(19)대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) 。Int. Cl.
H04L 12/22 (2006.01)
G06F 15/00 (2006.01)
H04L 9/00 (2006.01)

(45) 공고일자 2007년07월11일
(11) 등록번호 10-0738537
(24) 등록일자 2007년07월05일

(21) 출원번호 10-2005-0130889
(22) 출원일자 2005년12월27일
심사청구일자 2005년12월27일

(65) 공개번호 10-2007-0068845
(43) 공개일자 2007년07월02일

(73) 특허권자 삼성전자주식회사
경기도 수원시 영통구 매탄동 416

(72) 발명자 손태식
경기 수원시 팔달구 인계동 1135-8 삼부오피스텔 1415

(74) 대리인 박상수

(56) 선행기술조사문헌
한국항공대학교 석사학위 논문 (2003.02.28):Support Vector Machine을 이용한 호스트 기반의 침입 탐지에 관한 연구

심사관 : 김병균

전체 청구항 수 : 총 15 항

(54) 네트워크 침입 탐지 시스템 및 그 탐지 방법

(57) 요약

본 발명은 네트워크 상의 적어도 하나 이상의 패킷을 캡쳐하는 패킷 캡쳐부와, 패킷 캡쳐부에서 캡쳐된 각 패킷의 특성에 따른 특성 값을 제공하는 전처리부와, 전처리부로부터 제공되는 각 특성 값에 따른 패턴을 상이한 두 개의 패턴 집합으로 구분하고, 각 패턴 집합 중 원소가 많은 패턴 집합을 기준 집합으로 선택하여 네트워크 침입을 탐지하는 학습 엔진부를 포함하는 네트어크 침입 탐지 시스템을 개시함으로써, 이미 알려진 공격 패턴에 따른 사전 데이터에 의존하지 않음으로, 변형된 공격 패턴을 탐지함은 물론, 네트워크 침입을 효과적으로 탐지할 수 있도록 하는 것이다.

대표도

도 1

특허청구의 범위

청구항 1.

네트워크 침입 탐지 시스템에 있어서,

네트워크 상의 적어도 하나 이상의 패킷을 캡쳐하는 패킷 캡쳐부와,

상기 패킷 캡쳐부에서 캡쳐된 각 패킷의 특성에 따른 특성 값을 제공하는 전처리부와,

상기 전처리부로부터 제공되는 상기 각 특성 값에 따른 패턴을 상이한 두 개의 패턴 집합으로 구분하는 초월면을 생성하고, 상기 초월면의 바이어스 텀을 2차원 평면의 원점으로 수렴시켜, 상기 각 패턴 집합 중 원소가 많은 패턴 집합을 기준 집합으로 선택하여 네트워크 침입을 탐지하는 학습 엔진부를 포함하는 네트워크 침입 탐지 시스템.

청구항 2.

제 1 항에 있어서, 상기 전처리부는,

상기 패킷의 각 필드 값에 상응하는 특성 값을 제공하는 네트워크 침입 탐지 시스템.

청구항 3.

제 1 항에 있어서, 상기 학습 엔진부는,

상기 각 특성 값에 따른 패턴을 상이한 두 개의 패턴 집합으로 구분하는 초월면을 생성하고, 상기 초월면의 바이어스 텀을 2차원 평면의 원점으로 수렴시켜, 상기 기준 집합을 선택하고, 상기 기준 집합의 패턴에 따른 기준 프로파일을 생성하는 학습부와,

상기 기준 프로파일과, 네트워크 상의 패킷 특성 값을 비교하여 네트워크 침입을 탐지하는 탐지부를 포함하는 네트워크 침입 탐지 시스템.

청구항 4.

네트워크 침입 탐지 시스템에 있어서,

네트워크 상의 적어도 하나 이상의 패킷 특성 값에 따른 패턴을 SVM 기법에 따라 상이한 두 개의 패턴 집합으로 구분하고, 상기 복수개의 패턴 집합을 구분하는 초월면을 위치를 조절하여 하나의 기준 집합에 따른 기준 프로파일을 생성하는 학습부와,

상기 기준 프로파일과, 네트워크 상의 패킷 특성 값을 비교하여 네트워크 침입을 탐지하는 탐지부를 포함하는 네트워크 침입 탐지 시스템.

청구항 5.

제 4 항에 있어서, 상기 학습부는,

상기 각 패턴을 아래 수학식을 이용하여 두 개의 패턴 집합으로 구분하는 네트워크 침입 탐지 시스템.

$$\underset{w,b,\Xi}{\text{Minimize}} \quad \Phi(w,b,\Xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k$$

$$\text{subject to} \quad y_i(w^T\phi(x_i)+b) \geq 1-\xi_i, \quad \xi_i \geq 0, \quad i = 1,\ldots l$$

여기에서, 'ω'는 조절 가능한 가중치 벡터 변수이고, ' x_i'는 입력 패턴 벡터 변수, 'b'는 바이어스 텀 변수, ' ξ'는 에러 보정 변수이다.

청구항 6.

제 5항에 있어서, 상기 학습부는,

상기 두 개의 패턴 집합으로 구분하는 초월면($\omega^T x_i + b = 0$)의 바이어스 텀을 2차원 평면의 원점으로 수렴시켜 아래 수학식과 같이 기준 집합을 선택하는 네트워크 침입 탐지 시스템.

$$
\begin{array}{lcl}
\textit{soft margin SVM without a bias} & \cong & \textit{one-class SVM} \\
\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k & \cong & \frac{1}{2}\|w\|^2 + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i^k - \rho \\
y_i(w^T\phi(x_i)) \geq 1 - \xi_i & \cong & y_i(w^T\phi(x_i)) \geq \rho - \xi, \\
 & & 0 < \nu < 1,\ 1 < l,\ 0 \leq \rho
\end{array}
$$

여기서, 'ɣ'는 원점부터 초월면 사이까지의 거리(distance) 변수이고, ' l '은 패턴 집합의 최대 원소 개수를 나타내는 변수이다.

청구항 7.

제 4항에 있어서, 상기 학습부는,

아래 수학식을 이용하여 상기 각 패턴의 기준 집합으로 선택하는 네트워크 침입 탐지 시스템.

$$
\begin{array}{ll}
\textit{Minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k - E, & 0 < C < 1 \\
\textit{Subject to } \quad y_i(w^T\phi(x_i)) \geq E - \xi, & 0 < E < 1
\end{array}
$$

여기서, 'ω'는 조절 가능한 가중치 벡터 변수이고, ' x_i'는 입력 패턴 벡터 변수, 'b'는 바이어스 텀 변수, ' ξ'는 에러 보정 변수이다.

청구항 8.

제 4 항에 있어서, 상기 학습 엔진부는,

상기 각 패킷의 패턴을 고차원으로 사상시켜 SVM 기법에 따라 상기 각 패턴을 구분하는 초월면을 생성하고, feature mapping 함수를 이용하여 2차원으로 사상시킨 결과 원점에 분포하는 패턴을 아웃레이어로 처리하는 네트워크 침입 탐지 시스템.

청구항 9.

네트워크 침입 탐지 방법에 있어서,

네트워크 상의 적어도 하나 이상의 패킷을 캡쳐하는 단계;

상기 캡쳐된 각 패킷의 특성에 따른 특성 값을 도출하는 단계;

상기 각 특성 값에 따른 패턴을 상이한 두 개의 패턴 집합으로 구분하는 초월면을 생성하고, 상기 초월면의 바이어스 텀을 2차원 평면의 원점으로 수렴시켜, 패턴을 상이한 두 개의 패턴 집합으로 구분하는 단계;

상기 각 패턴 집합 중 원소가 많은 패턴 집합을 기준 집합으로 선택하여 기준 프로파일을 생성하는 단계; 및

상기 패킷의 특성 값과 상기 기준 프로파일을 비교하여 네트워크 침입을 탐지하는 단계를 포함하는 네트워크 침입 탐지 방법.

청구항 10.

제 9 항에 있어서, 상기 특성 값을 도출하는 단계는,

상기 패킷의 각 필드 값에 상응하는 특성 값을 도출하는 네트워크 침입 탐지 방법.

청구항 11.

삭제

청구항 12.

삭제

청구항 13.

네트워크 침입 탐지 방법에 있어서,

네트워크 상의 적어도 하나 이상의 패킷 특성 값에 따른 패턴을 SVM 기법에 따라 상이한 두 개의 패턴 집합으로 구분하는 단계;

상기 복수개의 패턴 집합을 구분하는 초월면을 위치를 조절하여 하나의 기준 집합을 선택하는 단계;

상기 기준 집합의 패턴에 따라 기준 프로파일을 생성하는 단계; 및

패킷의 특성 값과 상기 기준 프로파일을 비교하여 네트워크 침입을 탐지하는 단계를 포함하는 네트워크 침입 탐지 방법.

청구항 14.

제 13 항에 있어서, 상기 패턴 집합으로 구분하는 단계는,

상기 각 패턴을 아래 수학식을 이용하여 두 개의 패턴 집합으로 구분하는 네트워크 침입 탐지 방법.

$$\underset{w,b,\Xi}{\text{Minimize}} \quad \Phi(w,b,\Xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k$$

$$\text{subject to} \quad y_i(w^T\phi(x_i)+b) \geq 1-\xi_i, \quad \xi_i \geq 0, \quad i = 1,\ldots l$$

여기에서, 'ω'는 조절 가능한 가중치 벡터 변수이고, 'x_i'는 입력 패턴 벡터 변수, 'b'는 바이어스 텀 변수, 'ξ'는 에러 보정 변수이다.

청구항 15.

제 13항에 있어서, 상기 하나의 기준 집합을 선택하는 단계는,

상기 두 개의 패턴 집합으로 구분하는 초월면($\omega^T x_i + b = 0$)의 바이어스 텀을 2차원 평면의 원점으로 수렴시켜 아래 수학식에 이용하여 제 1 패턴 집합을 제 2 패턴 집합의 아웃레이어로 처리하는 네트워크 침입 탐지 방법.

$$\begin{array}{lcl} \textit{soft margin SVM without a bias} & \cong & \textit{one-class SVM} \\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k & \cong & \frac{1}{2}\|w\|^2 + \frac{1}{vl}\sum_{i=1}^{l}\xi_i^k - \rho \\ y_i(w^T\phi(x_i)) \geq 1 - \xi_i & \cong & y_i(w^T\phi(x_i)) \geq \rho - \xi, \\ & & 0 < v < 1,\ 1 < l,\ 0 \leq \rho \end{array}$$

여기서, 'ɣ'는 원점부터 초월면 사이까지의 거리(distance) 변수이고, ' l '은 패턴 집합의 최대 원소 개수를 나타내는 변수이다.

청구항 16.

제 13항에 있어서, 상기 기준 집합을 선택하는 단계는,

아래 수학식을 이용하여 상기 각 패턴의 기준 집합을 선택하는 네트워크 침입 탐지 방법.

$$\textit{Minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k - E, \qquad 0 < C < 1$$

$$\textit{Subject to } \quad y_i(w^T\phi(x_i)) \geq E - \xi, \quad 0 < E < 1$$

여기서, 'ω'는 조절 가능한 가중치 벡터 변수이고, ' x_i'는 입력 패턴 벡터 변수, 'b'는 바이어스 텀 변수, ' ξ'는 에러 보정 변수이다.

청구항 17.

제 13 항에 있어서, 상기 패턴을 구분하는 단계는,

상기 각 패킷의 패턴을 고차원으로 사상시켜 SVM 기법에 따라 상기 각 패턴을 구분하는 초월면을 생성하고, feature mapping 함수를 이용하여 2차원으로 사상시키는 단계를 포함하는 네트워크 침입 탐지 방법.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 네트워크 침입 탐지 시스템 및 그 탐지 방법에 관한 것이다.

오늘날 네트워크의 기술의 발전과 사용자의 증가로 정보화 사회로 발전하고 있는 반면, 네트워크를 통해 다른 사용자들에게 바이러스를 유포하거나, 공격하는 부정적인 측면도 증가하고 있다.

이러한, 네트워크의 침입을 탐지하는 위해 제시된 것이 침입 탐지 시스템(intrusion detection system)이다. 침입 탐지 시스템은 네트워크의 비정상적인 행위, 오용 등을 실시간으로 탐지하는 시스템이다.

네트워크 침입 탐지 기술은 크게 오용 탐지(misuse detection)와 비정상행위 탐지(anomaly detection)로 구분될 수 있다.

먼저, 오용 탐지 기술은 이미 알려진 공격 패턴에 대한 시그너쳐(signature)나 룰셋(rule set)을 생성하고, 그 생성된 시그너쳐 또는 룰셋에 일치하는 패턴을 확인함으로써, 공격을 탐지하는 기술이다. 이러한 오용 탐지 기술은 패턴 매칭, 전문가 시스템, 상태 전이 모델, 키 스토로크 모니터링 등이 있다.

그리고, 비정상행위 탐지 기술은 정상행위에 대한 정상 프로파일을 생성하고, 그 생성된 정상 프로파일을 벗어나는 행위들을 공격으로서 간주하는 기술로써, 통계적 방법, 신경망 기법, 예측 가능 패턴 생성 등의 기법 등이 존재한다.

그러나, 일반적인 침입 탐지 기술은 오용 또는 비정상행위를 탐지하기 위해서는 사전 데이터가 필요함은 물론, 사전 데이터에서 벗어나는 오용 또는 비정상행위를 탐지할 수 없다.

예를 들어, 오용 탐지 기술에는 이미 알려진 공격 패턴에 대한 시그너쳐 또는 룰셋을 생성하기 위한 사전 데이터가 필요하며, 또한, 시그너쳐 또는 룰셋을 벗어나는 패턴에 대해서는 탐지할 수 없다.

또한, 비정상행위 탐지 기술에서는 사전 데이터에 의존하여 비정상행위를 탐지하기 위한 정상 프로파일을 생성하기 때문에 사전 데이터에 탐지 기준이 의존적이며, 정상 프로파일을 생성하기 위한 학습 과정에 많은 학습 데이터가 필요하다.

발명이 이루고자 하는 기술적 과제

따라서, 본 발명은 상기와 같은 문제점을 해결하기 위하여 창안된 것으로, 이미 알려진 공격 패턴에 따른 사전 데이터에 의존하지 않음으로, 변형된 공격 패턴을 탐지함은 물론, 네트워크 침입을 효과적으로 탐지할 수 있는 네트워크 침입 탐지 시스템 및 그 탐지 방법을 제공하는 것에 그 목적이 있다.

발명의 구성

상기 목적을 달성하기 위한 본 발명의 일측면에 따른 네트워크 침입 탐지 시스템은, 네트워크 상의 적어도 하나 이상의 패킷을 캡쳐하는 패킷 캡쳐부와, 패킷 캡쳐부에서 캡쳐된 각 패킷의 특성에 따른 특성 값을 제공하는 전처리부와, 전처리부로부터 제공되는 각 특성 값에 따른 패턴을 상이한 두 개의 패턴 집합으로 구분하고, 각 패턴 집합 중 원소가 많은 패턴 집합을 기준 집합으로 선택하여 네트워크 침입을 탐지하는 학습 엔진부를 포함한다.

본 발명에 따른 전처리부는, 패킷의 각 필드 값에 상응하는 특성 값을 제공한다.

본 발명에 따른 학습 엔진부는, 각 특성 값에 따른 패턴을 상이한 두 개의 패턴 집합으로 구분하는 초월면을 생성하고, 초월면의 바이어스 텀을 2차원 평면의 원점으로 수렴시켜, 기준 집합을 선택하고, 기준 집합의 패턴에 따른 기준 프로파일을 생성하는 학습부와, 기준 프로파일과, 네트워크 상의 패킷 특성 값을 비교하여 네트워크 침입을 탐지하는 탐지부를 포함한다.

본 발명의 다른 측면에 따른 네트워크 침입 탐지 시스템은, 네트워크 상의 적어도 하나 이성의 패킷 특성 값에 따른 패턴을 SVM 기법에 따라 상이한 두 개의 패턴 집합으로 구분하고, 복수개의 패턴 집합을 구분하는 초월면을 위치를 조절하여 하나의 기준 집합에 따른 기준 프로파일을 생성하는 학습부와, 기준 프로파일과, 네트워크 상의 패킷 특성 값을 비교하여 네트워크 침입을 탐지하는 탐지부를 포함한다.

본 발명에 따른 학습부는, 각 패턴을

$$\underset{w,b,\Xi}{\text{Minimize}} \quad \Phi(w,b,\Xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k$$

$$\text{subject to} \quad y_i(w^T\phi(x_i)+b) \geq 1-\xi_i, \quad \xi_i \geq 0, \quad i = 1,\ldots l$$

의 수학식으로 구분한다.

여기에서, 'ω'는 조절 가능한 가중치 벡터 변수이고, 'x_i'는 입력 패턴 벡터 변수, 'b'는 바이어스 텀 변수, 'ξ'는 에러 보정 변수이다.

본 발명에 따른 학습부는, 두 개의 패턴 집합으로 구분하는 초월면($\omega^T x_i + b = 0$)의 바이어스 텀을 2차원 평면의 원점으로 수렴시켜,

$$\begin{array}{lcl} \textit{soft margin SVM without a bias} & \cong & \textit{one-class SVM} \\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k & \cong & \frac{1}{2}\|w\|^2 + \frac{1}{vl}\sum_{i=1}^{l}\xi_i^k - \rho \\ y_i(w^T\phi(x_i)) \geq 1 - \xi_i & \cong & y_i(w^T\phi(x_i)) \geq \rho - \xi, \\ & & 0 < v < 1,\ 1 < l,\ 0 \leq \rho \end{array}$$

의 수학식으로 기준 집합을 선택한다.

여기서, 'ɣ'는 원점부터 초월면 사이까지의 거리(distance) 변수이고, ' l '은 패턴 집합의 최대 원소 개수를 나타내는 변수이다.

본 발명에 따른 학습부는,

$$\textit{Minimize}\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k - E, \qquad 0 < C < 1$$

$$\textit{Subject to}\quad y_i(w^T\phi(x_i)) \geq E - \xi, \quad 0 < E < 1$$

의 수학식으로 패턴의 기준 집합으로 선택한다.

여기서, 'ω'는 조절 가능한 가중치 벡터 변수이고, ' x_i'는 입력 패턴 벡터 변수, 'b'는 바이어스 텀 변수, ' ξ'는 에러 보정 변수이다.

본 발명에 따른 학습 엔진부는, 각 패킷의 패턴을 고차원으로 사상시켜 SVM 기법에 따라 각 패턴을 구분하는 초월면을 생성하고, feature mapping 함수를 이용하여 2차원으로 사상시킨 결과 원점에 분포하는 패턴을 아웃레이어로 처리한다.

본 발명의 다른 측면에 따른 네트워크 침입 탐지 방법은, 네트워크 상의 적어도 하나 이상의 패킷을 캡쳐하는 단계와, 캡쳐된 각 패킷의 특성에 따른 특성 값을 도출하는 단계와, 각 특성 값에 따른 패턴을 상이한 두 개의 패턴 집합으로 구분하는 단계와, 각 패턴 집합 중 원소가 많은 패턴 집합을 기준 집합으로 선택하여 기준 프로파일을 생성하는 단계와, 패킷의 특성 값과 기준 프로파일을 비교하여 네트워크 침입을 탐지하는 단계를 포함한다.

본 발명에 따른 특성 값을 도출하는 단계는, 패킷의 각 필드 값에 상응하는 특성 값을 도출한다.

본 발명에 따른 패턴 집합으로 구분하는 단계는, 각 패턴을 상이한 두 개의 패턴 집합으로 구분하는 초월면을 생성한다.

본 발명에 따른 기준 프로파일을 생성하는 단계는, 패턴을 구분하는 초월면의 바이어스 텀을 2차원 평면의 원점으로 수렴시켜, 기준 집합을 선택하는 단계와, 기준 집합의 패턴에 따른 기준 프로파일을 생성하는 단계를 포함한다.

본 발명의 또 다른 측면에 따른 네트워크 침입 탐지 방법은, 네트워크 상의 적어도 하나 이성의 패킷 특성 값에 따른 패턴을 SVM 기법에 따라 상이한 두 개의 패턴 집합으로 구분하는 단계와, 복수개의 패턴 집합을 구분하는 초월면을 위치를 조절하여 하나의 기준 집합을 선택하는 단계와, 기준 집합의 패턴에 따라 기준 프로파일을 생성하는 단계와, 패킷의 특성 값과 기준 프로파일을 비교하여 네트워크 침입을 탐지하는 단계를 포함한다.

본 발명에 따른 패턴 집합으로 구분하는 단계는, 각 패턴을

$$\underset{w,b,\Xi}{\text{Minimize}}\ \ \Phi(w,b,\Xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k$$

$$\text{subject to}\ \ y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1,\ldots l$$

으로 두 개의 패턴 집합으로 구분한다.

여기에서, 'ω'는 조절 가능한 가중치 벡터 변수이고, ' x_i'는 입력 패턴 벡터 변수, 'b'는 바이어스 텀 변수, ' ξ'는 에러 보정 변수이다.

본 발명에 따른 하나의 기준 집합을 선택하는 단계는, 두 개의 패턴 집합으로 구분하는 초월면($\omega^T x_i + b = 0$)의 바이어스 텀을 2차원 평면의 원점으로 수렴시켜,

$$\begin{array}{lcl} \textit{soft margin SVM without a bias} & \cong & \textit{one-class SVM} \\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k & \cong & \frac{1}{2}\|w\|^2 + \frac{1}{vl}\sum_{i=1}^{l}\xi_i^k - \rho \\ y_i(w^T\phi(x_i)) \geq 1 - \xi_i & \cong & y_i(w^T\phi(x_i)) \geq \rho - \xi, \\ & & 0 < v < 1,\ 1 < l,\ 0 \leq \rho \end{array}$$

으로 제 1 패턴 집합을 제 2 패턴 집합의 아웃레이어로 처리한다.

여기서, 'ɣ'는 원점부터 초월면 사이까지의 거리(distance) 변수이고, ' l '은 패턴 집합의 최대 원소 개수를 나타내는 변수이다.

본 발명에 따른 기준 집합을 선택하는 단계는,

$$\textit{Minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k - E, \qquad 0 < C < 1$$

$$\textit{Subject to } \quad y_i(w^T\phi(x_i)) \geq E - \xi, \quad 0 < E < 1$$

으로 각 패턴의 기준 집합을 선택한다.

여기서, 'ω'는 조절 가능한 가중치 벡터 변수이고, ' x_i'는 입력 패턴 벡터 변수, 'b'는 바이어스 텀 변수, ' ξ'는 에러 보정 변수이다.

본 발명에 따른 패턴을 구분하는 단계는, 각 패킷의 패턴을 고차원으로 사상시켜 SVM 기법에 따라 각 패턴을 구분하는 초월면을 생성하고, feature mapping 함수를 이용하여 2차원으로 사상시키는 단계를 포함한다.

이하 본 발명에 따른 네트워크 침입 탐지 시스템 및 그 탐지 방법을 첨부한 도면을 참조하여 상세히 설명한다.

도 1은 본 발명에 따른 네트워크 침입 탐지 시스템을 설명하기 위한 블록 도면이다.

도 1을 참조하면, 본 발명에 따른 네트워크 침입 탐지 시스템은, 패킷 캡쳐부(100), 전처리부(200), 및 학습 엔진부(300)를 포함하고, 학습 엔진부(300)는 학습부(310) 및 탐지부(320)를 포함한다.

패킷 캡쳐부(100)는 네트워크 상의 패킷을 무작위 또는 소정 시간동안 캡쳐한다. 즉, 패킷 캡쳐부(100)는 네트워크 침입 탐지 시스템의 대상이 네트워크 또는 호스트인지 여부에 따라 네트워크 상의 패킷을 캡쳐한다.

그리고, 전처리부(200)는 패킷 캡쳐부(100)에서 캡쳐한 패킷을 학습하기 위한 포맷으로 변환한다. 즉, 전처리부(200)는 학습 엔진부(300)가 캡쳐되는 패킷을 기반으로 학습 과정을 수행할 수 있도록 패킷의 정보를 전처리한다.

일례를 들어, 전처리부(200)는 캡쳐되는 TCP/IP 패킷의 각 필드 특성에 상응하는 특성 값으로 변환 처리한다.

학습 엔진부(300)는 전처리부(200)로부터 제공되는 각 패킷의 특성 값을 학습하여, 네트워크 침입을 탐지한다.

학습 엔진부(300)의 학습부(310)는 통계적 학습 이론(Statistical Learning Theory)을 기반으로 각 패킷의 특성 값에 따라 정상 집합과 비정상 집합으로 구분하고, 정상 집합으로부터 기준 프로파일을 도출한다.

이러한, 학습부(310)는 네트워크 침입을 탐지하기 위한 사전 데이터가 없는 상태에서 수신되는 패킷의 패턴에 따라 두 개의 상이한 집합인 정상 집합과 비정상 집합으로 구분하고, 이들 집합을 구분하는 초월면을 패턴 원소가 극히 적은 집합으로 수렵시키고, 하나의 집합으로부터 네트워크 침입을 탐지하는 기준 프로파일을 생성한다.

그리고, 탐지부는 캡쳐되는 패킷의 특성 값에 따른 패턴과 기준 프로파일을 비교하여 네트워크 침입 여부를 탐지한다.

이때, 학습 엔진부(300)는 초기 소정 시간동안 학습 과정을 거쳐 기준 프로파일을 도출하고, 지속적으로 캡쳐되는 패킷의 특성 값에 따라 기준 프로파일을 갱신하거나, 소정 주기마다 학습 과정을 거쳐 기준 프로파일을 갱신할 수 있다.

도 2는 본 발명의 바람직한 실시예에 따라 두 개의 집합으로 패턴을 구분하는 것을 설명하기 위한 도면이다.

도 2에 도시된 바와 같이 학습 엔진부(300)는 캡쳐되는 패킷의 특성 값에 따라 패킷의 패턴(x)을 도식한다.

그리고, 학습 엔진부(300)는 도식되는 패턴을 구분 가능한 분류 알고리즘을 이용하여 두 개의 상이한 패턴 집합, 즉, 정상 집합(Class 2)과, 비정상 집합(Class 1)으로 구분한다.

일례를 들어, 학습 엔진부(300)는 패턴을 상이한 두 개의 집합으로 분류하는 알고리즘인 SVM(Support Vector Machine) 알고리즘을 이용하여, 두 개의 상이한 패턴 집합으로 구분할 수 있으며, SVM 알고리즘을 이용하여 두 개의 패턴 집합으로 구분하는 초월면(hyperplane)(l)을 생성한다.

다음 수식 1은 SVM 알고리즘에 따라 두 개의 상이한 패턴 집합을 분류(classifier)하는 조건 수학식이다.

수학식 1

$$\underset{w,b,\Xi}{\text{Minimize}} \quad \Phi(w,b,\Xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k$$

$$\text{subject to} \quad y_i(w^T\phi(x_i)+b) \geq 1-\xi_i, \quad \xi_i \geq 0, \quad i = 1,\cdots l$$

상기 수학식 1은 SVM 알고리즘에서 이진 분류하는 분류기를 결정하는 조건 수학식이며, $'\omega^T x_i + b = 0'$를 두 개의 패턴 집합을 구분하는 초월면(l)이라고 가정하면, 'ω'는 조절 가능한 가중치 벡터이고, $'x_i'$는 입력 패턴 벡터, 'b'는 바이어스 텀이다.

상기 도 2에서와 같이, 두 개의 패턴 집합(Class1, Class 2)은 초월면($\omega^T x_i + b = 0$)(l)에 의해서 구분 가능하며, 에러는 'ξ'로 보정할 수 있다. 예를 들어, 입력되는 패턴(x_i, x_j)를 패턴 집합내에서 의미 없는 패턴으로 간주하면, 초월면(l)의 바이어스 텀(b)을 'ξ'만큼 보정함으로써, 초월면의 위치(l', l")를 조절할 수 있다.

즉, 학습 엔진부(300)는 입력되는 패킷의 패턴들을 SVM 알고리즘에 따른 교사 학습(Supervised Learning) 과정을 통해 두 개의 패턴 집합으로 구분한느 초월면(l)을 생성함으로써, 정상 집합(Class 2)과 비정상 집합(Class 1)의 두 개 클래스로 분류한다.

여기서, 두 개의 클래스를 분류해주는 결정 평면이 초월면(l)이며, 초월면을 결정하는 입력 패턴들이 결정 벡터(Support Vector : SV)이다.

그리고, 각 패턴이 두 개의 패턴 집합으로 구분 가능한 경우, 초월면(l)은 결정 벡터의 패턴까지의 거리를 최대화하며, 모든 결정 벡터의 패턴은 초월면으로부터 같은 최소 거리에 위치하게 된다.

그러나, 패킷의 패턴이 선형으로 구분되는 경우는 극히 드물기 때문에 패킷의 패턴 집합은 비선형의 특징을 가지게 된다. 따라서, 입력 패턴을 커널 트릭과 같은 기법을 이용하여 고차원으로 사상시킨 후에 feature mapping 함수를 이용하여 다시 2차원으로 사상시킨다.

그리고, SVM 기법 중 one-class SVM 기법은 하나의 클래스, 즉, 하나의 패턴 집합에 대한 패턴만을 가지고 학습이 이루어지는 비교사 기반의 방식으로, 패턴 집합에 포함되지 않는 패턴(아웃 레이어)들은 feature mapping 함수를 이용하여 고차원에서 2차원으로 사상시키면, 2차원의 원점 근처에 위치하게 된다.

또한, 상기 도 2에서와 같이, 초월면(l)을 $'\omega^T x_i + b = 0'$와 같은 1차 함수로 가정하면, 바이어스 텀(b)을 조절함으로써 초월면(l)이 생성되는 위치(l', l")를 조절할 수 있다.

만약, 초월면(l)의 바이어스 텀(b)이 '0'에 가까워지는 경우, 즉, 초월면(l)으로 구분되는 두 개의 패턴 집합 중 어느 한 패턴 집합에 포함되는 특성 값은 매우 적어지는 경우, 초월면(l")으로 구분되는 비정상 집합의 크기는 매우 작아진다.

도 3은 본 발명의 바람직한 실시예에 따라 하나의 집합으로 패턴을 구분하는 것을 설명하기 위한 도면이다.

도 3에 도시된 바와 같이, 초월면(l)의 바이어스 텀(b)이 '0'에 매우 가깝다고 가정하면, 비정상 집합(Class 1)에 포함되는 패킷의 패턴은 매우 적어진다.

따라서, 학습 엔진부(300)는 비정상 집합(Class 1)에 포함되는 패턴이 매우 적어짐으로, 상대적으로 정상 집합(Class 2)을 제외한 패턴을 에러 또는 아웃레이어(outlier)로 간주할 수 있다.

즉, 학습 엔진부(300)는 SVM 기법에 따라 두 개의 패턴 집합으로 구분하고, 상이한 두 개의 패턴 집합 중 중심이 되는 하나의 패턴 집합을 학습하고, 그 패턴 집합을 제외한 패턴을 에러 또는 아웃레이어로 간주함으로써, 하나의 패턴 집합만을 남긴다.

따라서, 학습 엔진부(300)는 캡쳐되는 패킷의 특성에 따라 구분되는 비정상 집합(Class 1)을 정상 집합(Class 2)의 아웃레이어로 간주할 수 있음으로, 정상 집합(Class 2)을 기반으로 기준 프로파일을 도출할 수 있다.

여기서, SVM 알고리즘의 soft-margin SVM 기법과 one-class SVM 기법을 잠시 살펴보면, soft-margin SVM 기법은 교사 학습의 정확성과 빠른 학습률을 가지지만 학습 단계에서 두 클래스에 대한 명확한 규정이 필요한 문제점을 가지고 있다.

그리고, One-class SVM 기법은 단일 클래스에 대한 학습 가능성으로 인해 비정상행위 탐지이 효과적이나, 단일 클래스 학습이라는 고유 특성에 의해 높은 false positive와 정확도가 낮은 문제점을 가지고 있다.

또한, 일반적으로 네트워크 상의 패킷 중 정상적인 패킷 양보다 비정상적인 패킷 양이 극히 적으므로, 학습 엔진부(300)는 각 패킷의 특성 값 패턴에 따라 soft-margin SVM 기법을 기반으로 두 개의 상이한 패턴 집합으로 구분하고, 비정상적인 패턴을 정상 집합(Class 2)의 에러 또는 아웃레이어로 간주할 수 있음으로, 하나의 패턴 집합인 정상 집합(Class 2)만을 남기로, 정상 집합(Class 2)을 기반으로 네트워크 침입을 탐지하는 기준이 될 수 있는 기준 프로파일을 생성할 수 있다.

다음 수학식 2는 soft-margin SVM 기법에 따른 두 개의 패턴 집합을 one-class SVM 기법에 따른 하나의 패턴 집합으로 사상시키기 위한 비교 수학식이다.

수학식 2

$$\begin{array}{lcl} \textit{soft margin SVM without a bias} & \cong & \textit{one-class SVM} \\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k & \cong & \frac{1}{2}\|w\|^2 + \frac{1}{vl}\sum_{i=1}^{l}\xi_i^k - \rho \\ y_i(w^T\phi(x_i)) \geq 1-\xi_i & \cong & y_i(w^T\phi(x_i)) \geq \rho - \xi, \\ & & 0<v<1,\ 1<l,\ 0 \leq \rho \end{array}$$

상기 수학식 2의 One-class SVM 기법에서 'ϒ'는 원점부터 초월면(l) 사이까지의 거리(distance)와, 초월면(l)에 의해 분리되는 패킷의 특성 값 개수와 트레이드-오프(trade-off) 관계에 있다.

또한, ' l '은 전체 패턴 집합의 패턴 개수를 나타낸다. 즉, 입력 패킷 패턴(x_i)은 최대 ' l ' 만큼의 개수를 가질 수 있다.

초월면(l)과 패턴 집합에 속한 패턴과의 거리와 에러 수용 변수(ξ)와의 트레이드-오프 관계에 있는 soft-margin SVM 기법의 'C' 값을 one-class SVM 기법의 제약 조건에 적합하도록 1보다 작은 값으로 유지하고, 'ρ'(수식 3에서의 변수 E로 전환)의 값을 0과 1사이의 극히 작은 정수로 유지하면, soft-margin SVM 기법의 제약 조건이 다음 수식 3과 같이 변형되며, soft-margin SVM 기법과 one-class SVM 기법의 특성을 가진 이진 분류기를 생성할 수 있다.

수학식 3

$$\textit{Minimize}\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i^k - E, \qquad 0 < C < 1$$

$$\textit{Subject to}\quad y_i(w^T\phi(x_i)) \geq E - \xi, \qquad 0 < E < 1$$

상기 수학식 3에서 설명되어지는 바와 같이, soft-margin SVM 기법에 따른 두 개의 패턴 집합을 one-class SVM 기법에 따른 하나의 패턴 집합으로 사상될 수 있다.

즉, soft-margin SVM 기법에 따라 두 개의 상이한 패턴 집합으로 구분한 초월면($\omega^T x_i + b = 0$)의 바이어트 텀(b)을 2차원의 원점으로 수렴시키면, one-class SVM 기법의 하나의 패턴 집합으로 사상시킬 수 있으며, 하나의 패턴 집합으로부터 기준 프로파일을 생성할 수 있다.

도 4는 본 발명의 바람직한 실시예에 따른 네트워크 침입 탐지 방법을 설명하기 위한 플로챠트 도면이다.

도 4를 참조하면, 네트워크 침입 탐지 시스템은 네트워크 상의 패킷을 캡쳐한다(S 100).

그리고, 네트워크 침입 탐지 시스템은 패킷의 특성 값에 따라 패턴을 구분하여 두 개의 상이한 집합으로 구분한다(S 110).

네트워크 침입 탐지 시스템은 상기 도 2에서 설명되어지는 바와 같이, 패킷의 특성 값에 패턴을 도식하고, 각 패턴을 통계적 학습 이론에 따라 학습하여, 패턴을 상이한 두 개의 패턴 집합, 즉 정상 집합과, 비정상 집합으로 구분한다.

이때, 패턴 집합은 soft-margin SVM 기법에 따른 초월면을 생성하여 구분할 수 있다.

일반적으로 네트워크 상에서 정상적인 패킷 양이 비정상적인 패킷 양보가 매우 큼으로, 두 개의 상이한 패턴 집합, 즉 정상 집합(Class 2)에 속하는 패턴이 비정상 집합(Class 1)에 속하는 패턴보다 매우 많음으로, 초월면($\omega^T x_i + b = 0$)의 바이어스 텀(b)을 원점으로 수렴시키실 수 있다.

따라서, 두 개의 상이한 패턴 집합 중 비정상 집합(Class 1)에 속한 패턴을 정상 집합(Class 2)의 에러 또는 아웃레이어로 간주할 수 있음으로, 정상 집합(Class 2)에 속한 패턴을 이용하여 기준 프로파일을 도출한다(S 120).

즉, 네트워크 침입 탐지 시스템은 네트워크 상의 패킷 중 정상적인 패킷 양보다 비정상적인 패킷 양이 극히 적으므로, 각 패킷의 특성 값 패턴에 따라 soft-margin SVM 기법을 기반으로 두 개의 상이한 패턴 집합으로 구분하고, 비정상적인 패턴을 정상 집합(Class 2)의 에러 또는 아웃레이어로 간주할 수 있음으로, 하나의 패턴 집합인 정상 집합(Class 2)만을 남기로, 정상 집합을 기반으로 네트워크 침입을 탐지하는 기준이 될 수 있는 기준 프로파일을 생성할 수 있다.

네트워크 침입 탐지 시스템은 기준 프로파일을 이용하여 네트워크 상의 패킷이 비정상적인 패킷인지 여부를 탐지한다(S 130). 즉, 네트워크 침입 탐지 시스템은 기준 프로파일을 이용하여 네트워크 침입을 탐지한다.

상술한 본 발명의 상세 설명에서는 일례를 들어, SVM 알고리즘의 soft-margin SVM 기법을 이용하여 두 개의 패턴 집합으로 구분하고, 하나의 패턴 집합의 패턴에 따라 기준 프로파일을 생성하여 네트워크 침입을 탐지하는 경우에 대하여 설명하였으나, 기타 학습 알고리즘을 이용하여 사전 데이터 없이 네트워크 탐지를 위한 기준 프로파일을 생성하는 경우도 이와 동일하게 적용될 수 있다.

발명의 효과

상기한 바와 같이, 본 발명에 따르면, SVM 기법에 따라 이미 알려진 사전 데이터가 없이 패킷의 패턴을 학습하여, 하나의 기준 프로파일을 생성할 수 있음으로, 사전 데이터에 의존적이지 않으면서 침입 탐지의 정확성 및 빠른 학습율을 가질 수 있다.

도면의 간단한 설명

도 1은 본 발명에 따른 네트워크 침입 탐지 시스템을 설명하기 위한 블록 도면.

도 2는 본 발명의 바람직한 실시예에 따라 두 개의 집합으로 패턴을 구분하는 것을 설명하기 위한 도면.

도 3은 본 발명의 바람직한 실시예에 따라 하나의 집합으로 패턴을 구분하는 것을 설명하기 위한 도면.

도 4는 본 발명의 바람직한 실시예에 따른 네트워크 침입 탐지 방법을 설명하기 위한 플로챠트 도면.

<도면의 주요 부분에 대한 부호의 설명>

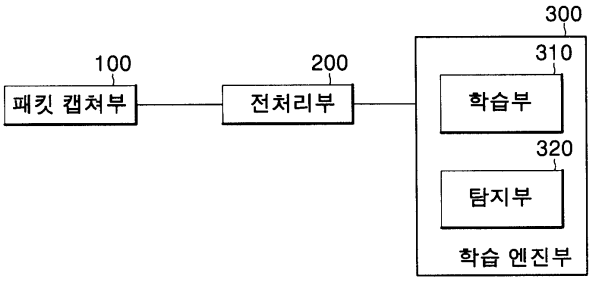
100 : 패킷 캡쳐부 200 : 전처리부

300 : 학습 엔진부 310 : 학습부

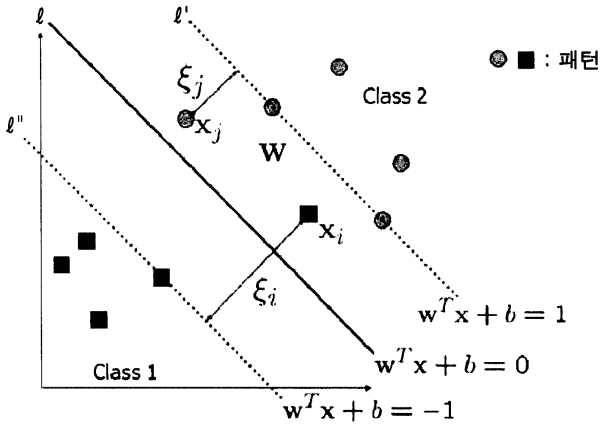
320 : 탐지부

도면

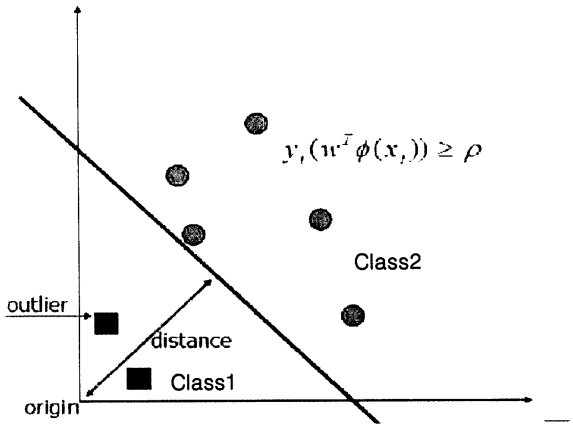
도면1



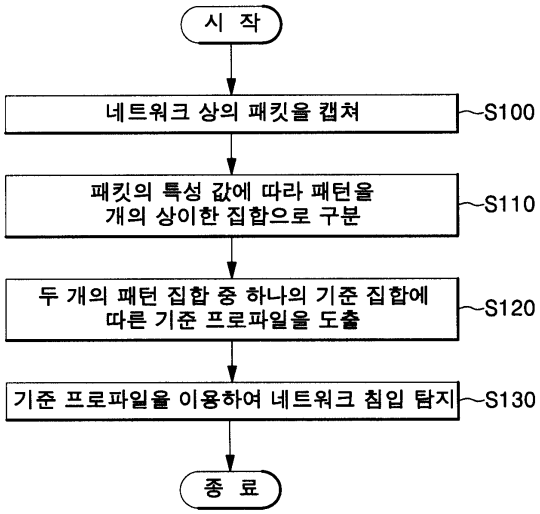
도면2



도면3


$y_i(w^T\phi(x_i)) \geq \rho$
Class2
outlier
distance
Class1
origin

도면4


시 작
네트워크 상의 패킷을 캡쳐
S100
패킷의 특성 값에 따라 패턴을 개의 상이한 집합으로 구분
S110
두 개의 패턴 집합 중 하나의 기준 집합에 따른 기준 프로파일을 도출
S120
기준 프로파일을 이용하여 네트워크 침입 탐지
S130
종 료