ISSN 1598-0170 (Print)
ISSN 2287-1136 (Online)
http://www.jksii.or.kr

추론 및 비교사학습 기법 기반 레이블링을 적용한 탐지 모델☆

A Detection Model using Labeling based on Inference and Unsupervised Learning Method

홍 성 삼[1] 김 동 욱[1] 김 병 익[2] 한 명 묵[1]*
Sung-Sam, Hong Dong-Wook, Kim Byungik Kim Myung-Mook, Han

요 약

탐지 모델은 인공지능 기법들이나 데이터 마이닝 기법, 또는 지능형 알고리즘들을 이용하여 어떠한 목적에 맞는 결과를 찾고자 하는 모델들이다. 사이버 보안에서는 주로 침입탐지, 악성코드 탐지, 침해사고 탐지, 공격 탐지로 활용되고 있다. 보안데이터와 같은 실제 환경에 수집되는 데이터들을 레이블이 되지 않은 데이터들이 많다. 클래스 레이블이 정해지지 않아 유형을 알 수 없는 데이터가 많아 정확한 탐지 및 분석을 하기 위해서는 레이블 결정과정이 필요하다. 본 논문에서 제안하는 방법은 레이블 결정을 위해 D-S 추론 알고리즘과 비교사 방법인 k-means 알고리즘을 적용하여 각 데이터의 레이블을 융합하여 결정할 수 있는 KDFL(K-means and D-S Fusion based Labeling)제안하였으며 이를 적용한 탐지 모델 구조를 제안하였다. 제안하는 방법은 실험을 통해 기존의 방법에 비해 탐지율, 정확도, F1-measure 성능 지표에서 우수한 성능을 나타냈다. 또한 오류율도 크게 개선된 결과를 나타내어 제안하는 방법의 성능을 검증할 수 있었다.

☞ 주제어 : 레이블링, 분류기반 탐지 모델, 데이터 마이닝, 추론, 교사 및 비교사 학습, 보안

ABSTRACT

The Detection Model is the model to find the result of a certain purpose using artificial intelligent, data mining, intelligent algorithms In Cyber Security, it usually uses to detect intrusion, malwares, cyber incident, and attacks etc. There are an amount of unlabeled data that are collected in a real environment such as security data. Since the most of data are not defined the class labels, it is difficult to know type of data. Therefore, the label determination process is required to detect and analysis with accuracy. In this paper, we proposed a KDFL(K-means and D-S Fusion based Labeling) method using D-S inference and k-means(unsupervised) algorithms to decide label of data records by fusion, and a detection model architecture using a proposed labeling method. A proposed method has shown better performance on detection rate, accuracy, F1-measure index than other methods. In addition, since it has shown the improved results in error rate, we have verified good performance of our proposed method.

☞ keyword : Labeling, Detection Model based on classification, Data Mining, Inference, Supervised/Unsupervised Learning, Security

1. 서 론

탐지 모델은 인공지능 기법들이나 데이터 마이닝 기법, 또는 지능형 알고리즘들을 이용하여 어떠한 목적에 맞는 결과를 찾고자 하는 모델들이다. 사이버 보안에서는 주로 침입탐지, 악성코드 탐지, 침해사고 탐지, 공격 탐지, 사기 탐지로[1, 2] 활용되고 있으며 공격 또는 악성 패턴이나 시그니처들을 미리 학습하여 탐지하는 오용탐지(misuse detection)과 정상상태를 학습하여 발견되는 이상치를 탐지하는 이상탐지(anomaly)이 있다. 최근 연구들을 이 두가지 방법을 혼합하는 hybrid 형태의 연구들이 진행되고 있다.

데이터 마이닝 기반의 탐지 모델들은 정확도와 성능을 높이기 위해 교사학습(supervised learning)방법을 적극적으로 활용한다. 교사학습 방법은 미리 정해진 클래스의 레이블(label)들을 학습하여 학습 모델을 생성하고, 분류기에 의해 데이터를 분류하는 방법들이다. 학습된 모델에 의해 데이터를 분류하기 때문에 정확도가 높고, 신뢰도가 높다. 탐지 모델에서는 분류 결과를 기반으로 공격, 정상,

1 Department of Computer Engineering, Gachon University, Seongnam-si, 13120, Korea.
2 Department of Security R&D Team 1, Korea Internet& Security Agency, Seoul, 05717, Korea.
* Corresponding author (mmhan@gachon.ac.rk)
[Received 14 December 2016, Reviewed 15 December 2016, Accepted 22 January 2017]
☆ 이 논문은 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.B0101-16-0300, 사이버 공격의 사전 사후 대응을 위한 사이버 블랙박스 및 통합 사이버보안 상황분석 기술 개발)

공격 유형 등을 판단하게 된다.

보안데이터와 같은 실제 환경에 수집되는 데이터들을 레이블이 되지 않은 데이터들이 많다. 즉, 해당 데이터는 클래스 레이블이 정해지지 않아 유형을 알 수 없는 데이터인 경우가 많다. 이러한 데이터들을 학습 및 분류를 통해 더 정확한 탐지 및 분석을 하기 위해서는 레이블 결정 과정이 필요하다. 레이블 결정은 전문가와 경험에 의해 수동으로 수행할 수 있지만 데이터량이 많고 전체 데이터 셋을 직관적으로 분석하기 어려운 경우 데이터 셋일 경우에는 이 방법을 활용하기는 어렵다.

이를 해결하기 위해서 레이블링 알고리즘 연구가 진행되고 있다. 먼저 [3]연구와 같이 비교사 학습(unsupervised learning) 방법으로 데이터를 군집화한 뒤 군집 클러스터 결과(대표값, 중심값 등)를 기반으로 레이블링하는 연구들이 진행되고 있다. 또한 [4]에서는 익명성이 존재하는 트래픽 데이터에 대해 레이블링하기위해 D-S 추론 알고리즘을 사용하기도 하였다. 각 방법들은 레이블을 결정해 줄 수 있지만 결과들을 학습하여 분류 및 탐지를 수행했을 경우 정확도 성능과 오류 성능에서 부족한 부분이 나타났다. 추론 중 D-S를 사용한 방법은 미탐율이 높고, 비교사 학습 중 k-means를 사용한 방법은 정확도가 낮고, 오탐율이 높게 나타나 각각을 단일로 사용하기에는 부족한 부분이 있어 이를 개선할 방법이 필요하다. 이를 개선하기 위해 알고리즘간 단점을 상호보완할 수 있는 융합형태의 레이블링 방법이 필요하다.

본 논문에서 제안하는 방법은 레이블 결정을 위해 D-S(Dampster-Shafer) 추론 알고리즘과 비교사 방법인 k-means 알고리즘을 적용하여 각 데이터의 레이블을 융합하여 결정해주는 방법을 제안하였다. D-S 레이블링을 단독으로 사용한 방법은 탐지율이 높지만 오탐율이 너무 높아 오류가 크며, k-means 레이블링을 단독으로 사용한 방법은 오탐율은 낮지만 탐지율이 낮아 정확도가 떨어진다는 단점이 있다. 이를 개선하기 위해 두 결과를 융합하는 KDFL(K-means and D-S Fusion based Labeling algorithm)제안하였으며 이를 적용한 탐지 모델 구조를 제안하였다. 제안하는 방법은 각 알고리즘의 단점을 상호보완하여 레이블링을 수행하였으며, 레이블링 결과를 분류기반 탐지를 수행하였을 때 정확도와 오류 성능을 개선할 수 있었다. 성능 평가를 위해 naive bayes 분류기 기반의 탐지 모델을 이용하여 kdd cup'99(이미 클래스 레이블의 답이 알려져 있는)에 제안하는 방법을 적용하여 실험을 수행하였다. 제안하는 방법은 실험을 통해 기존의 방법에 비해 탐지율, 정확도, F1-measure 성능 지표에서 우수한 성능을 나타냈으며, 오류율도 크게 개선한 결과를 나타내어 성능을 검증할 수 있었다. 또한 레이블 되지 않은 데이터에 대해 제안하는 방법에 의해 자동으로 레이블링한 결과를 탐지모델에서 사용할 수 있는 가능성을 확인할 수 있었다.

본 논문의 구성은 2장에서 관련 연구를 소개하고, 3장에서는 제안하는 레이블링 방법과 이를 적용한 탐지 모델을 서술하였다. 4장에서는 제안하는 방법 및 기존 방법들에 대해 실험을 통해 성능을 평가하였으며 5장에서 결론을 맺는다.

2. 관련 연구

2.1 데이터 마이닝 기반의 공격탐지

비교사 기반의 알고리즘 중 밀도 기반 클러스터링 스키마는 단순 로그파일 클러스터링 툴(SLCT)[3]이라고 불리는데, 이 툴은 정상이거나 악의적인 네트워크 정상과 비정상적인 트래픽의 차이점을 구별하기 위하여, 이 연구에서는 클러스터에 포함된 고정(fix) 특징들의 퍼센트를 설정하기 위한 매개변수 M을 사용했으며, 클러스터를 수행하기 위한 고정 특징은 변하지 않는 값을 의미한다. M이 0이 되면, 데이터는 모두 군집을 이루고, M의 값이 높게 나오면, 악의적인 것들만 있는 군집이라고 생각할 수 있다. 이러한 방식을 사용하게 되면 이전에 알려지지 않은 공격을 막을 수 있으며, 지정된 매개변수들을 이용한 클러스터링 단계를 지난 후, 모든 클러스터들은 공격으로 간주된다. 다만 최적화된 매개변수 M를 찾는 작업이 필요하다. 이 시스템은 두 개의 클러스터링 스키마로 구성되게 하여 정상 또는 비정상을 탐지하는 것에 사용한다. 정상 트래픽을 결정하기 위한 탐지 모델에서는 supervised 방식으로 사용되게 된다. 이 둘의 차이점은 설정된 매개변수의 상태(modification)이며, 두 클러스터링 스키마는 병렬방식으로 정상과 비정상 트래픽을 탐지한다[1].

교사학습 기반의 공격탐지 연구들에서는 기존의 알려진 데이터의 레이블을 학습하여 분류를 기반으로 공격을 탐지하는 방법을 연구하였다. [5]에서는 Naive Bayes 분류기로 알려진 베이지안 네트워크 형식의 모델을 사용하였다. 이 실험에서는 KDD 1999 데이터 셋을 사용하였고 3개로 분류한 카테고리를 공격 시나리오와 성능 측정을 반영하기 위해서 그룹화 하였다. 하나의 공격과 정상적인 데이터는 첫 번째 데이터 셋에 포함되었으며, 두 번째 데이터 셋에는 KDD 1999 데이터 셋에 있는 4가지 공격 타

입을 포함하였고, 오용탐지를 위한 다중 클래스 분류를 수행하였다. 세 번째 셋에는 정상 데이터와 첫 번째 카테고리를 포함한 4가지 공격 타입을 포함하였고, 여기에선 이상 탐지를 수행하였다. 이 실험에서는 Normal, DoS, R2L, U2R, Probe or Scan에 대해서 97%, 96%, 9%, 12%, 88%의 정확성을 보여줬으며, 오경보도 나타나지 않았다. 하지만 일반적으로 97%정도만이 수행되기 때문에 FAR가 3%다 작다고는 확실하게 이야기 할 수 없으며, 이상탐지 실험은 정상과 비정상에 대해서 98%, 89%의 정확성을 보였다.

2.2 클러스터링 기반의 클래스 레이블 결정

비교사 알고리즘은 레이블되지 않은 데이터들을 유사하거나 관련있는 것들끼리 분리할 수 있기 때문에 레이블을 설정하는데 활용할 수 있다. 비교사 알고리즘으로 정해진 개수 또는 알고리즘에 의해 일정 개수로 데이터가 분리되면 각 집단의 대표 값이나 특성들을 분석하여 공격을 정의할 수 있다. 이를 통해 데이터를 레이블링 할 수 있다.

앞서 소개된 연구 [3]에서 나온 결과물은 다른 시스템이나 사이버 보안 전문가가 사용하게 될 규칙 기반 모델의 시그니처들로 활용이 되고, 이 연구방식의 특별한 점은 각각의 비정상적인 클러스터 centroid가 시스템 상에서 필러링된 시그니처들로 이루어져 있다는 것이다[1]. 다만 단점은 불균형한 클래스를 가진 데이터 셋을 레이블하여 분류할 경우 규모가 작은 클래스의 경우 데이터 셋에 포함된 데이터가 적기 때문에 정확한 클러스터를 찾아내기가 어렵다. 따라서 False Positive 오류가 크고, 전체적인 정확도 낮게 나올 수 있다. [18] 논문에서 다양한 비교사 학습 방법들을 이용한 탐지 모델들(k-means, k-medoids, Expectation Maximization) 및 거리기반 이상탐지 모델의 성능을 비교 분석하였는데 그 결과 낮은 정확도(0.5781, 0.654, 0.7671, 0.7806, 0.8015)와 False Positive 오류가 높게(0.2295, 0.2152, 0.2183, 0.2074, 0.2114)나오는 것을 볼 수 있었다. 이는 본 논문에서 실험한 결과와 유사하며 레이블링의 정확도와 관계성이 높다.

2.3 추론 알고리즘 : Dempster-Shafer

데이터 융합 문제의 Dempster-Shafer 이론은 1981년에 소개된 것으로, 불확실성 처리에 관한 연구로 많이 진행되어 왔다. Dempster-Shafer는 많은 응용 분야에 있어서 유효성이 입증된 효과적인 방법으로 사용되었다. 그러나 계산이 비효율적인 단점을 가지고 있고, 속성에 대한 여러 가설이 주어질 때마다, 여러 개의 가설들이 모여 하나의 가설을 이루는 multiple hypothesis이 허용되는 알고리즘이기 때문에 가능한 가설의 개수는 지수적으로 증가하게 된다. 지수적으로 불어나는 가설의 개수를 처리해야하기 때문에 복잡도가 늘어나 전체적인 시스템의 효율성에 영향을 주어 성능 저하를 나타낼 수 있다. 그러나 Dempster-Shafer 이론은 베이지안과 달리 하나의 수치로 표현하기 보다는 신뢰구간을 지정하여 유효한 표현이 가능하며, 정보의 부족에서 오는 불확실성에 대한 차이를 말할 수 있다[6]. Dempster-Shafer 증거이론(Theory of Evidence)은 θ의 참값이 알려지지 않는 확률변수의 분별 프레임이라는 가설의 집합을 정의한다. D-S프레임에 대한 완전한 확률할당은 기본확률할당(basic probability assignment, BPA)으로서 정의되며 이산 BPA 함수(discrete BPA function) m(A)는 다음의 식(1)을 만족하여야 한다.

$$0 \le m(A) \le 1 \quad \text{for all } A \subset \theta \qquad (1)$$
$$\sum_{A \subset \theta} m(A) = 1.0, \quad m(\varnothing) = 0$$

어떤 부분집합의 $A \subset \theta$에 대한 믿음함수(belief function) Bel(A) 및 가능성 함수(plausibility function) Pl(A)가 θ에 대한 BPA로부터 식(2)와 정의될 수 있다.

$$Bel(A) = \sum_{B|B \subseteq A} m(B),$$
$$Pl(A) = \sum_{B|B \cap A \neq \varnothing} m(B) \qquad (2)$$

bel(A)와 Pl(A)는 수식(3)처럼 각각 A에 대한 확률의 상한과 하한이 정의된다.

$$Bel(A) \le P(A) \le Pl(A), \quad A \subset \theta \qquad (3)$$

이러한 이론을 통해 수집된 데이터의 증거는 일반적으로 Dempster의 규칙을 사용하여 융합된다. belief의 A는 bel(A)=$\sum_{B \subseteq E} m(B)$로 정의되며 타당성의 pl(A)=$\sum_{B \cap C = E \neq \varnothing}$으로 정의된다. $m_{1,2}(A) = (m_1 \oplus m_2)(A)$ 정보의 두 가지 정보를 고려하여 다음과 같은 수식(4)로 나타내며, (5)에서의 포함되지 않을 경우 0으로 할당하며,

질량함수 m1, m2가 계산된다[7].

$$m_{1,2}(A) = (m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{B \cap C = A \neq \varnothing} m_1(B) m_2(C) \quad (4)$$

$$m_{1,2}(\varnothing) = 0 \quad (5)$$

이러한 정보간의 충돌의 양을 수식(6)으로 나타낼 수 있다.

$$K = \sum_{B \cap C = \varnothing} m_1(B) m_2(C) \quad (6)$$

2.4 Dempster-Shafer 추론을 이용한 클래스 레이블 결정

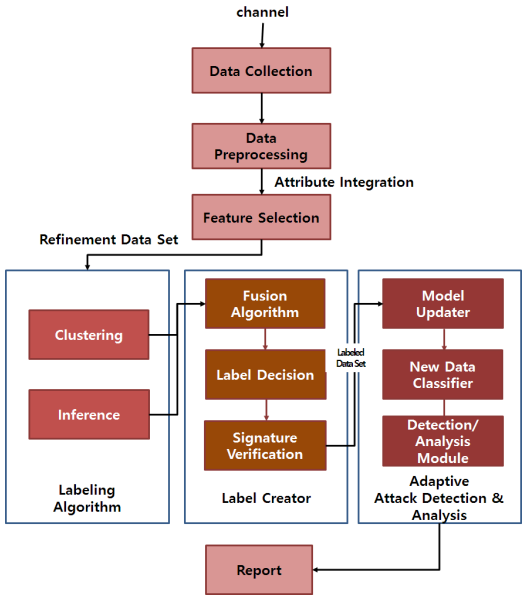
불확실성에 대한 처리는 Bayesian 방법과는 달리 각 소스는 다른 수준의 정보를 기여할 수 있고, 가설 집단의 모든 부분집합에 대해 신뢰도를 배정할 수 있도록 할 수 있어 모든 부분집합에 대한 분포를 형성할 수 있게 된다[8, 11].

데이터 레이블링에서는 주어진 데이터의 클래스가 불확실하기 때문에 불확실성 데이터의 처리가 가능한 D-S 추론은 사용하기 적합하다. 주어진 데이터들이 갖고 있는 정보들로부터 BPA를 정의하고, 이로부터 새로 유입되는 데이터에 대해 새로운 레이블을 추론할 수 있다. 여기서 보유한 데이터에 정보가 많을수록 레이블 추론이 정확해질 수 있다. 다만 클래스가 불균형한 데이터 셋일 경우, 특히 이상탐지와 같이 이진분류에서는 규모가 작은 클래스의 데이터들의 정보가 적어 일방적으로 한쪽의 클래스 결과로 추론되게 되기 때문에 규모가 작은 클래스에 대한 분류를 하지 못하게 되는 단점이 있다.

3. 제안하는 KDFL 기법 및 Supervised 기반 공격탐지 모델

제안하는 KDFL(K-means and D-S Fusion based Labeling)를 적용한 공격 분류 및 탐지 시스템의 구조는 (그림 1)과 같다. 레이블이 알려지지 않은 데이터로부터 공격 행위를 탐지가 가능하며, 새로운 데이터에 대해서도 탐지가 가능하다. 또한 새로운 공격 유형에 대한 대처도 가능하도록 적응형 구조로 구성하였다.

제안하는 알고리즘을 적용한 공격 분류 및 탐지 시스템 구조는 CRISP-DM의 모델[10]을 기반하여 데이터 마이닝 기반의 공격 탐지 프레임워크를 구성하였다. Data Collection 단계에서는 분석 및 비즈니스 목적에 맞게 각 채널로부터 데이터를 수집하는 과정이다. 수집되는 채널 및 센서들은 공격 탐지를 위한 정보들을 제공해준다. Data Preprocessing 과정에서는 데이터 탐색 과정에서 나타난 데이터 셋의 특성에 따라 데이터 전처리를 수행한다. 속성 및 데이터들을 분석 모델의 입력으로 사용하기 위해 클랜징하는 과정으로 핵심 요소는 주요 속성 선택, 이상치 및 결측치 제거, 값 변환 및 정규화이다. 범주형 속성과 수치형 속성은 각 유형에 맞게 전처리과정이 다르게 적용된다. 본 논문의 실험에서는 이미 정규화까지 완료된 데이터셋으로부터 7개의 특징을 선택하여 전처리를 하였다.



(그림 1) KDFL을 적용한 공격 탐지 및 분석 모델
(Figure 1) An attack detection and analysis model using KDFL

특징 선택은 특징 선택 과정은 통합된 속성으로 만들어진 데이터 셋에서부터 차원 축소 및 핵심 특징들을 선택하기 위한 알고리즘을 적용하는 과정이다. 이 과정을 통해 데이터 분석 성능 및 속도를 개선할 수 있다. 특징 선택 알고리즘은 PCA[11]와 같은 차원 축소 알고리즘이나 [12, 13] 등의 다양한 알고리즘들을 활용할 수 있다. Association Rule을 이용한 분석이나 진화 프로그래밍 기

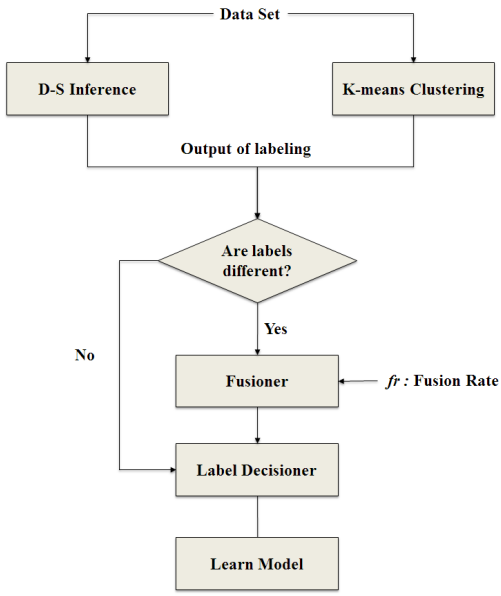
반의 분석의 경우 데이터 군집화나 분류에서 특징선택이 같이 수행되도록 활용할 수 도 있다.

3.1 Inference와 비교사 Algorithm을 이용한 데이터 레이블링 방법

레이블되지 않은 데이터들로부터 공격 및 정상 시그니처들을 추출하기 위한 과정으로 본 논문에서는 추론을 사용하여 알지 못하는 데이터의 레이블을 결정할 수 있다.실제 환경에서 수집되는 데이터들을 보통 레이블되지 않은 데이터인 경우가 보통이며 이를 학습 및 분류를 통해 더 정확한 탐지를 하기 위해서는 레이블 결정과정이 필요하다. 레이블 결정은 전문가에 의해 수동으로 수행할 수 있지만 데이터량이 많고 직관적으로 분석하기 어려운 경우 이 방법을 활용하기는 어렵다.

따라서 본 논문에서는 레이블 결정을 위해 D-S 추론 알고리즘과 비교사 방법인 k-means 알고리즘을 적용하여 각 데이터의 레이블을 융합하여 결정해주는 알고리즘을 제안하였다. 제안하는 방법은 두개의 알고리즘에서 각각 추론과 비교사 방법으로 결정된 데이터의 레이블 값들을 융합하여 더 좋은 레이블 값을 얻고자 하였다. 그림 2는 제안하는 알고리즘 과정을 나타내고 있다. 데이터 셋으로부터 각 추론알고리즘인 D-S알고리즘과 비교사 방법인 k-means clustering 방법이 각각의 레이블링 결과를 발생시키고, 그 결과 값들을 비교하여 서로 다른 결과를 산출하였을 경우 fusioner를 통해 결과를 융합한다. 여기서 fr(fusion rate)는 두 개의 알고리즘의 결과값을 융합할 때 비율을 조정하기 위한 파라미터 값으로 값이 높아질수록 추론 결과가, 낮아질수록 비교사 방법의 결과가 레이블을 결정하는데 더 반영되도록 하였다. 예를 들어 fr이 0.3이면 비교사 결과가 70%, 추론 결과가 30%로 반영되는 것으로, 전체 데이터 셋에 대한 비교사 레이블 결정 결과 중 무작위로 70%의 데이터 레코드 레이블 결과를 최종 레이블로 결정하고, 그 외의 30% 데이터 레코드들은 추론 결과로 최종 레이블을 결정하는 것이다. 이는 도메인 환경이나 사용 목적에 따라 탐지율, 정확도, 오탐율, 미탐율에 대한 요구사항이 다르기 때문에 이를 적절하게 충족할 수 있는 융합결과를 얻기 위한 파라미터이다. 융합된 결과는 Label Decision 모듈로 전달되어 레이블을 결정하고 결정된 레이블은 학습모델에 적용되어 학습되어진다.

D-S 추론 알고리즘을 데이터들의 최대, 최소, 평균값들을 파라미터로 입력하여 이를 기반으로 데이터가 공격 데이터인지 아닌지를 확률적으로 추론할 수 있다. 100%



(그림 2) 제안하는 레이블링 방법 순서도
(Figure 2) A flowchart of proposed labeling method

정확한 정답을 찾아주지는 못하지만 어느 정도 예상되는 데이터 레이블을 식별해주기 때문에 레이블이 없는 데이터의 레이블을 결정하는데 적용할 수 있다. 레이블 추론 단계에서는 기존의 공격 또는 정상 데이터 등을 토대 초기 D-S의 BPA를 할당한 뒤 레이블이 필요한 데이터를 알고리즘에 적용하여 각 데이터의 레이블을 추론한다. D-S 알고리즘은 구조상 데이터가 계속적으로 알고리즘의 입력으로 사용되고 레이블을 추론함에 따라 점차 BPA와 조합규칙이 데이터 셋의 특성을 반영하여 세밀하게 확률구간이 조정될 수 있다. 따라서 데이터가 많을수록 좀 더 정교한 추론이 가능하게 된다. 특히 목표하는 레이블 즉, 클래스별로 특정 클래스의 특성을 갖는 데이터 개수가 많을수록 해당 클래스의 레이블을 추론하는 정확도가 높아진다. 특정 클래스의 데이터가 집중되어 있거나 이진화된 분류(이상탐지와 같은)에 적합한 특징이다. 이러한 특징으로 인해 D-S 알고리즘을 적용한 레이블 추론 방법은 편중된(불균형한) 데이터 형태를 주로 갖고 있는 보안 데이터 셋들에 적용하기 적합하다. 다만 상대적으로 데이터 개수가 적은 클래스에 대한 오류를 줄이기 위한 방법이 필요하다.

비교사 방법인 k-means 클러스터링을 이용한 레이블링 방법은 데이터 포인트 간 유사도를 계산하여 데이터 포

인트 벡터가 유사한 레코드들 끼리 군집하여 해당 군집에 따라 레이블을 결정해준다. D-S 알고리즘과 k-means 알고리즘은 목적하는 클래스의 개수를 정해줄 수 있기 때문에 목표하는 클래스 개수를 동일한 값으로 지정하여 레이블을 결정할 수 있다. k-means을 통한 레이블링의 경우 데이터에 편중에 상관없이 데이터 형태에 따라 적절한 클러스터를 선정해주게 된다. 하지만 D-S와는 반대로 많은 비중을 차지하는 클래스에 해당하는 데이터 포인터들이 산개되어 있으면 해당 클래스를 군집했을 때 오류가 발생하게 된다. D-S에 비해 비중이 높은 데이터에 대한 분류율이 낮다. 따라서 제안하는 방법은 D-S 알고리즘 결과와 k-means 결과의 장점들을 반영할 수 있는 적절한 융합알고리즘을 적용하여 레이블을 결정하고, 이를 통해 각 알고리즘의 레이블링 결정방법이 갖고 있는 단점들을 개선하여 정확도를 높이고 오류율을 낮출 수 있었다.

추론된 레이블 값들을 데이터의 레이블로 결정하여 데이터의 클래스를 선정한다. 이를 기반으로 수집된 데이터 셋에서 테스트 셋을 만들고 선택된 특징들을 기반으로 학습을 통해 공격 탐지 모델을 구축한다. 이를 통해 새로 유입되는 데이터에 대해 공격탐지를 수행한다.

3.2 분류 기반 공격 탐지

이 과정에서는 실질적인 데이터 분류 및 분석이 수행되는 과정으로 주로 교사학습기반의 분류기가 모델을 적용하여 데이터들을 분류해낸다. 분류해낸 결과는 공격인지 아닌지 판단하는 기준이 되어 탐지 결과를 도출해낸다. 기존의 수집된 데이터들로부터 생성된 레이블들을 기반으로 학습을 하여 모델을 생성하고 분류기들을 통해 새로운 데이터들에 대해 분류 및 분석을 수행할 수 있다. 적응형 프로세스를 수행하기 위해 새로운 데이터들로부터 알려지지 않은 레이블들을 생성한 경우에는 updater를 통해 레이블 및 학습셋 들을 업데이트하여 모델을 갱신한다. 이를 통해 새로운 데이터에 대한 공격 판단이 가능하며 새로운 레이블들을 생성하여 모델을 갱신할 수 있는 적응형 공격 탐지 시스템을 구현할 수 있다. 본 논문에서는 분류기로 naive bayes를 사용하였다.

4. 실험 및 결과

4.1 실험 환경

실험에 사용된 하드웨어 및 OS 환경은 다음과 같다.

- CPU : Intel Core i7 6700k 4.00Ghz
- RAM : 16GB
- OS : Windows 10 64bit

실험을 위해 사용된 tool은 open software R[14]을 사용하였으며 버전은 3.02이다. 분류기인 naive bayes 알고리즘은 각각 R 패키지로 구현되어 있는 e1071 [15] 패키지를 사용하여 실험을 수행하였다.

4.2 데이터 셋 및 특징 선택

실험을 위해 사용된 데이터 셋은 anomaly detection dataset 중 가장 대표적인 KDD‘ cup 99 데이터 셋[16] 을 사용하였다. 이 데이터 셋은 정형화된 데이터로 MIT Lincoln Lab에서 이상탐지를 실험하기 위해 미 공군 LAN으로부터 수집한 공격 및 정상 데이터이다. 많은 IDS 연구에서 실험용 데이터 셋으로 널리 사용되고 있는 데이터 셋이다. 약 490만개의 레코드가 존재하며, 41개의 특징(feature)으로 구성되어있다. 대부분 10% 데이터 셋을 활용하여 실험을 수행하며, 본 실험에서도 10% 데이터셋인 494,021개의 레코드를 사용하였다. 본 논문에서 KDD‘ cup 데이터 셋을 사용한 이유는 이 데이터 셋의 경우 이미 레이블이 결정이 되어 있어 답을 알 수 있는 데이터로 레이블링 알고리즘으로 결정된 레이블 값으로 분류 기반 탐지를 수행한 결과 값을 원본의 답과 비교하여 성능을 검증할 수 있기 때문에 실험에 사용하였다. 특징 선택의 경우 41개의 모든 특징을 사용하지 않고 특징선택을 수행한 결과로 결정된 7개의 주요 수치형 특징들만을 사용하여 실험하였다. 주요 특징 선택 기준은 [13]에서 제안한 방법에 의해 결정된 특징들을 사용하였다. 실험에 사용한 데이터의 클래스 분포는 공격 396,743개, 정상 92,278개의 레코드를 사용하였다.

4.3 성능 측정 방법

본 논문에서는 탐지성능을 측정하기 위해 F1-measure는 데이터 분류, 문서 분류, 분류탐지에서 단순 정확도나 탐지율 등의 성능평가 방법을 개선한 방법이다. TP(True Positive), TN(True Negative), FP(False Positive), FN(False Negative)으로 precision과 recall값을 구하면 각 값의 비중을 동일하게 하여 조화 평균을 구한다[17]. 높을수록 분류 탐지기의 성능이 높다고 평가한다. 본 평가에서는 탐지 성능을 체크하기 위해서 positive 클래스를 attack로 정하

고 성능평가를 한다. P는 precision, R은 recall이며 각 식 (7)에 의해 구해지며 F1-measure는 식 (8)과 같다. 먼저 supervised 방법에 대한 결과를 평가하는 방법이다.

- TN: Normal data correctly classified as normal.
- TP: Anomalous data correctly classified as anomalous.
- FP: Normal data classified as anomalous.
- FN: Anomalous data classified as normal.

$$P = \frac{TP}{(TP+FP)} \quad R = \frac{TP}{(TP+FN)} \tag{7}$$

$$F1-measure = \frac{2 \times P \times R}{P+R} \tag{8}$$

또한 탐지 시스템의 전체적인 정확도, 탐지율(DR: Detection Rate), 잘못허용율(FAR:False Accept Rate)=미탐율, 거절 실패율(FRR:False Reject Rate)=오탐율을 측정하여 성능 지표를 확인하도록 한다.

4.4 레이블링 및 분류 탐지 실험 결과

레이블링 알고리즘의 성능을 검증하기 위해 각 레이블링 알고리즘의 결과를 학습하여 분류 기반 탐지를 수행하였다. 먼저 naive bayes 분류기 자체 성능을 검증하기 위해 실험에 사용한 kdd 데이터 셋의 원본 클래스 레이블을 학습하여 naive bayes 모델로 분류 기반 탐지를 수행한 실험 결과는 아래와 같다. 학습셋으로 394,021개, 테스트 셋으로 100,000개의 레코드를 사용하였다.

(표 1)에 나타나있는 실험결과를 보면 kdd 데이터 셋에 대해 naive bayes 분류기는 정확도는 0.979, f1-measure 지표는 0.9871, 탐지율은 0.9805로 전체적인 지표에서 우수한 성능을 나타내고 오탐과 미탐 오류도 적게 나타난 것을 볼 수 있다. 따라서 탐지 시스템에서 사용하기 적합한 분류 모델이라고 볼 수 있어 본 실험에 검증 모델로 활용할 수 있음을 판단할 수 있다.

4.4.1 추론 : D-S 레이블링 실험 결과

D-S 알고리즘으로 레이블을 결정한 결과를 학습하여 naive bayes 분류 기반으로 탐지를 수행한 결과는 다음과 같다. 레이블 결정 과정에서는 총 데이터 셋인 494,021개를 입력하여 모든 레코드에 대해 레이블을 결정하였다. 결정된 레이블을 학습셋으로 394,021개, 테스트 셋으로 100,000개로 분리하여 학습을 통해 분류 기반 탐지 실험을 수행하였다. D-S의 초기 파라미터 값으로 특징 선택으로 선택된 각 특징들의 최소값, 최대값, 평균값을 설정하여, 레코드별 추론을 수행하였다. (표 2)는 D-S 레이블링 방법 기반의 분류 기반 탐지 성능 실험 결과를 나타내고 있다.

실험 결과를 보면 탐지율(DR)은 0.9387로 높게 나타났으며 그에 따라 FRR, 즉 미탐율은 0.0608로 낮게 나타났다. 반대로 공격이 아닌 것에 대한 탐지는 상당히 낮게 오탐율도 0.9482로 높게 나왔다. 전체적인 정확도도 0.7636으로 낮게 나와서 단독으로 탐지 시스템에 적용하기는 어렵다는 것을 알 수 있었다. 또한 실험 데이터 셋의 경우 공격 클래스가 더욱 많이 분포되어 있는데, D-S의 경우 클래스 비중에 높은 데이터에 대해 더욱 정확한 레이블을 결정을 하지만 적은 데이터에 대한 레이블 추론은 정확도가 떨어진다는 것을 볼 수 있었다.

(표 1) 원본 kdd cup'99 데이터셋에 대한 naive bayes classifier 성능
(Table 1) The naive bayes classifier performance using original kdd cup' 99 data set

원본 정답 label + naive bayes	
측정지표	value
P	0.9785
R	0.9958
F1-measure	0.9871
Accuracy	0.9794
DR	0.9805
FRR	0.0215
FAR	0.0168

(표 2) D-S 레이블링 방법의 분류 기반 탐지 성능 실험결과
(Table 2) A result of detection performance based on classification of D-S labeling

D-S labeling + naive bayes	
측정지표	value
P	0.9391
R	0.8006
F1-measure	0.8644
Accuracy	0.7636
DR	**0.9387**
FRR	**0.0608**
FAR	0.9482

(표 3) k-means 레이블링 방법의 분류 기반 탐지 성능 실험 결과
(Table 3) A result of detection performance based on classification of k-means labeling

k-means labeling + naive bayes	
측정지표	value
P	0.7063
R	1.0000
F1-measure	0.8279
Accuracy	0.7644
DR	0.7060
FRR	0.2937
FAR	**0.0001**

4.4.2 비교사 학습 : k-means 레이블링 실험결과

비교사 방법인 k-means 알고리즘으로 레이블을 결정한 결과를 학습하여 naive bayes 분류 기반으로 탐지를 수행한 결과는 아래와 같다. 마찬가지로 전체 데이터 셋에 대해 레이블을 결정하고, 학습셋으로 394,021개, 테스트 셋으로 100,000개의 레코드로 분리하여 분류 기반 탐지 실험을 수행하였다. k값은 2로 지정하였으며 군집에 사용된 특징들은 특징 선택으로 선택된 각 특징들이 입력으로 사용되었다. (표 3)은 k-means 레이블링 방법 기반의 분류 기반 탐지 성능 실험 결과를 나타내고 있다. 실험 결과를 보면, 오탐율(FAR)은 0.0001로 낮게 나타나는 것을 볼 수 있는데 이는 정상에 대한 분류는 정확히 하는 것이라고 판단할 수 있다. 반대로 공격에 대한 탐지율은 0.706, 미탐율(FRR)은 0.2937 많은 부분 탐지하지 못하는 것을 볼 수 있다. 또한 전체 정확도도 0.7644로 낮으며 F1-measure 값도 0.8279로 D-S 방법보다 낮게 측정되었다. 이를 통해 k-means 레이블링 방법은 알고리즘에 의해 데이터가 정확히 분리되지 않는 경우가 많기 때문에 레이블 결과가 비중이 높은 클래스의 데이터들일수록 분산되어 있을 수 있으므로 이 부분에서 정확도가 낮게 나오는 것을 볼 수 있었다.

4.4.3 제안하는 방법 : KDFL실험 결과

위 D-S와 k-means 기반의 레이블링 결과를 학습하여 분류 기반 탐지를 수행한 실험결과들을 보면 D-S와 k-means은 상호보완적인 관계를 나타낸다고 볼 수 있다. 제안하는 방법은 각각의 장점을 활용하여 결과를 융합하여 레이블링을 수행하면 단점을 보완하고, 더 좋은 성능

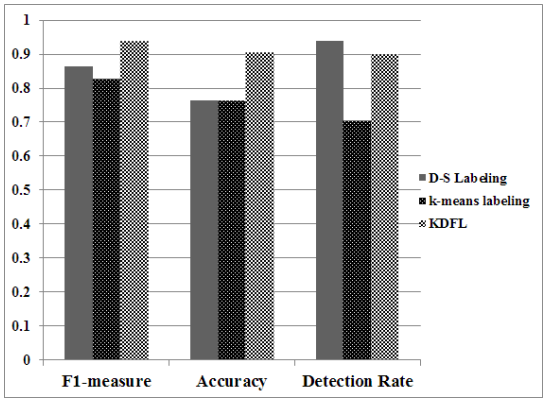
(표 4) 제안하는 방법의 분류 기반 탐지 성능 실험결과
(Table 4) A result of detection performance based on classification of proposed method

KDFL + naive bayes	
측정지표	value
P	0.9779
R	0.9007
F1-measure	**0.9378**
Accuracy	**0.9040**
DR	**0.9007**
FRR	**0.0993**
FAR	**0.0826**

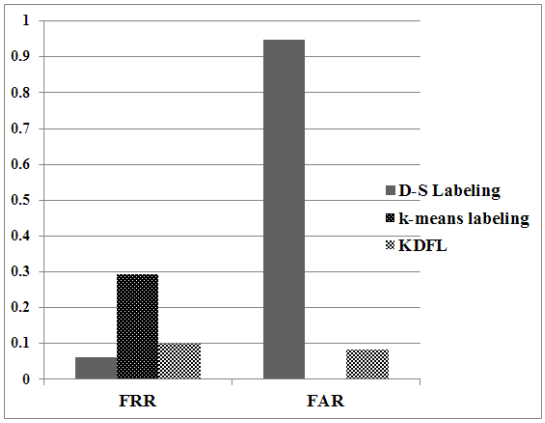
(표 5) 알고리즘별 성능 지표 결과 비교
(Table 5) A comparison of label algorithms performance

	D-S Labeling	k-means Labeling	**KDFL**
F1 -measure	0.86442	0.8279	**0.9378**
Accuracy	0.7636	0.7644	**0.9040**
Detection Rate	0.9387	0.7060	**0.9007**
	D-S Labeling	k-means Labeling	**KDFL**
FRR	0.0608	0.2937	**0.0993**
FAR	0.9482	0.0001	**0.0826**

결과를 얻을 수 있을 것으로 판단된다. 이를 검증하기 위해 각 D-S와 k-means 알고리즘이 전체 데이터 셋에 대해 결정한 각 레코드에 대한 레이블들에 대해 결합하는 방법 적용하여 새로운 레이블들을 결정하였고, 결정한 레이블로 구성된 데이터 셋을 학습 셋으로 394,021개, 테스트 셋으로 100,000개의 레코드로 분리하여 분류 기반 탐지 실험을 수행하였다. 알고리즘에 사용되는 fusion rate 파라미터 값은 0.44로 설정하였으며, rate는 높아질수록 D-S의 레이블 결과의 중요도가 높아져 D-S 결과가 더 크게 반영되었다. 본 실험에서는 두 방법의 비중을 같게 하여 레이블을 결합하여 분류 기반 탐지를 수행하였다. 실험 결과는 위 (표 4)와 같다. 실험 결과를 보면, 보편적 성능 지표들이 F1-measure는 0.9378, 정확도는 0.904, 탐지율이 0.9007로 좋은 결과를 나타내었다. 또한 오류관련 지표인 FRR과 FAR도 각 0.0993과 0.0826으로 비교적 낮게 나타나는 것을 볼 수 있었다. 위 (표 5)는 각 알고리즘과 제안



(그림 3) 알고리즘 성능 비교 : 정확도 성능 지표
(Figure 3) A comparison of algorithms performance : Accuracy, F1-measure, Detection Rate



(그림 4) 알고리즘 성능 비교 : 정확도 성능 지표
(Figure 4) A comparison of algorithms performance : Error Performance Indexes

하는 방법과의 성능 지표 결과를 종합적으로 나타내고 있다.

(그림 3)은 D-S와 k-means 방법과 제안하는 방법과의 정확도 관련 성능 수치를 비교한 결과이다. 결과에서 볼 수 있듯이 제안하는 방법은 전체적인 성능이 F1-measure는 0.9378, 정확도는 0.9040, 탐지율은 0.9007로 다른 방법들에 비해 높게 나타난 것을 볼 수 있다. 특히 정확도 성능은 다른 알고리즘에 비해 상당히 개선된 것을 볼 수 있다.

(그림 4)는 k-means 방법과 제안하는 방법과의 오류 관련 성능 수치를 비교한 결과이다. 오탐과 미탐율의 평균적인 성능 수치에서는 제안하는 방법이FRR(오탐율)은 0.0993, FAR(미탐율)은 0.0826으로 낮게 나타났다. D-S는 FRR, k-means은 FAR이 제안하는 방법보다는 낮게 나타났지만 다른 오류 수치에서 아주 좋지 않은 성능 지표를 보여주고 있기 때문에 단독으로 사용하기에는 부족하다는 것을 실험을 통해 알 수 있었다.

실험 결과들을 분석해본 결과, D-S와 k-means을 단독으로 사용한 것에 비해 상당한 성능 개선이 되었다는 것을 알 수 있다. 제안하는 알고리즘은 fusion 방식을 통해 각 알고리즘의 단점이 개선되어 종합적인 성능을 개선시켰음을 판단할 수 있었다.

결론적으로, 레이블을 알고 있는 경우에 비해서는 성능이 비교적 낮게 보일 수 있으나 레이블이 되지 않은 데이터에 대해 레이블을 결정하고, 이를 분류 기반으로 탐지하는 성능에 대한 평가는 충분히 공격 탐지 또는 분석 시스템에 적용할만한 성능 결과를 나타낸 것으로 판단된다. 제안하는 방법은 레이블이 결정되지 않은 데이터에 대해서 자동으로 레이블을 결정할 수 있으며, 적응형으로 레이블 학습과 새로운 데이터에 대한 분류, 새로운 레이블 생성, 학습과 모델 갱신들이 가능하기 때문에 점진적인 성능 개선이 나타날 수 있을 것으로 판단된다. 또한 fusion rate 파라미터를 최적화된 값을 찾아내거나 fusion 방식을 추가적으로 개선하면 더 좋은 결과를 얻을 수 있을 것이다.

5. 결 론

본 논문에서는 탐지모델에서 분류 기반 모델을 사용하기 위해 레이블 되지 않은 데이터에 대해 레이블을 결정해줄 수 있는 방법을 제안하였다. 탐지 및 분석 모델에서 비교사 방법은 정확도 및 오류 성능을 향상시키는데 주요한 방법이다. 비교사 방법을 활용하기 위해서는 입력 데이터의 클래스 레이블이 필요하기 때문에 자동으로 레이블을 결정할 수 있는 방법을 제안하였다.

제안하는 방법은 추론과 비교사 학습 기반의 레이블링 방법들의 결과를 융합하는 형태로 더 좋은 레이블 값을 얻을 수 있었다. 이는 실험을 통해 분류 및 탐지 성능이 우수하게 나타난 것을 검증할 수 있었으며, 기존 방법들의 단점들을 개선하고 상호 보완할 수 있는 방법임을 알 수 있었다. 이미 레이블을 알고 있는 데이터 셋을 적용한 결과와는 성능이 비교적 좋지 않게 보일 수 있으나 레이블이 되지 않은 데이터들을 사용해야하는 경우를 가정할 때, 레이블을 결정하고 이를 분류 기반으로 탐지하는 성

능에 대한 평가는 충분한 성능 결과를 나타낸 것으로 판단된다. 또한 이를 적용한 공격 탐지 및 분석 모델을 제시하여 실제적인 적용 방안을 모색하였다.

향후 연구로는 탐지 및 분석 성능 개선을 위해 제안한 알고리즘의 파라미터 값 등의 레이블링 성능을 개선하고, 연속적으로 유입되는 데이터 및 시계열 분석 처리가 가능한 탐지 및 분석 모델에 대해 연구하도록 하겠다.

참고문헌(Reference)

[1] Anna L. Buczak, Erhan Guven, “A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection,” *IEEE COMMUNICATIONS SURVEYS & TUTORIALS*, Vol.18, No.2, 2016. https://doi.org/10.1109/comst.2015.2494502

[2] Sannasi Ganapathy, Kanagasabai Kulothungan, Sannasy Muthurajkumar, Muthusamy Vijayalakshmi, Palanichamy Yogesh, and Arputharaj Kannan, “Intelligent feature selection and classification techniques for intrusion detection in networks: a survey,” *EURASIP Journal on Wireless Communications and Networking (open access)*, 2013. https://dx.doi.org/10.1186/1687-1499-2013-271

[3] R. Hendry and S. J. Yang, “Intrusion signature creation via clustering anomalies,” *Proc. SPIE Defense Secur. Symp. Int. Soc. Opt. Photonics,* pp.69730C–69730C, 2008. https://doi.org/10.1117/12.775886

[4] Claudio Mazzariello, “Multiple classifier Systems for Network Security from data collection to attack detection,” *Universitа degli Studi di Napoli Federico Il Open Archive*, Doctor Thesis, 2008.

[5] N. B. Amor, S. Benferhat, and Z. Elouedi, “Naïve Bayes vs. decision trees in intrusion detection systems,” *in Proc ACM Symp. Appl. Comput.,* pp.420–424, 2004. https://doi.org/10.1145/967900.967989

[6] Bass, Tim, “Intrusion detection systems and multisensor data fusion,” *Communications of the ACM,* Vol.43, No.4, pp.99-105, 2000. https://doi.org/10.1145/332051.332079

[7] MLA Deng, Xinyang, and Yong Deng, “Multisensor Information Fusion Based on Dempster-shafer Theory and Power Average Operator,“ *Journal of Computational Information Systems*, Vol.9, No.16 pp.6417-6424, 2013. https://doi.org/10.12733/jcis7841

[8] Seo, Young Mi Jee, Hong Ke and Soontak Lee, “Rainfall Frequency Analysis and Uncertainty Quantification Using Dempster-Shafer Theory, ” Korea Water Resources Association 2010 KWRA conference, pp.1390-1394, 2010.

[9] Burroughs, Daniel J., Linda F. Wilson and George V. Cybenko, “Analysis of distributed intrusion detection systems using Bayesian methods. Performance,” *The 21st IEEE International Computing, and Communications*, 2002. https://doi.org/10.1109/ipccc.2002.995166

[10] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R, “CRISP-DM 1.0 Step-by-step data mining guide”, IBM, 2000.

[11] Soukaena Hassan Hashem, “Efficiency of SVM and PCA to Enhance Intrusion Detection System,” *Journal of Asian Scientific Research,* Vol.3, No.4, pp.381-395, 2013.

[12] Hong, Sung-Sam, Wanhee Lee, and Myung-Mook Han, “The Feature Selection Method based on Genetic Algorithm for Efficient of Text Clustering and Text classification,” *International Journal of Advances in Soft Computing & Its Applications*, Vol.7, No.1, 2015.

[13] Rampure, Vinod, and Akhilesh Tiwari. ″A Rough Set Based Feature Selection on KDD CUP 99 Data Set.″ *International Journal of Database Theory and Application,* Vol.8, No.1, pp.149-156, 2015. https://doi.org/10.14257/ijdta.2015.8.1.16

[14] http://www.r-project.org/

[15] https://cran.r-project.org/package=e1071

[16] KDD‘ cup 99, “Knowledge discovery in databases DARPA archive,” http://www.kdd.ics.uci.edu/databases/kddcup99/task.html, 1999.

[17] Monowar H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, “Network Anomaly Detection: Methods, Systems and Tools,” *IEEE Communications Surveys & Tutorials,* Vol.16, No.1, pp.303-336, 2014. https://doi.org/10.1109/surv.2013.052213.00046

[18] Syarif, A. Prugel-Bennett, G. Wills, “Unsupervised clustering approach for network anomaly detection,” Networked digital technologies communications in computer and information science, Vol.293, Springer, pp.135-145, 2012.
https://doi.org/10.1007/978-3-642-30507-8_13

◐ 저 자 소 개 ◑

홍 성 삼(Sung-Sam Hong)
2009년 가천대학교 전자거래학과(공학사)
2011년 가천대학교 일반대학원 전자계산학과(공학석사)
2016년 가천대학교 일반대학원 전자계산학과(공학박사)
2016년~현재 가천대학교 컴퓨터공학학과 연구교수
관심분야 : 정보보호, 인공지능, 데이터 마이닝, 데이터 분석, 지능형 시스템
E-mail : sungsamhong0@gachon.ac.kr

김 동 욱(Dong-Wook Kim)
2015년 가천대학교 컴퓨터공학과 공학사
2015년~현재 가천대학교 일반대학원 IT융합공학과 석사과정
관심분야 : Data Mining, data fusion
E-mail : kog7306@naver.com

김 병 익(Byungik Kim)
2010년 2월 아주대학교 컴퓨터공학과 학사
2010년 7월~현재 한국인터넷진흥원 선임연구원
관심분야 : 네트워크보안, 악성코드 탐지 및 분석, CTI분석기술
E-mail : kbi1983@kisa.or.kr

한 명 묵(Myung-Mook Han)
1980년 연세대학교 요업공학과(공학사) (8.5pt)
1987년 뉴욕공과대학교 대학원 컴퓨터공학과(공학석사)
1997년 오사카시립대학교 대학원 정보공학부(이학박사)
1998년~현재 가천대학교 컴퓨터공학과 교수
관심분야 : 정보보호, 알고리즘, 데이터 마이닝
E-mail : mmhan@gachon.ac.kr