por trabalhos na área de detecção de trafego anômalo e fusão de dados, assim como novas idéias que possam contribuir com o desenvolvimento do ADS-Fusion.

Definição de escopo e cronograma: Essa atividade consiste da definição do escopo do nosso trabalho de graduação baseado no tempo que dispomos e nos objetivos que almejamos obter no fim do projeto.

Confecção da proposta do TG: Essa atividade compreende a elaboração do documento que contém uma breve descrição do trabalho que desejamos desenvolver, os objetivos esperado e o cronograma que iremos seguir durante o período de desenvolvimento do trabalho.

Elaboração e comparação de técnicas de fusão de dados: Nesta etapa iremos desenvolver e comparar variações de técnicas de fusão de dados para que possamos utilizar a que mais se adéqua a nossos objetivos.

Implementação e teste do protótipo: Essa etapa consiste na implementação do modulo de fusão de dados e teste do módulo visando quantificar a eficiência da detecção de trafego.

Escrita do Documento: Consiste na elaboração do relatório do trabalho de graduação. Neste documento estarão presentes todos os passos e dificuldades que surgiram durante a elaboração do trabalho de graduação, bem como os materiais de pesquisa utilizados nesse período.

Revisão do Documento: Etapa final da confecção do relatório do trabalho de graduação onde realizaremos as correções necessárias e alteração sugeridas pelos orientadores.

Confecção da Apresentação: Essa atividade consiste na elaboração da apresentação final, baseada no relatório de trabalho de graduação, responsável por descrever o projeto e os resultados alcançados.

Referências

[1] Sundaram, A. “An Introduction to Intrusion Detection”. ACM Crossroads 2.4, 1996.

[2] Mahoney, M., P. K. Chan. “PHAD: Packet Header Anomaly Detection for identifying Hostile Network traffic”. Florida Tech. Technical report 2001-04. http://cs.fit.edu/~tr/.

[3] Mahoney, M., P. K. Chan. “Learning Models of Network traffic for Detecting Novel Attacks”. Florida Tech. Technical report 2002-08. http://cs.fit.edu/~tr/.

[4] Mahoney, M., P. K. Chan. “Learning Nonstationary Models of Normal Network traffic for Detecting Novel Attacks”. Edmonton, Alberta. In Proceedings of the SIGKDD, 2002

[5] Pearl, J. Reverend Bayes on inference engines a distributed hierarchical approach. In Proceedings of the Second National Conference on Artificial Intelligence, Pittsburgh, PA, 1982, p. 133-136.

[6] Zadeh, L. A. Fuzzy Sets as a Basis for a Theory of Possibility. In: Fuzzy Sets and Systems 1, 1978, pp 3-28.

[7] G. Shafer. “A Mathematical Theory of Evidence”, Princeton University Press, 1976.

ADS-Fusion

Fusão de dados para detecção de anomalias baseada na teoria da evidência de Dempster-Shafer

Assinaturas

Recife, 19 de Agosto de 2008

Orientador

Prof. Dr. Djamel Sadok

Co-orientador

Prof. MSc. Eduardo Feitosa

Aluno

Bruno Filipe de Oliveira Lins