International Commerce and Information Review
Volume 9, Number 3, Sept, 2007: pp. 305~320
통상정보연구 제9권 3호 (2007년 9월 27일)

논문투고일자 : 2007. 07. 28305
게재확정일자 : 2007.

전자무역의 베이지안 네트워크 개선방안에 관한 연구*

A Study on the Improvement of Bayesian networks in e-Trade

정분도(Boon-Do Jeong)
조선대학교 경상대학 무역학과 교수

목 차

Abstract

With expanded use of B2B(between enterprises), B2G(between enterprises and government) and EDI(Electronic Data Interchange), and increased amount of available network information and information protection threat, as it was judged that security can not be perfectly assured only with security technology such as electronic signature/ authorization and access control, Bayesian networks have been developed for protection of information. Therefore, this study speculates Bayesian networks system, centering on ERP(Enterprise Resource Planning).

The Bayesian networks system is one of the methods to resolve uncertainty in electronic data interchange and is applied to overcome uncertainty of abnormal invasion detection in ERP. Bayesian networks are applied to construct profiling for system call and network data, and simulate against abnormal invasion detection. The host-based abnormal invasion detection system in electronic trade analyses system call, applies Bayesian probability values, and constructs normal behavior profile to detect abnormal behaviors. This study assumes before and after of delivery behavior of the electronic document through Bayesian probability value and expresses before and after of the delivery behavior or events based on Bayesian networks. Therefore, profiling process using Bayesian networks can be applied for abnormal invasion detection based on host and network. In respect to transmission and reception of electronic documents, we need further studies on standards that classify abnormal invasion of various patterns in ERP and evaluate them by Bayesian probability values, and on classification of B2B invasion pattern genealogy to effectively detect deformed abnormal invasion patterns.

Key Words : e-Trade, Bayesian networks, ERP, FTP

* 이 논문은 2007년도 조선대학교 학술연구비의 지원을 받아 연구되었음.

Ⅰ. 서 론

정보기술의 급속한 발전은 사회 전반에 걸쳐 막대한 영향을 미치고 있고 B2B, B2G 등의 전자무역 거래에서도 활용 영역이 빠르게 확장되고 있다. 글로벌체재하의 통상환경을 마비시키기 위한 악의적인 바이러스의 감염과 해킹이 갈수록 증가하고 있는 실태이며, 기업간 네트워킹에 대한 공격 방법이 교묘해지면서 새로운 형태의 공격 기법들이 발견되고 있다. 이러한 추세에 반하여 비검증된 사용자로부터 정보의 조작과 접근을 방지하기 위한 전자서명, 전자인증, 접근제어 등의 보안 기술이 개발되고 있다.

그러나 인터넷의 급격한 사용 확산에 따라 인터넷에 연결된 기관 및 기업의 인트라넷에 손쉽게 접근하여 필요한 정보를 획득 및 이용하는 것이 가능해졌지만 역기능으로 네트워킹에 대한 해커들의 불법적인 침입도 날로 증가되고 있다. 특히 윈도우XP의 주요 구성요소인 유닉스 운영체제와 TCP/IP/가 정보보호측면에서 많은 취약점을 가지고 있고 인트라넷 및 익스트라넷 등에 연결된 모든 인터페이스가 비검증된 공격으로부터 위험에 노출된 실정이다.

현재의 글로벌 기업 환경하에서는 전자서명·인증 확인 및 접근제어 통제방식만으로는 보안 문제를 해결하기에 충분하지 못하고 정보 보호를 위한 2차 방어선으로 침입탐지시스템(IDS)등이 계속 개발되고 있다.[1] 침입탐지모델은 오용침입탐지(Misuse Intrusion Detection)와 이상침입탐지(Anomaly Intrusion Detection)로 분류가 된다. 오용탐지는 알려진 침입방법들을 수집하여 지식 베이스에 저장하고, 동일한 침입 유형을 지식 베이스 검색을 통한 비교에 의해 침입을 탐지하는 방법이다. 이상탐지는 정상 행위로부터 벗어나는 주목할 만한 특이한 행위 패턴을 침입으로 규정하여 침입을 탐지한다.

이들 방법들은 ERP 활용 기업들에게 상업화되어 이미 사용되고 있지만 새로운 침입 패턴과 변형된 침입 패턴을 탐지할 수 없는 문제점이 있으며 오용 탐지를 위한 공격 유형을 분석, 전자무역서류 인코딩 작업에 시간과 비용이 많이 소요되는 문제점들을 아직도 해결하지 못하고 있다.[2]

호스트 기반의 이상 탐지 기법은 열거형, 빈도기반, 전자문서 데이터 마이닝 접근, 유한상태 기계 방법으로 분류할 수 있다. 열거형 순차방법은 전자무역 네트워크상의 인증된 문서를 경험적으로 추적하여 알려지지 않은 패턴을 모니터링하여 이상 유무를 탐지한 후, 기업 통합시스템인 ERP에 통보한다. 빈도기반 방법은 다양한 이벤트의 빈도 분포를 기준으로 하여 침입을 탐지하며, 데이터 마이닝 접근 방법은 정상행위 데이터로부터 발생하는 공통의 원소로부터 특징을 추출하여 규칙집합으로 기술함으로서 침입탐지가 가능하도록 한다. 또한 유한상태 기계방법은 기계학습기법으로 프로그램을 추적하여

1) 침입탐지시스템은 단순한 접근 제어 기능을 넘어서서 전자문서 네트워크나 시스템의 사용을 실시간 모니터링하여 침입탐지패턴을 이용하여 침입을 탐지하는 시스템으로서 악의적인 사용자로부터 접속, 정보의 조작, 오용, 남용 등 컴퓨터 시스템 또는 네트워크 상에서 시도되었거나 진행 중인 불법적인 행위를 탐지하는 시스템이다.

2) Steven Noel, Duminda Wijesekera, Charles Youman, "Modern Intrusion Detection, Data Mining, and Degrees of Attack Guilt", Applications of Data Mining in Computer Security, Daniel Barbara and Sushil Jajodia (eds.), Kluwer Academic Publishers, 2002. pp.302-324.

인식하는 유한상태기계를 구축하여 이상침입을 탐지하는 방법들이다.[3)]

전자무역에서 네트워크 기반의 이상 침입 탐지 시스템은 네트워크 상의 패킷 데이터를 수집하여 이상 침입을 탐지하는 시스템들이다. ADAM, NIDES, SPADE 등의 네트워크 이상 탐지 시스템들은 감시 데이터로는 패킷의 헤더 정보인 IP 주소, 포트, TCP/IP 상태 등을 이용한다. Matthew는 네트워크 이상 탐지 시스템에 PHAD와 ALAD의 두 요소로 구성하여, 패킷 헤더 데이터의 이상 탐지와 응용 계층의 이상 탐지를 수행하고 있다.[4)]

본 연구논문은 전자무역에서 기업간에 상호 제시 할 문서 전달과정 중 호스트 기반과 네트워크 기반의 불확실성한 이상 침입 탐지를 위하여 베이지안 기법을 적용하였고, 이러한 불확실성을 처리하는 베이지안 이론을 이상 침입 탐지영역에 도입하여 적용함으로써 오용 탐지의 한계성을 극복하여 알려지지 않은 전자서류 해킹 탐지방법을 제시하고자 하였다.

Ⅱ. 전자무역 베이지안 네트워크

1. 전자무역의 베이지안 이론

디지털 환경에서 요구되는 기업간 협업과 네트워크화를 통한 e-트랜스포메이션은 중요한 수단이 될 수 있다. BPO(기업 프로세스 아웃소싱)을 활용한 e-트랜스포메이션을 통해 경쟁력을 강화하기 위해서는 먼저 기능적 조직을 프로세스 중심으로 재정립할 필요가 있다.

현재 대부분 기업들이 기능별 조직형태를 취하고는 있으나, 이러한 기능별 조직에서는 타 기능 부서와의 잦은 정보 및 인력교류가 없고 기능부서 내의 수직적인 업무 흐름만 존재하고 있다.

실제로 중요한 물류정보의 경우 기능별 조직을 관통하여 수평적으로 원활히 움직여야 하는데, 기능별, 수직적 조직에서는 이러한 수평적 정보이동이 어렵다. 즉, 물류부문 내에서 존재하는 프로세스도 많고, 물류부문과 수송개발 부문을 관통하는 프로세스, 혹은 유통, 애프터서비스, 제고관리, 네트워킹을 연결하는 프로세스 등도 여러 가지가 있다.

우리나라 경우에 많은 기업들이 BPR(기업 프로세스 재계획)을 실시하여 조직의 많은 부문이 이미 프로세스화 되었기 때문에 조직의 기능별 관점과 프로세스 관점에서 조직의 경쟁력을 높일 수 있는 BPO(기업 프로세스 아웃소싱)를 활용한 e-트랜스포메이션은 그 의미가 크다고 할 수 있다.

e-트랜스포메이션은 보이지 않게 여러 가지 변화를 가져오고 있다. 특히 인터넷 버블경제의 붕괴와

3) Christina Warrender, Stephanie Forrest, Barak Pearlmutter, "Detecting Intrusion Using System Calls : Alternative Data Models", 2004. pp.146-167.

4) Matthew V. Mahoney and Philip K. Chan, "Learning Nonstationary Models of Normal Network Traffic for Detecting Novel Attacks", 2003. pp.117-132.

함께 전자화폐, 전자결재 등의 빠른 인터넷 확산과 더불어 파생되는 많은 현상들도 접하게 하였고, 그에 따른 기술적인 노력도 필요하게 만들었다.

e-트랜스포메이션은 새로운 기술 환경에서 제공하는 가능성과 이것이 기업들에 미치는 효과는 시공간의 제약을 초월함으로써 통상적인 물류비를 절감하고 거래비용을 줄임으로써 판매 및 마케팅 비용도 감소시킬 수 있다. 매출원가를 낮추게 할 수도 있고 신규 비즈니스 기회도 제공하게 된다.

따라서 기술적 부문에 기반을 둔 개별 브랜드와 전체 포트폴리오를 재검토하는 작업을 e-트랜스포메이션을 통하여 적극적으로 수행해야만 물류유통 제품군의 유형 변화에 성공적으로 대비하고 주도할 수 있게 된다. 이러한 여러 가지 과정 등을 통하여 혁신주도형 사업과 원가주도형 사업을 적절히 통합하여 운영함으로써 수익성을 극대화하는 사업으로도 발전시킬 수 있게 된다[5].

새로운 가치 제안(Value Proposition)과 판매채널 및 시장 확대도 동시에 가능하게 하여 준다.

결론적으로 e-트랜스포메이션의 네트워크 향상과 전자문서교환 방법 등의 기술적 향상은 모든 기업들에게 있어서는 직·간접적인 기회를 제공하게 되며 이를 잘 활용하는 기업은 글로벌 환경하에서 경쟁우위에 설 수 있다. 이 같은 이유로 전자무역의 네트워크 베이지안 이론은 확률, 템스트·쉐퍼 이론, 퍼지 이론 등과 같이 불확실성을 처리하기 위한 하나의 방법임을 알 수 있다.

불확실성이란 기업간(B2B) 최종 의사결정을 하기 위해 필요한 데이터 정보가 부족한 상황을 의미하며 불확실성의 원인으로는 정보의 유실에 의한 부분 정보만의 존재, 정보들 간의 충돌, 정보에 대한 신뢰성의 부족, 지식표현 언어의 한계 등을 들 수 있다.

전자무역에서 베이지안 네트워크 이론을 정의하기에 앞서, 확률의 정의, 기본 성질, 독립 사건과 종속 사건을 언급 해 보면, 확률의 정의는 어떤 시행에서 모든 경우의 수를 N, 그 중에서 사건 A가 일어나는 경우의 수를 a라 할 때, 사건 A가 일어날 확률은 $P(A) = \frac{a}{N}$ 라고 할 수 있다.

확률의 기본 성질은 $0 \le P(A) \le 1$, $P(U) = 1$, $P(\varnothing) = 0$, $P(A^C) = 1 - P(A)$ 이다.

독립과 종속 사건은 “어느 한쪽이 일어나는 것이 다른 쪽의 영향을 받지 않을 때 A와 B는 독립, 그렇지 않을 때는 종속” 이라 한다.

덧셈의 정리에 의해서 사건 A와 B가 독립이면, $P(A \cup B) = P(A) + P(B)$이고,

사건 A와 B가 종속이면, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$가 된다.

곱셈의 정리 또한, 사건 A와 B가 독립이면, $P(A \cap B) = P(A) \cdot P(B)$이고,

사건 A와 B가 종속이면, $P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$가 된다.

조건부 확률 $P(B|A)$은 사건 A를 전제로 사건 B가 일어날 확률로써,

5) 에릭 조아킴스탈러外, 현대경제연구원 역편,“브랜드경영”, 21세기북스, 2003. p.277.

$P(B|A) = \frac{P(A \cap B)}{P(A)}$ 로 정의된다.

전자무역의 베이지안 이론은 무역 서류 전달과정 중에 사건 A가 발생한 후 사건 B가 발생할 확률인, 조건부 확률 $P(B|A)$의 역 확률 $P(A|B)$를 간단하게 산출할 수 있다. 역 확률 $P(A|B)$는 나중에 발생한 사건 B에 대하여 먼저 발생한 사건 A의 확률을 의미한다. 전자무역에서 전자문서 상호간 전달 중에 해킹이나, 버퍼링으로 인하여 송수신이 어려울 때 그 증상이 나타나고 있다.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)},$$

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{P(E|H_i)P(H_i) + P(E|H_2)P(H_2) + \cdots + P(E|H_n)P(H_n)}$$

따라서 전자무역에서 베이지안 네트워크 이론을 적용하기 위해서는 먼저 각 사건이 독립적이고 명확한 사전 확률이 요구된다. 매우 복잡한 문제에는 적합하지 않지만 잘 정의된 좁은 영역의 문제 해결에는 매우 유용하다.

2. 전자무역의 베이지안 네트워크

전자무역 베이지안 네트워크 이론의 조건부 독립을 그래프의 네트워크 형태로 표현하고 있는데, 실세계의 지식을 확률이 부여된 방향성 비순환 그래프로 표시한다.

전자무역의 베이지안 네트워크를 인과 네트워크(Causal Network) 또는 신뢰 네트워크(Belief Network)라고도 한다. 베이지안 네트워크에 의해서 표현된 지식을 이용하여 추론이 가능하다. 추론의 종류는 인과 추론, 분석 추론 그리고 상호 인과 추론이다.

인과 추론은 사건 A에 의해 사건 B가 발생한다고 할 때, 사건 A의 값을 알면 사건 B의 확률을 계산할 수 있다. 분석 추론은 사건 A에 의해 사건 B가 발생한다고 할 때, 사건 B 값을 알면 사건 A의 확률값을 계산할 수 있다.

상호 인과 추론은 사건 A와 B 모두 사건 C의 원인이 된다고 할 때 사건 C를 알면 사건 A의 확률 변화가 사건 B의 확률에 미치는 영향을 계산할 수 있다. 전자무역의 베이지안 네트워크의 문제점은 모든 가능한 사건들이 상호 배타적이며, 조건부 확률을 사용하므로, 사전 확률을 모두 알고 있어야 한다. 계산 복잡도가 사건의 개수에 지수적으로 비례하여 증가한다.

3. 전자무역의 베이지안 기법 적용

전자무역에서 베이지안 기법을 이용한 네트워크 이상 탐지는 통계적 기법의 이상 탐지이다.

그렇기 때문에 베이지안 기법의 이상 탐지를 위해서는 먼저 정상적인 송수신 전자문서 교환 서류 등의 데이터로부터 사전 확률 정보를 획득하여야 한다.

이상 행위를 탐지하기 위해서는 정상 행위 데이터를 이용한 정상 행위에 대한 프로파일링 구축과 행위나 이벤트를 기술하는 행위 패턴 생성과정이 필요하다.

정상 행위가 이루어진 송수신 네트워크 시스템의 로그파일로부터 정상 행위 데이터를 수집한다. 일반적으로 정상 행위 데이터는 침입을 탐지하려는 영역에 의해서 호스트 기반과 네트워크 기반으로 분류된다. 호스트 기반의 이상탐지는 데몬 프로그램 수행 중에 호출되는 시스템 호출 순서나, 일반 사용자 행위를 이용하고, 네트워크 기반은 네트워크의 패킷 데이터 등을 이용한다.

	공격 탐지 & 차단	취약점 분석	암호화, Anti-virus
Network	• Firewall (Network-based Access Control System) • Network-based IDS	• Network Vulnerability Scanner	• IP Tunneling • VPN Appliance • Virus Filtering Gateway
System (Host)	• Host-based IDS • Host-based Access Control System	• Host(OS) Vulnerability Scanner	• File System Encryption • File System Integrity • Anti-virus System (for mail server, etc.)
App.	• PC(Client) Security with strong authentication	• Database Vulnerability Scanner • PC Vulnerability Scanner	• Data Encryption (Email, File, etc.) • Public Key Infrastructure • Anti-virus System

[그림 1] 전자무역의 침입 탐지 구성도

전처리 과정으로 정상 행위 데이터를 세션(Session)단위로 구분 한다. 세션에 의해서 행위의 시작과 끝을 구분하고 하나의 행위 단위로 간주한다. 세션의 구분을 위한 척도로는 정상 행위의 데이터 종류에 따라 다르겠지만 데몬 프로그램의 시스템 호출은 프로세스 아이디, 사용자 행위는 로그인 아이디, 네트워크 패킷 데이터는 발신지와 수신지의 IP와 포트 번호에 의해서 세션을 구분한다. 세션 단위로 구분된 정상 행위 데이터로부터 행위 또는 이벤트에 대한 사전 확률 정보를 획득한다.

각각의 행위와 이벤트들 간의 관계를 베이지안 확률에 의해서 산출한다. 세션 단위로 구분된 정상 행위 데이터를 클러스터링하여 유사도가 가까운 행위를 전자무역 베이지안 네

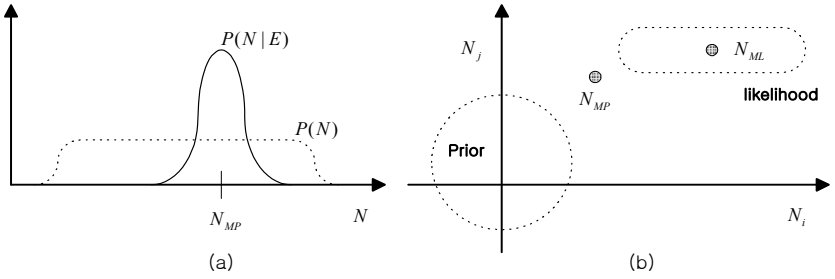
트워크를 이용하여 정상 행위를 프로파일링한다.

새로운 행위나 이벤트에 대해 사전 확률과 사후 확률 값 계산에 의해서 이상 침입 행위를 탐지한다. 이상 침입 탐지를 위한 베이지안 확률값 계산은 사전 확률과 사후 확률, 우도 함수 등을 이용하여 다음과 같은 식으로 제시할 수 있다.

$$P(N|E) = \frac{P(E|N)P(N)}{P(E)} = P(N)\frac{P(E|N)}{P(E)}$$

사전확률 $P(N)$는 정상 행위나 이벤트의 발생 빈도에 의해 좌우되며, 정상 행위나 이벤트에 대한 완전한 정보를 제공하지 못한다.

그러나 사후 확률 $P(N|E)$는 이벤트 E 라는 조건에 의하며, 가장 유력한 정상 행위 이벤트는 $P(N|E)$부분에서 집중적인 확률 분포를 보인다.[6] 사후 확률 $P(N|E)$은 사전 확률 $P(N)$와 우도 함수 $P(E|N)$의 곱에 이벤트의 확률 $P(E)$로 나눔으로써 계산되며, 위의 관계를 그림으로 나타내면 다음과 같다.



[그림 2] 전자문서 교환의 정상 행위에 대한 확률 관계

Ⅲ. 전자무역 호스트 기반

전자무역에서 호스트 기반의 이상 침입 탐지 시스템은 호스트로부터 생성, 수집된 감사 데이터를 근거로 이상 침입을 탐지하는 시스템이다. 감사 데이터로는 프로세스를 이용하며, 프로세스에 대한 원래의 소유자와 그룹, 수행되고 있는 프로세스의 현재 사용자와 그룹, CPU 사용량, I/O 할당량, 이 프로세스가 사용하는 파일, 이 프로세스가 호출하는 시스템 호출 등을 수집하여 정상 패턴과 이상 패턴을 구축하여 여러 가지 이상 탐지 방법을 통하여 이상 침입을 탐지하게 된다.

6) Christoper M. Bishop, "Neural Networks for Pattern Recognition", Oxford Press, 2001, pp.385-433.

1. EDI 프로그램 행위 프로파일링

EDI 프로그램 행위를 분석하여 프로파일을 구축하는 기법들은 사용자 행위 침입 탐지 기법의 대안으로 연구되어 왔다. 프로그램 행위 프로파일은 정상적인 프로그램이 수행되면서 발생시키는 시스템 호출들을 수집 및 분석하여 구축한다. 시스템 호출을 이용한 이상 탐지 기법들은 열거형 순차 방법, 빈도 기반의 방법, 데이터 마이닝 접근 방법 그리고 유한 상태 기계방법 등이 있다.[7)]

열거된 순차에 의존하는 방법들은 lookahead pairs, tide, stide 등이 있다. 이 방법들은 정상 행위를 경험적으로 추적하여 알려지지 않은 패턴을 모니터링한다. 전자무역 활용 초기에 이 기법들은 패턴에 대한 통계적 분석이 적용되지 않았다. 빈도 기반의 방법들은 다양한 이벤트의 빈도 분포를 모델로 하며, 텍스트 문서를 분류하는데 사용된 n-그램 벡터가 여기에 속하였다. 네트워크 시스템 호출 추적의 방법으로 이 기법은 프로그램이 종료되어야 추적 벡터를 계산할 수 있기 때문에 온라인 테스팅에서는 부적절하다.

또한 벡터의 크기를 결정하는데 어려움이 있으며, 동일한 프로그램의 정상 행위와 이상 행위를 추적을 위한 충분한 정밀도를 제공하지 못한다. 데이터 마이닝 접근법은 많은 수집된 데이터로부터 가장 중요한 특징을 결정하기 위해 설계되었다. 이상 침입 탐지에서는 발생된 정상 행위의 모든 패턴을 단순하게 나열하여 얻기보다는 간결하게 정의할 수 있는 정상 행위 패턴을 발견하는데 있다.

데이터 마이닝 접근법으로 RIPPER는 정상 행위 데이터로부터 발생하는 공통의 요소를 작은 규칙 집합으로 특징을 기술하는 능력을 제공한다. 기계 학습 접근법으로 프로그램을 추적하여 인식하기 위하여 유한 상태 기계(Finite State Machines)를 구축하여 이상을 탐지한다. 매우 강력한 유한 상태 기계로는 은닉 마코프 모델이 있으며, 이 모델은 이중 추정 통계 과정으로 기술된다. 여러 모델 중에서 가장 이상탐지 능력이 뛰어난 것으로 판명되었으나 계산하는데 복잡한 단점을 갖고 있다. 이상 탐지 모델에서 발생하는 문제점들은 통계적 분석의 필요, 실시간(Real-Time) 처리의 어려움, 정상 행위의 간결한 정의, 계산 복잡도 문제 등이다.

본 논문에서는 각각의 시스템 호출 정보를 이용한 베이지안 네트워크를 구축하여 프로세스간의 관련성을 표현함으로써 프로그램 수준의 변형된 침입 탐지를 증명한다.[8)]

2. ERP 프로파일 활용과 N-gram 기법

전자무역에서 프로그램 행위 기반 침입 탐지 기법의 전제는 대부분의 공격은 프로그램 결함이나 버그로 인하여 발생할 수 있으며 프로그램의 정상적인 사용과는 그 행위가 다르다는데 있다.

---

7) Christina Warrender, Stephanie Forrest, Barak Pearlmutter, "Detecting Intrusion Using System Calls : Alternative Data Models", 2004. pp.146-167.

8) Mehdi Nassehi, "Characterizing Masqueraders for Intrusion Detection", Computer Science/Mathematics, 2001. pp.268-277.

프로그램의 행위가 적합하게 표현될 수 있다면 침입 탐지를 위한 행위 특성으로 활용될 수 있다.[9] 전자무역에서 N-gram 기법은 프로그램에 의해 발생되는 일정 길이의 순차 시스템 호출들, 즉 N-gram 또는 스트링(string)으로 프로파일 데이터베이스를 구축한다. 프로파일 데이터베이스가 구축된 후, 프로그램이 발생시킨 시스템 호출들 중에서, 특정 길이의 일련의 시스템 호출들이 프로파일에 존재하지 않는다면 비정상 행위로 간주하여 개수를 센다.

세션내의 총 스트링 개수에 대해 비정상 행위로 간주된 스트링의 개수의 비율이 매우 크다면 그 세션을 비정상으로 판정한다. 그러나 이 기법은 프로그램마다 매우 큰 프로파일이 필요하다는 문제점이 있다. N-gram 기법을 기반으로 탐지율을 더 높이기 위한 방법들로는 Teiresias 알고리즘을 이용하여 가변 길이의 스트링을 발굴하고 프로파일 데이터베이스를 구축하여 이상 행위를 탐지하는 방법과 RIPPER를 이용하여 스트링의 각 위치마다 시스템 호출의 발생 확률을 측정하여 이상 행위를 탐지하는 방법 등이 있다.

또한 로그인 세션 중에서 비정상적인 세션을 효율적으로 판정하기 위해 침입 탐지 기법에 다변량 분석 기법을 도입한 방법이 있다.[10] 전자무역에서 N-gram 기법을 활용할 때 단순한 알고리즘과 높은 탐지율을 보이지만 프로파일 데이터의 크기 및 오버헤드가 매우 크다는 단점을 동시에 갖고 있다.

3. 네트워크 기반의 이상탐지

네트워크 기반의 이상 침입 탐지 시스템은 네트워크 상의 패킷 데이터를 수집하여 이상 침입을 탐지하는 시스템이다. ADAM, NIDES, SPADE 등의 네트워크 이상 탐지 시스템들은 감사 데이터로는 패킷의 헤더 정보인 IP 주소, 포트, TCP/IP 상태 등을 이용한다.

Matthew는 네트워크 이상 탐지 시스템에 PHAD와 ALAD의 두 요소로 구성하여 패킷 헤더 데이터의 이상 탐지와 응용계층의 이상 탐지를 수행한다.[11]

네트워크 기반의 이상 침입을 탐지하기 위해서는 베이지안 네트워크를 적용한 패킷 데이터의 정상 행위를 프로파일링 해야 한다.

4. 정상 행위 프로파일링

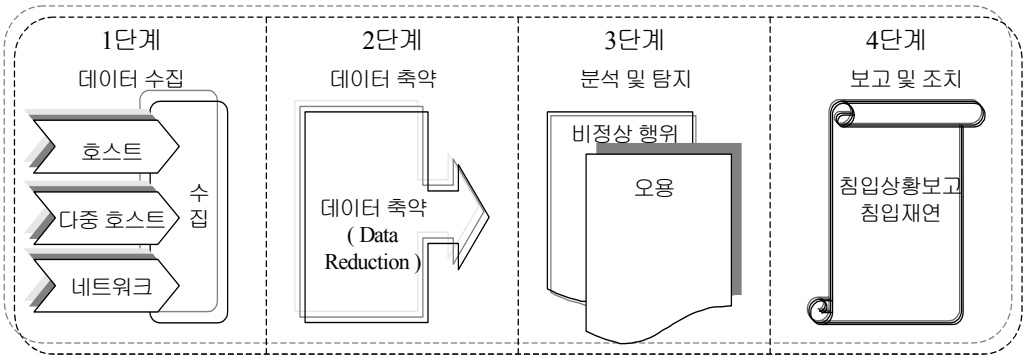
네트워크 기반의 침입 탐지에는 네트워크 데이터인 패킷의 헤더 정보를 이용하여 이상이나 오

9) S. A. Hofmeyr, A. Somayaji and S. Forrest, "Intrusion Detection using Sequences of System Calls", Journal of Computer Security, Vol.6, 1998. pp.151-180.

10) N. Ye, Q. Chen, S. Vibert, "Multivariate Statistical Analysis of Audit Trails for Host-Based Intrusion Detection", IEEE Transactions of computers, Vol. 51, No. 7, 2002. pp.808-828.

11) Matthew V. Mahoney and Philip K. Chan, "Learning Nonstationary Models of Normal Network Traffic for Detecting Novel Attacks", 2003. pp.117-132.

용 침입을 탐지한다. 전자무역에서는 TCP/IP/ 기반의 서비스에 대한 EDI/ebXML의 패킷 헤더 정보를 이용하여 서비스별로 분류하며, 다양한 서비스별로 정상 행위를 프로파일링 한 후 이상 침입을 탐지한다.



[그림 3] 침입 탐지 시스템의 일반적인 구조

1단계 : 데이터 수집(Raw data collection) 단계⇒ 침입탐지 시스템이 대상 시스템에서 제공하는 시스템 사용 내역, 컴퓨터 통신에 사용되는 패킷 등과 같은 탐지대상으로부터 생성되는 데이터를 수집하는 감사 데이터(audit data) 수집 단계이다.

2단계 : 데이터 가공 및 축약(Data reduction and filtering) 단계⇒ 수집된 감사데이터가 침입 판정이 가능할 수 있도록 의미 있는 정보로 전환시킨다.

3단계 : 분석 및 침입탐지 (Analysis & Detection) 단계⇒ 수집된 데이터를 분석하여 침입 여부를 판정하는데, 이 단계는 침입탐지 시스템의 핵심 단계이며, 시스템의 비정상적인 사용에 대한 탐지를 목적으로 하는지, 시스템의 취약점이나 응용 프로그램의 버그를 이용한 침입에 대한 탐지를 목적으로 하는지에 따라 비정상적 행위 탐지 기술과 오용 탐지 기술로 나뉘어 진다.

4단계 : 보고 및 대응(reporting and response) 단계⇒ 침입탐지 시스템이 시스템의 침입 여부를 판정한 결과 침입으로 판단된 경우 이에 대한 적절한 대응을 자동으로 취하거나, 보안관리자에게 침입 사실을 보고하여 보안관리자에 의해 조치를 취하게 한다. 즉 침입 탐지 시스템은 보호하고자 하는 시스템으로부터 침입을 판단하기 위한 데이터를 수집하고 중복된 데이터나 쓸모없는 데이터를 필터링하고 탐지 기법을 사용해 침입을 탐지하고 그에 해당하는 응답을 하는 시스템이다.[12)]

EDI 과정 중 네트워크 침입 탐지는 단지 TCP/IP 패킷의 이상 유무와 침입시의 패킷의 여러 특징에 의해서 이상 침입을 탐지한다. 본 논문은 전자무역 패킷 정보 서비스에 특정한 제약을 적용함으로써 좀더 네트워크 이상 침입을 명확히 구분하고자 하였다.

12) Edward G. Amoroso, "Intrusion Detection: An Introduction to Internet Surveillance, Correlation, Traps, Trace Back, and Response", 1999. pp.168-195.

5. 전자무역의 FTP 서비스 분석

전자무역에서 FTP는 가장 오래된 TCP/IP 프로토콜 중 하나이다. 이 프로토콜은 호스트들 사이에서 텍스트 또는 바이너리 파일들을 전송하기 위해 사용된다. FTP는 접근을 제어하는 패스워드를 사용하거나 익명 접근을 허용할 수 있다. 익명 접근을 허용하면 계정이나 패스워드 없이도 유용한 정보를 공공에게 공개할 수 있다. 전자무역의 서류전달 과정 중 인터넷상의 해커들에게 서버가 노출되기 때문에 익명 접근은 조심스럽게 관리해야 한다. 단순한 FTP의 파일 전송 절차는 FTP의 시작, 파일 데이터의 전송, 연결종료 등으로 이루어진다.

FTP시작은 ftp<hostname> or ftp open <hostname>, <IP_address> 명령에 의해서 FTP의 시작과 동시에 호스트에 접속이 이루어진다. 파일 전송은 PC에서 호스트로 파일을 전송하는 Put과 반대 과정의 Get 명령으로 구성된다. 연결 종료는 FTP 프롬프트 상태에서 Quit 명령으로 FTP 연결을 종료한다. FTP 서비스는 제어 포트와 데이터 포트라는 두 채널을 구축하며 두 채널에 의해서 제어 신호와 파일 전송이 같이 이루어진다.

FTP 서비스의 단순한 파일 전송을 위해서는 FTP 시작과 동시에 제어 포트를 통한 제어 연결이 설정된다. 제어 연결을 통해서 ID와 패스워드의 접근 제어 과정을 거친 뒤에 전송 받을 파일을 검색한다. 파일 전송을 위해서는 데이터 포트를 통한 데이터 연결이 설정된다. 데이터 연결을 통해서 파일들이 전송되고 전송 완료되면 데이터 연결은 종료된다. 제어 연결은 FTP 종료 전까지는 설정되어 있다.[13]

6. 문제점

베이지안 네트워크는 서브네트 상의 모든 네트워크 패킷을 수집하여 이를 감시 자료로 사용하기 때문에 보안 영역으로 지정된 호스트들의 감시가 가능하고 실시간으로 침입을 탐지하는 것이 가능하다. 베이지안 네트워크는 크게 감시자료 수집 및 탐지모듈로 이분화 되어 있다.

감시자료 수집 모듈은 보안영역으로 설정된 특정 호스트의 패킷만을 포착하여 감시 자료를 생성하여 탐지 모듈에서 분석하게 된다.[14] 베이지안 네트워크를 호스트 기반과 네트워크 기반의 침입 탐지 시스템으로 분류하여 살펴보았을 때 호스트 기반은 침입 탐지 시스템을 목적지 호스트에 설치해야 하므로 해당 호스트의 성능이 저하되고 데이터를 얻기 위해 로깅 등에 대한 설정이 번거로우며 목적지 호스트가 있는 네트워크 내의 다른 호스트들이 공격을 당해도 알 수가 없다.

네트워크 기반 침입탐지 시스템은 네트워크 패킷을 캡쳐하여 프레임 패킷 중에서 데이터 부분을 검

13) Eleazar Eskin, “Anomaly Detection over Noisy Data using Learned Probability Distributions”, In Proceedings of the 17 th International Conference on Machine Learning (ICML-2000), 2000. pp.151-172.

14) Edward G. Amoroso, “Intrusion Detection: An Introduction to Internet Surveillance, Correlation, Traps, Trace Back, and Response”, 1999.

출하여 IP 데이터그램과 TCP/IP, UDP 각각의 세그먼트에 대한 패킷을 필터링한다. 네트워크 트래픽이 감시자료 수집모듈의 캡쳐링 속도 보다 빠르다면 감시자료 적체 현상과 사용자의 입력과 침입유형을 비교하는 침입 판정 엔진에 과도한 오버헤드를 초래하게 되므로 침입탐지 시스템의 성능에 악영향을 줄 수 있다.

대부분의 침입탐지 시스템의 구조는 감시자료 수집 모듈에서 캡쳐한 프레임을 버퍼에 저장하고 침입탐지 모듈을 버퍼에 있는 내용을 탐지하는 구조를 가지고 있다.

이 경우에 버퍼의 용량이 남아있으면 탐지 모듈의 과부하에 상관없이 패킷을 캡쳐하는 구조를 가지게 된다. 고속 네트워크에서는 보다 고성능의 시스템이 필요하다. 감시기록 추적모델 설계 및 구현을 이용하면 침입탐지시스템의 감시 기록 수집 모듈과 탐지 모듈간의 속도 차이 문제를 해결할 수 있고, 침입을 미리 차단하므로 과부하 문제도 해결되므로 감시기록 추적 모델 설계 및 구현이 필요하다.

Ⅳ. 전자무역 TCP/IP 패킷

전자무역에서 TCP/IP는 연결지향 프로토콜이다. 연결지향 프로토콜은 발신지와 목적지간에 가상 경로를 설정한다. 메시지에 속하는 모든 세그먼트들은 설정된 가상 경로를 통하여 전송된다. 전체 메시지를 하나의 단일 가상 경로를 이용하여 전송되므로 손상 또는 손실된 프레임의 재전송뿐만 아니라 확인응답 프로세스도 가능하게 된다. TCP/IP에서 연결지향 전송은 연결 설정과 연결 종료의 두 가지 절차를 통해서 이루어진다.[15)]

1. 연결 설정 및 종료

전자무역에서 TCP/IP는 전자전이(e-Transformations) 중 모드로 전송된다. 두 호스트 사이의 TCP/IP가 연결되면 서로 세그먼트를 주고받을 수 있어야 한다.

데이터 교환이 이루어지기 전에 한 편에서는 통신을 개시하고 다른 편에서는 통신 개시의 요구에 대한 승인이 먼저 이루어져야 한다는 것을 의미한다. 그러므로 데이터 전송이 이루어지기 전에 두 호스트는 3단계 핸드셰이크라는 절차를 수행하여야 한다.

핸드셰이크 절차는 서버에서부터 시작한다. 서버 프로그램은 자신의 TCP/IP에게 연결을 수락할 준비가 되어 있다는 것을 알린다. 이것을 수동 개방(passive open)을 위한 요구라고 한다. 수동 개방이라는 것은 비록 TCP/IP가 다른 시스템으로부터 어떠한 연결도 수락할 수 있지만 자신이 먼저 연결을 개설

15) Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim, "Efficient Algorithms for Mining Outliers from Large Data Sets", Technical report, Bell Laboratories, Murray Hill, 1998. pp.87-106.

할 수 는 없다는 것을 의미한다.

클라이언트 프로그램은 능동 개방(active open)을 위한 요구를 실행한다. 서버와 연결하고자 하는 클라이언트는 자신의 TCP/IP에게 특정한 서버와 연결을 설정할 필요가 있다고 알린다. 3단계 핸드셰이크 절차는 첫째, 클라이언트는 첫 번째 SYN 세그먼트를 전송한다. 둘째, 서버는 두 번째 SYN+ACK 세그먼트를 전송한다. 셋째, 클라이언트는 세 번째 ACK 세그먼트를 전송한다.

두 프로세서가 동시에 서로에게 능동 개방을 요구하는 상황이 일어날 수도 있다. 이 경우, 양쪽 TCP/IP는 서로에게 SYN+ACK 세그먼트를 전송하게 되고 하나의 단일 연결이 두 TCP/IP 사이에 설정된다. 데이터를 교환하는 어느 쪽(클라이언트와 서버)도 연결을 종료할 수 있다. TCP/IP 연결은 양 방향으로 이루어져 있으며, 한 방향의 연결이 종료되더라도 다른 시스템은 다른 방향을 통하여 데이터 전송을 계속할 수 있다.

양 방향 연결이 종료되기 위해서는 4단계 핸드셰이크가 필요하다. 4단계 핸드셰이크는 일반적으로 클라이언트 프로그램이 연결 종료를 요구한다. 클라이언트 프로그램은 자신의 TCP/IP에게 데이터 전송이 종료되었고 따라서 연결을 종료하고자 한다는 것을 알린다.

이것을 능동 종료(active close) 요구라고 한다. 능동 종료 요구를 수신한 후에, 클라이언트 TCP/IP는 클라이언트-서버 방향의 연결을 종료한다. 그러나 다른 방향으로의 통신은 여전히 개방되어 있다. 서버 프로그램이 서버-클라이언트 방향으로의 데이터 전송을 끝마치게 되면, 서버는 서버-클라이언트 방향의 연결 해지를 자신의 TCP/IP에게 요구할 수 있다. 이것은 보통 수동 종료(passive close)라고 한다.

연결 종료를 위한 4단계는 첫째, 클라이언트 TCP/IP는 첫 번째 FIN 세그먼트를 전송한다. 둘째, 서버 TCP/IP는 FIN 세그먼트의 수신 확인을 위하여 두 번째 ACK 세그먼트를 전송한다. 셋째, 서버 TCP/IP는 세 번째 FIN 세그먼트를 전송한다. 넷째, 클라이언트 TCP/IP는 FIN 세그먼트의 수신 확인을 위하여 네 번째 ACK 세그먼트를 전송한다.[16]

2. 연결 리셋

전자무역에서 TCP/IP는 연결 리셋(connection reset)을 요구할 수 있다. 여기에서 리셋이라는 것은 연결이 파기되었다는 것을 의미한다. 리셋은 다음과 같은 3가지의 경우에 발생할 수 있다.

첫째, 한 쪽의 TCP/IP가 존재하지 않는 포트로 연결을 요구하면, 다른 쪽의 TCP/IP는 연결을 파기하기 위하여 RST 비트를 설정한 세그먼트를 전송한다.

둘째, 한 TCP/IP는 비정상적인 상황으로 인하여 연결의 중단을 요구할 수 있다. 이 경우 TCP/IP는 연결을 종료하기 위하여 RST 세그먼트를 전송할 수 있다.

16) W. Lee and S. Stolfo, "Learning Patterns from Unix process Execution Traces for Intrusion Detection", AAAI Workshop : AI Approaches to Fraud Detection and RISK management, July, 1997. pp.50-56.

셋째, 한 쪽의 TCP/IP는 다른 편의 TCP/IP가 긴 시간동안 휴지 상태에 있다는 것을 확인한 후, 연결을 파기하기 위하여 RST 세그먼트를 전송한다.[17)]

3. 전자무역의 TCP/IP 상태 전이 다이어그램

연결설정, 연결 종료, 그리고 데이터 전송 기간 동안 발생하는 여러 가지 이벤트들을 관리하기 위하여, TCP/IP 프로토콜은 유한 상태 기계(Finite state machine)를 이용하여 구현된다.

유한 상태 기계는 제한된 수의 상태를 가지는 기계이다. 어느 순간에 기계는 여러 개의 상태 중의 하나에 있을 수 있으며 이벤트가 발생하기 전까지는 그 상태에 머무른다.

만일 이벤트가 발생하면 기계는 자신의 상태를 새로운 상태로 바꾸거나 또는 어떠한 행동을 수행한다. 이벤트는 상태에 적용되는 입력이다. 이벤트는 상태를 변화시킬 수 있고 또한 출력을 발생할 수도 있다.

V. 연구요약 및 결론

전자무역 B2B, B2G 거래의 급속한 확대 보급에 따라 가용 네트워크 정보량의 증가와 정보 보호 위협 요인의 증가로 인하여 전자서명/인증과 접근제어의 보안기술만으로는 보안 문제 해결에 충분치 못하여 정보 보호를 위한 방법으로 베이지안 네트워크 기법들이 개발되고 있다.

본 논문은 전사적기업자원관리(ERP) 시스템을 중심으로 베이지안 네트워크 기법을 연구하였다. 베이지안 기법은 불확실성 문제를 해결하기 위한 기법 중의 하나이며, ERP 시스템에서 이상 침입 탐지의 불확실성을 해결하기 위하여 적용된다.

베이지안 네트워크를 적용하여 기업간에 시스템 호출 그리고 네트워크 데이터에 대해 프로파일링을 구축하고 이상 침입 탐지를 위한 시뮬레이션을 수행 할 수 있다.

전자무역에서 호스트 기반의 이상 침입 탐지 기법은 데몬 프로그램이나 루트 권한으로 실행되는 프로그램이 발생시키는 시스템 호출들을 분석하여 베이지안 확률 값을 적용하며 정상 행위 프로파일을 구축하여 비정상 행위를 탐지한다.

전자무역에서 베이지안 이론은 패턴인식 등의 여러 학문 분야에서 불확실성 문제의 해결 및 의사결정 문제에 우수성을 보이고 있다. 본 논문에서 베이지안 기법의 연구 결과는 전자문서 전달 행위의 전후 관계를 베이지안 확률 값으로 추정한 후, 전자문서 전달 행위 또는 이벤트의 전후 관계를 베이지안 네트워크로 표현하였다.

---

17) Susan M. Bridges, Rayford B. Vaughn, “Fuzzy Data Mining And Genetic Algorithms Applied to Intrusion Detection”, 23 rd National Information Systems Security Conference October, 2000. pp.16-19.

전자문서 전달행위의 전후 관계를 이용한 정상 행위를 간결하게 프로파일링하며 변형되거나 새로운 행위에 대해서도 탐지가 가능하다. 베이지안 기법 네트워크를 이용한 정상행위 프로파일링 과정을 호스트 기반과 네트워크 기반의 이상 침입 탐지에 적용 할 수 있다.

향후, 전자무역 거래에 ERP의 다양한 데이터에 대한 이상 침입 패턴 분류와 베이지안 확률값에 의한 이상 침입 패턴을 평가하는 기준을 제시하고, 변형된 이상 침입 패턴을 효과적으로 탐지하기 위한 침입 패턴 계보 분류에 대한 연구가 필요 할 것이다.

참 고 문 헌

김선숙, “인터넷쇼핑몰 성공의 열쇠”, 21세기사, 2002.

김준한, “일본의 인터넷 이용현황 및 전망”, 정보통신정책, 정보통신정책연구원, 2002. 6.

김재전외 3인, “성공적인 SCM을 위한 공급사슬 파트너쉽의 구조적 관계 모형에 관한 연구”, 「한국정보전략학회지」, 6권 1호, 2003, pp. 36-52.

김학민, “전자무역의 학제적 특성 및 실행체계에 관한 연구”, 「무역학회지」, 제30권 제5호, 한국무역학회, 2005. 10. pp. 114-139.

김태환 외 3인, “한국의 전자무역네트워크 구축사업의 현황과 개선방안에 관한 연구”, 「관세학회지」, 제8권 제1호, 한국관세학회, 2007. pp. 116-145.

문희철외 2인, “우리나라 무역업체의 EDI 도입 및 구현에 관한 실증적 연구”, 「무역학회지」, Vol 20, No 2, 한국무역학회, 1995. pp. 96-124.

변대호, “믿을 수 있는 전자상거래 쇼핑몰”, 진한도서, 2001.

손용엽·이상호, “사이버 시장의 경쟁원리”,시그마인사이트컴, 2001.

산업자원부 · 한국전자거래진흥원, “2004 e-비즈니스 백서”, 2004.

서아영 · 신경식, “공급자-구매자 관계유형에 따른 공급사슬관리 성공요인에 관한 실증연구”, Information System Review, Vol. 3. No. 1, 2001. 11. pp.136-152.

이동만 외 2인, “정보기술목표와 정보기술 기업가치간의 관련성”, 「경영교육논총」, 제24권, 2001. pp.104-138.

에릭 조아킴스탈러外, 현대경제연구원 역편, “브랜드경영”, 21세기북스, 2003.

한국전자통신연구원, “50대 전략품목 기술/시장보고서”, Telematics, 2002.

황수성, “인터넷 전자상거래 세계동향과 대응방안”, 통상산업부 산업표준과, 2001.

Bobby Vandalore, Sonia Fahmy, Raj Jain, Rohit Goyal, Mukul Goyal, " A Definition of General Weighted Fairness and its Support in Explicit Rate Switch Algorithms, "Proceedings of ICNP'98, 1998. pp.204-229.

Christoper M. Bishop, Neural Networks for Pattern Recognition, Oxford Press, 2001. pp.385-433.

Christina Warrender, Stephanie Forrest, Barak Pearlmutter, "Detecting Intrusion Using System Calls : Alternative Data Models", 2004. pp.146-167.

DeLone, W. H. and McLean, E.R., "The DeLone and McLean-Model of Information Systems Success: A Ten-Year Update", Journal of MIS, Vol. 19. No. 4. 2003. pp.49-61.

Eleazar Eskin, “Anomaly Detection over Noisy Data using Learned Probability Distributions”, In Proceedings of the 17 th International Conference on Machine Learning (ICML-2000), 2000. pp.151-172.

Gray, P., "The effects of knowledge management systems on emergent teams : Towards a research model", Journal of Strategic Information Systems, Vol. 9, No. 2-3, 2000, pp.146-161.

Jiawei Han, Micheline Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2001. pp.86-105.

Krubel, Karl, "Benefits and Shortcomings of Business Internet Use: Conclusions from an Empirical Study of German Companies", Proceedings of International Conference on Electronic Commerce '98, April 6-9, 1998. pp.283-302.

Matthew V. Mahoney and Philip K. Chan, "PHAD : Packet Header Anomaly Detection for Identifying Hostile Network Traffic", Florida Institute of Technology Technical Report CS-2001-2004, 2001. pp.163-195.

Matthew V. Mahoney and Philip K. Chan, "Learning Nonstationary Models of Normal Network Traffic for Detecting Novel Attacks", 2003. pp.117-132.

Mehdi Nassehi, "Characterizing Masqueraders for Intrusion Detection", Computer Science/Mathematics, 2001. pp.268-277.

N. Ye, Q. Chen, S. Vibert, "Multivariate Statistical Analysis of Audit Trails for Host-Based Intrusion Detection", IEEE Transactions of computers, Vol. 51, No. 7, 2002. pp.808-828.

S. A. Hofmeyr, A. Somayaji and S. Forrest, "Intrusion Detection using Sequences of System Calls", Journal of Computer Security, Vol.6, 1998. pp.151-180.

Steven Noel, Duminda Wijesekera, Charles Youman, "Modern Intrusion Detection, Data Mining, and Degrees of Attack Guilt", Applications of Data Mining in Computer Security, Daniel Barbara and Sushil Jajodia (eds.), Kluwer Academic Publishers, 2002. pp.302-324.

Susan M. Bridges, Rayford B. Vaughn, “Fuzzy Data Mining And Genetic Algorithms Applied to Intrusion Detection”, 23 rd National Information Systems Security Conference October, 2000. pp.16-19.

Tallon, P.P., Kraemer, K.L. & Gurbaxani, V., "Executive's Perceptions of the Business Value of Information Technology: A Process-Oriented Apporach", Journal of MIS, Vol. 16. No. 4. 2000. pp.116-132.

W. Lee and S. Stolfo, "Learning Patterns from Unix process Execution Traces for Intrusion Detection", AAAI Workshop : AI Approaches to Fraud Detection and RISK management, July, 1997. pp.50-56.