Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.28, NO.2, Apr. 2018 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2018.28.2.385

기계학습 기반 IDS 보안이벤트 분류 모델의 정확도 및 신속도 향상을 위한 실용적 feature 추출 연구*

신 익 수,[1,2†] 송 중 석,[1,2] 최 장 원,[2] 권 태 웅[2‡]
[1]과학기술연합대학원대학교, [2]한국과학기술정보연구원

A Practical Feature Extraction for Improving Accuracy and Speed of IDS Alerts Classification Models Based on Machine Learning*

Iksoo Shin,[1,2†] Jungsuk Song,[1,2] Jangwon Choi,[2] Taewoong Kwon[2‡]
[1]University of Science & Technology,
[2]Korea Institute of Science & Technology Information

요 약

인터넷의 성장과 함께 각종 취약점을 악용한 사이버 공격들이 지속적으로 증가하고 있다. 이러한 행위를 탐지하기 위한 방안으로 침입탐지시스템(IDS; Intrusion Detection System)이 널리 사용되고 있지만, IDS에서 발생하는 많은 양의 오탐(정상통신을 공격행위로 잘못 탐지한 보안이벤트)은 여전히 해결되지 않은 문제로 남아있다. IDS 오탐 문제를 해결하기 위한 방법으로 기계학습 알고리즘을 통한 자동분류 연구가 진행되고 있지만 실제 현장 적용을 위해서는 정확도와 데이터 처리속도 향상을 위한 연구가 더 필요하다. 기계학습 기반 분류 모델은 다양한 요인에 의해서 그 성능이 결정된다. 최적의 feature를 선택하는 것은 모델의 분류 성능 및 정확성 향상에 크게 영향을 미치기 때문에 기계학습에서 매우 중요한 부분을 차지한다. 본 논문에서는 보안이벤트 분류 모델의 성능 향상을 위해 기존 연구에서 제안한 기본 feature에 추가로 10종의 신규 feature를 제안한다. 본 논문에서 제안하는 10종의 신규 feature는 실제 보안관제센터 전문 인력의 노하우를 기반으로 고안된 것으로, 모델의 분류 성능을 향상시킬 뿐만 아니라 단일 보안이벤트에서 직접 추출 가능하기 때문에 실시간 모델 구축도 가능하다. 본 논문에서는 실제 네트워크 환경에서 수집된 데이터를 기반으로 제안한 신규 feature들이 분류 모델 성능 향상에 미치는 영향을 검증하였으며, 그 결과, 신규 feature가 모델의 분류 정확도를 향상시키고 오탐지율을 낮춰주는 것을 확인할 수 있었다.

ABSTRACT

With the development of Internet, cyber attack has become a major threat. To detect cyber attacks, intrusion detection system(IDS) has been widely deployed. But IDS has a critical weakness which is that it generates a large number of false alarms. One of the promising techniques that reduce the false alarms in real time is machine learning. However, there are problems that must be solved to use machine learning. So, many machine learning approaches have been applied to this field. But so far, researchers have not focused on features. Despite the features of IDS alerts are important for performance of model, the approach to feature is ignored. In this paper, we propose new feature set which can improve the performance of model and can be extracted from a single alarm. New features are motivated from security analyst's know-how. We trained and tested the proposed model applied new feature set with real IDS alerts. Experimental results indicate the proposed model can achieve better accuracy and false positive rate than SVM model with ordinary features.

Keywords: Network security, IDS, false alarm, machine learning, SVM

Received(02. 01. 2018), Modified(03. 05. 2018), Accepted(03. 05. 2018)

* 본 연구는 2018년도 한국과학기술정보연구원(KISTI) 주요사업 과제로 수행한 것입니다.

† 주저자, iksooman@kisti.re.kr
‡ 교신저자, taewoong.kwon@kisti.re.kr(Corresponding author)

I. 서 론

최근 몇 십년간 인터넷은 급격히 성장해왔고, 인간의 삶에 있어 없어서는 안 될 한 부분으로 자리매김하였다. 인터넷의 발달과 함께 인터넷 자원의 취약점을 악용한 범죄 형태인 사이버 공격이 모습을 보였고, 점차 그 형태는 다양해졌다. 사이버 공격으로부터 정보 자산을 보호하기 위해 보안 시스템들이 필요하게 되었고, 방화벽(Firewall), 침입탐지시스템(IDS; Intrusion Detection System), 안티바이러스(AV; Anti-Virus) 등 다양한 종류의 보안 시스템들이 만들어졌다.

여러 보안 시스템 가운데 IDS는 많은 기관에서 활용하는 기본적인 보안장비로, 통신 패킷 분석을 통해 사이버 공격을 탐지하고 보안이벤트를 발생시키는 시스템이다[1]. IDS가 이러한 행위를 탐지하는 방법에는 크게 2가지가 있다. 첫 번째는 오용탐지로, 오용탐지는 사용자가 등록한 문자열이 패킷 상에서 발견되었을 때, 보안이벤트를 발생시켜 사용자에게 알려주는 방식이다. 두 번째는 이상탐지로, 이상탐지는 정상적인 통신 방식과 다른 형태의 통신이 나타났을 때, 보안이벤트를 발생시켜 사용자에게 알려주는 방식이다.

IDS는 세계적으로 널리 사용되는 시스템이지만 오탐이 많다는 단점을 가지고 있다[2]. 오탐은 정상적인 통신을 공격행위로 잘못 탐지한 보안이벤트를 말하며, 반대로 공격 행위를 정확히 탐지한 보안이벤트는 정탐이라 한다. 정탐과 오탐을 구분해내기 위해서는 추가적인 분석 작업이 필요하며, 주로 네트워크 운영 기관의 보안관제센터(SOC; Security Operating Center)에서 이러한 업무를 수행한다. SOC에서는 실시간으로 발생하는 IDS 보안이벤트를 분석하고, 공격 상황으로 판단 시, 즉각적인 대응을 수행하고 있다. 하지만 IDS에서 발생하는 대량의 오탐은 모든 보안이벤트에 대한 분석을 불가능케 하고 공격에 대한 즉각적이고 정확한 대응을 어렵게 한다.

대량의 IDS 오탐 문제를 해결하기 위해 다양한 방면으로 연구가 진행되고 있으며, 그 중 하나로 기계학습을 이용한 방법이 있다[3]. 기계학습은 컴퓨터가 기존 데이터에 대한 학습을 통해 신규 데이터에 대한 분류작업을 수행할 수 있도록 해주는 기술이다. 기계학습을 이용하는 이 연구들의 목표는 기계학습 모델을 학습시켜 신규 발생 보안이벤트를 자동으로 분류하게 하는 것이다. 기계학습 알고리즘을 통해 복잡한 데이터의 패턴을 파악하고 자동으로 보안이벤트를 처리할 수 있기 때문에 그 활용이 매우 기대되고 있다.

하지만 기계학습 알고리즘을 실시간 보안이벤트 분류에 적용하기 위해서는 몇 가지 문제점들을 해결해야 한다. 보안이벤트 분류 모델로서 가장 중요한 것은 정확한 보안이벤트 분류 능력이다. 알고리즘이 분류한 결과에 대해 신뢰할 수 있어야만 실전 적용이 가능하기 때문이다. 또한, 변화하는 공격 방법에 대한 지속적인 학습과 사고에 대한 즉각적인 대응이 가능해야하기 때문에 실시간 학습 및 분류를 위한 데이터 처리 속도도 해결해야할 과제 중 하나이다.

본 논문에서는 이러한 문제점들을 극복하고 향후 실시간 분류가 가능한 기계학습 모델 구축을 위해 feature에 주목한다. feature는 기계학습 모델이 데이터를 분류하기 위해 사용하는 데이터의 정보 혹은 속성으로, 사용하는 feature에 따라 기계학습 알고리즘의 분류 성능이 크게 달라진다. 본 논문에서는 기존 논문에서 사용하지 않았던 새로운 feature를 제안한다. 제안하는 feature는 실시간 단일 보안이벤트에서도 쉽게 추출이 가능하도록 고안되었고, 분류 성능을 높일 수 있도록 보안관제요원의 노하우를 기반으로 만들어졌다. 제안 feature의 모델 개선 효과를 증명하기 위해 새로운 feature를 적용한 SVM 모델과 기존 feature만을 적용한 SVM 모델의 비교 결과도 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 기계학습 기반 IDS 보안이벤트 분류 관련 연구들에 대해 소개하고, 3장에서는 본 논문에서 제안하는 feature에 대해 소개한다. 4장에서는 제안된 feature의 성능을 평가하기 위한 실험 방법에 대해 설명하고, 5장에서 실험 결과를 분석한다. 마지막으로 6장에서는 본 논문의 결론과 향후 연구방향에 대해 기술한다.

II. 관련 연구

지난 몇 십년간 IDS의 오탐 문제를 해결하기 위해 다양한 방면의 시도가 있어왔다[4]. IDS 탐지룰 개선을 통해 오탐을 줄이는 연구, 공격 시나리오를 이용하여 보안이벤트 간의 상관관계를 찾아내는 연구, 공격의 성공 여부를 확인하여 정탐을 찾아내는 연구, 보안이벤트를 시각화하여 관제요원의 분석을

돕는 연구 등 다양한 종류의 연구가 수행되었다[5][6][7][8].

기계학습을 적용한 보안이벤트 분류 연구도 그 중 하나이다. 보안이벤트 분류를 위해 다양한 기계학습 모델들이 적용되어 왔으며, 관련 연구 초기인 2000년대 초중반에는 Decision Tree와 클러스터링 알고리즘을 적용한 연구들이 주를 이루었다[9][10]. 그 이후, Neural network, Bayesian network, SVM을 적용한 방법들이 시도되었으며, 2010년부터는 더욱 정확한 보안이벤트 분류를 위해 기존 모델에 준지도 학습 알고리즘, active learning과 같은 방법 등을 적용한 연구들이 시도되고 있다[11][12][13][14][15][16][17][18][19].

기계학습을 적용한 최신연구 중 일부를 살펴보면, Meng과 Kwok[15]은 기존 연구된 기계학습 모델들을 상황에 따라 적합한 것으로 바꿔가며 사용하는 방식을 제안하였다. 그들은 먼저 6가지 기계학습 모델을 선택하여 모델들의 보안이벤트 분류 성능을 테스트하였다. 6가지 모델 중 보안이벤트 분류에 가장 좋은 성능을 보인 kNN, Decision Tree, SVM 3가지 지도학습 모델을 사용하여 지속적인 성능 모니터링을 통해 상황별로 가장 성능이 우수한 알고리즘을 선택하는 방식을 사용하였다. 실험을 통해 그들은 상황별로 적합한 모델을 선택하는 것이 전체적인 분류 성능을 유지하는 데 도움을 준다는 것을 보여주었다.

또한, Meng과 Kwok[17]은 그들의 다음 연구에서 준지도 학습 알고리즘을 활용한 방법도 제안하였다. 기계학습 모델 중 지도학습 모델은 클래스가 미리 분류된 데이터만 학습에 사용할 수 있다. 하지만 IDS 보안이벤트 데이터의 클래스는 사람이 직접 분류해야하기 때문에 활용할 수 있는 자료가 매우 제한적이다. 한편, 준지도 학습 모델은 클래스가 분류된 데이터와 분류되지 않은 데이터를 모두 사용할 수 있기 때문에 많은 데이터를 학습에 사용할 수 있다는 장점이 있다. 그들의 실험 결과는 클래스가 분류된 데이터를 많이 사용하지 않는 준지도 학습 알고리즘이 오탐 감소 측면에서 지도학습 모델보다 좋은 성능을 보일 수 있다는 가능성을 보여주었다. 준지도 학습 방법은 Li 등[19]에 의해 다시 활용되었으며, multi-view 방식과 결합한 그들의 준지도 학습 모델은 기존 모델들에 비해 향상된 분류 정확도를 보여주었다.

한편, Liang 등[20]은 기계학습 알고리즘 중 클러스터링 알고리즘을 적용한 오탐 감소 연구를 수행하였다. 그들이 적용한 클러스터링 알고리즘은 k-means와 FCM 알고리즘이다. 그들은 DARPA 2000 LLDOS 1.0 데이터와 snort를 사용하여 실험 데이터를 구축하였고, 2가지 클러스터링 알고리즘을 테스트하였다. 그들의 실험 결과는 클러스터링 알고리즘의 활용이 오탐 감소에 매우 효과적임을 보여주고 있다.

Teemu 등[21]도 오탐 감소를 위해 클러스터링 방법을 사용하였다. 그들은 k-means 알고리즘을 사용하여 보안이벤트 클러스터를 만들고, 신규 보안이벤트가 큰 규모의 클러스터에 포함되면 오탐으로 분류하는 방식을 제안하였다. 이는 다량으로 발생하는 보안이벤트는 오탐이라는 것을 가정으로 했기 때문이다. 그들은 '보안이벤트 종류별 발생 횟수', '보안이벤트 종류별 관련 IP 개수' 2가지 feature를 가지고 모델을 구축하였으며, 이를 통해 관제요원이 관심을 가져야할 보안이벤트를 분류할 수 있다고 제안한다.

이처럼 다양한 방면으로 보안이벤트 분류를 위한 기계학습 알고리즘 연구가 진행되고 있지만 그들이 모델에 사용하고 있는 feature는 단순히 IDS 보안이벤트가 가지고 있는 기본적인 feature이거나 보안이벤트 간의 상관관계 정보와 같이 실시간 보안이벤트 분류에 적합하지 않은 feature들이었다. IDS 보안이벤트의 기본적인 feature는 보안이벤트의 클래스 분류를 위해 고려된 정보가 아니기 때문에 분류 모델에 적합하지 않다. 또한 일부 연구에서 사용한 보안이벤트 간 상관관계 관련 feature는 사후 분석으로 사용해야하기 때문에 보안사고 발생에 즉각 대응할 수 없다. 따라서 본 논문에서는 IDS 보안이벤트의 feature에 집중하여 실시간 보안이벤트 분류에 적합하면서 기계학습 알고리즘의 분류 정확도를 높일 수 있는 feature를 제안하고 성능을 확인해보고자 한다.

III. 신규 feature 10종

기계학습 모델의 분류 정확도 향상과 신속한 분석을 위해 본 논문에서 제안하는 신규 feature 10종에 대해 소개한다. 10종의 feature 목록은 Table 1에서 확인할 수 있으며, 각 feature에 대한 설명은 아래에서 확인할 수 있다.

Table 1. 10 new features list

No	Feature
1.	Is source IP in the target network?
2.	Is destination IP in the target network?
3.	Does the payload have 'Referer'?
4.	Does the payload have '200 OK'?
5.	How many does the payload have security-related strings?
6.	The TTL value in the payload
7.	The length of the payload
8.	Does web-server use common port?
9.	Which form does the payload use for 'Host'?
10.	What kind of 'User-agent' does the payload use?

1. Is source IP in the target network?

출발지 IP가 내부 IP인지 여부를 나타내주는 feature이다. 예를 들어, 악성코드에 감염되어 감염 신호를 전송하는 패킷을 탐지하는 보안이벤트의 경우, 감염된 PC에서 외부로 보내는 패킷을 탐지하기 때문에 출발지 IP가 내부기관이어야 정탐이다. 출발지가 내부기관인지 여부에 따라 0 혹은 1로 표기된다.

2. Is destination IP in the target Network?

도착지 IP가 내부 IP인지 여부를 나타내주는 feature이다. 예를 들어, SQL Injection 공격의 경우, 외부에서 내부 기관 서버로 SQL 쿼리문이 포함된 페이로드를 보낸다. 따라서 도착지 IP가 내부기관인 패킷을 탐지한 보안이벤트가 정탐이라고 볼 수 있다. 내부기관인지 여부에 따라 0 혹은 1로 표기된다.

3. Does the payload have 'Referer'?

http 프로토콜 관련 페이로드에 Referer가 존재하는지 여부를 나타내주는 feature이다. 내부기관에서 악성 URL로 접근하는 것을 탐지하는 보안이벤트의 경우, Referer가 존재하지 않는다면 악성코드에 의한 직접적인 접근일 가능성이 높다. 하지만 Referer가 존재한다면 사용자의 웹서핑 도중 탐지되었을 가능성이 높다. Referer가 존재하는지 여부에 따라 0 혹은 1로 표기되며. http 프로토콜 통신이 아니면 0으로 표기된다.

4. Does the payload have '200 OK'?

http 프로토콜 관련 페이로드에 200 OK 문자열이 존재하는지 여부를 나타내주는 feature이다. IDS 보안이벤트의 경우, 사용자가 등록한 악성행위 관련 내용이 패킷 상에 포함되어 있다. 200 OK 문자열이 포함되었을 시에 공격이 성공했을 가능성이 높다. 200 OK 문자열이 포함되어 있는지 여부에 따라 0 혹은 1로 표기되며, http 프로토콜 통신이 아니면 0으로 표기된다.

5. How many does the payload have security-related strings?

페이로드 내에 포함된 보안관련 문자열 개수를 보는 feature이다. ID, PASSWORD와 같은 개인정보나 CPU, 운영체제 버전 정보 등의 시스템 정보는 해커가 해킹을 하기 위해 필요로 하는 정보이다. 이러한 문자열이 많이 포함될수록 악성행위로 인해 정보가 유출되는 패킷일 확률이 높아진다. 문자열 포함 개수에 따라 0 이상의 정수로 표기된다.

6. The TTL value in the payload

탐지된 패킷의 TTL 값을 나타내주는 feature이다. 악성코드의 경우 운영체제에 따라 코드의 실행이 가능하지 않을 수 있다. 각 운영체제는 서로 다른 최대 TTL 값을 사용하기 때문에 이를 통해 운영체제에 대한 정보를 얻을 수 있다. 또한, 과거 연구에서 TTL 값을 이용한 분류가 IDS 오탐을 줄이는 데 있어 효과가 있음을 보여준 사례가 있다[22]. 페이로드의 TTL 값에 따라 0 이상의 정수로 표기된다.

7. The length of the payload

탐지된 패킷의 페이로드 길이를 나타내주는 feature이다. 버퍼 오버플로우 공격의 경우, 페이로드에 일반적이지 않은 길이의 문자열을 다수 포함하여 전달한다. 이 정보를 통해 페이로드 길이가 일반적이지 않은 패킷을 분류할 수 있다. 패킷길이에 따라 0 이상의 정수로 표기된다.

8. Does web-server use common port?

웹서버가 일반적인 포트로 통신하는지 여부를 나타내주는 feature이다. 웹서버의 경우, 외부로 정보를 제공하는 서버이기 때문에 일반적으로 80 혹은 8080 포트를 사용한다. 웹서버가 이외에 다른 포트를 사용하여 통신한다면 의심스러운 행위가 된다. 웹서버가 80 혹은 8080포트로 통신하는지 여부에 따라 0 혹은 1로 표기되며, http 프로토콜 통신이 아니면 0으로 표기된다.

9. Which form does the payload use for

'Host'?

http 프로토콜 관련 페이로드에 Host가 IP 형태인지 여부를 나타내주는 feature이다. 일반적인 사용자는 웹브라우저 사용 시에 URL을 통한 웹서버 접근을 시도한다. 하지만 IP를 통한 직접적인 서버 접근을 시도한다면, 악성코드에 의한 악성 서버로의 접근을 의심할 수 있다. Host 정보가 IP인지 URL인지 여부에 따라 0 혹은 1로 표기되며, http 프로토콜 통신이 아니면 0으로 표기된다.

10. What kind of 'User-agent' does the payload use?

http 프로토콜 관련 페이로드의 User-agent 종류에 대한 정보이다. http 프로토콜 통신 시에 다양한 종류의 User-agent가 사용된다. 정상적인 브라우저와는 달리 악성코드에서는 일반적으로 사용되지 않는 User-agent를 사용한다. User-agent별로 번호를 할당하고 할당된 번호가 없는 새로운 User-agent는 새로운 번호를 부여한다.

이상의 10가지 feature는 보안관제요원의 보안이벤트 분류 업무 수행 시 사용되는 정보로 이루어져 있다. 분류 업무 수행 시 사용하는 정보 중 단일 보안이벤트에서 추출 가능한 정보로 feature를 구성했고, 모델에 입력 가능하도록 수치적으로 표현하였다. 제안한 feature의 성능을 확인하기 위해 직접 모델에 적용하고 분류 성능을 실험하였으며, 실험 방법과 결과를 4절과 5절에서 기술한다.

IV. 실험 방법

4.1 실험 개요

제안 feature의 성능을 확인하기 위한 본 논문의 실험 개요는 Fig. 1과 같다. 먼저 수집된 IDS 보안이벤트 데이터를 모델 학습용 데이터와 모델 테스트용 데이터로 분류한다. 이후 학습 및 테스트 데이터에서 feature를 추출한다. 이 때 feature는 IDS 기본 feature 7개에 신규 feature 10개를 추가하여 총 17개를 사용한다. 각 데이터에서 추출한 feature는 정규화를 거친다. 정규화 작업이 끝난 학습 데이터는 기계학습 모델에 입력되어 모델을 학습시킨다. 학습이 완료된 모델에 테스트 데이터를 넣고 분류된 결과를 확인한다. 최종적인 분류 결과는 정확도, 오탐지율 등의 평가지수를 통해 기존 feature만을 적용한 모델 결과와 비교한다. 4.2절부터는 각 단계에 대해 자세히 설명한다.

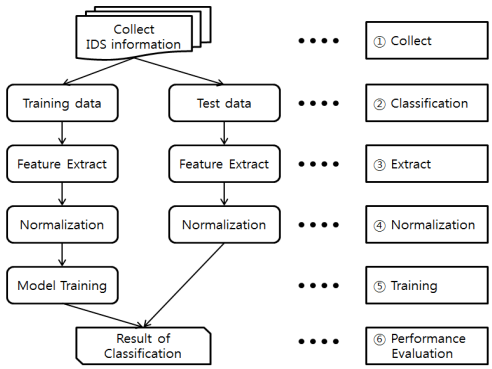

Fig. 1. Experiment process

4.2 수집/분류

수집/분류 단계에서는 IDS에서 탐지된 raw 데이터를 확보하고 모델 학습에 사용할 데이터와 테스트에 사용할 데이터로 분류한다. 이 때 raw 데이터는 실제 네트워크에서 탐지된 데이터를 사용하였다. 데이터에 대한 자세한 내용은 5.1절에서 설명한다.

4.3 추출

추출 단계에서는 각각의 보안이벤트에서 보안이벤트가 가지고 있는 feature를 추출해낸다. 추출할 feature는 앞서 3절에서 설명한 10개 신규 feature와 보안이벤트가 기본적으로 가지고 있는 feature 7개이다. 기본 feature 7개는 Table 2에서 확인할 수 있다.

4.4 정규화

기계학습 모델 중 다수는 데이터 간의 거리를 통해 클래스를 분류한다. 따라서 데이터 간의 거리가

Table 2. 7 basic features list

1. Source IP	5. Event class
2. Destination IP	6. Priority
3. Source Port	7. Protocol
4. Destination Port	

클래스 분류에 있어 중요한 역할을 한다. 값의 범위가 서로 다른 feature들을 통해 분류를 수행하기 위해서는 feature 간의 영향을 맞춰주기 위한 정규화를 수행해주어야 한다. 이번 실험에 사용하는 기계학습 모델은 SVM으로, SVM 또한 데이터 간의 거리를 통해 각 데이터의 클래스를 나누는 모델이기 때문에 데이터를 모델에 적용하기 전 정규화는 필수적인 작업이다. 본 논문에서 데이터 정규화를 위해 사용한 방식은 standard score이다. standard score 방식은 임의의 변수가 해당 feature 집합의 평균에서 얼마나 떨어져 있는지를 보여주는 지수로, 계산식은 식(1)과 같다. 여기서 x_i는 standard score를 적용할 변수이고, m은 해당변수가 포함된 feature 집합의 평균, σ는 해당변수가 포함된 feature 집합의 표준편차이다.

$$Standard\,Score = \frac{x_i - m}{\sigma} \quad (1)$$

4.5 학습

본 논문에서는 신규 feature의 성능을 확인하기 위한 모델로 SVM(Support Vector Machine)을 사용하였다. SVM은 데이터를 두 가지 클래스로 분류하는 모델로, Vapnik에 의해 처음 제안되었다[23]. SVM은 다양한 분야에서 널리 사용되는 기계학습 모델이며, IDS 오탐 분류 수행 시 타 모델에 비해 우수한 성능을 보여주었다[15]. 모델의 세부구조에 대한 설명은 이 논문의 주제에서 벗어난다고 판단되기 때문에 생략한다.

SVM 모델을 학습시키기 위해서는 커널과 파라미터를 먼저 선택하여야 한다. 본 논문에서 사용한 커널은 RBF(Radial Basis Function) 커널이다. RBF 커널은 전반적인 분야에서 타 커널에 비해 좋은 성능을 가지고 있으며, 사용하기도 용이한 편이다[24]. 파라미터인 c와 gamma 값은 모델 학습 과정 중 n-fold cross validation을 적용한 여러 번의 실험을 수행하여 가장 성능이 우수한 값으로 선택하였다.

본 논문에서는 SVM 모델을 사용하기 위해 LIBSVM 라이브러리를 참고하였다[25]. LIBSVM은 다양한 형태의 SVM 모델을 사용할 수 있는 소프트웨어로 클래스 분류를 위한 SVC(Support Vector Classification)나 회귀분석을 위한 SVR(Support Vector Regression) 등 SVM 관련 여러 가지 기능을 제공한다.

4.6 성능 평가

성능 평가 단계에서는 학습된 모델에 테스트 데이터를 입력하여 분류된 결과를 확인한다. 학습된 모델의 최종적인 분류 성능을 평가하기 위해 정확도, 탐지율, 오탐지율, F1 score 총 4가지 평가지수를 활용하였다. 각 데이터는 데이터의 실제 클래스와 학습된 모델이 분류한 클래스에 따라 Table 3과 같이 구분된다. 실제 정탐에 대해 모델이 정탐으로 분류하면 TP(True Positive), 오탐으로 분류하면 FN(False Negative), 실제 데이터가 오탐인데 모델이 정탐으로 분류하면 FP(False Positive), 오탐으로 분류하면 TN(True Negative)로 구분된다.

다음은 각 평가지수와 평가지수 산출 방법에 대한 설명이다. 산출식에 사용된 변수는 Table 3의 내용을 기준으로 하고 있다.

Table 3. Types of classification of data

	True Alarm (Classified)	False Alarm (Classified)
True Alarm (Actual)	TP	FN
False Alarm (Actual)	FP	TN

4.6.1 정확도

정확도는 정탐과 오탐을 합친 전체 테스트 데이터 가운데 정확히 분류된 데이터의 비율을 의미하며 높을수록 좋은 성능을 나타낸다. 계산식은 식(2)와 같다.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

4.6.2 탐지율

탐지율은 정탐 가운데 정확하게 정탐으로 분류된 데이터의 비율을 의미하고 높을수록 좋은 성능을 나타낸다. 계산식은 식(3)과 같다.

$$True\ Positive\ Rate = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

4.6.3 오탐지율

오탐지율은 오탐 가운데 정탐으로 잘못 분류된 데이터의 비율을 의미하고 낮을수록 좋은 성능을 나타낸다. 계산식은 식(4)와 같다.

$$False\ Positive\ Rate = \frac{FP}{FP+TN} \times 100\% \quad (4)$$

4.6.4 F1 score

F1 score는 정탐과 오탐 모두에 대한 분류 성능을 나타내주는 지표이며, 높을수록 좋은 성능을 나타낸다. 계산식은 식(5)와 같다.

$$F1\ score = \frac{2*TP}{2*TP+FP+FN} \quad (5)$$

V. 실험 결과

5.1 실험 데이터

본 논문에서 사용한 IDS 보안이벤트 데이터는 과학기술사이버안전센터에서 탐지하고 분석한 결과 데이터이다. 과학기술사이버안전센터는 약 3,000개의 탐지룰을 적용하여 국내 연구기관들에서 발생하는 보안이벤트를 24시간 모니터링하고 있으며 TMS(Threat Management System)라는 IDS를 사용하고 있다. 과학기술사이버안전센터에서 분석한 실제 IDS 보안이벤트 중 분석이 완료된 2017년 7월 1일과 2일 보안이벤트 데이터 일부를 이번 실험에 사용하였다. 과학기술사이버안전센터에서는 매일 탐지룰이 업데이트되기 때문에 비교적 동일한 탐지룰이 적용된 1일과 2일 데이터를 사용하였다. 일자별로 각각 학습용, 테스트용 데이터로 사용하였고, 각 일자별 정탐과 오탐 개수는 Table 4와 같다. 학습 및 테스트 전체 데이터는 각 보안이벤트별로 앞서 설명하였던 17개의 feature를 추출한 후, 정규화를 시켜주었고, 정규화 과정에 필요한 평균과 표준편차는 학습 데이터와 테스트 데이터 모두 동일하게 7월 1일 데이터 세트의 값을 사용하였다.

Table 4. Experiment Data for training and test

	Training	Test
Date	2017-07-01	2017-07-02
# of True Alarms	2,320	2,462
# of False Alarms	49,456	96,446
# of Total Alarms	51,776	98,908

5.2 최적 파라미터 분석결과

모델의 최적 파라미터를 선택하기 위해 학습 데이터와 n-fold cross validation 방법을 사용하여 c와 gamma를 바꿔가며 반복적으로 정확도를 확인하였다. 이때 n은 10으로 하였다. c는 최소 2^{-5}부터 최대 2^{15} 사이의 값을 사용하였고, 지수를 2씩 높여가며 테스트하였다. gamma는 최소 2^{-15}부터 최대 2^{3} 사이의 값을 사용하였고, 마찬가지로 지수를 2씩 높여가며 테스트하였다. 수행한 결과는 Fig. 2와 같다.

Fig. 2의 x축은 c를 log 스케일로 나타내었고, y축은 gamma를 log 스케일로 나타내었다. 각 선은 두 개 변수에 의해 동일 정확도가 나타나는 지점을 선으로 연결한 것이다. 전체적인 그림은 그래프의 중앙과 중앙 우측에서 높은 정확도를 보이고, 외부로 갈수록 정확도가 저하되는 형태를 보였다. c와 gamma를 변형해가며 정확도를 확인해본 결과, c가 2048, gamma가 0.0078125일 때 정확도가 99.1231%로 가장 높았다.

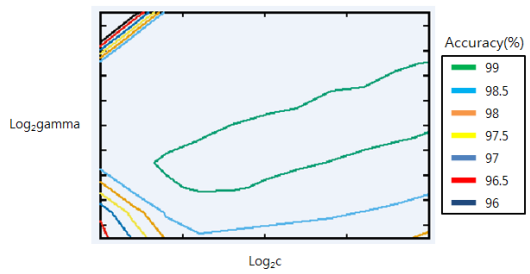

Fig. 2. The accuracy graph according to c and gamma

5.3 테스트 데이터 분류 결과

본 논문에서 제안한 feature와 최적 파라미터 분석을 통해 나온 파라미터를 적용하여 학습된 모델이 최종적으로 테스트 데이터를 분류한 결과는 Table 5에서 볼 수 있다. 학습된 모델은 2,462개 정탐 중에서 2,314개를 정탐으로 정확히 분류했고, 148개를 오탐으로 잘못 분류하였다. 96,446개 오탐 중에서는 95,934개를 오탐으로 정확히 분류했고, 512개를 정탐으로 잘못 분류하였다.

평가지수 산출결과는 Table 6에서 볼 수 있다. 평가지수 분석 결과, 정확도는 99.33%로 나타났고, 탐지율은 93.99%, 오탐지율은 0.53%로 나타났다. F1 score는 0.88로 나타났다. 정탐과 오탐 각각에 대한 분류 성능을 비교해보면, 제안 모델은 정탐보다 오탐에 대한 분류 성능이 더 좋았다.

제안모델의 최종결과를 기본 feature만 사용한 SVM 모델과 비교하였다. 비교에 사용된 자료는 Meng과 Kwok〔15〕이 그들의 논문에서 제시한 결과이고, 그들의 논문에서 사용한 지수를 통해 성능을 비교하였다. 비교결과는 Table 7에서 확인할 수 있다. 평가지수를 통해 성능을 비교해보았을 때 10가지 신규 feature를 적용한 모델의 분류 성능이 기존 feature만을 적용한 모델보다 정확도 및 오탐지율에서 더 향상된 결과를 보여주었다. 이로써 제안한 feature가 모델의 분류 성능을 향상시키는 것을 알 수 있다.

Table 5. Evaluation results

Accuracy	99.33%
True Positive Rate	93.99%
False Positive Rate	0.53%
F1 score	0.88

Table 6. Result of classification of test data

	True Alarm (Classified)	False Alarm (Classified)
True Alarm (Actual)	2,314	148
False Alarm (Actual)	512	95,934

Table 7. Comparison with other result

	Meng and Kwok〔15〕	Proposed Method
Accuracy	88.21%	99.33%
False Positive Rate	13.4%	0.53%

VI. 결 론

본 논문에서는 기계학습 기반 실시간 보안이벤트 분류 모델을 위한 새로운 feature 10종을 제안하였다. 제안한 feature는 IP, Port, 프로토콜, 위험도 등 기존 연구에서 사용하였던 feature 외에 단일 보안이벤트에서 직접 추출 가능하며 모델의 분류 성능도 높일 수 있도록 고안되었다. 제안 feature의 분류 성능을 확인하기 위해 기존 사용되던 feature에 새로운 feature를 추가 적용하여 SVM 모델을 구축하고 분류 실험을 수행하였다. 모델 실험 결과를 통해 기존 feature만 사용한 모델보다 정확도와 오탐지율 측면에서 성능이 향상된 것을 확인할 수 있었다.

제안한 feature는 모델의 분류 성능을 높일 뿐만 아니라 단일 보안이벤트에서 추출되었기 때문에 다른 기계학습 기반 보안이벤트 분류 모델 연구에도 쉽게 적용 가능하며 다양하게 활용될 수 있다. 따라서 여러 종류의 모델에 적용해보고 성능을 확인해볼 필요가 있을 것이다. 또한, 보안이벤트 분류 완전 자동화를 위한 정확한 보안이벤트의 분류를 위해서는 새로운 feature들의 지속적인 제안과 feature별 영향 분석을 수행하여 최적의 feature들을 선택하는 연구가 수행되어야할 것이다. 비록 이번 실험의 결과가 99% 이상의 정확도를 보이긴 했지만, 실제 현장에서는 1건의 보안이벤트가 치명적인 사고를 초래할 수 있기 때문에 이에 대한 지속적인 연구를 수행할 계획이다.

References

〔1〕 K. Scarfone and M. Peter, “Guide to intrusion detection and prevention systems (IDPS),” NIST Special Publication-800-94, Feb. 2007.

〔2〕 T. Pietraszek, “Using adaptive alert

classification to reduce false positive in intrusion detection," Recent Advances in Intrusion Detection, pp. 102-124, 2004.
[3] N. Hubballi and S. Vinoth, "False alarm minimization techniques in signature-based intrusion detection systems: A survey," Computer Communications, vol. 49, pp. 1-17, Aug. 2014.
[4] G. Spathoulas and K. Sokratis, "Methods for post-processing of alerts in intrusion detection: A survey," International Journal of Information Security Science, vol. 2, no. 2, pp. 64-80, June 2013.
[5] R. Sommer and P. Vern. "Enhancing byte-level network intrusion detection signatures with context," Proceedings of the 10th ACM conference on Computer and communications security. ACM, pp. 262-271, Oct. 2003.
[6] S.J. Yang, A. Stotz, J. Holsopple, M. Sudit, and M. Kuhl, "High level information fusion for tracking and projection of multistage cyber attacks," Information Fusion, vol. 10, issue. 1, pp. 107-121, Jan. 2009.
[7] E. Raftopoulos and D. Xenofontas, "Detecting, validating and characterizing computer infections in the wild," Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, pp. 29-44, Nov. 2011.
[8] G. Spathoulas and K. Sokratis, "Enhancing IDS performance through comprehensive alert post-processing," Computers & Security, vol. 37, pp. 176-196, Sep. 2013.
[9] M.S. Shin, E.H. Kim, and K.H. Ryu, "False alarm classification model for network-based intrusion detection system," International Conference on Intelligent Data Engineering and Automated Learning, pp. 259-265, Aug. 2004.
[10] T. Pietraszek and A. Tanner, "Data mining and machine learning-Towards reducing false positives in intrusion detection," Information Security Technical Report, vol. 10, pp. 169-183, 2005.
[11] C. Thomas and N. Balakrishnan, "Performance enhancement of intrusion detection systems using advances in sensor fusion," pp. 1-7, July 2008.
[12] G. Tjhai, S. Furnell, M. Papadaki, and N. Clarke, "A preliminary two-stage alarm correlation and filtering system using SOM neural network and K-means algorithm," Computers & Security, vol. 29, pp. 712-723, Sep. 2010.
[13] N. Hubballi, S. Biswas, and S. Nandi, "Network specific false alarm reduction in intrusion detection system," Security and Communication Networks, vol. 4, pp. 1339-1349, Nov. 2011.
[14] C. Chiu, Y. Lee, C Chang, W. Luo, and H Huang, "Semi-supervised learning for false alarm reduction," Industrial conference on data mining, pp. 595-605, 2010.
[15] Y. Meng and L. Kwok, "Adaptive false alarm filter using machine learning in intrusion detection," Practical applications of intelligent systems, pp. 573-584, 2011.
[16] S. Benferhat, A. Boudjelida, K. Tabia, and H. Drias, "An intrusion detection and alert correlation approach based on revising probabilistic classifiers using expert knowledge," Applied Intelligence, vol. 38, pp. 520-540, 2013.
[17] Y. Meng and L. Kwok, "Intrusion detection using disagreement-based semi-supervised learning: detection enhancement and false alarm reduction," Cyberspace Safety and Security, pp. 483-497, 2012.
[18] Y. Meng and L. Kwok, "Enhancing false alarm reduction using pool-based active learning in network intrusion detection,"

International Conference on Information Security Practice and Experience 2013, pp. 1-15, 2013.

[19] W. Li, W. Meng, X. Luo, and L. Kwok, "MVPSys: Towards practical multi-view based false alarm reduction system in network intrusion detection," Computers & Security, vol. 60, pp. 177-192, 2016.

[20] H. Liang, L. Taihui, X. Nannan, and H. Jiejun, "False positive elimination in intrusion detection based on clustering," 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp. 519-523, Aug. 2015.

[21] T. Alapaholuoma, J. Nieminen, J. Ylinen, T. Seppälä, and P. Loula, "A behavior-based method for rationalizing the amount of ids alert data," ICCGI 2012, The Seventh International Multi-Conference on Computing in the Global Information Technology, June 2012.

[22] J.O. Nehinbe, "Automated method for reducing false positives," 2010 International Conference on Intelligent Systems, Modelling and Simulation, pp. 54-59, Jan. 2010.

[23] V. Vapnik, "The nature of statistical learning theory," Springer science & business media, 2013.

[24] C.W. Hsu, C.C. Chang, and C.J. Lin, "A practical guide to support vector classification," pp. 1-16, 2003.

[25] C.C. Chang and C.J. Lin, " LIBSVM: a library for support vector machines," ACM transactions on intelligent systems and technology, vol. 2, issue. 3, Apr. 2011.

〈저 자 소 개〉

신 익 수 (Iksoo Shin) 학생회원
2012년 8월: 경북대학교 천문대기과학과 졸업
2016년 9월~현재: 과학기술연합대학원대학교 과학기술정보과학 통합과정
〈관심분야〉 네트워크 보안, 보안관제, 데이터마이닝, 머신러닝

송 중 석 (Jung-suk Song) 정회원
2003년 2월: 한국항공대학교 통신정보공학 졸업
2005년 2월: 한국항공대학교 정보공학 석사
2009년 3월: 교토대학교(일본) 지능정보학 박사
2009년 4월~2010년 9월: 일본정보통신연구원 정보통신 보안연구소 전문연구원
2010년 10월~2011년 9월: 일본정보통신연구원 네트워크 보안연구소 선임연구원
2011년 10월~현재: 한국과학기술정보연구원 첨단연구망정보보호실 선임연구원
2012년 9월~현재: 과학기술연합대학원대학교 과학기술정보과학 부교수
〈관심분야〉 보안관제, 침해사고대응, 악성코드 분석, 네트워크 보안

최 장 원 (Jang-won Choi) 정회원
1996년 8월: 홍익대학교 전자공학과 졸업
1998년 8월: 홍익대학교 전자공학 석사
2009년 2월: 고려대학교 전산학 박사
2000년 1월~2014년 12월 : 한국과학기술정보연구원 슈퍼컴퓨팅본부 선임연구원
2015년 1월~현재: 한국과학기술정보연구원 첨단연구망정보보호실 실장(책임연구원)
〈관심분야〉 보안관제, 침대사고대응, 악성코드 분석, 네트워크 보안, 분산시스템

권 태 웅 (Tae-woong Kwon) 정회원
2012년 2월: 숭실대학교 컴퓨터학부 졸업
2014년 8월: 고려대학교 정보보호대학원 정보보호학과 석사
2014년 12월~현재: 한국과학기술정보연구원 첨단연구망정보보호실 연구원
〈관심분야〉 정보보호, 보안관제, 네트워크 보안, 네트워크 가시화