인공지능 보안

이수미

이상탐지와 비지도 학습

순서

1. 이상 탐지 이해
2. 비지도 학습과 이상탐지
3. 이상 탐지와 정보 보안
4. 정리

이상 탐지 이해

이상 탐지란

✓ 대다수의 데이터와 다른 양상을 보이는 특이하고 이상한(비정상으로 의심이 되는) 데이터를 찾아내는 기술을 의미

✓ 이상 금융 거래 탐지, 특정 질병의 발생, 생산 라인의 결함 관리, 네트워크 침입 탐지 등 다양한 분야에서 이상 탐지 기술을 적용

✓ 'Anomaly Detection'용어, outlier, exception, aberration, surprise detection
→ 근본적으로 '정상 범주에서 벗어나는 무언가'를 표현하기 위한 용어

✓ 'novelty detection'도 이상 탐지와 동일한 목표

- 정상과 이상을 학습해 새로운 데이터의 이상 여부를 판단하는 이상탐지와는 달리
- 지금까지 한 번도 관찰하지 않은 새로운 패턴을(학습 데이터에 포함되지 않는 이상 데이터도 찾아내는)탐지한다는 측면에서 약간의 차이가 있다.

Novelty Detection(이상치 탐지)

✓ Novel Data란?

"Observations that deviate so much from other observations as to arouse suspicions that they were generated by a different mechanism(Hawkins, 1980)" "Instances that their true probability density is very low(Harmelinget al., 2006)"

→ 다른 관측치랑 비교해서 많이 벗어나 있는 관측치가 이상치

✓Novel data vs noise data

Noise는 random error로서 이상치 탐지 전에 데이터 전처리 과정에서 제거해줘야 할 부분이며 outlier는 우리가 찾고자 하는 관측치

Novel Data의 유형 3가지

① Global outlier

✓ 일반적인 관측치들과 많이 동떨어진 관측치.

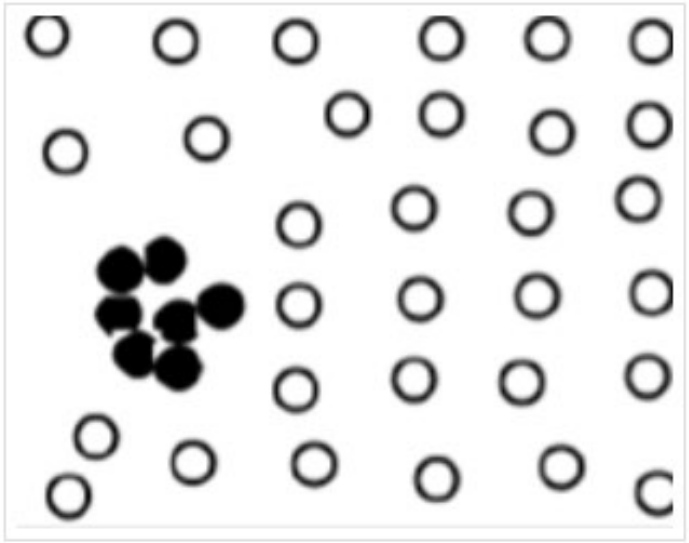
✓ 이러한 이상치들을 찾을 때는 얼마나 떨어져 있는가 측정을 어떻게 하는지가 중요

이미지 출처: https://jayhey.github.io/novelty%20detection/2017/10/18/Novelty_detection_overview/

Novel Data의 유형 3가지

② Contextual outlier(local outlier)

- ✓ 특정 부분에서 다른 부분과는 다른 양상을 띄는 이상치
- ✓ 예를 들어 사하라 사막의 온도를 측정하는데 어떤 한 부분의 온도가 영상 5도라면 이 관측치는 이상치
- ✓ 관측치들의 context를 어떻게 설정하는가(사막의 온도는 몇도부터 몇도까지인가)가 중요



이미지 출처: https://jayhey.github.io/novelty%20detection/2017/10/18/Novelty_detection_overview/

Novel Data의 유형 3가지

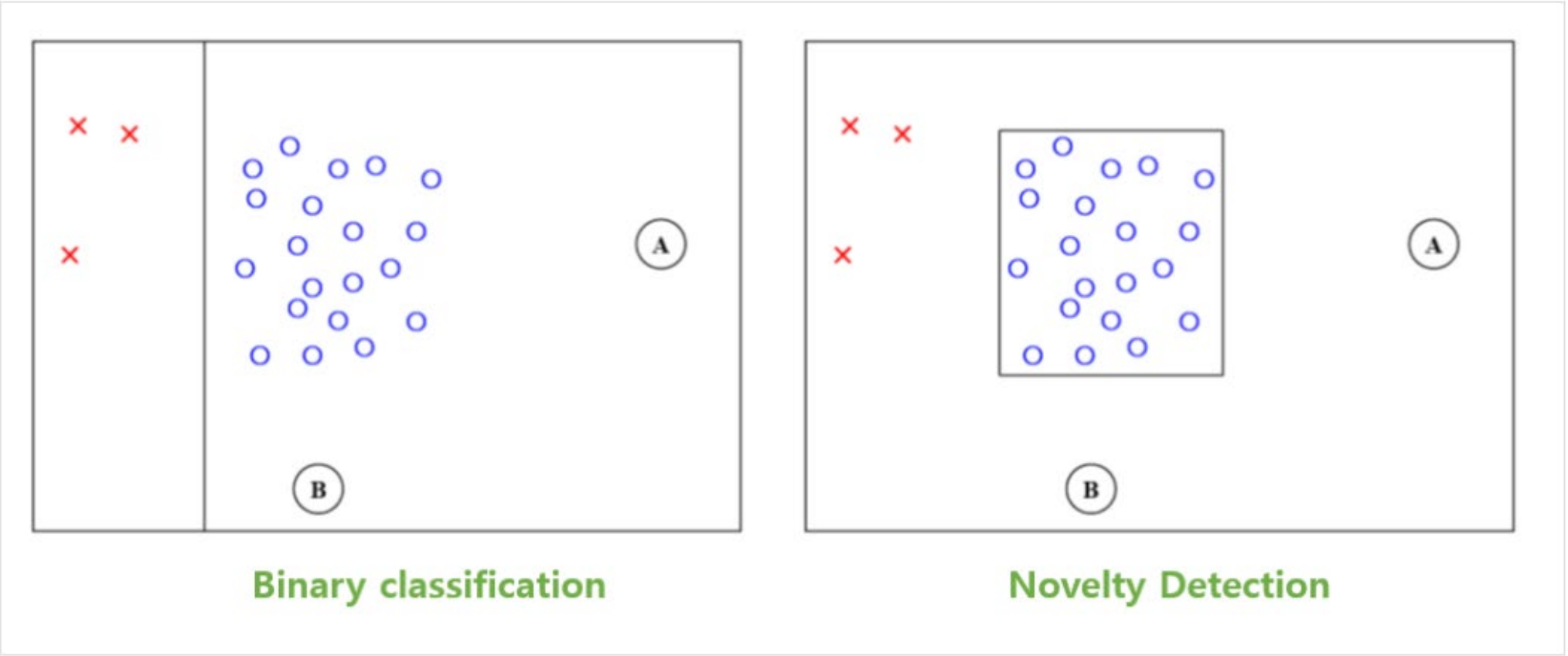
③ Collective outlier

- ✓ 관측치 하나하나가 outlier가 아닐지라도 전체 데이터를 봤을 때 편차가 심하게 난다면 collective outlier
- ✓ 만약 서버가 DDos 공격을 받는다면 전체적으로 봤을 때는 접속자 하나하나의 패킷 흐름은 전부 동일
- ✓ 그러나 동일한 패킷이 갑자기 엄청 많은 수가 한 번에 들어온다면 이 부분을 DDos 공격

  → DDos 공격도 collective outlier 종류 중 하나

Novelty Detection(이상치 탐지)

✓ Classification vs. Novelty Detection – 1



이미지 출처: https://jayhey.github.io/novelty%20detection/2017/10/18/Novelty_detection_overview/

Novelty Detection(이상치 탐지)

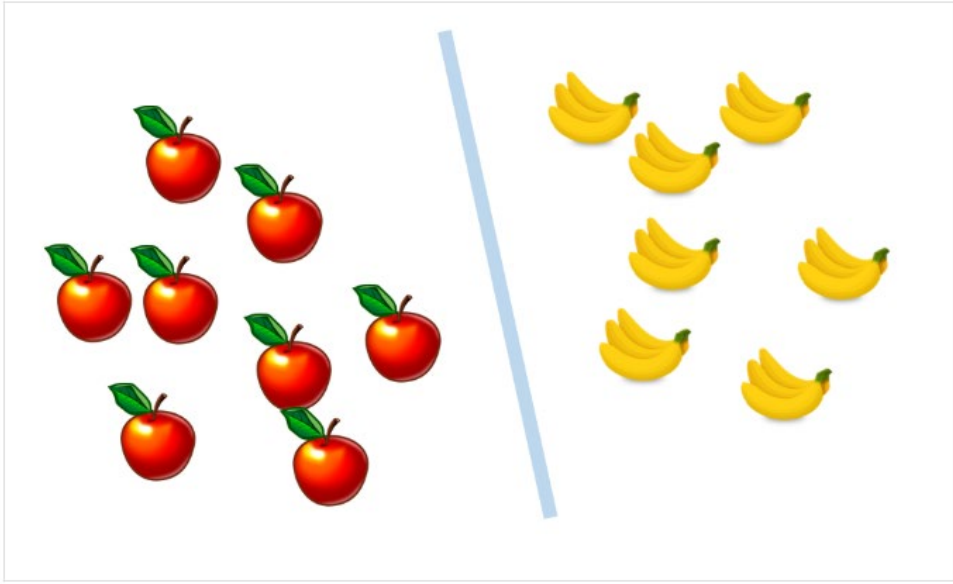
✓ Classification vs. Novelty Detection – 1

- classification → 경계면을 찾는 것.
- 하지만 novelty detection은 **다수 범주 데이터만 가지고 접근 →** 이상치가 아닌 데이터들의 영역을 칠해주는 것
- 예를 들어 100만 건 중 3건이 불량이라면 classification으로는 접근이 불가능
- 하지만 이상치 탐지 기법으로는 접근이 가능하기 때문에 분류기법으로 접근 못했을 때 사용하는 대안

Novelty Detection(이상치 탐지)

✓ Classification vs. Novelty Detection – 2

- 아래 그림은 사과와 바나나를 분류하는 것으로, 비슷한 수의 사과와 바나나가 존재하고 이 데이터를 통해서 분류 모델(가운데 하늘색 선)을 학습
- 분류 문제 두 범주에 대한 데이터가 어느정도 존재하고 “사과와 바나나” 이렇게 간단하게 나눌 수 있다

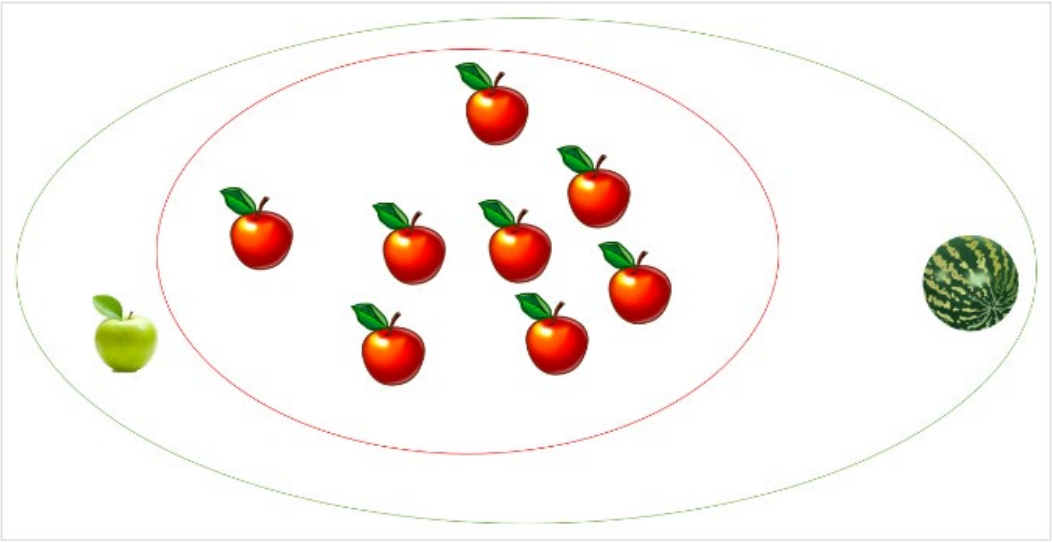


이미지 출처: https://jayhey.github.io/novelty%20detection/2017/10/18/Novelty_detection_overview/

Novelty Detection(이상치 탐지)

✓ Classification vs. Novelty Detection – 2

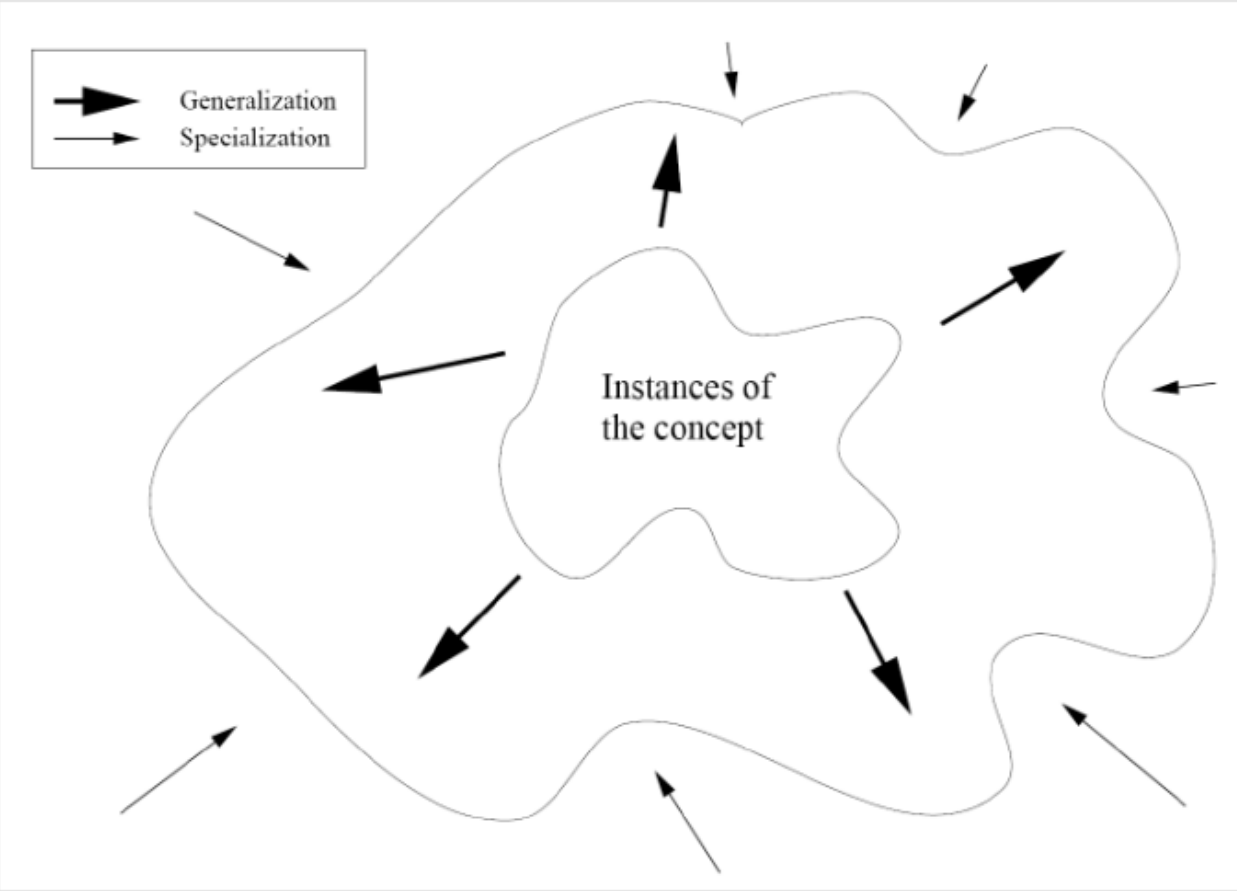
- 아래의 그림에서 사과의 경계에 빨간 원을 쳤습니다. '어? 분류가 잘 되었네..?'라고 생각할 수 있지만 여기서 문제가 발생 → 과연 수박도 모양만 봤을 때는 동그란데 사과라고 해야 하나? 조금 더 자세히 들여다 보니까 청사과는 빨간색이 아닌데 사과의 범주에 넣어야 하나..? 계속 경계를 어떻게 설정할지 생각을 해줘야 한다.



이미지 출처: https://jayhey.github.io/novelty%20detection/2017/10/18/Novelty_detection_overview/

Novelty Detection(이상치 탐지)

- ✓ Generalization vs. Specialization
  - ✓ Generalization을 더 한다면 청사과는 포함시킬 수 있지만 너무 과도하면 수박까지 사과로 취급,
    하지만 specialization이 과하면 청사과를 사과로 생각하지 않게 될 수도 있다.
  - ✓ 이를 잘 조절하려면 도메인에 대한 지식 그리고 이상치 탐지 모델에 대한 이해도가 높아야 할 것



이미지 출처: https://jayhey.github.io/novelty%20detection/2017/10/18/Novelty_detection_overview/

이상 탐지 사례

✓ 이상 금융 거래 탐지 시스템(FDS, Fraud Detection System)
월요일부터 금요일까지 매일 같은 시간에 출퇴근을 하는 김모씨는 퇴근길에 슈퍼에 들러 아이들 과자를 사 오는 것이 유일한 낙인 평범한 직작인이다. 어느 날 갑자기 미국 뉴욕에 있는 애플 매장에서 1,000달러가 결제되었다는 문자 메시지가 전송되었다. 당황한 김모씨는 카드사에 신고해 해당 결제가 자신이 한 것이 아님을 신고했다. 그 동안의 카드 사용 패턴을 분석한 카드사는 개인 정보 유출로 인한 부당 결제로 결론 짓고 정상 취소 처리와 함께 지역 경찰에 정보를 넘겼다

✓ 설비 운용 중 발생하는 저주파 신호와 진동을 측정하는 센서가 구축된 경우, 우선 대상 설비가 정상적으로 동작할 때 발생하는 저주파 신호와 진동을 분석해 패턴 정보를 구축한다. 이 패턴 정보를 토대로 설비 상태를 지속적으로 모니터링하고, 평상시 패턴과 다른 양상을 보이는 신호 발생 여부를 찾아 고장을 미리 예측하고 빠르게 대응할 수 있다.

✓ 이 밖에도 질병 발병 예측, 이상 행위자 탐지(영상 기반), 네트워크 이상 탐지 등 다양한 분야에서 이상 탐지 기술을 사용할 수 있다.

이상 탐지 사례

✓위 사례의 공통점은 바로 ‘이상하다’라는 결론을 도출하는 방식

✓‘이상’을 주장하려면 먼저 ‘정상’에 대한 기준과 근거 데이터 필요

- 첫 번째 사례: 사용자의 평소 지출 패턴이 정상의 기준
- 두 번째 사례: 평시 운용 중 발생하는 센서 데이터 패턴이 정상의 기준

✓네트워크 이상 징후 탐지 분야에서도 마찬가지로 ‘정상’에 대한 명확한 정의가 필요

이상탐지는 어렵다

✓이상 탐지 시스템 구축 시 반드시 고려해야 할 제약사항

- 정상과 이상을 결정하는 기준
- 모든 분야에 적용 가능한 만능 이상 탐지 기술은 존재하지 않는다
- 이상과 노이즈의 경계선은 종이 한 장 차이다
- 이상 데이터를 구하는 것은 어렵다
- 수집한 데이터에 정확한 레이블을 지정하는 것이 어렵다
- 정상과 이상에 대한 기준은 시간이 지나면서 변화할 수 있다.
- 악의적인 사용자가 이상 데이터를 변조할 수도 있다

✓첫 번째부터 세 번째 항목은 가장 어렵고 중요한 항목이자 이상 탐지 시스템의 근본적인 목적과도 밀접하게 관련

이상탐지는 어렵다

✓매일 동일한 패턴으로 소비하는 사람도 가끔 기분이 우울할 때 과소비를 할 수도 있고,

✓해외 출장을 가서 새롭게 출시된 노트북을 구매할 수도 있다

✓단순히 평소와 다른 패턴을 보인다고 해서 '비정상 또는 이상'으로 판단할 수 있을까?

✓카드사 입장에서는 특정 행동의 '결과'에 해당하는 결제 금액, 매장 위치, 시간 등만 파악할 수 있을 뿐 특정 결제를 한 상황과 문맥은 알 수 없다.

✓'평소와 다른 패턴이 보인다'라고 말할 수 있지만 '이 패턴은 이상하다'라는 결론을 섣불리 내릴 수 없다.

이상탐지는 어렵다

✓정상의 '정도'도 성능에 매우 중요한 영향

- '이상'을 잡아내는 정도를 민감하게 설정할 경우, 과도한 이상 탐지 경보를 받아보게 된다.
- 이상의 정도가 극명하게 나타나는 데이터만 탐지하도록 시스템을 설계할 경우 시스템에 치명적인 영향을 주는 이상 데이터를 잘 잡아내지 못할 수도 있다.

이상탐지는 어렵다

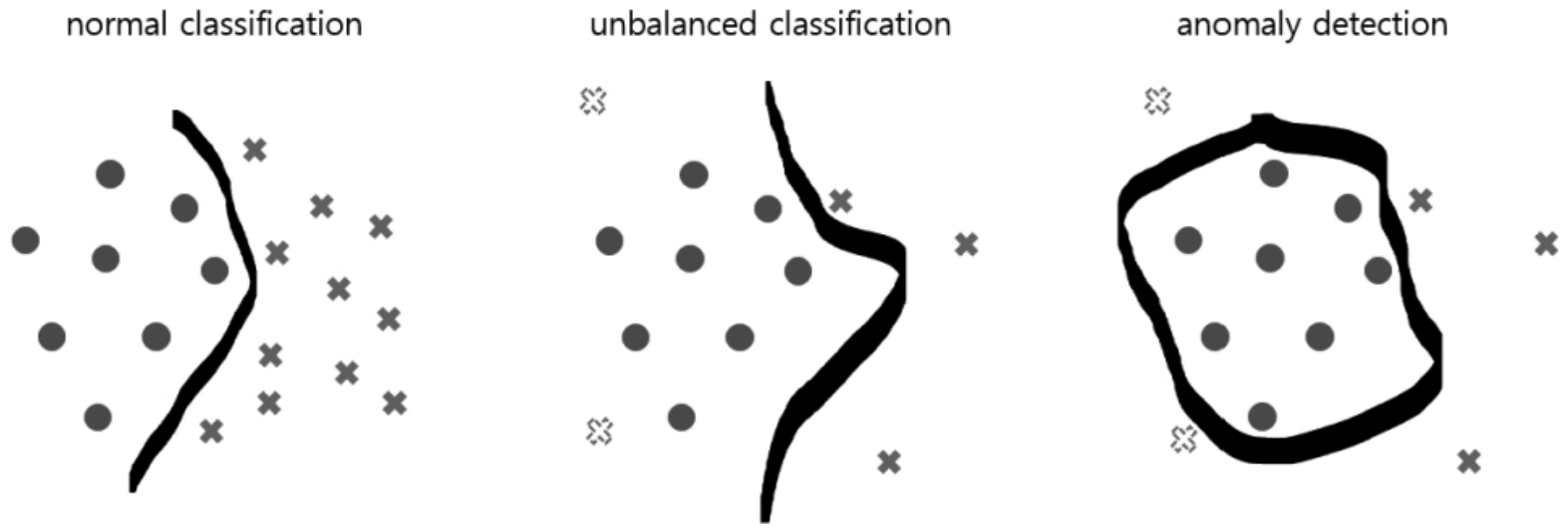
✓ 이상 탐지 시나리오에서 사용하는 데이터는 일반적인 분류(classification) 시나리오와 달리 이상 데이터를 확보하기 어렵다

- 지난 20년 동안 운용해온 발전 설비의 이상 징후를 찾아내는 시스템
  20년(17만 5,200 시간) 동안 발전 설비가 돌아가면서 고장이 몇 번이 났을까?
  전체 운용 시간 대비 실제 고장 난 시간은 17만 시간 중 1,000시간(41일) 채 되지 않을 것
  즉, 20년 동안 수집한 데이터에서 이상 데이터의 비율은 0.01%도 안된다
- 네트워크 침입 탐지 분야: 하루에도 수십억 건이 넘는 네트워크 트래픽이 발생하는데, 이 중에서 의심이 되는 이상 데이터 또는 악성 공격 데이터의 비중은 매우 적다
  비율이 극단적으로 차이가 날 경우 일반적인 분류 알고리즘을 사용해 모델을 만들기 어렵다.

이상탐지는 어렵다

✓분류 알고리즘은 충분한 정상 데이터의 패턴과 이상 데이터의 패턴이 갖춰진 상황에서 최적의 선을 찾게 된다.

✓이상 데이터가 매우 적은, 그리고 모든 이상 데이터를 다 확보하지 못한 상황에서 분류 모델을 적용할 경우 한 번도 보지 못한 새로운 유형의 이상 데이터를 정상으로 분류할 확률이 매우 높아짐



[그림 12-1] 일반적인 분류 모델(첫 번째)과 이상 데이터가 부족한 상황에서 만든 분류 모델(두 번째), 그리고 이상 탐지 모델(세 번째) 예시

이상탐지는 어렵다

- ✓수집한 이상 데이터가 확실히 이상 데이터인지 확인하는 일 → 무한대의 시간과 인력이 있다면 가능하겠지만 개별 데이터를 일일이 검사해 정확한 레이블을 지정하는 것은 불가능
- ✓수집한 데이터를 완벽히 분석해서 레이블링 한 후, 명확한 정상과 이상의 기준을 파악하고 이를 토대로 모델을 만든 경우, 일정 기간 동안에는 탐지 시스템을 활용할 수 있다.
- ✓데이터의 속성은 시간이 지나면서 변화하기 마련이고, 정상과 이상의 기준이 달라지기 때문에 데이터 속성과 이상 기준의 변화를 고려해 시스템을 설계해야 함

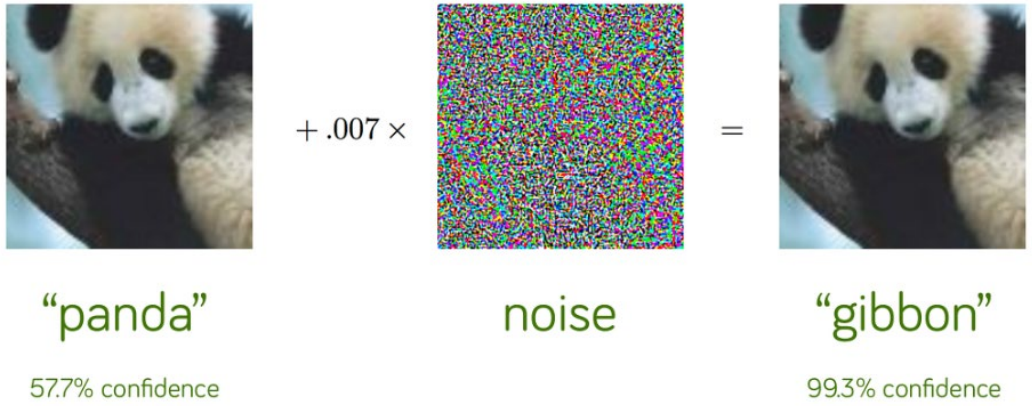
이상탐지는 어렵다

✓악의적인 사용자 또는 공격자들이 자신이 만든 악성 데이터가 정상처럼 보이도록 가장하는 방법

✓적대적인 머신러닝(Adversarial Machine Learning)

- 인공지능 기반 공격 탐지 모델을 공격하는 기법
  - 데이터에서 얻은 정보로 공격 여부를 판별하는 탐지 모델의 탐지 성능을 떨어뜨리는 것을 목표로 함
  - 인공지능 보안 모델 구축 시 시큐어 모델링도 함께 고려해야 함
- Adversarial Threat
  - 딥러닝 모델의 내부적 취약점을 이용하여 만든 특정 노이즈 값을 이용해 의도적으로 오분류를 이끌어내는 입력 값(AEs: Adversarial Examples)을 만들어 내는 것

[출처: https://rain-bow.tistory.com/entry/Adversarial-Attack]



이상 탐지 기법

(1) 학습 방식과 데이터 유형에 따른 분류

✓학습 방식: 이상 탐지는 지도 학습, 준지도 학습, 비지도 학습 모두와 관련이 있다.

- 지도학습: 학습 전에 전체 데이터를 정상과 이상으로 명확히 구분이 가능한 경우
- 준지도 학습: 정상 데이터만 명확히 구분이 가능한 경우
- 비지도 학습: 정상과 이상 데이터 중 어떠한 데이터도 명확히 분류가 어려운 상황

이상 탐지 기법

(1) 학습 방식과 데이터 유형에 따른 분류

✓데이터 유형

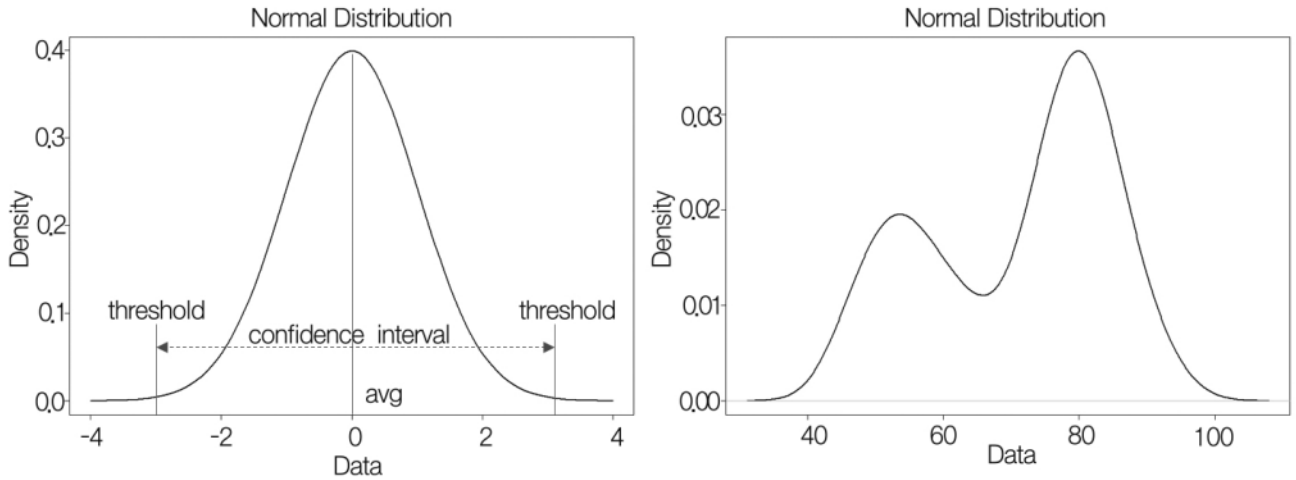
- 데이터의 형식이 수치형이자 동시에 정적인 데이터일 경우 데이터가 좌표상에 위치하는 점을 기준으로 정상과 이상의 영역을 구분할 수 있다.
- 데이터가 1차원이고(특징이 1개) 중심 값이 하나인 경우 특정 데이터가 전체 데이터의 중심으로부터 얼마나 떨어져 있는 지를 '이상'데이터 판단 기준으로 사용할 수 있다.

이상 탐지 기법

(1) 학습 방식과 데이터 유형에 따른 분류

✓데이터 유형

- [좌측 이미지]는 데이터 중심을 기준으로 양 끝 임계점을 결정한 후, 임계점 안쪽(신뢰구간) 부분에 위치한 데이터는 정상, 바깥 쪽에 위치한 데이터는 이상으로 판단할 수 있다.
- [우측 이미지] 여러 개의 중심, 여러 개의 분포가 혼합된 형태의 경우 단순히 중심 경향만으로 이상 여부를 판단하기 어렵다. → 클러스터를 기반 방식으로 사용(어떠한 클래스에도 속하지 않는 데이터, 특정 거리 이상 떨어진 데이터)



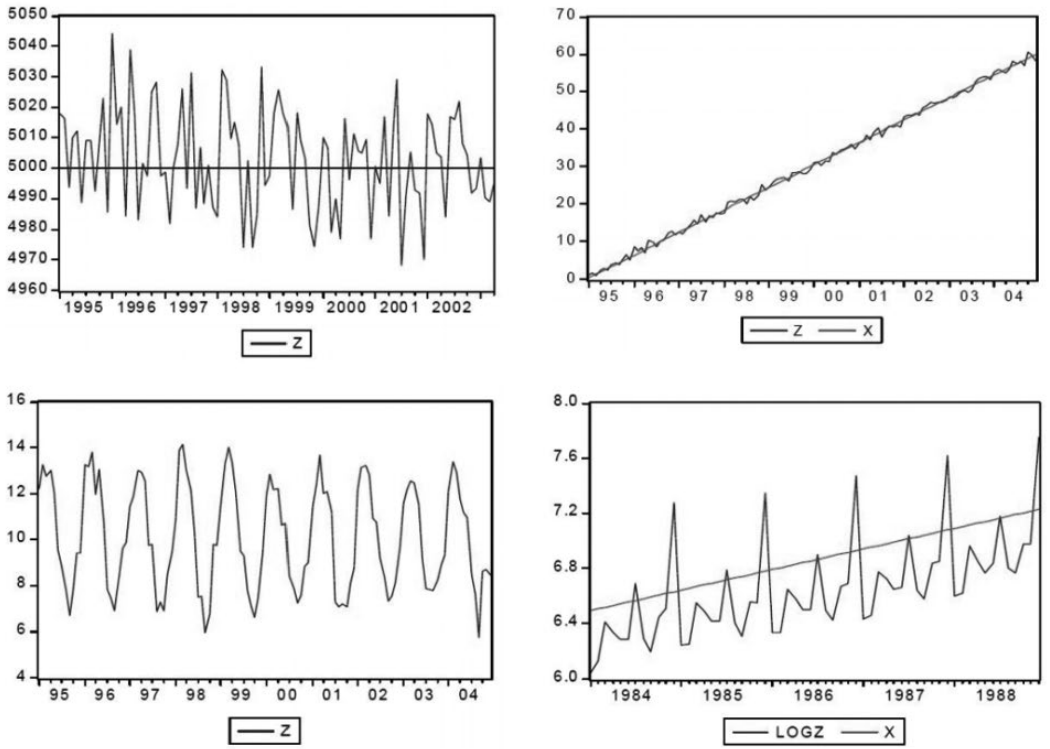
[그림 12-2] 정규 분포(좌)와 다봉 분포(우) 예시

이상 탐지 기법

(1) 학습 방식과 데이터 유형에 따른 분류

✓데이터 문맥(contextual): 인과 관계와 시간 흐름을 함께 고려해 이상 데이터를 판별

- 시계열 데이터의 이상탐지
  - Trend, cycle, seasonality, irregularity 형태
  - 실제 분석 대상이 되는 시계열 데이터는 개별 형태보다 규칙성을 가지는 패턴과 불규칙한 패턴이 혼합된 형태
  - 시계열 데이터 분석의 가장 어려운 점은
  - ‘어느 정도의’ 데이터를 판단의 단위로 사용할지 결정하는 것



[그림 12-3] 시계열 데이터의 형태 예시(좌측 상단 이미지부터 시계 방향으로 불규칙 변동, 추세 변동, 계절 변동, 추세+계절 변동)

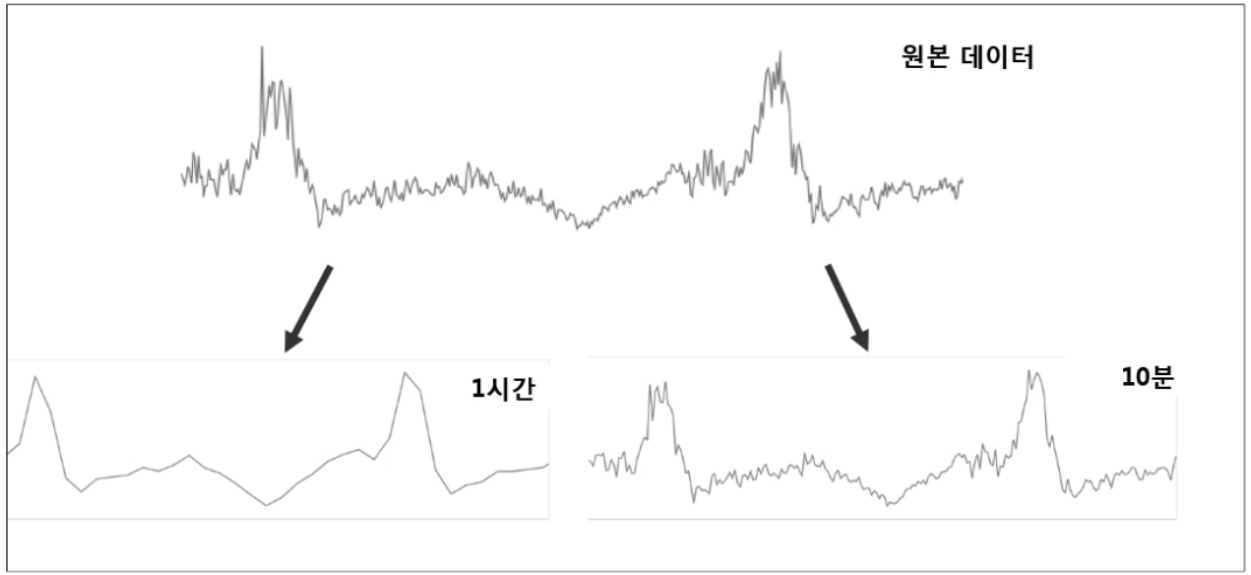
이상 탐지 기법

(1) 학습 방식과 데이터 유형에 따른 분류

✓데이터 문맥(contextual): 인과 관계와 시간 흐름을 함께 고려해 이상 데이터를 판별

- 내부 네트워크로 유입되는 트래픽 모니터링
  - 1시간 단위로 잘라 샘플링 분석할 경우 10분 단위로 데이터를 자르는 것보다 완만한 추이선 확인

    → 탐지를 원하는 '이상 데이터'의 특성을 고려해 적절한 크기의 샘플링 단위를 결정해주는 과정 선행!



[그림 12-4] 샘플링 단위에 따라 달라지는 데이터의 형태[78]

이상 탐지 기법

(2) 이상 탐지 방식에 따른 분류

✓통계 기반, 근접도(proximity) 기반, 밀도(density) 기반, 클러스터링 기반, 트리 기반, 주성분 분석 기반, 확률 기반 방식으로 구분

✓정상과 이상을 표현하는 방식과 기술적인 기준의 차이가 있을 뿐
'정상을 규정한 후 정상을 벗어나는 데이터를 이상으로 판정'하는 궁극적인 목표는 모두 동일

이상 탐지 기법

(2) 이상 탐지 방식에 따른 분류

✓통계 기반방식: 통계 모델(평균과 분산)을 벗어나는 이상치(outlier)를 찾는 방식

- 정규 분포 형태를 띄는 데이터의 이상치를 찾는 방법
  → 간단하고 직관적이지만, 데이터가 정규 분포 형태를 띄고 있어야 한다는 점, 고차원 데이터의 이상치를 찾기 어렵다는 단점

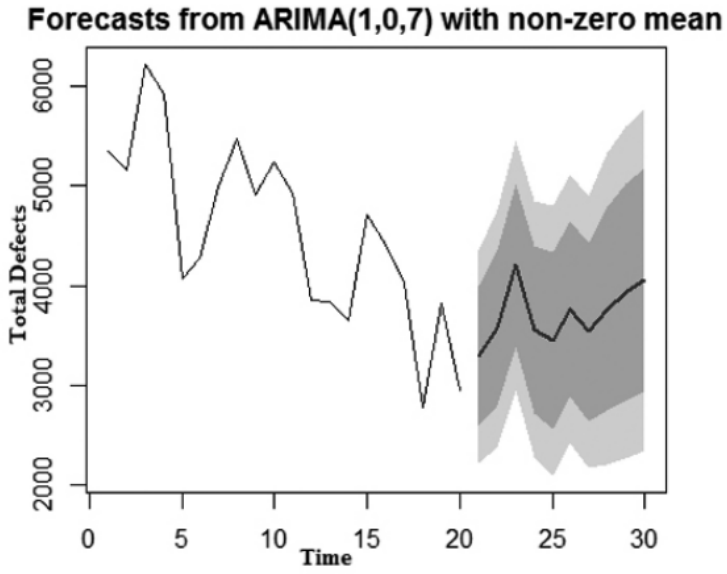
- 시계열 데이터의 경우 주어진 학습 데이터를 분석해 예측 모델을 만드는 방식으로 이상 데이터 탐지
  → 대표적인 모델로 ARIMA 모형이 있으며, 최근에는 딥러닝 기반 모델은 RNN의 원리를 이용한 연구 진행
  - ARIMA 모형은 AR(이전의 결과와 이후 결과 사이에서 발생하는 자기 상관성을 기반)모형 + MR (이전에 생긴 불규칙한 사건이 이후의 결과에 편향성을 초래하는 이동 평균 기반) 모형
    ➔ 누적된 데이터의 패턴과 데이터의 경향성 변화(이동 평균) 모두 고려해 이상 데이터를 찾는 방식

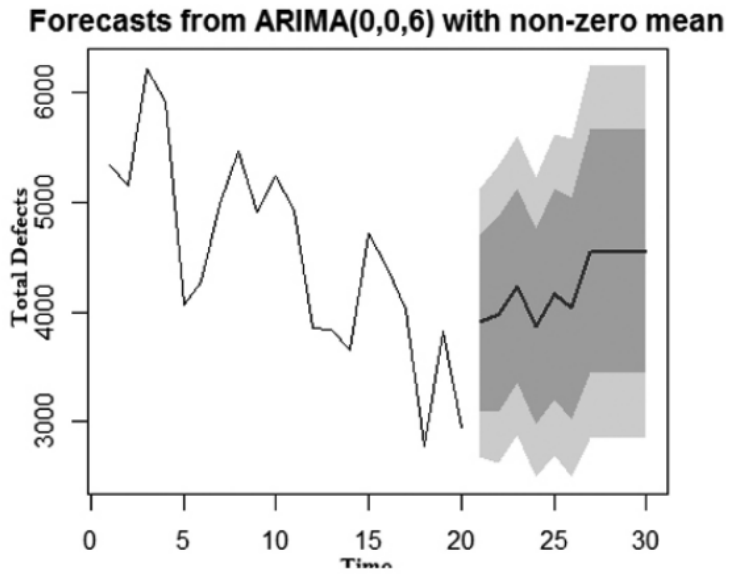
이상 탐지 기법

(2) 이상 탐지 방식에 따른 분류

✓통계 기반방식: 통계 모델(평균과 분산)을 벗어나는 이상치(outlier)를 찾는 방식

- ARIMA 모형을 사용한 시계열 데이터 이상 탐지 예시
  - 음영 안의 선은 실제 데이터의 추이, 음영 영역은 지금까지의 데이터 패턴과 이동 평균을 고려해 '정상'으로 간주 가능한 영역을 의미 → 음영을 벗어난 패턴이 보일 경우 '이상' 데이터로 판단하는 방식





이상 탐지 기법

(2) 이상 탐지 방식에 따른 분류

✓근접도, 밀도, 클러스터링 기반 방식:

- 데이터가 충분히 밀집되어 있는 많은 데이터로부터 어느 정도 떨어져 있는지를 판단해 이상여부를 결정
- 통계 기반 방식과 마찬가지로 거리를 계산한다는 점에서는 동일, 단순히 데이터의 중심으로부터의 거리가 아닌 복잡하게 얽힌 다차원 평면상에서 밀집도와 거리를 이용한다는 점에서 차이가 있다.
- 장점: 데이터의 이상 정도를 정량적으로 평가할 수 있고, 다차원 데이터에도 적용할 수 있다.
- 단점: 데이터 밀집도가 충분히 높지 않은 경우에 적용하기 어렵다. 모델 복잡도가 높고 초기값(예를 들어, KNN의 K값, DBSCAN의 epsilon)을 필요로 한다는 점에서 사용에 제약이 따른다.

이상 탐지 기법

(2) 이상 탐지 방식에 따른 분류

✓주성분 분석 기반 방식: 데이터의 분산을 최대화하는 주성분을 찾는 분석 기술

- 주성분: 데이터의 패턴을 가장 잘 나타내 주는 핵심 축으로, 이 축을 이용해 이상데이터를 탐지할 수 있다.
- 주성분 분석의 경우 데이터에 이상치가 포함되어 있을 경우 주성분을 제대로 찾아내지 못한다.
  → 최소한 정상 데이터에 대한 레이블링이 확실하게 정리된 상태에서 수행해야 한다.
- PCA 기반 이상 탐지는 정상 데이터와 이상 데이터가 서로 다른 주성분을 가지게 된다는 것을 이용해 이상 데이터를 탐지하는 방법으로 다음과 같은 과정을 수행한다.

  ① 정상으로 분류된 학습 데이터를 대상으로 주성분 분석을 수행한다.

  ② 가장 높은 비율을 차지하는 상위 성분을 선택한다

  ③ 상위 성분을 이용해 마할라노비스 거리를 계산한다.

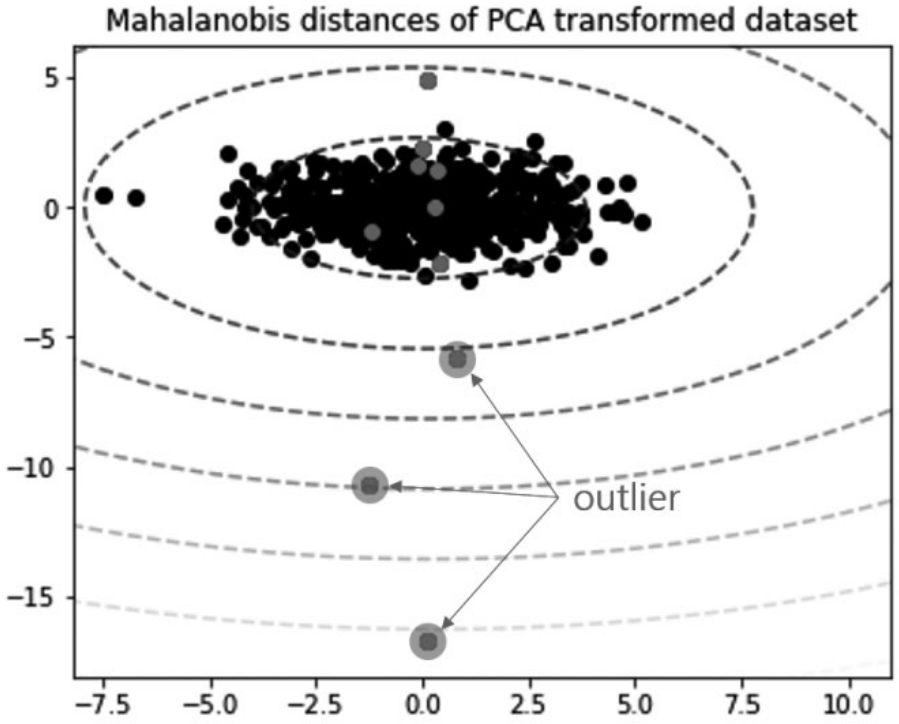
  ④ 새롭게 유입되는 데이터의 마할라노비스 거리 값이 크다면 이상 데이터로 판단한다.

이상 탐지 기법

(2) 이상 탐지 방식에 따른 분류

✓주성분 분석 기반 방식: 데이터의 분산을 최대화하는 주성분을 찾는 분석 기술

- 주성분: 데이터의 패턴을 가장 잘 나타내 주는 핵심 축으로, 이 축을 이용해 이상데이터를 탐지할 수 있다.



[그림 12-6] PCA와 마할라노비스 거리를 이용한 이상 탐지

비지도 학습과 이상 탐지

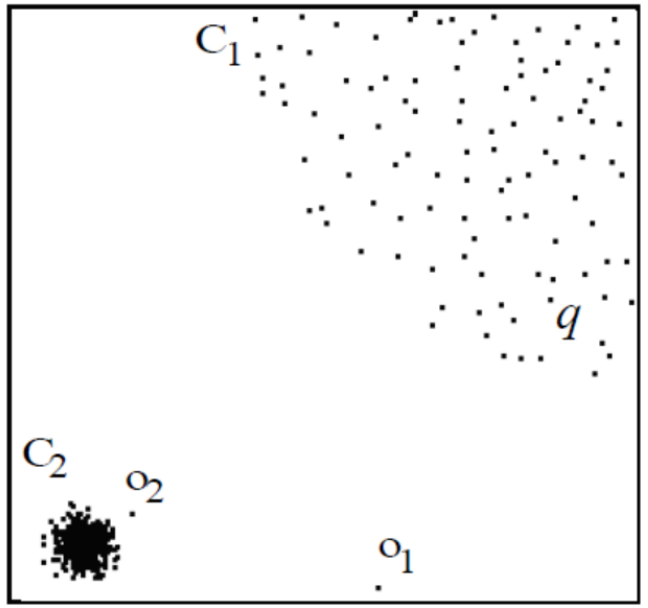
LOF(Local Outlier Factor) 소개

- ✓ Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000, May). LOF: identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data (pp. 93-104). 논문
- ✓ 고려대학교 강필성 교수님의 자료 참고

LOF(Local Outlier Factor) Motivation

✓ 대부분의 이상치 탐지 알고리즘은 전체 데이터와 비교하여 각각의 관측치가 이상치인지 아닌지 판단한다. 이러한 알고리즘은 아래 그림에서 O_2에 대해 이상치라고 판단하지 않는다.

- ✓ C_1 – 밀도가 낮은 그룹, C_2 – 밀도가 높은 그룹
- ✓ O_1 ,O_2 – 이상치



이미지 출처: https://velog.io/@vvakki_/LOFLocal-Outlier-Factor

✓ C_2 와 O_2 사이의 거리는 C_1그룹내 관측치 간의 거리와 유사하기 때문에, 전체적인 데이터 관점에서 보면 O_2 는 이상치로 보기 어렵다. 이러한 단점을 극복하기 위해, LOF는 국소적인(local) 정보를 이용하여 이상치 정도를 나타내는 것을 목적으로 한다.

LOF(Local Outlier Factor) 개념

✓LOF는 각각의 관측치가 데이터 안에서 얼마나 벗어나 있는가에 대한 정도(이상치 정도)를 나타낸다.

✓중요한 특징은 모든 데이터를 전체적으로 고려하는 것이 아니라, 해당 관측치의 주변 데이터(neighbor)를 이용하여 국소적(local) 관점으로 이상치 정도를 파악하는 것

✓주변 데이터를 몇개까지 고려할 것인가를 나타내는 k라는 하이퍼-파라미터(hyper-parameter)만 결정하면 되는 장점이 있다

LOF(Local Outlier Factor) 수식

✓K-distance:

- k-distance(A) 는 A로부터 k번째 근접 이웃까지의 거리. k-distance안에 들어오는 오브젝트의 집합을 $N_k(A)$라고 정의 → k-distance보다 작거나 같은 거리를 가지는 수

1st	2nd	3rd	4th	5th	6th	7th	3-distance	$N_3(A)$
1	2	3	3	3	4	5	3	5
1	2	2	2	3	4	5	2	4
1	1	1	1	2	3	4	1	4

LOF(Local Outlier Factor) 수식

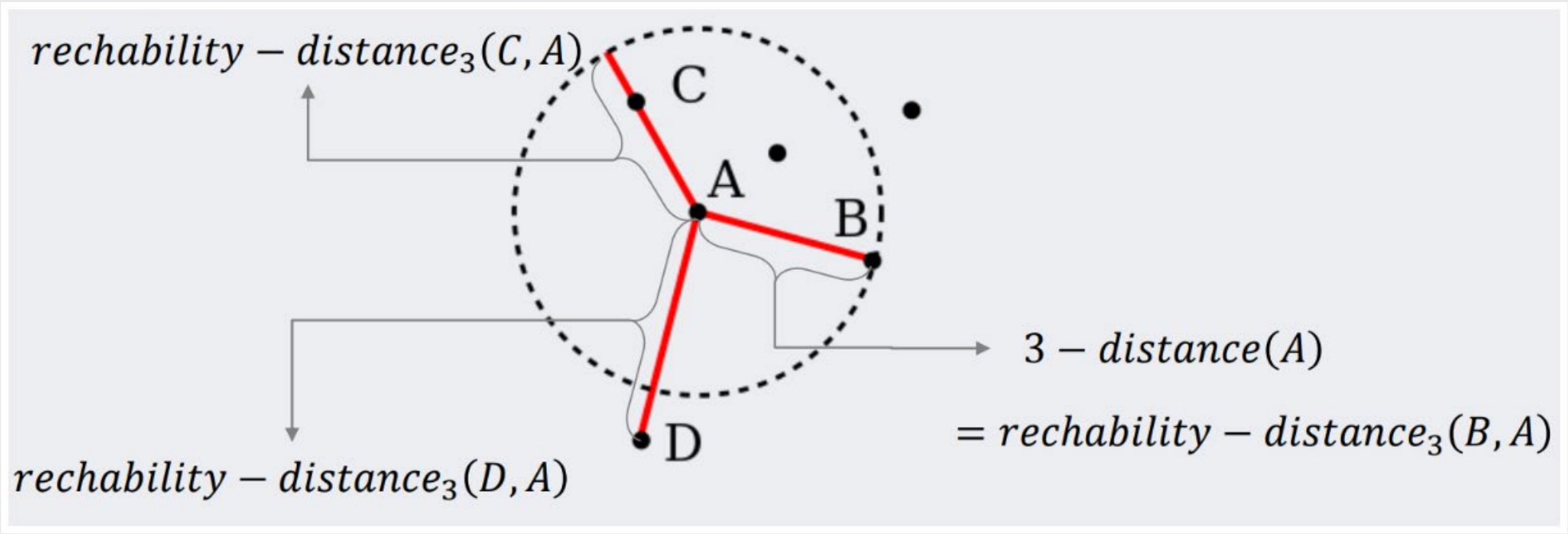
✓reachability distance

$$reachability - distance_k(A, B) = max\{k - distance(B), dist(A, B)\}$$

- A와 B까지의 거리 그리고 k-distance중 큰 값을 사용.
- k-distance 안에 들어오는 object들은 전부 원 위로 밀어내고 원 밖은 그대로 거리 값을 사용 하는게 reachability distance
- 만약 항상 k-distance를 사용한다면, LOF가 아닌 Simplified-LOF

LOF(Local Outlier Factor) 수식

✓reachability distance



- A를 기준으로 봤을 때, B와 C까지의 reachability-distance는 원 위로 밀어내서 3-distance(A)와 같아진다. 그리고 D의 경우는 원 밖에 있으니까 그대로 거리 값을 사용

LOF(Local Outlier Factor) 수식

✓Local reachability density

- 오브젝트 A에 대한 local reachability density는 다음과 같이 구할 수 있다.

$$lrd_k(A) = \frac{|N_k(p)|}{\sum_{O \in N_k(A)} reachability - distance_k(A, B)}$$

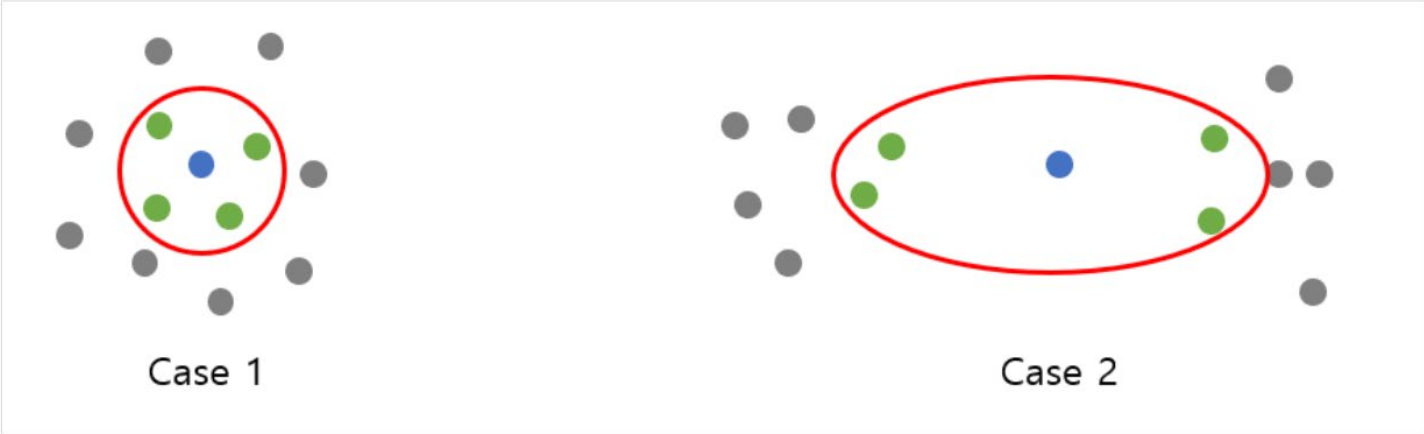
- 분자는 k-distance 안의 개체 수이고 분모는 A에서 다른 오브젝트들까지의 reachability-distance의 입니다. A로부터 다른 오브젝트들 까지의 reachability distance들의 평균 식을 거꾸로 뒤집은 것과 같다
  - Case 1 : 만약 A가 밀도가 높은(dense area) 지역에 위치한다면 분모가 작아질 것이고 $lrd_k(A)$값이 커짐
  - Case 2 : 반대로 A가 밀도가 높지 않은(sparse area) 지역에 위치한다면 분모가 커지게 되고 $lrd_k(A)$값은 작아짐

LOF(Local Outlier Factor) 수식

✓Local reachability density

$$lrd_k(A) = \frac{|N_k(p)|}{\sum_{O \in N_k(A)} reachability - distance_k(A, B)}$$

- Case 1 : 만약 A가 밀도가 높은(dense area) 지역에 위치한다면 분모가 작아질 것이고 $lrd_k(A)$값이 커짐
- Case 2 : 반대로 A가 밀도가 높지 않은(sparse area) 지역에 위치한다면 분모가 커지게 되고 $lrd_k(A)$값은 작아짐



이미지 출처: https://jayhey.github.io/novelty%20detection/2017/11/10/Novelty_detection_LOF/

LOF(Local Outlier Factor) 수식

✓Local Outlier Factor

$$LOF_k(A) = \frac{\sum_{B \in N_k(A)} \frac{lrd_k(B)}{lrd_k(A)}}{|N_k(A)|} = \frac{\frac{1}{lrd_k(A)} \sum_{B \in N_k(A)} lrd_k(B)}{|N_k(A)|}$$

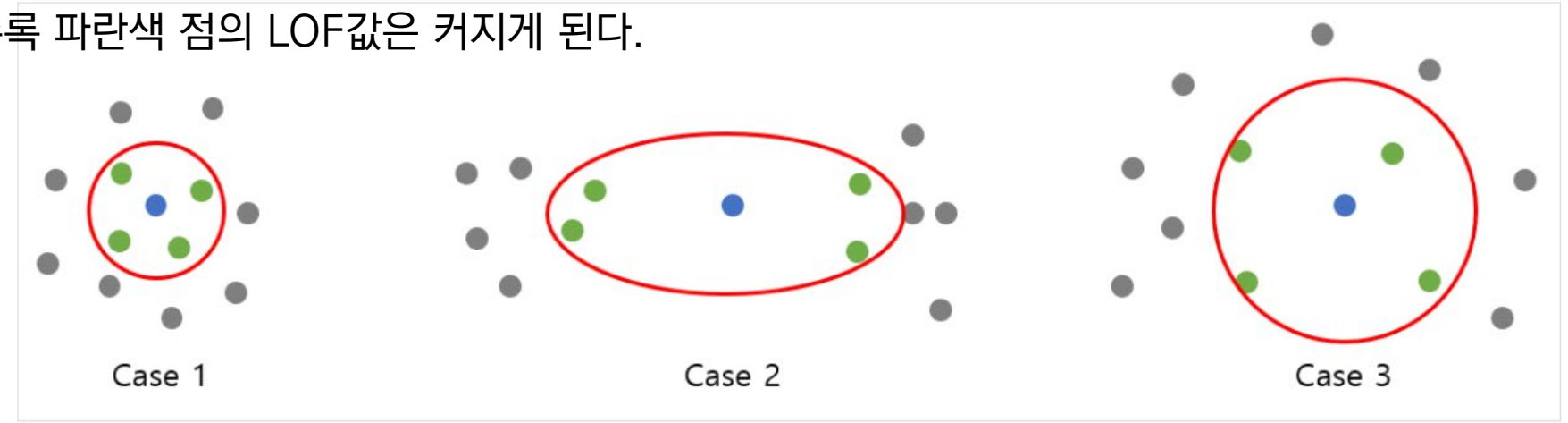
- 위 식에서 분자를 보면 A에 속한 B의 local reachability density의 평균을 $|N_k|$로 나눴다는 것을 확인
- 여기서 산출된 LOF score는 결국 A가 얼마나 이상치인가를 나타내는 정도

LOF(Local Outlier Factor) 수식

✓Local Outlier Factor

$$LOF_k(A) = \frac{\sum_{B \in N_k(A)} \frac{lrd_k(B)}{lrd_k(A)}}{|N_k(A)|} = \frac{\frac{1}{lrd_k(A)} \sum_{B \in N_k(A)} lrd_k(B)}{|N_k(A)|}$$

- 파란색 점이 A이고 초록색 점이 B입니다. LOF(A)값이 크다는 것은, 초록색 점들의 lrd(B)가 높고 파란색 점의 lrd(A)가 낮다. → 파란색 점이 밀도가 낮은 지역에 있을수록, 초록색 점들이 밀도가 높은 지역에 위치할수록 파란색 점의 LOF값은 커지게 된다.



Case	$ldr_k(A)$	$ldr_k(B)$	$LOF_k(A)$
Case 1	Large	Large	Small
Case 2	Small	Large	Large
Case 3	Small	Small	Small

이미지 출처: https://jayhey.github.io/novelty%20detection/2017/11/10/Novelty_detection_LOF/

LOF(Local Outlier Factor) 정리

✓장점: 밀집된 클러스터에서 조금만 떨어져 있어도 이상치로 탐지

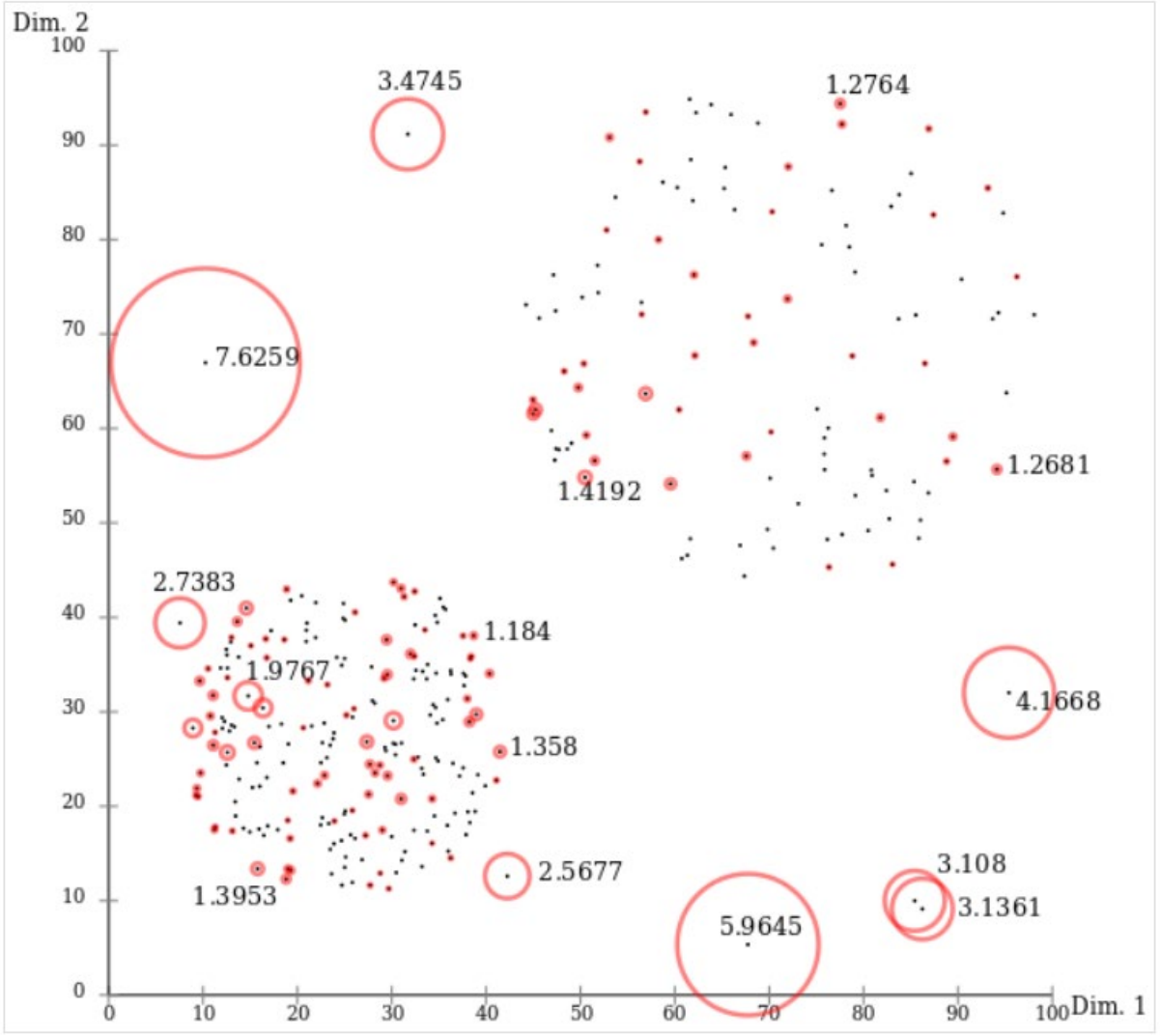
✓단점:

- 이상치라고 판단하는 기준을 어디에 잡아야 할 지 정해줘야 한다.
  - 다음 장의 그림의 경우 2차원 데이터라서 쉽게 시각적인 도움을 받을 수 있지만 차원이 늘어나면 판단하기 어렵다.
  - 어떤 데이터셋에서는 1.1이라는 값이 이상치 이지만 어떤 데이터셋에서는 2라는 값을 가지고 있어도 정상 데이터일 수 있다.

LOF(Local Outlier Factor) 정리

✓아래 그림에서 숫자들이 LOF 스코어를 나타내고 있다.

괴장히 빽빽한 곳에 가까운 이상치들은 확실히 더 높은 LOF값을 가지는 것을 확인할 수 있다



이미지 출처: https://jayhey.github.io/novelty%20detection/2017/11/10/Novelty_detection_LOF/

Isolation Forest 개념

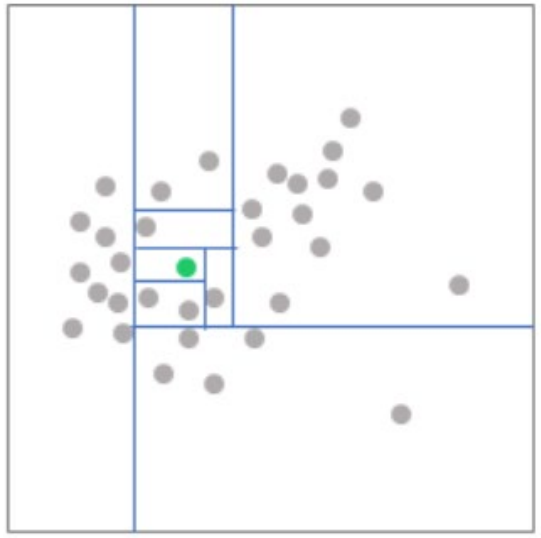
✓Isolation Forest는 Unsupervised Anomaly Detection 중 하나로 현재 갖고 있는 데이터 중 이상치를 탐지할 때 주로 사용. 이름에서 볼 수 있듯이 tree 기반으로 구현되는데, 랜덤으로 데이터를 split하여 모든 관측치를 고립시키며 구현

✓특히, 변수가 많은 데이터에서도 효율적으로 작동할 수 있는 장점이 있다

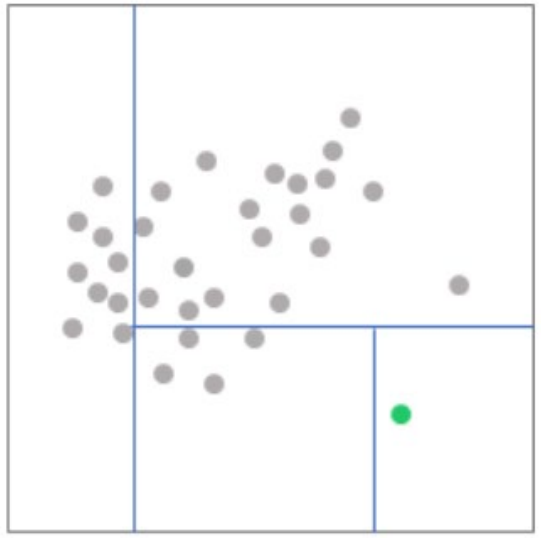
> The term *isolation* means 'separating an instance from the rest of the instances'.
>
> - Isolation Forest original paper

Isolation Forest 개념

✓Isolation Forest의 컨셉은 "각 관측치를 고립(=분리)시키기는 것은 이상치가 정상 데이터보다 쉽다.“



(1) 정상 데이터를 분리하는 경우



(2) 이상치를 분리하는 경우

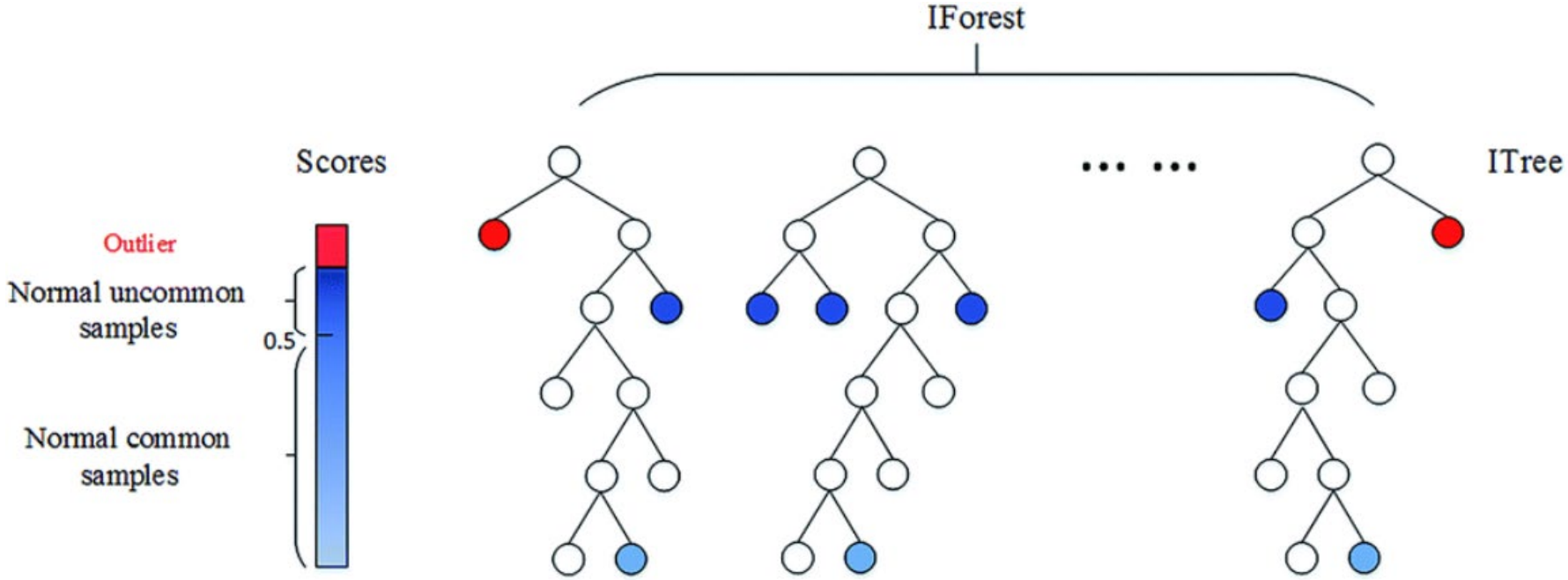
✓위의 예제는 다음과 같이 해석됩니다.

- (1) 정상 데이터를 분리하는 경우, 약 7번의 split 필요
- (2) 이상치를 분리하는 경우, 약 3번의 split 필요

Isolation Forest 학습방법

✓정상 데이터는 tree의 terminal node와 근접하며, 경로길이가 큼

✓이상치는 tree의 root node와 근접하며, 경로길이가 작음



Isolation Forest 학습방법

✓랜덤포레스트가 의사결정나무를 여러번 반복하여 앙상블 하듯이,
Isolation Forest는 iTree를 여러번 반복하여 앙상블

✓iTree

1. **Sub-sampling** : 비복원 추출로 데이터 중 일부를 샘플링
2. **변수 선택** : 데이터 X의 변수 중 q를 랜덤 선택
3. **split point 설정** : 변수 q의 범위(max~min) 중 uniform하게 split point를 선택
4. 1~3번 과정을 모든 관측치가 split 되거나, 임의의 split 횟수까지 반복(=**재귀 나무**)하며, 경로길이를 모두 저장

✓Isolation Forest

5. 1~4번 과정(iTree)을 여러번 반복

Isolation Forest Scoring

✓랜덤포레스트가 의사결정나무를 여러번 반복하여 앙상블 하듯이,
Isolation Forest는 iTree를 여러번 반복하여 앙상블

✓iTree

1. **Sub-sampling** : 비복원 추출로 데이터 중 일부를 샘플링
2. **변수 선택** : 데이터 X의 변수 중 q를 랜덤 선택
3. **split point 설정** : 변수 q의 범위(max~min) 중 uniform하게 split point를 선택
4. 1~3번 과정을 모든 관측치가 split 되거나, 임의의 split 횟수까지 반복(=**재귀 나무**)하며, 경로길이를 모두 저장

✓Isolation Forest

5. 1~4번 과정(iTree)을 여러번 반복

Novelty 이상탐지
https://jayhey.github.io/novelty%20detection/2017/10/18/Novelty_detection_overview/

실습: https://partrita.github.io/posts/isolation-forest/

https://donghwa-kim.github.io/iforest.html

이상 탐지와 정보 보안

이상탐지와 정보 보안

✓정보 보안 분야에서는 보안 위협으로 의심되는 데이터를 찾아내는 하위 분야에 이상탐지 기법을 활용

✓정상과 악성 데이터 모두 충분히 확보 가능한 악성 코드 분야보다는 이상 데이터 확보가 어렵고 정상과 이상에 대한 기준이 모호한 네트워크 침입 탐지, 내부 정보 유출 탐지, 침해 사고 탐지 분야에 활용

✓지금까지 한 번도 보지 못한 새로운 공격 기법 또는 제로 데이 코드를 포함하고 있는 악성코드를 탐지하는 것이 목표라면 이상 탐지 기술을 충분히 활용

이상탐지 실행단계

✓데이터 수집 -〉 데이터 가공 및 축약 -〉 침입 분석 및 탐지 -〉 보고 및 대응

① 데이터 수집 : 탐지 대상(시스템 사용내역 및 패킷)으로부터 생성되는 데이터를 수집하는 감사 데이터 수집

② 데이터 가공 및 축약 : 수집된 감사 데이터를 침입 판정이 가능하도록 의미 있는 정보로 전환시키는 단계

③ 침입 분석 및 탐지 : 데이터를 분석하여 침입여부를 판단하며, 비정상적 행위 탐지 기법(비정상적인 행위) , 오용 탐지 기법(취약점 버그) , 하이브리드 탐지 기법 등등이 있다.

④ 보고 및 대응 : 침입으로 판정된 경우 이에 대한 적절한 대응을 자동으로 취하거나 보안 관리자에게 침입 사실을 보고하여 조치.

설치 위치에 따른 이상 탐지 분류

✓설치 위치가 네트워크냐 호스트냐에 따라 네트워크 기반 침입탐지시스템(NIDS : Network-Based IDS)과 호스트 기반 침입탐지시스템(HIDS : Host-Based IDS)으로 분류

✓설치 위치에 따라서 침입을 판단하기 위해 분석하는 대상에 차이

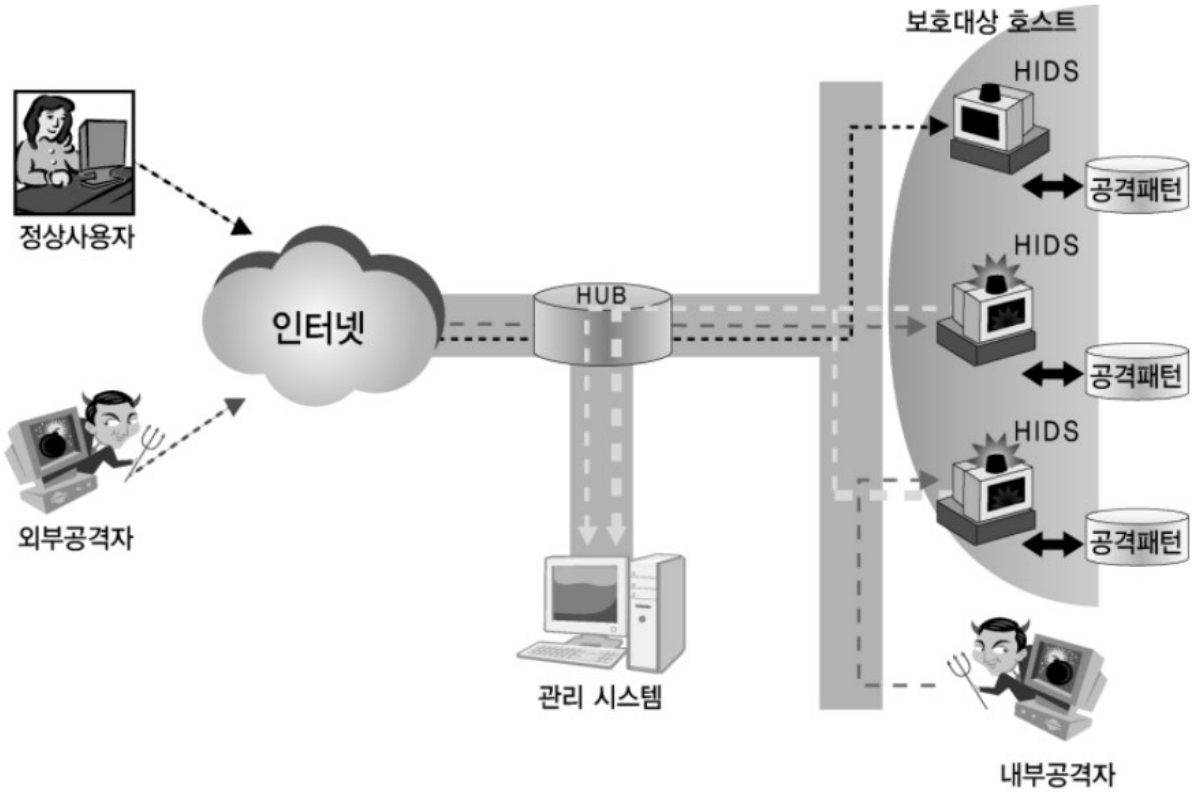
✓침입의 과정 중에서 공격 대상을 찾아내고 취약점을 스캐닝 하는 행위, 그리고 공격 대상 시스템 내부로 들어가기 위한 원격 공격 등의 공격들은 주로 NIDS 에서 탐지

✓시스템 내부로 들어온 이후에 관리자의 권한을 획득하거나 시스템을 변조하는 등의 내부 공격들은 주로 HIDS 에서 탐지

설치 위치에 따른 이상 탐지 분류

✓호스트 기반 IDS(HIDS : Host-Based IDS)

- 각 호스트 내에서의 운영체제 감사 자료와 시스템 로그 분석, 프로세스 모니터링을 통해 침입 탐지를 하는 시스템. 즉, 감시 대상이 되는 서버에 각각 설치



호스트 기반 IDS(HIDS : Host-Based IDS)

서버에 직접 설치하므로 네트워크 환경과 무관

이미지 출처: https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=wnrjsxo&logNo=221068231482

호스트 기반 이상 탐지

✓장점

- 정확한 탐지 가능, 침입 방지 가능 다양한 대응책 수행
- 암호화 및 스위칭 환경에 적합
- 추가적인 하드웨어가 필요하지 않음.
- 트로이 목마, 백도어, 내부 사용자에 의한 공격 탐지 가능

✓단점

- 각각의 시스템마다 설치해야 하므로 다양한 OS를 지원해야 함
- 해커에 의한 로그 자료의 변조 가능성 존재 및 DOS 공격으로 IDS 무력화 가능
- 구현이 용이하지 않음
- 호스트 성능에 의존적이며, 리소스 사용으로 서버 부하 발생

EDR(Endpoint Detection & Response)

✓단말 시스템 수준의 모니터링과 대응 솔루션

✓보안 침해 탐지, 침해 조사, 보안 통제, 치료 등에 EDR 솔루션의 핵심 기능

→ 단말 시스템에서 일어나는 행위들을 빠르게 파악하고 대응할 수 있도록 데이터 가시성 확보하는 것이 핵심 기능

✓EDR 솔루션이 설치된 환경이라면 어렵지 않게 이상 탐지에 활용 가능한 데이터 확보 가능

→ But! 이상 탐지 모델 연구 단계에서 솔루션을 구축하는 것은 현실적으로 어려움

EDR(Endpoint Detection & Response)

✓단말 시스템 수준의 모니터링과 대응 솔루션

1. 단말 행위 모니터링/수집

· File, Module, Process, Network, Registry 정보
· 사용자 및 단말에서 발생하는 이상 행위
· 외부 저장매체의 파일 정보
· 윈도우 이벤트 수집(옵션)
· 다양한 대시보드 제공

2. 위협의 탐지

· 침해사고지표(IOC*) 기반의 알려진 위협 탐지
· 머신러닝(ML)기반의 알려지지 않은 위협 탐지
· 행위 기반의 File-less 위협 탐지
· 야라(YARA)를 이용한 사용자 설정 기반의 심층조사

3. 위협의 대응

· 탐지된 위협 대상의 고지, 종료, 삭제, 고립, 네트워크 격리
· 알려진 위협 사전 대응
· 분석 후 대응(대응 시 동일 이벤트 자동 대응)
· 샌드박스, SIEM 등 기존 보안 솔루션 연동

4. 탐지 위협의 조사/분석

· 탐지된 위협의 상세 정보 제공, 의심 파일 수집
· 통합 검색 및 연관 검색
· 이벤트 타임라인 및 연관 분석(Chain of Event)
· Ecosystem(평판 서비스) 제공

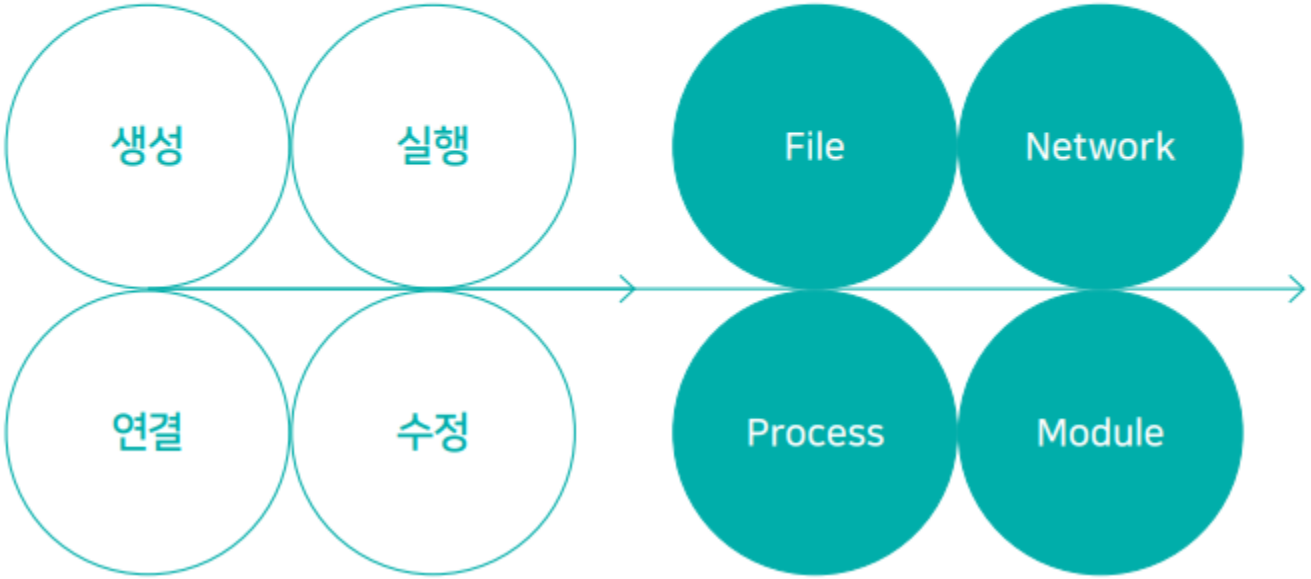
* IOC: Indicators of Compromise, 악성코드 및 접속 C&C 등 침해 사고의 흔적들에 대한 정형화된 데이터

이미지 출처: https://22120960.fs1.hubspotusercontent-na1.net/hubfs/22120960/Genians_July2022/PDF/Genian_EDR_brochure_2022.pdf

EDR(Endpoint Detection & Response)

단말 행위 모니터링

단말에서 발생하는 주요 행위를 모니터링하고 실시간 저장 후 분석합니다. 이를 통해 지능형 위협 등을 사전에 탐지/예방하고, 사후 감사 증적(Audit)이 가능합니다.



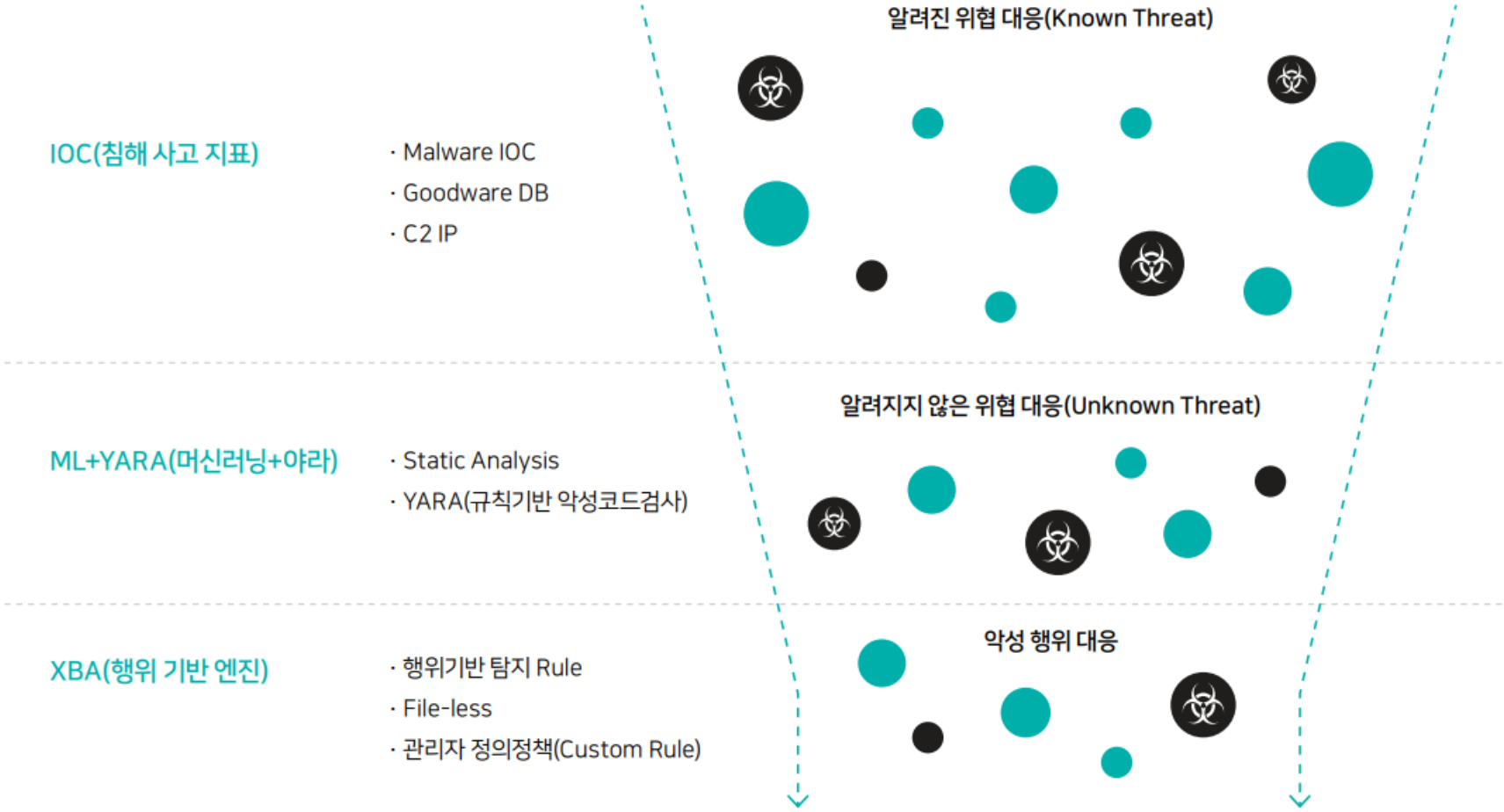
- 사용자, 단말의 행위 분석 및 이상행위(Anomaly) 탐지
- APT, 랜섬웨어 등 지능형 위협의 탐지
- 탐지된 위협 및 이상행위 대응(Response)
- 침해 사고 및 이상행위의 감사 증적(Audit)

이미지 출처: https://22120960.fs1.hubspotusercontent-na1.net/hubfs/22120960/Genians_July2022/PDF/Genian_EDR_brochure_2022.pdf

EDR(Endpoint Detection & Response)

위협(Threat) 탐지

IOC(침해 사고 지표), 머신 러닝, YARA를 이용하여 단계별로 위협을 탐지하며 최고 수준의 정탐률(악성파일+정상파일 탐지)을 제공합니다. XBA(행위 기반 엔진)을 통해 File-less를 포함한 다양한 형태의 악성행위를 탐지합니다.

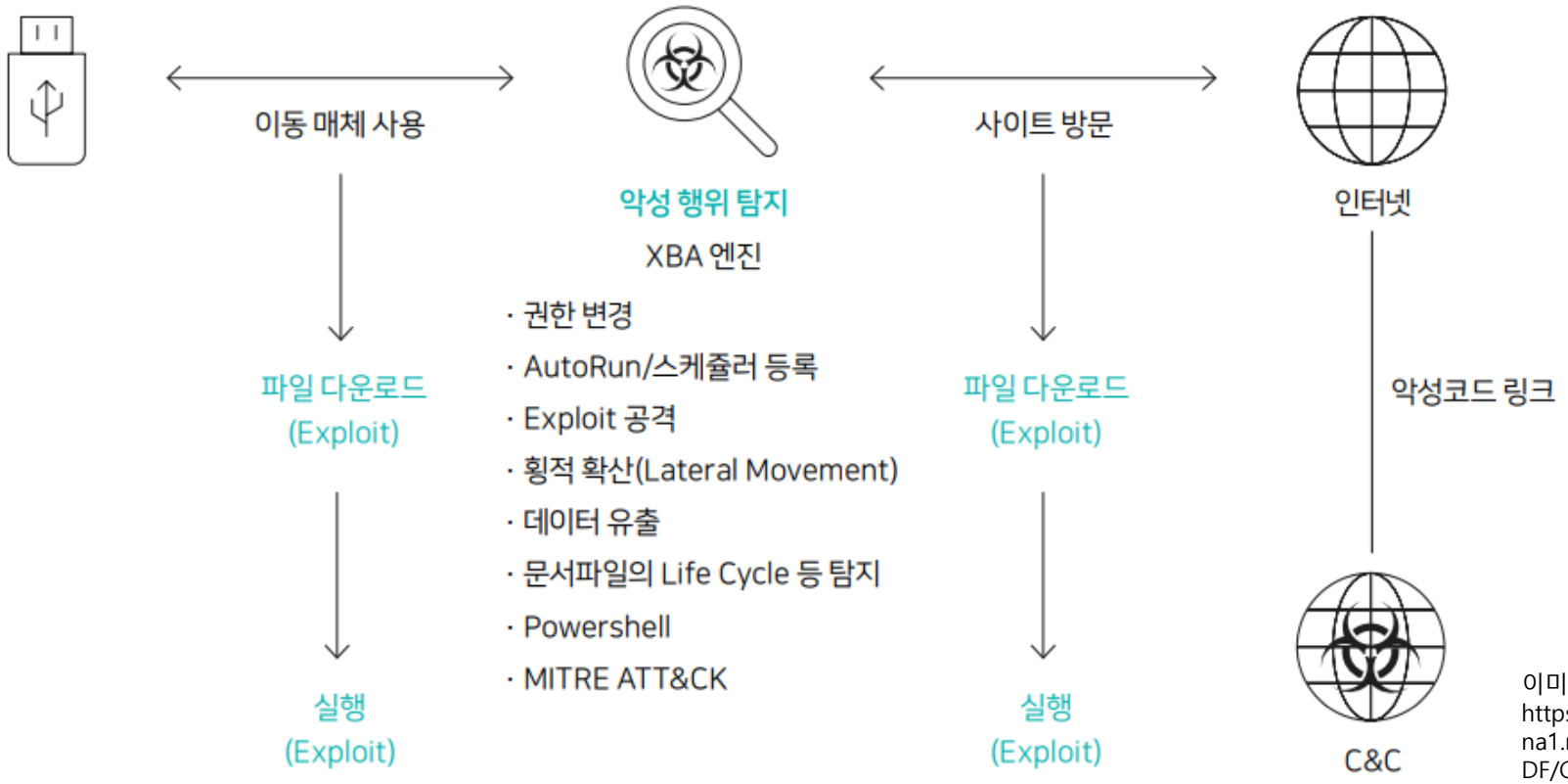


이미지 출처:
https://22120960.fs1.hubspotusercontent-na1.net/hubfs/22120960/Genians_July2022/PDF/Genian_EDR_brochure_2022.pdf

EDR(Endpoint Detection & Response)

이상행위 탐지

사용자 행위 및 단말의 이벤트를 감시하여 이상행위를 탐지합니다. 이상행위 여부 분석 후 위협의 조기 발견 및 IOC 등으로 대응하기 어려운 이상 행위를 탐지할 수 있습니다.



이미지 출처:
https://22120960.fs1.hubspotusercontent-na1.net/hubfs/22120960/Genians_July2022/PDF/Genian_EDR_brochure_2022.pdf

EDR(Endpoint Detection & Response)

Endpoint Discovery

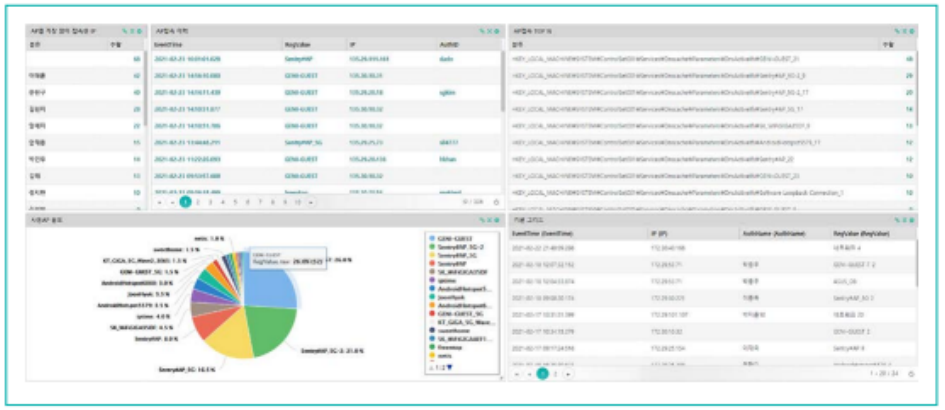
위협 탐지를 위해 실시간으로 수집한 로그는 다시 활용하여 기존에 알 수 없었던 단말에서 행해지는 상황을 파악할 수 있습니다. 문서 유출(업로드), 네트워크 접속 현황, AP접속 현황, 외장저장장치 사용, 메신저, 클라우드 서비스 사용자, DNS 변경 외 다수

문서 유출 모니터링

- 문서 업로드(Web, SNS 등)
- 외장 저장장치(USB, HDD 등) 복사/이동
- 문서 압축
- 확장자 변경 등

네트워크 접속 현황

- 단말의 유/무선 접속 현황
- SoftAP(테더링) 접속
- 외부 AP 접속
- 원격데스크탑/원격터미널 접속
- Putty, Telnet, FTP 등 접속 등

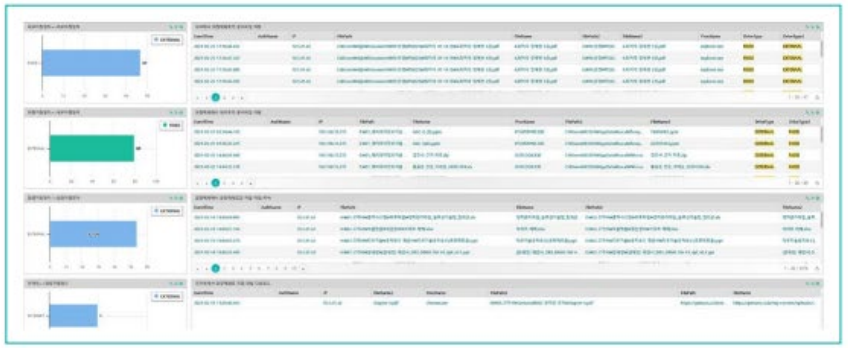


이미지 출처:
https://22120960.fs1.hubspotusercontent-na1.net/hubfs/22120960/Genians_July2022/PDF/Genian_EDR_brochure_2022.pdf

EDR(Endpoint Detection & Response)

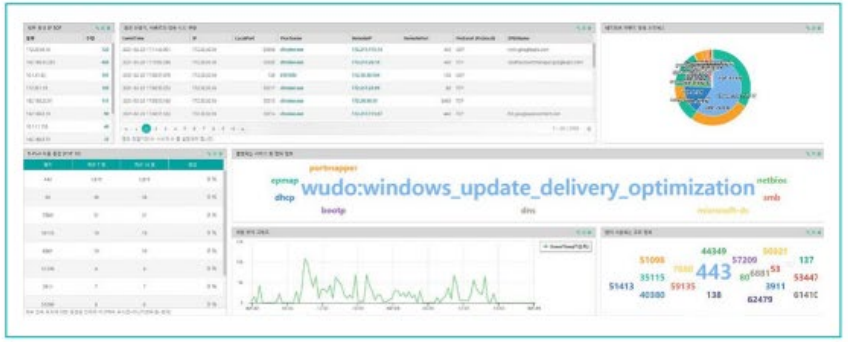
외장 저장장치 세부 사용 현황

- 외장 저장장치로 복사/이동에 대한 세부 내용
- PC → 외장 저장장치로 복사/이동
- 외장 저장장치 → PC로 복사/이동
- 외장 → 외장 저장장치로 복사/이동
- Internet → 외장 저장장치로 다운로드



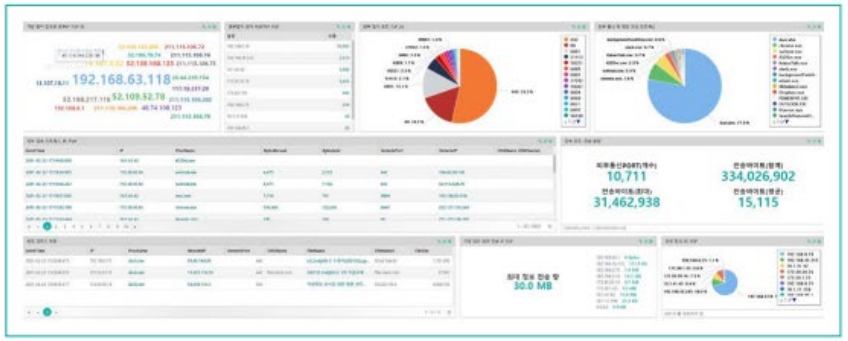
네트워크 이상 분석

- 네트워크 추이 현황(전주/금주 등)
- 오픈 포트 및 서비스 종류
- 네트워크 사용 프로세스 목록
- 특정 단말(관리자 PC 등)로의 접속 현황 등



특정 외부 IP 통신 분석

- 오픈 포트
- 세션을 통해 전송한 Byte 수
- 외부 IP 접속 프로세스 등



이미지 출처:
https://22120960.fs1.hubspotusercontent-na1.net/hubfs/22120960/Genians_July2022/PDF/Genian_EDR_brochure_2022.pdf

Sysmon

✓ 이벤트 로그 강화 도구로 기본 윈도우 로그에서 잡아내지 못하는 프로세스 생성, 네트워크 연결 등 자세한 시스템 행위 정보를 수집해 로그로 기록

✓ Sysmon에서 기록하는 로그 유형과 설명 →

이벤트 ID	설명
1-ProcessCreate	새롭게 생성된 프로세스 정보 기록
2-FileCreateTime	특정 프로세스가 파일 생성 시간을 명시적으로 변경한 내용을 기록
3-NetworkConnect	시스템에서 사용한 TCP/UDP 연결을 기록(연결된 프로세스 정보)
4-N/A	Sysmon 서비스 상태 변화 기록
5-ProcessTerminate	프로세스 종료 정보 기록
6-DriverLoad	시스템에 로드된 드라이브 정보를 기록(해시와 서명 포함)
7-ImageLoad	특정 프로세스 내에 로드되는 새로운 모듈 정보 기록
8-CreateRemoteThread	다른 프로세스 내에 스레드를 생성하는 프로세스 정보 기록
9-RawAccessRead	₩₩.₩ 구문을 사용해 파일 읽기 작업을 수행하는 프로세스 기록
10-ProcessAccess	다른 프로세스에 접근하는 프로세스 정보를 기록
11-FileCreate	파일이 생성 또는 덮어쓰기 되는 것을 기록
12-RegistryEvent	레지스트리 키 값이 새롭게 생성되고 삭제되는 것을 기록
13-RegistryEvent	레지스트리 값 변경을 기록
14-RegistryEvent	레지스트리 키와 값 이름이 변경되는 것을 기록
15-FileCreateStreamHash	명명된 파일 스트림 생성과 스트림에 할당된 파일 내용 변경 기록
17-PipeEvent	명명된 파이프 생성을 기록
18-PipeEvent	명명된 파이프 연결 수립 정보를 기록
19-WmiEvent	새로운 WMI 이벤트 필터 등록을 기록
20-WmiEvent	WMI 소비자 등록 정보를 기록
21-WmiEvent	WMI 소비자가 필터와 연결되는 것을 기록
255-Error	Sysmon에 에러 발생 시 이를 기록

[표 12-1] Sysmon 로그 유형

호스트 기반 이상 탐지 모델

✓단순히 특정 이벤트 로그에 담긴 정보만으로 침해 사고나 정보유출 여부를 판단할 수 없다

✓시스템에서 발생하는 연속적인 사건을 토대로 시나리오 또는 종합적인 문맥을 반영할 수 있는 모델 구축이 필요

1. 악성코드 파일 자체가 가지는 속성을 사용해 모델 구축

예) 최초 실행 시 특정 시스템 내부에 악성코드를 삽입해 공격을 수행하고, 재부팅 시에도 지속적으로 영향력을 행사하기 위해 레지스트리의 시작 프로그램에 자기 자신을 등록하는 악성코드의 경우 →

단순히 하나의 이벤트 발생이 아닌 악성코드 또는 침해사고와 연결된 일련의 이벤트 흐름 기록을 통해 시스템 이상 징후를 찾아낼 수 있다.

호스트 기반 이상 탐지 모델

2. 동일 네트워크에 연결된 모든 단말 시스템의 로그를 종합해서 보는 방법

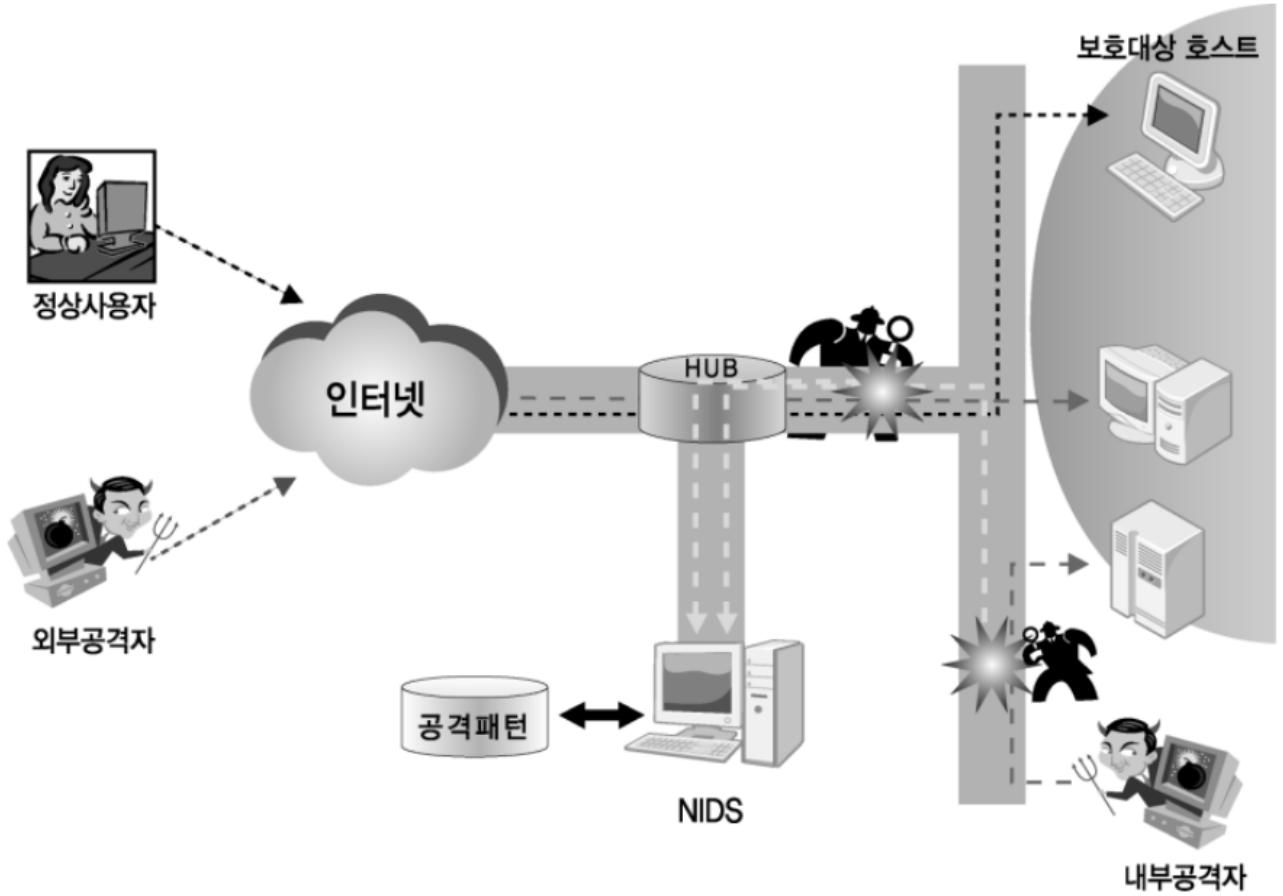
‘정상’에 대한 규정은 단말 시스템 사용자의 업무 영역과 패턴에 따라 조금씩 달라질 수 있다.

‘의심되는’ 이상 행동 패턴이 여러 시스템에 걸쳐 발생할 경우 침해사고를 의심해 볼 수 있다.

예) 특정 악성코드가 WMI(Windows Management Instrumentation)을 이용해 악성 기능을 시스템에 등록하는 경우, 일반적으로 명확한 관리 목적이 아닌 한 일반 사용자가 WMI를 쓰는 경우는 거의 없다. 하지만 내부 네트워크에 연결된 모든 시스템에서 일련의 행동 패턴과 연결된 WMI 이벤트가 기록될 경우, WMI 기반 악성코드 감염 전파를 의심해 볼 수 있다.

네트워크 기반 이상탐지

✓프러미스큐어스(Promiscuous) 모드로 동작하는 NIC 를 통해 네트워크 패킷을 캡처 후 분석을 통해 침입 탐지를 하는 시스템



이미지 출처: https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=wnrjsxo&logNo=221068231482

네트워크 기반 이상탐지

✓웹 어플리케이션 공격의 경우 HTTP 요청과 관련된 페이로드 부분에 빨간색으로 강조된 문자열 패턴이 발견될 경우 공격이 의심된다는 의미를 가진다. 이러한 공격의 경우 서버 프로그램 또는 데몬 코드에 존재하는 취약점과 관련된 것으로 명확한 패턴을 정의할 수 있다.

✓But! 웹 어플리케이션 공격에는 매우 다양한 유형이 있으며 고정되어 있지 않다. 단순히 하나의 유형이 아닌 '의심되는 웹 요청'을 잡아내려면 → HTTP 통신 방식과 탐지하려는 공격 유형에 따라 서로 다른 특징 추출 방식을 필요로 함

네트워크 기반 이상탐지

✓예) 현재 운영 중인 웹 서버로 들어오는 요청 중 공격으로 의심되는 요청을 찾아내려면 HTTP request 데이터를 관심 있게 살펴봐야 한다. 하지만 내부망에 연결된 사용자들의 웹 통신 내용을 토대로 웹 브라우저 기반 취약점 공격을 당하고 있는지를 찾아내려면 HTTP response 데이터에서 실제 html, javascript 코드를 추출한 후 코드 분석 관점의 특징을 추출해야 한다.

네트워크 기반 이상탐지

✓장점

- 트래픽을 몇몇 위치에만 설치하므로 초기 구축 비용이 저렴
- 운영체제에 독립적이므로 구현 및 관리 쉬움
- 캡처된 트래픽에 대해 침입자가 제거하기가 어려움.
- 네트워크에서 발생하는 여러 유형의 침입을 탐지
- 네트워크에서 개별 실행되어 개별 서버의 성능 저하가 없음.

✓단점

- 암호화된 패킷을 분석할 수 없음
- 고속 네트워크 환경에서는 패킷 손실률이 많아 탐지율이 떨어짐
- 호스트 상에서 수행되는 세부 행위에 대해 탐지할 수 없음
- 오탐률이 높음(False Positive)

〈참고〉침입탐지시스템(Intrusion Detection System; IDS)의 위치

1. 패킷이 라우터로 들어오기 전 :
모든 공격을 탐지할 수 있지만 네트워크에
치명적인 공격에는 대처가 어렵다는 단점이 존재

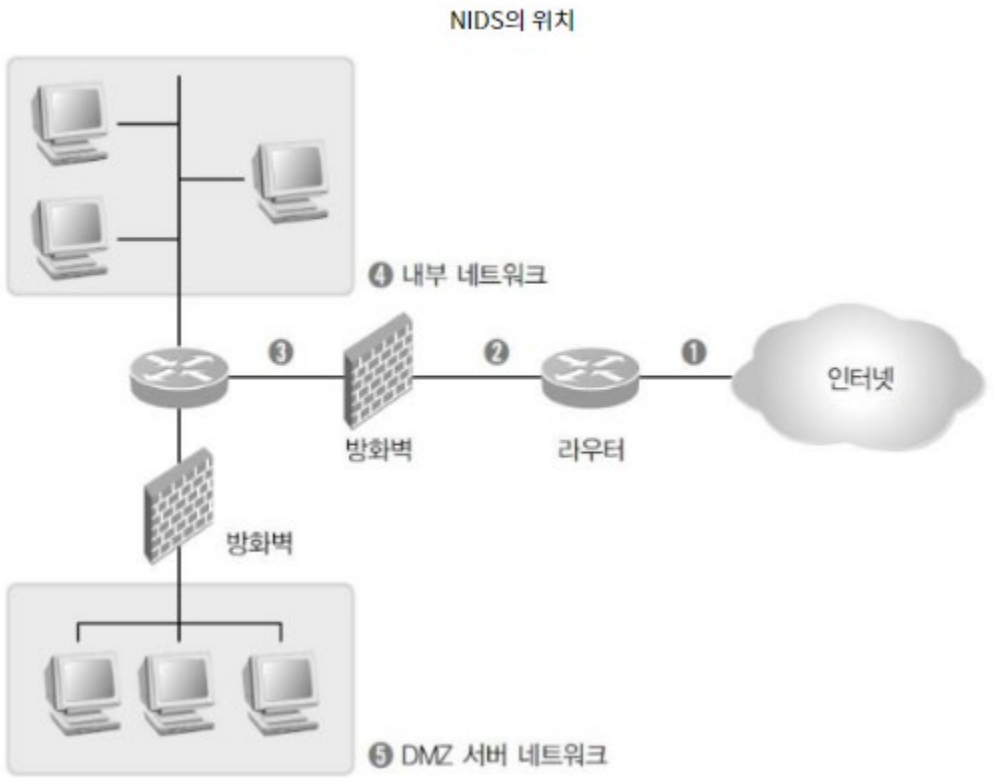
2. 라우터 뒤 : 패킷 필터링을 거친 후의 패킷 탐지,
좀 더 강력한 의지가 있는 공격자 탐지 가능

3. 방화벽 뒤 : 만약 침입 탐지 시스템을 설치한다면 or
만약 한대만 설치할 수 있다면 이곳에 설치해야 함.

4. 내부 네트워크 : 방화벽은 침입을 일차적으로 차단하지만
내부에 대해서는 무방비 상태

5. DMZ : 아주 능력이 뛰어난 외부 공격자와 내부 공격자에
의해 중요 데이터 손실이나 서비스 중단을 막기 위함

✓설치 우선순위로는 3 -〉 5 -〉 4 -〉 2 -〉 1



이미지 출처: https://didimdol20.tistory.com/61

Q&A