(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0126814
(43) 공개일자 2013년11월21일

(51) 국제특허분류(Int. Cl.)
G06F 21/30 (2013.01) *G06F 11/30* (2006.01)
G06F 17/00 (2006.01)
(21) 출원번호 10-2012-0043733
(22) 출원일자 2012년04월26일
심사청구일자 없음

(71) 출원인
한국전자통신연구원
대전광역시 유성구 가정로 218 (가정동)
(72) 발명자
유재학
충청북도 옥천군 옥천읍 죽향리 옥향아파트 101동 1304호
이병복
대전광역시 유성구 장대동 344 드림월드 108동 201호
방효찬
대전광역시 유성구 노은동 열매마을 9단지 902동 1401호
(74) 대리인
특허법인무한

전체 청구항 수 : 총 12 항

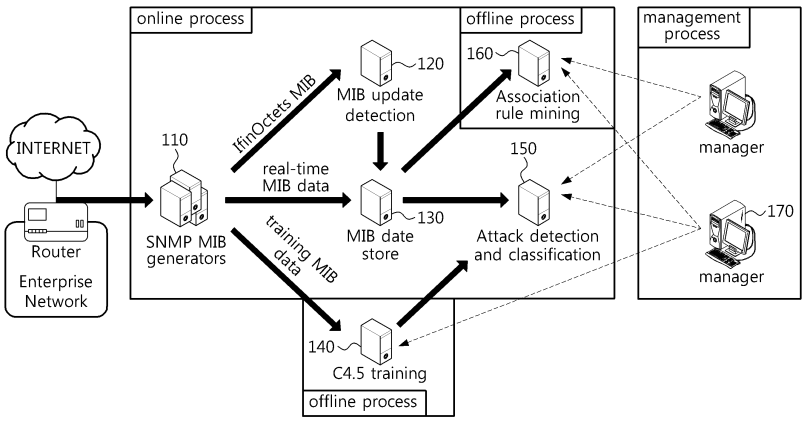
(54) 발명의 명칭 **데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치 및 방법**

(57) 요 약

서비스 거부 공격(Denial of Service, 이하 DoS) 공격을 보다 다양하고 견고하게 발전시킨 분산 서비스 거부(Distributed Denial of Service, 이하 DDoS/트래픽 폭주 공격) 공격에 대한 신속한 탐지 및 공격유형별 분류, 공격에 대한 의미적 해석하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치 및 방법을 개시한다.

본 발명의 일실시예에 따른 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치 및 방법은 데이터 마이닝의 예측 및 분석기법을 기반으로 트래픽 폭주 공격에 대한 신속한 탐지와 공격유형별 분류, 그리고 의미적 해석으로 보다 안정적인 서비스 제공 및 시스템 운용을 지원한다.

대 표 도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호	10035310
부처명	지식경제부
연구사업명	일반회계사업(산업원천)
연구과제명	차세대 USN기반의 스마트 사회안전 프레임워크 기술 개발
기 여 율	1/1
주관기관	전자부품연구원
연구기간	2010.04.01 ~ 2013.03.31

특허청구의 범위

청구항 1

네트워크 트래픽 데이터로부터 관리 정보 베이스(MIB: Management information base)를 생성하는 생성 모듈;

상기 관리 정보 베이스를 수집하여 탐지 시스템의 동작 시점을 결정하는 감지 모듈;

상기 관리 정보 베이스를 분석하는 상기 탐지 시스템에서 결정된 관리 정보 베이스를 저장하는 저장 모듈; 및

상기 결정된 관리 정보 베이스에 기반하여 공격유무와 공격유형을 판단하는 공격 판단 모듈

을 포함하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치.

청구항 2

제1항에 있어서,

상기 탐지 시스템은,

다양한 트래픽 공격을 임의로 발생시켜 의사결정나무 기반의 학습을 실시하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치.

청구항 3

제1항에 있어서,

상기 저장 모듈에 저장된 데이터의 특징을 규칙의 형태로 추출하고 분석하는 의미론적 심층해석을 실시하는 연관관계규칙 모듈

을 더 포함하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치.

청구항 4

제3항에 있어서,

상기 공격 판단 모듈에 대한 실시간 탐지 및 분류유형에 대한 상세 정보를 모니터링하고 상기 탐지 시스템과 상기 연관관계규칙 모듈에서 제공하는 규칙 및 의미론적 해석정보로 침입탐지 및 대응시스템의 정책 수립에 활용하는 관리자 모듈

을 더 포함하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치.

청구항 5

제1항에 있어서,

상기 공격 판단 모듈은,

공격트래픽을 탐지하면 침입 사실을 관리자 모듈에 실시간으로 보고하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치.

청구항 6

제1항에 있어서,

상기 공격 판단 모듈은,

공격트래픽을 TCP-SYN 플러딩(flooding), UDP 플러딩, ICMP 플러딩으로 분류하고 공격유형에 대한 추가적인 정보를 제공하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치.

청구항 7

네트워크 트래픽 데이터로부터 관리 정보 베이스(MIB: Management information base)를 생성하는 단계;

상기 관리 정보 베이스를 수집하여 탐지 시스템의 동작 시점을 결정하는 단계;

상기 관리 정보 베이스를 분석하는 상기 탐지 시스템에서 결정된 관리 정보 베이스를 저장하는 단계; 및

상기 결정된 관리 정보 베이스에 기반하여 공격유무와 공격유형을 판단하는 단계

를 포함하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 방법.

청구항 8

제7항에 있어서,

상기 탐지 시스템은 다양한 트래픽 공격을 임의로 발생시켜 의사결정나무 기반의 학습을 실시하는 단계

를 더 포함하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 방법.

청구항 9

제7항에 있어서,

상기 저장된 관리 정보 베이스의 특징을 규칙의 형태로 추출하고 분석하는 의미론적 심층해석을 실시하는 단계

를 더 포함하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 방법.

청구항 10

제9항에 있어서,

상기 판단하는 단계에 대한 실시간 탐지 및 분류유형에 대한 상세 정보를 모니터링하고 상기 탐지 시스템과 상기 의미론적 심층해석을 실시하는 단계에서 제공하는 규칙 및 의미론적 해석정보로 침입탐지 및 대응시스템의 정책 수립에 활용하는 단계

를 더 포함하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 방법.

청구항 11

제7항에 있어서,

상기 판단하는 단계는,

공격트래픽을 탐지하면 침입 사실을 실시간으로 보고하는 단계

를 포함하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 방법.

청구항 12

제7항에 있어서,

상기 판단하는 단계는,

공격트래픽을 TCP-SYN 플러딩(flooding), UDP 플러딩, ICMP 플러딩으로 분류하고 공격유형에 대한 추가적인 정보를 제공하는 단계

를 포함하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 방법.

명 세 서

기 술 분 야

[0001] 본 발명의 실시예들은 서비스 거부 공격(Denial of Service, 이하 DoS) 공격을 보다 다양하고 견고하게 발전시킨 분산 서비스 거부(Distributed Denial of Service, 이하 DDoS/트래픽 폭주 공격) 공격에 대한 신속한 탐지 및 공격유형별 분류, 공격에 대한 의미적 해석하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] DoS/DDoS로 대표되는 트래픽 폭주 공격은 대상이 되는 컴퓨터 시스템은 물론 네트워크의 자원을 고갈시킴으로써 정상적인 서비스를 수행하지 못하게 하는 공격으로 업무에 막대한 피해를 준다. 이러한 악의적인 접근이나 침입 등을 신속하게 탐지하고 대처할 수 있는 보안 기술이 요구된다.

[0003] DoS/DDoS 탐지에서의 전통적인 패킷 수집 방법들은 공격에 대한 상세한 분석은 가능하나, 고가의 고성능 분석시스템이 요구될 뿐만 아니라 설치 및 운영상의 확장성이 부족한 실정이다.

[0004] 이를 보완하기 위한 방법으로 최근 SNMP(Simple Network Management Protocol)에서의 MIB(Management Information Bases; 관리 정보 베이스) 정보를 이용한 침입탐지 방법론이 주목을 받고 있다. SNMP MIB 정보를 이용한 트래픽 폭주 공격 탐지는 MIB 정보 수집을 위한 시스템 및 네트워크 리소스의 사용이 적고, 표준화된 네트워크 성능 데이터를 제공 받을 수 있기 때문에 패킷 기반 탐지 방법에 비해 보다 빠르고 효과적인 탐지를 지원할 수 있다.

[0005] SNMP MIB 정보를 이용하는 DDoS 탐지 방법은 프로토콜별 추이분석, 일주 트래픽 추이분석, 그리고 MIB에서의 특정 속성과 속성 정보간의 상관관계를 이용하는 방법 등으로 분류된다. 그러나 이러한 방법론들은 대부분 테스트에 사용된 공격툴의 기능과 특성에 의존적으로 개발된 시스템으로, 새로운 공격 형태나 툴이 발견되면 그때마다 새롭게 알고리즘 전체를 수정해야하는 단점을 가지고 있다.

[0006] 최근의 연구문헌 조사에 의하면, 기계학습 기법과 SNMP MIB 정보를 이용한 매우 흥미로운 몇 개의 침입탐지 시스템이 발표되었다. 그 예로, SNMP MIB-II 데이터를 probability density function으로 변환한 후, backpropagation 기반의 인공신경망을 이용하여 침입 여부를 결정하는 시스템, SNMP MIB 정보를 Bayesian 분류기에 적용하여 Mobile Adhoc NETworks에서의 비정상 트래픽을 탐지 방법, Principle Component Analysis 기반의 비정상 탐지 알고리즘을 이용한 침입 탐지, Support Vector Machine을 이용하여 트래픽 폭주공격을 탐지하고 공격유형별 분류를 수행하는 시스템 등이 있다. 그러나 이러한 연구들은 모두 전통적인 DDoS 탐지 방법론의 단점을 해결한다는 기치아래 자칫 전통적 방법론의 장점을 간과할 수도 있다. 즉, 효율적인 시스템의 구축이라는 입장만을 견지하는 위의 기계학습론적 방법론은 시스템 작동 원리의 역학적 해석을 간과하여, 핵심 동작원리를 블랙-박스화하여 내부 메카니즘의 이해 및 해석, 규칙화가 어렵다.

[0007] 따라서, 본 발명에서는 휴리스틱한 방법론이긴 하나 전통적인 DDoS 탐지 방법론의 해석학적 장점도 고려할 수 있는 보다 포괄적인 시스템을 제안한다. 데이터마이닝의 대표적인 예측 및 분류 모델인 의사결정나무(decision tree) 중, C4.5 알고리즘을 기반으로 SNMP MIB 정보를 사용하여 트래픽 폭주공격을 탐지하고 각 공격유형별 분류를 수행하는 시스템을 설계 및 구현한다. 또한, 데이터의 전처리과정으로 SNMP MIB 정보에 대한 속성 부분집합의 선택 방법(attribute subset selection)을 사용하여 특징선택 및 축소(feature selection & reduction)를 실시한 후, 데이터마이닝의 대표적인 해석학적 분석 모델인 연관관계규칙기법(association rule mining)을 이용하여 트래픽 폭주 공격 및 공격유형별 SNMP MIB 정보에 내재되어 있는 특징들을 규칙의 형태로 추출하여 분석하는 의미론적 심층해석을 실시한다.

발명의 내용

해결하려는 과제

[0008] 본 발명의 일실시예는 DDoS 공격에 대한 의사결정나무 모델(C4.5) 기반의 신속한 탐지 및 공격유형별 분류, 연관관계규칙기법(association rule mining) 기반의 공격 탐지와 유형별 분류에 대한 의미적 심층 해석으로 과학적인 정책 근거를 제공함으로써, 보다 안정적인 네트워크 환경과 원활한 자원관리를 지원하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치 및 방법을 제공한다.

과제의 해결 수단

[0009] 상기의 일실시예를 이루기 위한, 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치는 네트워크 트래픽 데이터로부터 관리 정보 베이스(MIB: Management information base)를 생성하는 생성 모듈; 상기 관리 정보 베이스를 수집하여 탐지 시스템의 동작 시점을 결정하는 감지 모듈; 상기 관리 정보 베이스를 분석하는 상기 탐지 시스템에서 결정된 관리 정보 베이스를 저장하는 저장 모듈; 및 상기 결정된 관리 정보 베이스에 기반하여 공격유무와 공격유형을 판단하는 공격 판단 모듈을 포함한다.

[0010] 상기의 일실시예를 이루기 위한, 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 방법은 네트워크 트래픽 데이터로부터 관리 정보 베이스(MIB: Management information base)를 생성하는 단계; 상기 관리 정

보 베이스를 수집하여 탐지 시스템의 동작 시점을 결정하는 단계; 상기 관리 정보 베이스를 분석하는 상기 탐지 시스템에서 결정된 관리 정보 베이스를 저장하는 단계; 및 상기 결정된 관리 정보 베이스에 기반하여 공격유무와 공격유형을 판단하는 단계를 포함한다.

발명의 효과

[0011] 본 발명의 일실시예에 따르면, 데이터 마이닝의 예측 및 분석기법을 기반으로 트래픽 폭주 공격에 대한 신속한 탐지와 공격유형별 분류, 그리고 의미적 해석으로 보다 안정적인 서비스 제공 및 시스템 운용을 지원한다.

[0012] 또한, 본 발명의 일실시예에 따르면, 데이터 마이닝의 대표적인 예측 및 분류 모델인 의사결정나무의 C4.5 알고리즘을 기반으로 SNMP MIB 정보를 사용하여 트래픽 폭주 공격을 탐지하고, 각 공격유형별 분류를 수행하는 새로운 방안 제시한다.

[0013] 또한, 본 발명의 일실시예에 따르면, C4.5에서 추가적으로 제공하는 동작원리에 대한 규칙들을 추출하여 분석하고, IF-THEN 형태의 규칙 생성 및 의미론적 해석 방법을 제공한다.

[0014] 또한, 본 발명의 일실시예에 따르면, 연관관계규칙기법을 사용하여 공격 패턴 및 공격유형별 데이터 속에 내재되어 있는 유용한 지식의 발견과 심층적 분석으로 보다 안정적인 네트워크 환경과 원활한 자원관리를 지원한다.

[0015] 또한, 본 발명의 일실시예에 따르면, 트래픽 폭주 공격 탐지 및 공격유형별 분류, 공격 정보에 내재된 자동적 규칙 추출 및 의미론적 심층해석으로 침입탐지 및 침입대응 시스템을 위한 새로운 모멘텀을 제시할 수 있다.

도면의 간단한 설명

[0016] 도 1은 본 발명의 일실시예에서 제안하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치의 구성도이다.

도 2는 본 발명의 일실시예에 따른 의사결정나무 중 C4.5 기반의 트래픽 폭주 공격 탐지 및 공격 유형별 분류 방법을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0017] 이하에서, 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다. 그러나, 본 발명이 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.

[0018] 이하, 도면을 참조하여 본 발명의 실시 예에 따른 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치 및 방법에 대하여 설명한다.

[0019] 도 1은 본 발명의 일실시예에서 제안하는 데이터마이닝을 이용한 트래픽 폭주 공격 탐지 및 심층적 해석 장치의 시스템 구성도로써, 계층적 구조의 트래픽 폭주공격 탐지 및 분석을 위해 총 3개의 모듈로 구성된다. 온라인 처리 모듈에는 생성 모듈(110)(SNMP MIB generators), 감지 모듈(120)(MIB update detection), 저장 모듈(130)(MIB data store), 공격 판단 모듈(150)(Attack detection and classification), 오프라인 모듈에는 탐지 시스템(140)(C4.5 training)과 연관관계규칙 모듈(160)(Association rule mining), 마지막으로 관리자 모듈(170)에는 시스템 관리자로 구성된다.

[0020] 생성 모듈(110)에서는 인터넷의 네트워크 트래픽 데이터로부터 SNMP 프로토콜에 해당하는 MIB 정보를 생성한다.

[0021] 감지 모듈(120)은 생성 모듈(110)로부터 SNMP 프로토콜의 ifInOctets MIB 정보를 수집하여 탐지 시스템(140)의 동작 시점을 결정하고 저장 모듈(130)을 실행시킨다.

[0022] 저장 모듈(130)은 수집된 MIB 정보에서 타깃 시스템으로부터 탐지 시스템(140)(C4.5 training) 에서 결정된 MIB 정보만을 선택하여 저장한다.

[0023] 수집된 MIB 정보는 공격 판단 모듈(150)로 전달되고, 공격 판단 모듈(150)은 MIB 정보에 기반하여 실시간으로 공격유무와 공격유형을 판단한다.

[0024] 탐지 시스템(140)에서는 다양한 트래픽 공격을 임의로 발생시켜 의사결정나무 C4.5 기반의 학습을 실시한다.

[0025] 연관관계규칙 모듈(160)은 저장 모듈(130)에 저장된 MIB 정보의 특징들을 규칙의 형태로 추출하고 분석하는 의미론적 심층해석을 실시한다.

[0026] 관리자 모듈(170)은 공격 판단 모듈(150)을 통해 트래픽 폭주공격에 대한 실시간 탐지 및 분류유형에 대한 상세 정보를 모니터링하고, 탐지 시스템(140)의 C4.5 학습 및 연관관계규칙 모듈(160)에서 제공하는 규칙 및 의미론적 해석정보로 침입탐지 및 대응시스템의 정책 수립 등에 활용한다.

[0027] 도 2는 본 발명의 일실시예에 따른 의사결정나무 중 C4.5 기반의 트래픽 폭주 공격 탐지 및 공격 유형별 분류 방법을 나타낸다.

[0028] 공격 판단 모듈(150)은 두 가지 계층으로 구성되어 있다. 먼저, 첫 번째 계층은 정상 트래픽과 공격 트래픽을 분류하는 계층으로써 공격 트래픽이 탐지되면 침입대응시스템에 침입 사실을 관리자 모듈(170)을 통해 시스템 관리자에게 실시간으로 보고한다. 두 번째 계층은 트래픽 폭주 공격으로 판단된 공격트래픽을 TCP-SYN 플러딩(flooding), UDP 플러딩, ICMP 플러딩으로 각각 분류하고 탐지 시스템(140)인 침입 대응 시스템에 공격유형에 대한 추가적인 정보를 제공한다. 공격 판단 모듈(150)은 트래픽 폭주 공격을 유형별로 분류함으로써 공격이 발생한 프로토콜에 대해서만 서비스를 제한하고 관리할 수 있으므로 보다 안정적인 네트워크 환경과 원활한 자원 관리를 지원할 수 있다.

[0029] 의사결정나무는 데이터 마이닝 분야에서 분류 및 예측문제에 자주 사용되는 기법으로 변수들 간에 미치는 영향이나 상호작용을 누구나 쉽게 이해할 수 있는 방법론이다. 의사결정나무는 신경망구조 분석과는 달리 얻어진 지식을 표현하는 것이 직관적이며 손쉽게 규칙을 생성하기 때문에 분류 및 예측 모형을 제시하는데 주로 사용된다. 대부분의 의사결정 알고리즘은 하향식의 분할-정복(divide and conquer) 방법을 따르고 있으며, 훈련 데이터와 연관된 클래스 레이블로부터 모형을 구축한다. 의사결정나무의 대표적 알고리즘인 ID3는 많은 범위의 값을 갖는 속성이 상위노드로 선택되는 단점을 가지고 있기 때문에 본 발명에서는 가장 진보되고 분류 및 예측 성능이 이미 검증된 C4.5 의사결정나무 알고리즘을 사용한다.

[0030] 각 속성에 대한 엔트로피와 정보량은 아래의 [수학식 1]에서와 같이 얻는다.

수학식 1

[0031]
$$H(Y) = -\sum_{y \in Y} p(y)\log_2(p(y))$$

[0032]
$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x)\log_2\left(p(y|x)\right)$$

[0033] 따라서, 정보이익(information gain)은 아래의 [수학식 2]로 정의된다.

수학식 2

[0034]
$$Gain(Y) = H(Y) + H(X) - H(X,Y)$$

[0035] 엔트로피와 유사하게 정의되는 분리정보(split information)를 사용하여 정보이익을 아래의 [수학식 3]과 같이 정규화 한다.

수학식 3

[0036]
$$SplitInfo(Y) = -\sum_{i=1}^{n} \frac{|T_i|}{|T|} \times \log_2\left(\frac{T_i}{T}\right)$$

[0037] 아래의 [수학식 4]에서와 같이 최대 이익비(gain ratio)를 갖는 속성을 분리속성으로 선택한다.

수학식 4

[0038] $$\text{GainRatio}(Y) = \frac{\text{Gain}(Y)}{\text{SplitInfo} * Y)}$$

[0039] 본 발명에서는 데이터의 전처리 과정으로 SNMP MIB 정보에 대한 속성 부분집합의 선택 방법(attribute subset selection)을 사용하여 특징 선택 및 축소(feature selection & reduction)를 실시한다. 또한, 데이터마이닝의 대표적인 해석학적 분석 모델인 연관관계규칙기법(association rule mining)을 이용하여 트래픽 폭주 공격 및 공격유형별 SNMP MIB 정보에 내재되어 있는 특징들을 규칙의 형태로 추출하여 분석하는 의미론적 심층해석을 실시한다. SNMP MIB 정보에 대한 속성 부분집합의 선택 방법(attribute subset selection) 중 그 성능이 이미 검증된 Hall의 방법(M. Hall, "Correlation-based Feature Selection for Machine Learning", PhD Diss. Department of Computer Science, 1998)을 사용한다. 이는 최적우선탐색(best first search) 방법과 속성 혹은 특징(attribute or feature) 값에 대한 엔트로피(entropy), 목표 클래스(target class)와 속성들 간의 피어슨 상관 계수(Pearson's correlation coefficient)를 이용한 조건부 확률(conditional probability)을 계산하여 전체 속성들의 확률 분포도를 가능한 가깝게 표현할 수 있는 최소 개수의 속성집합을 찾는 방법이다. 먼저 각 속성들에 대한 정보 이익(information gain)을 얻기 위해 임의의 속성에 대한 엔트로피와 속성 X와 Y사이의 관계는 X가 주어졌을 때 Y가 발생하는 조건부 확률로써 [수학식 5]에서와 같이 계산된다.

수학식 5

[0040] $$H(Y) = -\sum_{y \in Y} p(y) log_2(p(y))$$

[0041] 속성 X와 Y사이의 관계는 X가 주어졌을 때 Y가 발생하는 조건부 확률로써 [수학식 6]과 같이 계산된다.

수학식 6

[0042] $$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) log_2 (p(y|x))$$

[0043] 각 특징에 대한 정보 이익은 [수학식 5]와 [수학식 6]을 이용하여 [수학식 7]로 정의된다.

수학식 7

[0044] $$Gain = H(Y) + H(X) - H(X, Y)$$

[0045] [수학식 7]에서 얻은 정보 이익을 기반으로 [수학식 8]에서와 같이 symmetrical uncertainty를 이용하여 임의의 두 속성 X와 Y의 분포와 상관관계를 계산한다. 이때 속성 X를 기준으로 Y 가 높은 분포와 상관관계를 보이면 전체 속성들을 효율적으로 표현할 수 있는 부분집합에 속성 X는 포함되지만 Y는 포함되지 않는다. 마찬가지로 목표 클래스와 속성들 간의 분포와 상관관계를 계산하여 부분집합을 구성한다.

수학식 8

[0046] $$Symmetrical\ uncertainty\ coefficient = 2.0 \times \left[\frac{Gain}{H(Y)+H(X)}\right]$$

[0047] 각각의 부분집합 $F_S \subset F F_S \subset F$ 가 전체 속성들을 얼마나 효율적으로 표현하는지를 평가하기 위하여 메리트 함수(merit function)([수학식 9])를 사용한다. 메리트 함수의 값이 가장 큰 부분집합이 전체 속성들을 최적으로 표현할 수 있는 부분집합으로 결정된다

수학식 9

[0048] $$Merit(F_S) = \frac{k\overline{r_{cf}}}{\sqrt{k+k(k-1)\,\overline{r_{ff}}}}$$

[0049] 여기서, k는 부분집합 $F_S F_S$ 에서의 속성들의 개수를 의미하며, $\overline{r_{cf}}\,\overline{r_{cf}}$ 는 $F_S F_S$ 에 포함된 속성의 평균 분포(contribution), $\overline{r_{ff}}\,\overline{r_{ff}}$ 는 속성의 평균 상관관계 값을 의미한다.

[0050] 연관관계규칙 마이닝(association rule mining)이란 데이터 안에 존재하는 각 객체들 간의 의미 있는 연관관계를 찾아내는 방법론으로, 연관관계규칙은 $A \& B \Rightarrow C A \& B \Rightarrow C$ 와 같이 조건 명제의 형태로 표현된다. 전체 데이터 셋 D 가 있다고 가정할 때, 연관관계규칙 마이닝을 그대로 수행하게 되면 데이터 셋 D 로부터 거의 사용되지 않거나 중복된 성질을 갖는 데이터를 모두 포함하면 의미없는 많은 규칙들을 생성하게 된다. 때문에 연관관계규칙 마이닝을 수행하기 전 상기에서 언급한 속성 부분집합의 선택 방법(attribute subset selection)으로 사용되지 않거나 중복된 성질의 특징을 제거한 속성 부분집합 d 를 찾은 후 연관관계규칙 마이닝을 수행하여 데이터에 내재된 규칙 추출 및 심층적인 해석을 시도하게 된다. 먼저 항목들의 집합 $I = \{I_1, I_2, I_3, \ldots, I_m\} I = \{I_1, I_2, I_3, \ldots, I_m\}$ 와 각각의 트랜잭션 T 는 $T \subseteq I T \subseteq I$ 의 관계를 가진 항목들의 집합이 있을 때, 각각의 트랜잭션 T 는 고유한 트랜잭션 구분자(transaction identifier)를 갖는다. A 를 항목들의 집합이라고 하면, 트랜잭션 T 가 필요충분조건으로 $A \subseteq T \subseteq T$ 를 만족하는 경우에만 트랜잭션 T 가 항목 A 를 포함한다고 한다. 여기서 $A \subset I, B \subset I, A \cap B = \emptyset. A \subset I, B \subset I, A \cap B = \emptyset.$ 을 만족하는 경우, 연관규칙은 $R: A \Rightarrow B R: A \Rightarrow B$ 의 형식으로 표현되며, A 를 규칙의 조건부(antecedent), B 를 결과부(consequent)라 한다. 추출된 연관규칙의 평가 기준으로는 지지도(support)와 신뢰도(confidence)를 사용한다. 규칙 $A \Rightarrow B A \Rightarrow B$ 는 트랜잭션 집합 D에서 집합 A 와 B 를 동시에 포함하는 트랜잭션의 백분율이 S 경우 지지도 S 를 갖는다고 표현한다. 이는 확률 $P(A \cup B). P(A \cup B).$ 를 계산함으로써 얻을 수 있다. 집합 A 를 포함하는 트랜잭션 중에서 집합 B 도 포함하고 있는 트랜잭션의 백분율이 C 인 경우, 규칙 $A \Rightarrow B A \Rightarrow B$ 는 신뢰도 C 를 갖는다고 표현한다. 신뢰도는 조건부확률 $P(B|A)\ P(B|A)$ 를 계산함으로써 얻을 수 있다([수학식 10] 참조).

수학식 10

[0051] $$support(A \Rightarrow B) = P(A \cup B)$$

[0052] $$confidence(A \Rightarrow B) = P(B|A)$$

[0053] 여기서, 최소 지지도 임계값(minimum support threshold)과 최소 신뢰도 임계값(minimum confidence

threshold)을 동시에 만족하는 규칙을 강한(strong) 규칙이라고 한다. 이때 최소 지지도 값 이상을 갖는 항목집합(itemset)을 빈발항목집합(frequent itemset)이라 하고 K 개의 항목들로 이루어진 빈발항목집합을 K-빈발항목집합이라고 한다. 이진 연관관계규칙에 대한 빈발항목집합을 찾는데 유용한 Apriori 알고리즘은 K 번째 항목 집합이 $(K+1)$ 번째 항목집합을 발견하기 위해 사용되는 반복적 접근방법을 사용하는데, 이는 수준별(level-wise) 방법으로 알려져 있다. 먼저, 빈발 1-항목집합을 L_1L_1로 나타내며, L_2L_2은 2-항목집합인 L_2L_2를 찾는데 사용되고 이것은 다시 L_3L_3를 찾는데 이용된다. 이러한 방법은 더 이상 빈발 K-항목집합이 없을 때까지 진행한다.

[0054] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다. 따라서 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 구성들은 본 발명 사상의 범주에 속한다고 할 것이다.

부호의 설명

[0055] 110 : 생성 모듈

120 : 감지 모듈

130 : 저장 모듈

140 : 탐지 시스템

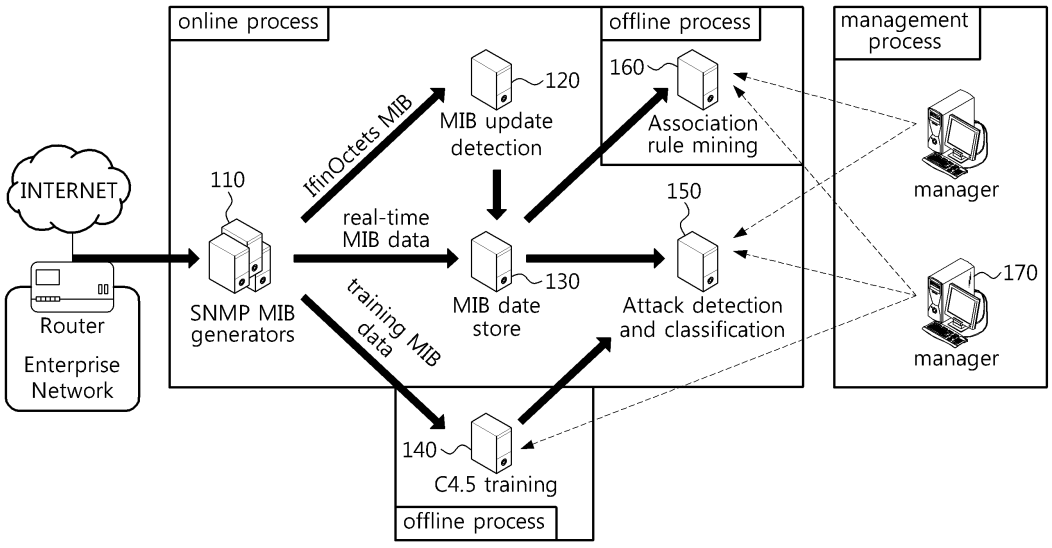
150 : 공격 판단 모듈

160 : 연관관계규칙 모듈

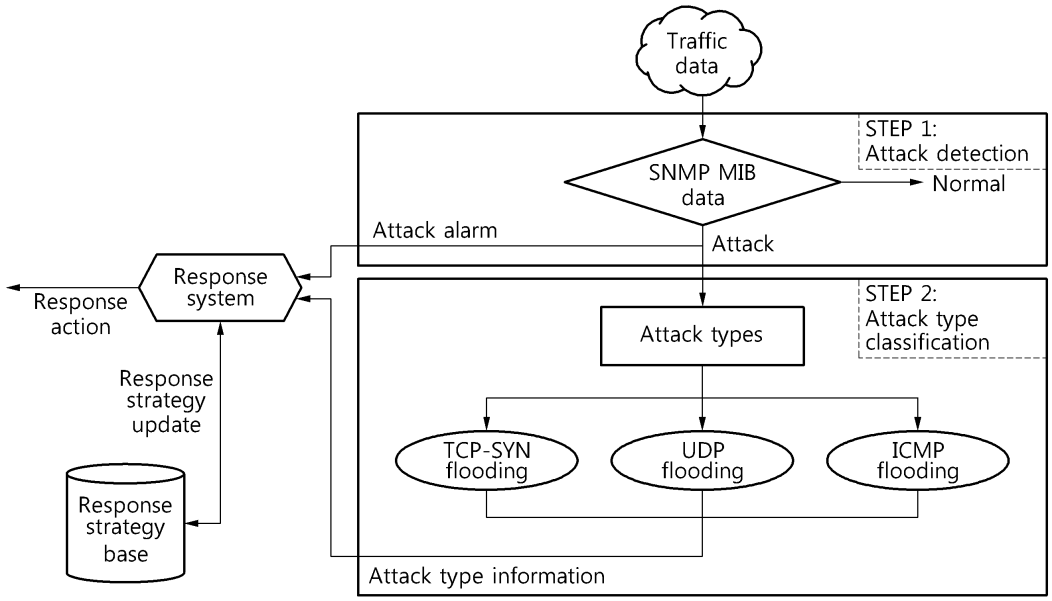
170 : 관리자 모듈

도면

도면1



도면2


Traffic data
STEP 1: Attack detection
SNMP MIB data
Normal
Attack alarm
Attack
Response system
Response action
STEP 2: Attack type classification
Attack types
Response strategy update
TCP-SYN flooding
UDP flooding
ICMP flooding
Response strategy base
Attack type information