韓國 컴퓨터情報學會 論文誌
第12卷 第1號, 2007. 3.

2007-12-1-2-3

# 세션화 방식을 통한 퍼지기반 네트워크 침입탐지시스템

박 주 기*, 최 은 복**

# A Fuzzy-based Network Intrusion Detection System Through sessionization

Ju-Gi Park*, Eun-Bok Choi**

## 요 약

인터넷의 광범위한 보급에 따라 컴퓨터를 이용한 불법적인 범죄가 증가하고 있고, 이러한 범죄를 막기 위한 정보보호 기술자체가 국가의 경쟁력이 되어 가고 있다. 본 논문에서는 퍼지 논리를 네트워크 침입탐지시스템에 적용하여 보안 전문가와 유사한 결과를 얻을 수 있는 자동화된 퍼지 논리기반의 침입탐지시스템을 제안한다. 프로토콜의 유사성과 시간적인 연속성을 통한 세션화된 패킷분류방식을 통한 퍼지 규칙을 본 시스템에 적용함으로서 다양하고 다변적인 공격패턴으로부터 신속한 침입 판정을 내릴 수 있다. 또한, 대용량의 네트워크 트래픽을 처리해야하는 현재의 네트워크 환경에서, 퍼지추론을 통한 자동화된 트래픽의 프로토콜별/세션별 분석결과를 보여 줌으로써 보안전문가들의 분석 시간과 비용을 절감할 수 있는 장점을 제공한다.

## Abstract

As the Internet is used widely, criminal offense that use computer is increasing, and an information security technology to remove this crime is becoming competitive power of the country. In this paper, we suggest network-based intrusion detection system that use fuzzy expert system. This system can decide quick intrusion decision from attack pattern applying fuzzy rule through the packet classification method that is done similarity of protocol and fixed time interval. Proposed system uses fuzzy logic to detect attack from network traffic, and gets analysis result that is automated through fuzzy reasoning. In present network environment that must handle mass traffic, this system can reduce time and expense of security

▸ Keyword : 침입탐지시스템(IDS), 전문가시스템(Expert System), 퍼지 이론(Fuzzy Theory), 트래픽분석(Traffic Analysis), 트래픽 모델링(Traffic Modeling)

• 제1저자 : 박주기, 교신저자 : 최은복
• 접수일 : 2006.11.28, 심사일 : 2007.1.17, 심사완료일 : 2007. 3.12.
* KT 책임연구원, ** 전주대학교 정보기술공학부 교수

## Ⅰ. 서 론

침입탐지시스템은 호스트기반의 침입탐지시스템과 네트워크기반의 침입탐지시스템으로 나뉜다. 네트워크기반의 침입탐지시스템은 광범위하고 복잡한 네트워크상에 흩어져 있는 많은 데이터를 수집하여 분석에 필요한 데이터만을 추출 또는 축약하기위한 과정이 요구되는데, 이 과정에서 많은 오버헤드가 발생할 뿐만 아니라, 보다 짧은 시간에 불법적인 행위를 검출하기 위해서는 보다 효과적이고 자동화된 분석 시스템이 요구되어진다[1,2].

그러나 트래픽을 통해 공격행위를 구별할 수 있는 완벽한 규칙은 존재하지 않기 때문에 네트워크기반의 침입탐지시스템은 여러 가지 문제점을 가지고 있는데 이중 오탐지(false positive detection) 문제는 가장 많이 지적 받는 부분이다. 오탐지율의 증가는 디스크 용량 점유, 디스크 과다 사용에 따른 성능 저하, 관리적 부담 증대 등의 과부하를 일으키기 때문에 네트워크 기반 침입탐지시스템이 반드시 해결해야할 사항이다.

다음으로 지적되는 부분은 미탐지(miss detection) 사항이다. 네트워크 기반 침입탐지시스템은 패턴(혹은 시그니처)이 없으면 분석 및 복구가 사실상 어렵기 때문에 이 문제를 해결하기 위해 여러 가지 비정상 탐지(anomaly detection) 기법(통계적인 방법, 특징 추출, 예측 가능한 패턴 생성, 신경망 등)을 제품에 적용시켜왔으나 오히려 오탐지율을 높인다는 단점이 발생되곤 하였다. 결국 미탐지 문제가 해결되지 않는다면 네트워크 기반 침입탐지시스템 도입의 장점은 퇴색될 수밖에 없는 셈이다. 마지막으로 자동화(automatic) 부분은 가장 많이 지적되는 점이다. 대용량의 네트워크 트래픽을 처리해야 하는 현재의 네트워크 환경에서 보안전문가의 분석시간과 비용을 줄일 수 있는 자동화된 시스템이 무엇보다 절실하다고 여겨진다[3].

이러한 네트워크기반의 침입탐지시스템이 갖는 여러 문제를 보완하기위한 방법으로 본 논문에서는 세션화방식의 퍼지 논리를 네트워크 침입탐지시스템에 적용하여 보안 전문가와 유사한 결과를 얻을 수 있는 세션화 방식을 통한 퍼지기반 네트워크침입탐지시스템을 제안하고 구현하였다.

본 논문의 구성은 다음과 같다. 2장에서는 침입탐지에 대한 관련 연구를 소개하고, 3장에서는 제안한 시스템의 구성과 각 모듈의 기능 및 특징에 대하여 서술한다. 4장과 5장에서는 제안한 시스템에 대한 실험 및 결과를 분석한다.

## Ⅱ. 관련연구

비정상행위탐지방법은 정상적인 행위에 대한 프로파일을 생성하고 정상행위 프로파일로부터의 편차에 의해 공격으로 판별될 수 있는지를 실험한다. 그러나 정상 파일을 정의하기 위한 탐지비용이 크고, 정상 파일을 유지관리 하는 것이 쉽지 않을 뿐 아니라 데이터베이스의 정확도에 따라 정상행위를 침입으로 분류하는 오탐지(False Positive Detection) 오류를 범하므로 인해 심각한 문제를 일으킬 수 있는 경향이 있다[4,5].

오용탐지방법은 침입으로 알려져 있는 행위, 비정상적인 행위, 또는 시스템의 취약점에 대한 패턴을 정의하고 수집된 감사사건이 미리 정의된 패턴과 일치하는 경우에 이를 침입(또는 오용)으로 판정한다. 오용탐지방법에서는 알려진 공격에 대한 패턴이 저장되면 패턴화된 이후의 공격들을 효율적으로 탐지해낼 수 있다는 장점을 가진다. 그러나 새로이 고안된 공격은 탐지되지 않을 수 있어서 용인될 수 없는 미탐지 에러율을 보일 수 있다.

인공지능 기법들은 오용탐지와 비정상행위탐지 모두에 적용되는데, 감사 자료에서 정상적인 패턴을 찾는 데이터마이닝 기법과 규칙기반 전문가 시스템이 대표적이다[6].

이중 규칙기반 전문가 시스템은 SRI와 같은 IDES (Intrusion Detection Expert System)의 기본이 되고 있으며 알려진 침입의 패턴을 IF-THEN 규칙으로 코드화 시킨다. 하지만 이러한 규칙을 얻어내는 과정은 쉽지 않으며 많은 시행착오를 발생시키므로 패턴을 규칙으로 바꾸는 과정을 자동화하기위한 노력을 하게 되었고, TIM (Time-based-Machine) 또는 신경망을 이용한 IDES가 나오게 되었다.

침입탐지의 목표는 시스템의 움직임에서 두개의 범주 즉 정상적인 작동의 패턴과 비정상적인 클래스에 속하는 이미 알려진 공격패턴을 이용하여 분류해야한다. 이러한 분류의 문제를 위한 해결방법으로 퍼지 규칙을 갖는 퍼지 로직을 이용할 수 있다[7,8].

퍼지 추론은 정량화를 하기 힘든 전문가의 지식을 반영하는데 효과적인 방법으로 알려져 있다. 침입탐지를 위한 다양한 척도들의 결과 값에 퍼지를 적용하여 관련된 정보를 반영한다면 더욱 정확하게 침입을 판정하리라 여겨진다.

침입탐지시스템에 퍼지 로직을 적용함으로서 얻을 수 있는 이점은 다음과 같다.

첫째, 침입탐지에는 많은 양적인 특징을 가지고 있다.

SRI의 차세대침입탐지전문가시스템의 경우는 4가지형태 측정값(서수, 범주, 이진범주 , 선형범주)의 보안과 관련된 통계적인 측정값을 가지는데, 기수(ordinal) 측정값의 예로는 CPU사용시간과 연결 기간(connection duration)을 들 수 있으며, 선형범주(linear Categorical) 측정값의 예로는 동일 소스 호스트에서 시작된 TCP/UDP의 수를 들 수 있다. 이러한 기수와 선형범주 측정값은 퍼지 변수화가 가능한 양적인 특징이며, 또한 퍼지로직은 여러 소스에서 발생한 입력변수들을 쉽게 조합하는 것이 가능하다.

둘째, 주어진 양적인 측정값에 대해서 정상적인 값임을 표시하기 위해 간격(interval)을 사용하는데, 간격에 해당하지 않는 모든 값은 간격에서 떨어진 정도에 관계없이 동일한 예외를 갖는 것으로 간주되며, 간격 내에 존재하는 값에 대해서도 동일하게 적용하여 정상적인 값으로 간주된다. 그러나 이러한 양적인 측정값을 표시하는데 퍼지화를 적용하면 정상과 비정상의 또는 비정상에서 정상으로 변화를 부드럽게 할 수 있다. 본 논문에서는 세션화 방식을 이용한 퍼지 논리를 네트워크 침입탐지시스템에 적용하여 보안 전문가와 유사한 결과를 얻을 수 있는 자동화된 세션화 방식을 통한 퍼지 논리기반의 네트워크 침입탐지전문가 시스템을 제안한다. 제안한 시스템은 퍼지 전문가 시스템에 초점을 두어 구현되었다.

## III. 세션화 방식을 통한 퍼지기반 네트워크 침입탐지 시스템

제안한 시스템은 트래픽 분석기(Traffic Analyzer), 퍼지화 모듈(Fuzzification), 퍼지 추론 엔진(Fuzzy Inference Engine), 지식베이스(Knowledge Base), 비퍼지화 모듈(Defuzzification), 침입탐지분석기의 6가지 컴포넌트로 구성되며, 시스템의 동작과정은 다음과 같다.

단계 1 : 트래픽 분석기는 네트워크 로그파일로부터 네트워크 트래픽을 읽는다. 트래픽 분석기는 네트워크 트래픽을 프로토콜 형태에 의해 3개의 클래스로 분류하고 이것들로부터 입력 변수들의 크리스프 값(crisp value)을 계산한다.

단계 2 : 퍼지화 모듈은 각각의 입력 변수의 크리스프값을 〈그림 1〉의 퍼지 소속 함수를 사용하여 퍼지화 한다. 이때 생성된 퍼지 술어 값(linguistic value)은 다음 단계 3에서 사용된다.

단계 3 : 모든 크리스프값들이 각각의 퍼지 술어 값들로 퍼지화되면, 퍼지 추론 엔진은 퍼지 술어 값과 출력 퍼지 술어 변수의 값을 추론하기 위해 지식베이스를 접근한다. 지식베이스에는 공격에 대한 퍼지규칙들이 저장되어 있다. 보통 각 규칙에서 구해진 적합도 값 중에서 가장 큰 값을 취하여 최종적인 추론결과를 구한다.

단계 4 : 비퍼지화 모듈은 단계 3의 추론결과인 퍼지집합을 비퍼지화시켜 확정치를 구한다.

단계 5 : 단계 4의 출력 값을 사용하여 침입탐지 분석기는 해당 패킷들이 공격 시그니쳐를 포함한 패킷들인지 아니면 정상패킷들인지를 판단한다. 만약 공격 시그니쳐를 포함한다고 판단되면, 침입탐지 분석기는 해당 패킷들로부터 증거를 수집하는 분석 작업을 통하여 적합한 보고서를 생성한다. 그리고 정상패킷들이라고 판단되면 분석 작업을 수행할 필요가 없으므로 단계 1을 다시 수행한다. 네트워크 트래픽이 더 이상 없을 때까지 위 과정을 반복하여 수행한다.

트래픽 분석기는 네트워크 로그파일 또는 실시간 네트워크 장비로부터 트래픽의 수집과 트래픽 분석의 두 가지 주요기능을 수행한다. 수집된 패킷은 각각의 패킷에 대한 분석 후 프로토콜 유사성과 시간적 연속성에 따라 패킷들을 같은 세션으로 분류하여 분석을 수행한다.

세션화 기능은 네트워크 침입탐지시스템에서 침입탐지 판별을 좀 더 효율적으로 하기위해 수행되는데, 대부분의 규칙기반 침입탐지시스템은 네트워크 트래픽에서 UDP와 같은 비연결형 프로토콜의 모든 패킷들을 패킷단위로 모든 규칙들과 비교함으로서 비효율성의 문제점을 갖는다.

본 논문에서 제안한 세션단위로 규칙들과 비교할 수 있는 세션화 방법은 다음과 같다.

패킷들을 같은 세션으로 분류하는 기준은 프로토콜 유사성(TCP, UDP, ICMP패킷 등)과 시간적인 연속성의 2가지 특성을 이용한다. 프로토콜 유사성은 네트워크 트래픽 중에서 패킷의 프로토콜 타입을 분석하여 같은 프로토콜의 패킷들을 같은 세션으로 분류하는 것이며, 시간적 연속성은 정해진 시간동안의 패킷들만 같은 세션으로 분류해 준다. 본 논문에서는 60초 ~ 120초를 사용하였다.

지식베이스는 퍼지 추론 엔진에서 추론이나 새로운 사실을 얻기 위해 사용되는 규칙들을 저장한다. 모든 퍼지논리 기반의 전문가 시스템은 IF-THEN 규칙들을 사용하는데 이 규칙은 일반적으로 다음과 같은 형식으로 구성된다.

IF $X_1 = A_1$ and $X_2 = A_2$ and ... and $X_n = A_n$
,THEN $Y = Z$

단, $X_i$ 와 Y는 퍼지 술어변수들이고, $A_i$ 와 Z 는 퍼지수(linguistic term)이다. 전반부 (IF part)는 전제를 의미하고 후반부 (THEN part)는 결론을 의미한다.

본 시스템의 네트워크 트래픽 분석에 사용될 공격에는 TCP포트 스캔, TCP SYN 플러딩, ICMP smurf , Land, Ping Of Death 의 5가지 공격이 사용되며 퍼지로직 규칙에 적용될 퍼지 술어 변수들은 다음과 같다.

【정의3.1】 Y = Si(Time α, Protocol p) = {$P_1$,...,$P_n$}는 시간 α동안 프로토콜 종류가 p인 패킷들의 집합이고, Protocol = { TCP, UDP, ICMP, IP}이다.

【정의3.2】 $X_1$ = N(Si(α,TCP), S, D, SYN)는 Si(α,TCP)에서 출발지는 S이고 목적지는 D인 TCP패킷으로, 제어(control) 플래그는 SYN을 갖는 패킷의 수를 표시한다.

【정의3.3】 $X_2$ = PN(Si(α,TCP),S,D)는 Si(α,TCP)에서 출발지는 S이고 목적지는 D인 TCP패킷으로 포트들의 수를 표시한다.

【정의3.4】 $X_3$ = N(Si(α,TCP),D,S, SYN/ACK)는 Si(α,TCP)에서 출발지는 S이 고 목적지는 D인 TCP패킷으로 제어(control) 플래그는 SYN/ACK을 갖는 패킷의 수를 표시한다.

【정의3.5】 $X_4$ = N( Si (α, ICMP), ANY, D, ICMP(8,0))는 Si(α, ICMP)에서 목적지가 D인 모든 ICMP패킷으로 ICMP echo request 패킷수

【정의3.6】 $X_5$ =IP(length)+IP(frag_offset)는 패킷의 IP 길이와 프래그멘트 오프셋 (offset)의 합

【정의3.7】 $X_6$ = N(Si(α ,IP),S,S)는 Si(α,IP)에서 출발지와 목적지가 S인 패킷들의 수

〈표 1〉은 퍼지 수, 〈표 2〉는 퍼지규칙들을 보여준다. 퍼지화 모듈은 각각의 퍼지 술어 변수에 대한 퍼지 소속함수를 사용하여 각 퍼지 집합의 크리스프값의 적합도가 결정되는데, 예를 들면, 수치 변수 $X_1$의 값이 30으로 주어진다면 퍼지 소속 함수 $\mu_{A1}(x)$에 의해 퍼지화되어 술어 변수 $X_1$에 대한 적합도는 0.5가된다. (표 1)에 정의된 모든 술어 변수들에 대한 크리스프값들이 퍼지화 모듈에 의해 퍼지화된다.

퍼지 추론 엔진은 모든 크리스프 입력 값들이 술어 값들로 퍼지화되면, 퍼지 추론 엔진은 퍼지 규칙을 사용하여 출력 술어 변수와 출력 술어 변수를 생성하기 위한 중간 변수들의 술어 값을 추론한다.

퍼지 추론 과정의 주요 단계는 결합과 합성이다. 결합은 퍼지 규칙의 전반부에 대한 계산이고 합성은 후반부 퍼지집합에 대한 계산이다. 퍼지 규칙의 전반부가 "$X_1$ is $A_1$ and $X_2$ is $A_2$ and ... and $X_n$ is $A_n$" 인 경우에 전반부 적합도는 가장 작은 값(MIN)을 취하므로 적합도: $min\ \{\ \mu_{A1}(x),\ \mu_{A2}(x),\ \ldots\ ,\ \mu_{An}(x)\ \}$ 가 된다.

퍼지 추론의 추론결과는 각 규칙에서 구해진 적합도 값 중에서 가장 큰 값(MAX)을 취한다.

추론결과로서 크리스프값이 필요한 경우 비퍼지화 모듈을 통해 출력 술어 변수 값들을 크리스프값으로 비퍼지화하며 대부분의 경우 무게중심법(Center of gravitiy)이 많이 사용된다.

본 연구에서는 면적을 계산하는 어려운 점 때문에 계산이 간단한 방법을 사용한다. 앞의 예에서, 출력 변수 Y의 크리스프값은 다음과 같이 계산된다.

$\min\{\max\{0.5, 0.0, 0.0, 0.0, 0.0\}\} = 0.5$

(단, 〈그림 1〉에서 $\mu_{z1}(0.6) = 0.5$).

침입탐지모듈은 출력변수에 대한 크리스프값을 사용하여 해당 세션들이 공격 시그니쳐를 포함하는지 결정하며 해

표 1. 퍼지수와 설명
Table 1. Linguistic term and explanation

| Input Linguistic term | Empirical value | Output Linguistic term | |
|---|---|---|---|
| $A_1$ | > 40 | $Z_1$ | TCP port SCAN Attack |
| $A_2$ | > 40 | $Z_2$ | TCP SYN Flooding Attack |
| $B_1$ | > 1500 | $Z_3$ | ICMP smurf Attack |
| $B_2$ | > 12 | $Z_4$ | Ping Of Death Attack |
| $C_1$ | > 1000 | $Z_5$ | Land Attack |
| $D_1$ | > 1 | | |
| $D_2$ | > 40000 | | |
| $E_1$ | > 1 | | |

표 2. 퍼지 규칙 표현
Table 2. Fuzzy Rule Representation

| Rule Name | Rule Representation | Membership functions |
|---|---|---|
| $R_1$ : TCP port SCAN | *if* $(X_1 = A_1)$ *and* $(X_2 = A_2)$ *then* $(Y = Z_1)$ | $\mu_{A1}(x)$, $\mu_{A2}(x)$, $\mu_{Z1}(x)$ |
| $R_2$ : TCP SYN Flooding | *if* $(X_1 = B_1)$ *and* $(X_3 = B_2)$ *then* $(Y = Z_2)$ | $\mu_{B1}(x)$, $\mu_{B2}(x)$, $\mu_{Z2}(x)$, |
| $R_3$ : ICMP smurf | *if* $(X_4 = C_1)$*then* $(Y = Z_3)$ | $\mu_{C1}(x)$, $\mu_{Z3}(x)$ |
| $R_4$ : Ping Of Death | *if* $(X_4 = D_1)$ *and* $(X_5 = D_2)$ *then* $(Y = Z_4)$ | $\mu_{D1}(x)$, $\mu_{D2}(x)$, $\mu_{Z4}(x)$ |
| $R_5$ : Land | *if* $(X_6 = E_1)$*t hen* $(Y = Z_5)$ | $\mu_{E1}(x)$, $\mu_{Z5}(x)$ |

표 3. 1998년 DARPA데이터의 실험결과 (S:snort, F:제안한 시스템)
Table 3. Experiment result of DARPA data(1998) (S:snort, F: Proposed system)

| | Tcp 포트스캔 | | Tcp Syn Flooding | | ICMP Smurf | | Ping of Death | | Land Attack | |
|---|---|---|---|---|---|---|---|---|---|---|
| | S | F | S | F | S | F | S | F | S | F |
| 실제공격 | 5 | 5 | 41 | 41 | 5 | 5 | 5 | 5 | 9 | 9 |
| 탐지 | 5 | 5 | 32 | 41 | 0 | 5 | 0 | 4 | 4 | 6 |
| 잘못된 탐지 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 0 |
| 탐지못함 | 0 | 0 | 9 | 0 | 5 | 0 | 5 | 1 | 5 | 3 |

표 4. 타임스탬프(120초) 패킷분석표(TCP SYN Flooding)
Table 4. Packet analysis table of timestamp(120 second) (TCP SYN Flooding)

| 일자 | 타임 스탬프 | 소스 IP주소 | 목적지 IP주소 | 패킷종류 | 패킷수 | 최초 패킷 발생시간 | 마지막 패킷 발생시간 | 소속 함수값($\mu$B1, $\mu$B2) | $\mu$Z2 | 비퍼지수 |
|---|---|---|---|---|---|---|---|---|---|---|
| 99/3/12 | 32 | 209.117.157.183 | 172.16.114.207 | Syn 패킷수 | 5373 | 99-03-12 01:04:12 | 99-03-12 01:05:59 | 1 | 1 | 1 |
| | | | | Ack/Syn 패킷수 | 161 | 99-03-12 01:04:16 | 99-03-12 01:05:55 | 1 | | |
| 99/3/13 | 40 | 204.97.153.43 | 172.16.114.50 | Syn 패킷수 | 5402 | 99-03-13 01:20:11 | 99-03-13 01:21:59 | 1 | 1 | 1 |
| | | | | Ack/Syn 패킷수 | 141 | 99-03-13 01:20:15 | 99-03-13 01:21:54 | 1 | | |

표 5. 타임스탬프(90초) 패킷분석표(TCP SYN Flooding)
Table 5. Packet analysis table of timestamp(90 second) (TCP SYN Flooding)

| 일자 | 타임 스탬프 | 소스 IP주소 | 목적지 IP 주소 | 패킷종류 | 패킷수 | 최초 패킷 발생시간 | 마지막패킷발생시간 | 소속 함수값($\mu$B1, $\mu$B2) | $\mu$Z2 | 비퍼지수 |
|---|---|---|---|---|---|---|---|---|---|---|
| 99/3/12 | 42 | 209.117.157.183 | 172.16.114.207 | Syn 패킷수 | 872 | 99-03-12 01:04:12 | 99-03-12 01:04:29 | 0.465066667 | 0.46507 | 0.1592 |
| | | | | Ack/Syn 패킷수 | 80 | 99-03-12 01:04:16 | 99-03-12 01:04:28 | 0.9 | | |
| 99/3/13 | 43 | 209.117.157.183 | 172.16.114.207 | Syn 패킷수 | 4501 | 99-03-12 01:04:30 | 99-03-12 01:05:59 | 1 | 1 | 1 |
| | | | | Ack/Syn 패킷수 | 81 | 99-03-12 01:04:32 | 99-03-12 01:05:55 | 1 | | |
| 99/3/13 | 53 | 204.97.153.43 | 172.16.114.50 | Syn 패킷수 | 2401 | 99-03-13 01:20:11 | 99-03-13 01:20:59 | 0.9802 | 0.9802 | 1 |
| | | | | Ack/Syn 패킷수 | 111 | 99-03-13 01:20:15 | 99-03-13 01:20:40 | 1 | | |
| 99/3/13 | 54 | 204.97.153.43 | 172.16.114.50 | Syn 패킷수 | 4501 | 99-03-13 01:21:00 | 99-03-13 01:22:29 | 1 | 0.4 | 0 |
| | | | | Ack/Syn 패킷수 | 30 | 99-03-13 01:21:54 | 99-03-13 01:21:54 | 0.4 | | |

당세션들로부터 관련정보를 자동화된 방법으로 수집하여 경보 메시지를 생성한다. 그리고 정상패킷이라고 판단되면 분석 작업을 수행할 필요가 없으므로 트래픽 분석기의 수행을 재개한다.

본 논문에서는 출력 크리스프값이 0.9~1.0 사이면 해당 세션이 공격 시그니쳐를 포함한다고 판단하며, 해당 세션들로부터 관련정보를 자동화된 방법으로 수집하여 증거를 생성한다.

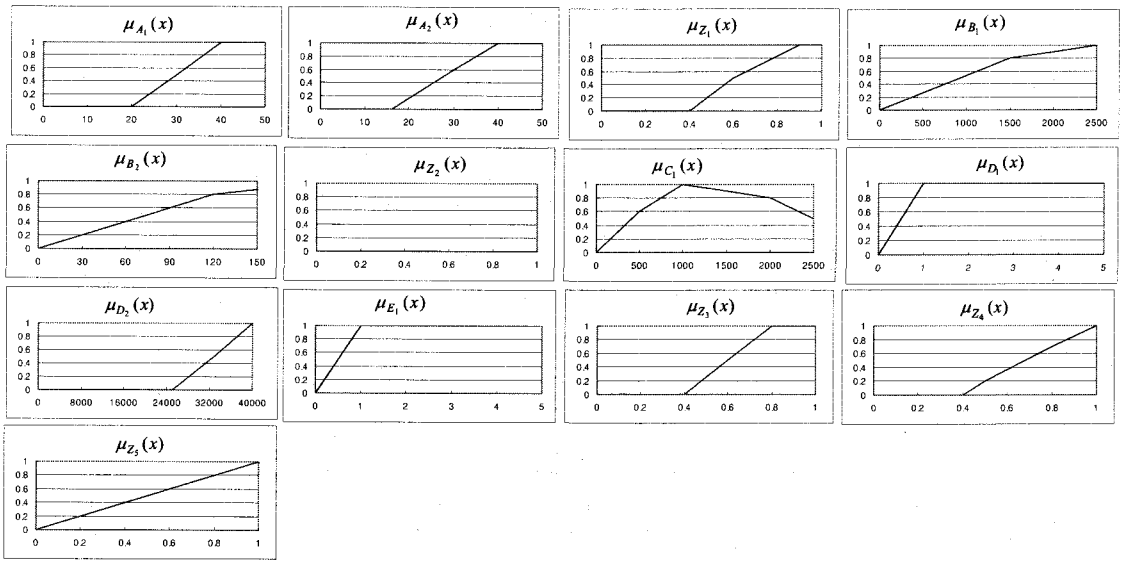


그림 1. 표 2의 퍼지 규칙에 따른 소속 함수들
Fig 1. Membership functions of table 2 by fuzzy rules

## IV. 실험 및 분석

본 논문에서는 1998년과 1999년 DARPA침입탐지 평가데이터[11,12]를 사용하여 제안한 시스템의 성능평가를 실시하였다. 1998년 DARPA데이터의 2주차, 4주차, 및 5주차 데이터를 제안한 시스템과 네트워크 침입탐지시스템으로 널리 알려진 snort를 사용하여 실험한 결과를 비교분석하고, 1999년 DARPA데이터의 2주차 데이터를 이용하여 제안한 침입탐지시스템의 성능평가를 위해 5가지 유형의 공격에 대해서 60초, 90초, 120초로 타임스탬프를 나누어 성능을 분석하였다.

〈표 3〉은 1998년 DARPA데이터의 2주차, 4주차, 및 5주차 데이터를 제안한 시스템과 네트워크 침입탐지시스템으로 널리 알려진 snort를 사용하여 실험한 결과이다. 실험결과에 대하여 분석의 정확도를 측정하가위하여 다음의 식 3.1을 사용하여 정확도를 측정하였다.

정확도 = 올바르게 분류된 공격의 수 /전체공격의 수 ····· (3.1)

식 3.1에 의해 제안한 시스템의 정확도는 93.8%(정확도 = 61/65)이고, snort의 정확도는 63.07%(정확도 =41/65)이다. 제안한 시스템은 Ping 공격(Ping of Death)와 ICMP 스머프 공격에 대해서 snort보다 월등한 성능을 보였다.

1999년 DARPA데이터의 2주차 데이터를 이용하여 제안한 침입탐지시스템의 성능평가를 위해 4가지 유형의 공격에 대해서 60초, 90초, 120초로 타임스탬프를 나누어 성능을 분석한 결과는 다음과 같다.

### 4.1 TCP SYN Flooding

120초를 타임스탬프로 이용한 경우 TCP SYN Flooding으로 판정하는 경우는 2개(209.117.157.183에서 172.16.114.207, 204.97.153.43에서 172.16.114.50)이다. 90초와 60초로 타임스태프를 나누어 분석한 결과와 비교시에도 동일한 결과가 나오지만, 120초 타임스탬프에 비해 짧은 시간간격으로 인해 패킷수가 작아지며, 퍼지함수에서 나오는 값도 작아진다. 120초 타임스탬프의 분석에 의하면 공격은 1999년 3월 12일 01시04분 12초에 시작으로 표시 되지만 90초 타임스탬프를 이용하는 경우에는 1999년 3월 12일 01시04분 30초에 시작한 것으로 표시한다. 그 이유는 90초 타임스탬프를 이용하

표 6. 타임스탬프(60초) 패킷분석표(TCP SYN Flooding)
Table 6. packet analysis table of timestamp(60 second) (TCP SYN Flooding)

| 일자 | 타임 스탬프 | 소스IP주소 | 목적지IP 주소 | 패킷종류 | 패킷수 | 최초패킷발생시간 | 마지막패킷발생시간 | 소속 함수값($\mu$B1, $\mu$B2) | $\mu$Z2 | 비퍼지수 |
|---|---|---|---|---|---|---|---|---|---|---|
| 99/3/12 | 64 | 209.117.157 .183 | 172.16.114 .207 | Syn 패킷수 | 2372 | 99-03-12 01:04:12 | 99-03-12 01:04:59 | 0.9744 | 0.9744 | 1 |
| | | | | Ack/Syn 패킷수 | 131 | 99-03-12 01:04:16 | 99-03-12 01:04:41 | 1 | | |
| 99/3/12 | 65 | 209.117.157 .183 | 172.16.114 .207 | Syn 패킷수 | 3001 | 99-03-12 01:05:00 | 99-03-12 01:05:59 | 1 | 0.4 | 0 |
| | | | | Ack/Syn 패킷수 | 30 | 99-03-12 01:05:54 | 99-03-12 01:05:55 | 0.4 | | |
| 99/3/13 | 80 | 204.97.153. 43 | 172.16.114 .50 | Syn 패킷수 | 2401 | 99-03-13 01:20:11 | 99-03-13 01:20:59 | 0.9802 | 0.9802 | 1 |
| | | | | Ack/Syn 패킷수 | 111 | 99-03-13 01:20:15 | 99-03-13 01:20:40 | 1 | | |
| 99/3/13 | 81 | 204.97.153. 43 | 172.16.114 .50 | Syn 패킷수 | 3001 | 99-03-13 01:21:00 | 99-03-13 01:21:59 | 1 | 0.4 | 0 |
| | | | | Ack/Syn 패킷수 | 30 | 99-03-13 01:21:54 | 99-03-13 01:21:54 | 0.4 | | |

표 7. 타임스탬프(120초) 패킷분석표(ICMP smurf)
Table 7. packet analysis table of timestamp(120 second) (ICMP smurf)

| 일자 | 타임스탬프 | 목적지 IP주소 | ICMP(8.0) 패킷수 | 최초패킷 발생시간 | 마지막패킷 발생시간 | 소속함수값($\mu$C1) | 비퍼지수 |
|---|---|---|---|---|---|---|---|
| 99/3/13 | 216 | 172.16.112.50 | 403 | 99-03-13 07:13:02 | 99-03-13 07:13:59 | 0.4836 | 0.209 |
| 99/3/13 | 217 | 172.16.112.50 | 850 | 99-03-13 07:14:00 | 99-03-13 07:15:59 | 0.88 | 1 |
| · | · | · | · | · | · | · | · |
| 99/3/13 | 221 | 172.16.112.50 | 810 | 99-03-13 07:22:00 | 99-03-13 07:23:59 | 0.848 | 1 |
| 99/3/13 | 222 | 172.16.112.50 | 443 | 99-03-13 07:24:00 | 99-03-13 07:25:06 | 0.5316 | 0.329 |

표 8. 타임스탬프(90초) 패킷분석표(ICMP smurf )
Table 8. packet analysis table of timestamp(90 second) (ICMP smurf)

| 일자 | 타임스탬프 | 목적지 IP주소 | ICMP(8.0) 패킷수 | 최초패킷 발생시간 | 마지막패킷 발생시간 | 소속 함수값($\mu$ C1) | 비퍼지수 |
|---|---|---|---|---|---|---|---|
| 99/3/13 | 289 | 172.16.112.50 | 637 | 99-03-13 07:13:30 | 99-03-13 07:14:59 | 0.7096 | 0.774 |
| 99/3/13 | 290 | 172.16.112.50 | 637 | 99-03-13 07:15:00 | 99-03-13 07:16:29 | 0.7096 | 0.774 |
| · | · | · | · | · | · | · | · |
| 99/3/13 | 295 | 172.16.112.50 | 606 | 99-03-13 07:22:30 | 99-03-13 07:23:59 | 0.6848 | 0.712 |
| 99/3/13 | 296 | 172.16.112.50 | 443 | 99-03-13 07:24:00 | 99-03-13 07:25:06 | 0.5316 | 0.329 |

는 경우 01시 04분 12초에서부터 01시 04분 28초 동안 발생한 패킷들에 대해서는 μZ2 값이 0.4651이 되어 공격으로 판정하지 못한다. 이러한 경우를 위해 타임스탬프를 줄이는 경우에는 퍼지함수를 바꾸어 공격으로 판정하도록 할수 있다. 그러나 이 경우에는 공격이 아닌 경우를 공격으로 판정하는 과탐의 우려가 있다.

### 4.2 ICMP Smurf

120초를 타임스탬프로 이용한 경우 ICMP smurf로 판정하는 경우는 1개이며, 목적지주소 172.16.112.50이다. 90초와 60초로 타임스태프를 나누어 분석한 결과와 비교 시에는 90초와 60초로 타임스탬프를 작게 주면 ICMP smurf검출 판정을 하지 못한다. 타임스탬프가 작아짐에 따라서 패킷수가 작아지고 소속 함수 값이 작아진다. 따라서 타임스탬프를 작게 주는 경우는 별도의 소속 함수 값을 정의하여야 하며, 120초와 비교했을 때 더 작은 패킷수에 대해서 ICMP smurf를 판정할 수 있도록 하여야 한다.

### 4.3 Land Attack

120초를 타임스탬프로 이용한 경우 Land Attack으로 판정하는 경우는2개이며, IP주소는 172.16.112.50이다. 연속된 타임스태프가 아니므로 2개의 Land Attack으로 판정하게 된다. 90초와 60초로 타임스태프를 나누어 분석한 결과와 비교 시에도 동일한 결과가 나온다.

### 4.4 TCP 포트 스캔

본 논문에서 제안한 퍼지시스템을 이용하여, 120초로 타

표 9. 타임스탬프(60초) 패킷분석표(ICMP smurf)
Table 9. packet analysis table of timestamp(60 second) (ICMP smurf)

| 일자 | 타임스탬프 | 목적지 IP주소 | ICMP(8.0) 패킷수 | 최초패킷 발생시간 | 마지막패킷 발생시간 | 소속 함수값(μC1) | 비퍼지수 |
|---|---|---|---|---|---|---|---|
| 99/3/13 | 433 | 172.16.112.50 | 403 | 99-03-13 07:13:02 | 99-03-13 07:13:59 | 0.4836 | 0.209 |
| 99/3/13 | 434 | 172.16.112.50 | 425 | 99-03-13 07:14:00 | 99-03-13 07:14:59 | 0.51 | 0.275 |
| . | . | . | . | . | . | . | . |
| 99/3/13 | 443 | 172.16.112.50 | 404 | 99-03-13 07:23:00 | 99-03-13 07:23:59 | 0.4848 | 0.212 |
| 99/3/13 | 444 | 172.16.112.50 | 402 | 99-03-13 07:24:00 | 99-03-13 07:24:59 | 0.4824 | 0.206 |

표 10. 타임스탬프(120초) 패킷분석표(Land Attack)
Table 10. packet analysis table of time stamp(120 second) (Land Attack)

| 일자 | 타임스탬프 | 소스 IP주소 | 목적지 IP주소 | 최초패킷 발생시간 | 마지막 패킷 발생시간 | 소속 함수값(μE1) | 비퍼지수 |
|---|---|---|---|---|---|---|---|
| 99/3/9 | 178 | 172.16.112.50 | 172.16.112.50 | 99-03-09 05:57:07 | 99-03-09 05:57:07 | 1 | 1 |
| 99/3/12 | 173 | 172.16.112.50 | 172.16.112.50 | 99-03-12 05:47:07 | 99-03-12 05:47:07 | 1 | 1 |

임스탬프로 나누어 1999년 DARPA데이터의 2주차를 분석하면 다음과 같이 침입이 탐지된다. 120초를 타임스탬프로 이용한 경우 TCP포트스캔으로 판정하는 경우는 22개이지만 이중 연속된 타임스탬프는 1개의 TCP포트스캔으로 판정해야하므로 총 6개의 TCP 포트스캔으로 판정한다. 이 과정에서 내부 통신망에서의 스캔과 외부 통신망에서의 포트스캔으로 분류할 수 있다. 90초를 타임스탬프로 이용한 경우 TCP포트스캔으로 판정하는 경우는30개이며, 120초를 타임스탬프로 하는 것에 비해 8개를 더 TCP 포트스캔으로 판정하지만 이중 연속된 타임스탬프는 1개의 TCP포트스캔으로 판정해야 하므로 120초를 타임스탬프로 이용하는 것과 동일한 결과를 얻는다. 60초를 타임스탬프로 이용한 경우 TCP포트스캔으로 판정하는 경우는40개이며, 120초를 타임스탬프로 하는 것에 비해 18개를 더 TCP 포트스캔으로 판정 하지만 이중 연속된 타임스탬프는 1개의 TCP포트스캔으로 판정해야 하므로 총 6개의 TCP 포트스캔으로 판정하며, 120초를 타임스탬프로 이용하는 것과 동일한 결과를 얻는다.

## V. 결론 및 향후 연구과제

1998년과 1999년 DARPA 데이터를 이용한 실험에서 시스템은 93.8%의 정확도 성능을 보였으며, 트래픽의 프로토콜별/세션별 분석결과를 제공하고 있다. 현재와 같은 빠른 인터넷의 확산으로 인해 대용량의 트래픽이 발생하는 상황에서 침입에 대한 빠르고 신속한 대응이 필수적이며, 이와 같은 상황에서 본 시스템이 대용량의 네트워크 트래픽을 처리해야하는 현재의 네트워크 환경에서 트래픽의 프로토콜별/세션별 분석결과를 보여 줌으로써 보안전문가들의 분석 시간과 비용을 절감할 수 있도록 해준다.

향후 연구 과제로는 본 논문에서 사용된 공격외에 다른 형태의 공격에 대해서 퍼지논리를 추가하고 기존 퍼지함수에 대한 지속적인 업데이트가 필요하다. 또한 트래픽 모델링과 퍼지논리에 쓰이는 소속함수의 연관성을 이용하여 시스템의 성능을 향상 시키는 부분과 특정 네트워크나 링크의 트래픽 특성의 분석이나 트래픽 모델링을 통해 감사자료의 생성에 필요한 트래픽 분석 시간을 줄이고, 세션의 길이나 퍼지함수의 값을 수정함으로써 오탐지율과 미탐지율을 줄일수 있도록 연구가 필요하다.

## 참고문헌

[1] Rebecca Gurley Bace, "Intrusion Detection," published by Macmillan Technical Publishing.
[2] Anonymous, "Maximum Security : A Hacker's Guide to Protecting Your Internet Site and Network -2e," published by Sams.
[3] Edward G. Amoroso, "Intrusion Detection : An

Introduction to internet Surveillance, Correlation, Traps, Trace Back, and Response," published by Intrusion.net books.

[4] Paul E. Proctor, "The practical Intrusion Detection handbook," published by Prentice Hall PTR.

[5] David J. Marchette, "Computer Intrusion Detection and Network Monitoring A Statistical Viewpoint," published by Springer

[6] Wenke Lee, Salvatore J. Stolfo, K, W. Mok, "Mining in a Data-flow Environment: Experience in Network Intrusion Detection", In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp.114-124 1999.

[7] 김상찬, 김용민, 김민수, 노봉남, "퍼지제어를이용한 다중탐지센서의 통합탐지방법", 한국정보과학회 2002 가을 학술발표논문집, 2002.

[8] David Moore, Geoffrey M. Voelker, et al, "Inferring Internet denial of service activity", ACM Transaction on Computer System, vol.24 No.2 pp.115-139, 2006.

[9] H. S. Vaccaro and G. E. Liepins, "Detection of anomalous computer session activity", Proc. of the 1989 IEEE Symposium on Research in Security and Privacy, pp.280-289, 1989.

[10] Cheri Dowell and Paul Ramstedt, "The Computer Watch data reduction tool", Proc. of the 13th National Computer Security Conference, pp.99-108, 1990.

[11] TCPDUMP public repository, http://www.tcpdump.org/

[12] 1999 Darpa Intrusion Detection Evaluation, http://www.ll.mit.edu/IST/ideval/docs/docs_index.html

## 저 자 소 개

**박 주 기**

1993년 전남대학교 대학원 전산학과 졸업(이학석사)

1993~현재 KT 책임연구원

관심분야 : 인터넷 보안, 인터넷트래픽 분석 및 모델링

**최 은 복**

2000년 전남대학교 전산학과 졸업(이학박사)

2002년- 현재 전주대학교 정보기술공학부 조교수

관심분야 : 통신망관리, 네트워크보안 etc.