퍼지제어를 이용한 다중 탐지센서의 통합탐지 방법

김상찬[0] 김용민 김민수 노봉남
전남대학교 정보보호 협동과정
{netzard[0], phoenix}@athena.chonnam.ac.kr {ymkim, bongnam}@chonnam.ac.kr

An Integrating Detection Method among Multi-sensors using Fuzzy Control

Sang-Chan Kim[0] Yong-Min Kim Min-Soo Kim Bong-Nam Noh
Dept. of Information Security, Chonnam National University

요 약

호스트나 네트워크에 여러 개의 탐지센서가 설치된 경우, 각 탐지센서들은 고유의 탐지영역과 타 센서들과의 중첩되는 탐지영역에 대한 각각의 탐지정도를 결합하여 각 센서의 최종 결과값으로 제시한다. 이러한 경우 여러 센서들이 동일한 판정의 결과를 제시하지 않고 판정이 모호한 결과를 갖는 경우 각 탐지센서 판정결과의 조율은 불가피하다. 본 논문에서는 이러한 다중 탐지센서들의 모호한 판정 값들에 대해서 퍼지제어를 이용한 통합된 형태의 탐지 판정값을 추론하는 방법을 제안한다.

1. 서 론

최근 정보보호의 중요성에 대한 인식이 높아지면서 일반 대중과 기업의 정보보호에 대한 관심이 점차 증대되고 있다. 더불어 정보보호를 위한 방안으로 기업차원에서 침입탐지시스템의 보급이 점차 증가하고 있는 추세이다. 이러한 침입탐지 시스템에 대한 연구는 크게 두 방향으로 나누어져 진행되고 있다. 그 하나는 규칙 기반에 의한 오용행위 탐지방법이고 다른 하나는 통계나 모델에 기반을 둔 이상행위 탐지방법이다.

현재는 알려지지 않은 공격(unknown attacks)을 탐지하려는 시도가 대학과 일부 보안관련 업체를 중심으로 활발히 진행 중인데 이러한 탐지방법의 경우 특정 탐지영역에 뛰어난 성능을 보이는 여러 개의 센서를 결합하여 전체적인 탐지성능을 높이는 것이 일반적인 방법이다.

하지만 이러한 경우 각 센서는 침입탐지를 위한 알고리즘과 침입의 특성에 대한 반응정도에 따라 서로 상이한 탐지결과를 나타내게 됨으로써 일관된 판정의 결과를 얻기가 어려워진다. 따라서 다중센서들이 서로 상이한 판정의 결과를 나타내는 경우 이를 통합탐지하기 위한 방법으로 퍼지제어의 사용을 제안한다.

본 논문의 구성은 2장에서 통합탐지 및 적용하는 퍼지제어에 대해 소개하고, 3장에서는 이 논문에서 제안하는 퍼지제어를 이용한 센서간 통합탐지 모델에 대해서 설명한다. 그리고 4장에서는 결론 및 향후 과제에 대해 기술한다.

2. 관련연구

2.1 기존의 통합탐지 방법

기존의 다중탐지센서 결과의 통합에 대한 연구는 크게 두 가지 목적을 가지고 진행되었다. 첫 번째는 이종의 각 센서에서 생성하는 대량의 로그들을 축소하려는 것이고 두 번째는 서로 연관된 사건들의 연관정도를 계산하고 이를 반영하여 오탐률(False Alarm Rate)을 줄임으로써 판정의 정확성을 높이기 위함이다.

통합탐지에 대한 연구는 DARPA에서 통계적인 방법으로 센서간의 연관정도를 계산하고 이를 반영해 로그를 통합하려는 연구를 진행하였다[1,2]. 통계적인 방법 외에도 퍼지를 적용하려는 시도도 이루어졌다[3]. 하지만 여전히 퍼지에 대한 활용은 기계제어분야[4]나 센서 수준의 탐지[5]에 활발히 응용되고 통합탐지를 위한 응용은 미비한 상태이다.

2.2 퍼지제어 시스템

퍼지에 관한 연구는 1965년 Zadeh 교수가 퍼지집합 이론을 제창한 이후 여러 분야에서 넓게 활용되고 있는데 그 활용범위가 가장 넓고 성공적인 분야가 퍼지제어 분야이다.

퍼지제어 시스템은 퍼지화부, 지식베이스, 추론(의사결정)부 그리고 비퍼지화부 등 크게 4부분으로 구분할 수 있다[6].

퍼지화부는 하나의 명확한 값(crisp value)으로 들어오는 입력에 대해서 적절한 퍼지값으로 변경하는 역할을 수행하게 되는데 만일 입력값에 대한 신뢰도가 보장된다면 입력값을 퍼지 단일값으로 변화시킴으로써 추론을 단순화 할 수 있다.

지식베이스는 크게 두 부분으로 나뉘어진다. 한 부분은 입력값에 대해서 제어 시스템의 제어특성을 반영할 수 있는 적절한 퍼지분할 방법과 각 퍼지분할에 대해서 소속

함수를 정의해 주는 부분이고 다른 부분은 " if-then" 형식의 언어적 형식으로 표현되는 규칙들의 저장소이다. 이러한 제어 규칙은 입출력 변수가 결정된 후, 제어시스템의 제어특성에 따라 제어규칙을 생성할 수 있다.

추론부는 퍼지제어기에서 언어적인 형태로 기술된 퍼지 제어규칙을 적용하기 위한 논리적인 실행 부분이다. 퍼지규칙으로부터의 논리적인 추론이 가능한 이유는 " if-then" 형식의 퍼지 제어규칙의 조건명제(①)는 퍼지관계(②)로 나타내어질 수 있으며 이로써 논리적인 추론이 가능하다.

① R_i : If x is A_i and y is B_i then z is C_i

② $\mu_{Ri} = \mu_{(Ai \text{ and } Bi \rightarrow Ci)}(U, V, W)$

$= [\mu_{Ai}(U) \text{ and } \mu_{Bi}(V)] \rightarrow \mu_{Ci}(W)$

x, y는 입력변수, z는 출력변수

U, V, W는 x, y, z의 전체집합

A_i, B_i, C_i는 U, V, W에서 정의된

x, y, z의 퍼지값(퍼지집합)

추론의 방법은 일반적인 추론이 퍼지로 확장된 1차적인 연역추론기로 동작하는 GMP(Generalized Modus Phonens)와 역방향 추론구조를 갖는 GMT(Generalized Modus Tollens)가 많이 사용되며 추론을 단순화 하기 위해서 몇 가지 성질을 이용하기도 한다[5]. 실제적인 추론 과정에서는 적합도라는 개념을 사용하는데 적합도(α_i)란 i번째 규칙이 제어에 관여하는 정도를 나타내는 값이다.

비퍼지화부는 퍼지제어의 추론결과로 나타나는 판정을 위한 퍼지집합에서 명확한 결과 값을 만들어내는 부분이다.

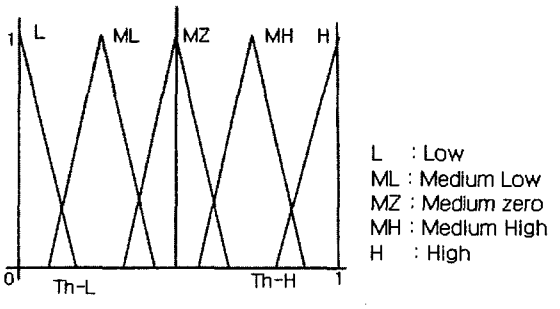
3. 퍼지 제어 통합탐지의 방법

3.1 통합탐지의 제어규칙

다중 탐지센서에서는 고유의 탐지영역을 갖는 각각의 탐지센서 결과가 추론을 이끌어낼 수 있는 근거 데이터가 되므로 제어 입력변수로 사용되고, 추론을 거쳐서 나온 퍼지집합 형태의 결과가 결론부 변수로 사용된다. 즉, 3개의 센서가 설치된 경우 S_a, S_b, S_c를 입력변수로 두고 이들의 추론결과인 Z를 출력부로 두게된다.

다중 센서의 입력변수 값은 통상적으로 각 센서가 탐지의 정도를 [0,1]사이의 폐구간 사이에서 표현하고 그 사이에서 최대값으로 1의 값을 갖을 수 있으므로 삼각숫자 개념의 도입이 가능하다. 즉, 입력변수의 전체집합을 퍼지분할 하고 각 분할에 대해서 적절한 퍼지소속함수를 부여하는 것이다.

다중 센서에서의 제어대상은 애매모호한 중간값들에 대한 제어이므로 퍼지분할 시 제어의 대상에 포함되지 않는 공격범주와 정상범주를 상한값과 하한값 형태로 지정해 두고, 실제 탐지 대상인 애매모호한 범주는 세분화해서 탐지의 정확성을 높일 수 있다.



(그림 1) 퍼지 소속함수

그림1은 공격범주, 정상범주, 모호한 범주의 퍼지분할에 의해 작성된 퍼지집합이 {L, ML, MZ, MH, H}인 소속함수이다. 여기서 각 센서의 탐지특성은 각기 다를 수 있으므로 범주의 크기도 달라 질 수 있다. 그림1 에서 정상 판정값인 Th-L(Threshold-Low)과 공격 판정값인 Th-H(Threshold-High)는 퍼지제어 시에는 포함되지 않지만 판정의 중요한 요소이다. 이 값은 센서의 신뢰성과 깊은 관련이 있는데 여러 번의 실험을 통해서 통계적인 절차를 거친 후 최적의 신뢰도를 적용해서 결정하게 된다.

퍼지 제어규칙은 침입탐지와 탐지센서의 특성에 의해 아래와 같은 형식의 규칙을 만들게 된다.

R_1 : IF (S_a is MH) and (S_b is MZ) and (S_c is MH) Then Z is MH

R_2 : IF (S_a is MZ) and (S_b is MH) and (S_c is MH) Then Z is H

R_3: IF (S_a is ML) and (S_b is ML) and (S_c is ML) Then Z is L

...

R_{20} : IF (S_a is MZ) and (S_b is MZ) and (S_c is MH) Then Z is MZ

R_1과 R_2는 MH가 2개 MZ가 1개로 서로 같지만 최종결과가 MH와 H로 서로 다름을 알 수 있다. 이것은 다중 탐지센서의 최종결과가, 센서가 나타내는 결과가 속하는 소속함수의 개수가 아닌 추론을 거친 결과에 의존함을 나타낸다.

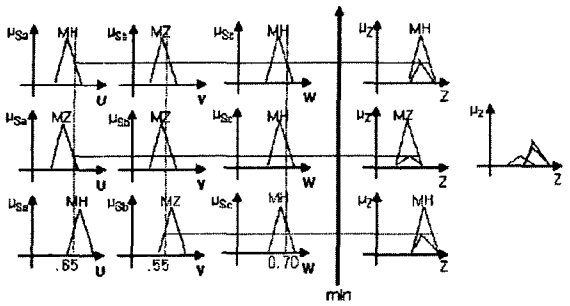
퍼지 제어규칙의 개수는 입력변수의 개수와 퍼지집합의 개수에 따라서 결정되는데 센서가 3개이고 퍼지집합이 3개인 경우 총 $27(3^3)$개가 가능하지만 R_3처럼 경험상 제외되는 규칙이 존재해 27개 모든 규칙이 필요한 것은 아니다.

3.2 통합판정의 추론

제어규칙이 생성되고 나면 센서를 통해 탐지된 사실(Facts)과 지식베이스의 제어규칙 간의 퍼지 연산방식을 적용해 새로운 결과를 도출해내는 추론이 이루어지게 된다.

다중 센서에 대한 추론의 경우 각 센서의 모호한 결과값을 기본 데이타로 1차적인 추론의 과정을 거쳐 최종 추

론 결과를 내어놓으므로 일반적인 연역추론인 GMP를 사용한 추론을 사용 한다. 추론시의 연산방식은 각 센서들의 입력값이 통합판정에 잘 전달될 수 있도록 GMP의 Larsen의 product 연산방식을 사용한다. 그림 2는 S_a가 0.65, S_b가 0.55 S_c가 0.7을 나타내었을 때의 적합도를 계산한 후 최종 퍼지집합을 추론하는 과정을 도시한 것이다.

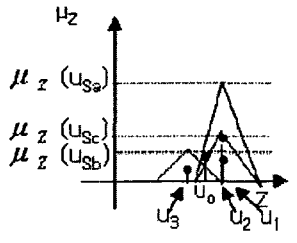


(그림 2) 추론의 과정

그림 2에서 좌측으로부터의 3개의 그래프는 각 센서별 퍼지정도의 표현이고 그 우측의 그래프는 입력값이 단일값인 경우 추론의 단순화를 위한 Larsen의 보조정리3[7]을 이용하여 적합도가 계산된 그림이며 맨 우측의 그림이 최종적인 추론의 결과인 퍼지집합을 표현한 그림이다.

3.3 통합판정의 비퍼지화

퍼지제어기를 사용한 추론의 결과는 센서의 전체 퍼지집합 안에서 정의된 퍼지집합이 된다. 따라서 최종 판정을 위해서는 이를 다시 비퍼지화 할 필요가 있다. 비퍼지화의 방법에는 이미 여러가지 방법들이 나와있는데 각 센서들의 상호 관련성을 잘 반영할 수 있는 무게중심법(center of gravity method)을 사용해 비퍼지화 작업을 수행한다.



$$U_O = \frac{\sum_{i=1}^{n} \mu z(ui) \bullet ui}{\sum_{i=1}^{n} \mu z(ui)}$$

(그림 3) 비퍼지화 작업

그림 3에서 $\mu_Z(u_i)$는 i번째 센서가 추론을 마친 다음 생성된 퍼지집합에서의 최대값을 나타내고, u_1, u_2, u_3은 각각 추론 후 생성된 퍼지집합(소속함수)에 대한 무게중심에서 x-축으로의 사영(Projection)을 한 값이다. 결과적으로 세 개의 센서에 대한 비퍼지화 값은 다음처럼 계산된다.

$$u_0 = \frac{[\mu_Z(u_{Sa}) \bullet u_1 + \mu_Z(u_{Sb}) \bullet u_3 + \mu_Z(u_{Sc}) \bullet u_2]}{[\mu_Z(u_{Sa}) + \mu_Z(u_{Sb}) + \mu_Z(u_{Sa})]}$$

4. 결론 및 향후 연구

본 논문에서는 다중탐지센서들이 모호한 값을 나타내는 경우, 이를 통합하고 최종 판정하기 위한 방법으로 전문가에 의해 생성된 제어규칙과 입력된 제어변수에 대해서 추론과정을 거침으로써 판정 퍼지함수를 생성하고 최종적인 판정을 위해서 비퍼지화 과정을 수행했다. 이러한 퍼지제어를 이용한 통합탐지는 제어규칙의 수가 증가함에 따라 연산시간이 늘어나는 문제가 있지만, 이 논문의 제안처럼 전문가의 도움을 얻어 적절한 제어규칙의 수를 유지한다면 기존의 방법에 비해서 빠른 처리시간과 높은 탐지율을 나타낼 수 있다.

이 논문에서는 전문가에 의해 처음 생성된 규칙들을 계속적으로 사용하지만 향후에는 동적으로 변화되는 상황을 반영할 수 있는 동적인 규칙의 추가가 가능하도록 연구가 진행되어야 한다.

향후 연구에서는 제어규칙의 동적인 추가가 가능한 알고리즘의 개발과 함께 이 논문에서 제안한 통합판정을 위한 퍼지제어 시스템을 구현할 계획이다.

< 참고문헌 >

[1] A. Valdes and K. Skinner. " Probabilistic Alert Correlation" , Fourth International Workshop on the Recent Advances in Intrusion Detection (RAID' 2001), October, 2001.

[2] Dan Andersson, Martin Fong, and Alfonso Valdes. " Heterogeneous Sensor Correlation: A Case Study of Live Traffic Analysis" , the 2002 IEEE Assurance and Security Workshop 2002.

[3] 한국정보보호진흥원, " 지역 호스트 및 전역 네트워크에서 통합 침입탐지 알고리즘 개발" , 최종보고서 2001.

[4] Freeman, J. "Artificial Intelligence: Fuzzy Systems for Control Applications: The Truck Backer-Upper", The Mathematica Journal 4, no. 1 1994 pp 64-69.

[5] J. Gomez, F.gonzalez, and D.Dasgupta, " Complete Expression Trees for Evolving Fuzzy Classifier Systems with Genetic Algorithms" , Evolutionary Computation Conference GECCO02, 2002.

[6] 이광형, 오길록 , 퍼지 이론 및 응용(I, II) 홍릉과학출판사, 1991.

[7] Lee, C. C., " Fuzzy logic in control systems: Fuzzy logic controller-part1," IEEE Tr.on systems, Man, and Cubernetics, Vol. 20, No.2, pp 404-418, March/April, 1990.