**제목: 위협정보 수집, 분석 및 기대 효과**

**㈜ 넥스투비 연구소**

**㈜ 넥스투비 정윤수 차장**

**요약**

기존의 침해지표 분석 활동은 침해를 확인하는 용도 또는 사고 대응 성격이 강하였기 때문에,

침해지표(IOC)를 분석해 공격 행위의 사전 탐지 및 예측에 한계가 있었습니다.

본 실험에서는 Threat hunting 을 통해 수집한 데이터로 학습 데이터를 생성하고 전통적 방식의

분석 한계를 극복하기 위해 기계학습을 사용한 결과로 새로운 통찰을 기대합니다.

**1. 소개**

기존의 침해지표 분석 활동은 침해를 확인하는 용도 또는 사고 대응 성격이 강하였기 때문에,

침해지표(IOC)를 분석해 공격 행위의 사전 탐지 및 예측에 한계가 있었습니다.

본 실험에서는 Threat hunting 을 통해 수집한 데이터로 학습 데이터를 생성하고 전통적 방식의

분석 한계를 극복하기 위해 기계학습을 사용하며 분석합니다. 또한 결과와 향후 과제에 대해

토의 합니다.

**2. 관련 연구**

- 침해지표(IOC)

정보보안에서 침해지표(IOC)란 디지털 침해사고를 분석하는데 사용되는 지표를 뜻합니다.

사이버 범죄자가 남긴 디지털 증거를 추출해 내는 ‘디지털 포렌식’ 과정에서, 그러한 증거의

유형들을 지표화한 것입니다.

이벤트 로그, 비정상적인 파일 시스템 기록 등 침해지표는 침해사고를 파악 및 분석하고 또

대응하는 기반으로 사용됩니다.

- Threat hunting

위협 사냥(Threat hunting)은 주로 보안기업 내에서 사이버 공격을 당하기 전 선제 방어

목적으로 자신들의 네트워크에 존재하는 사이버 위협 요소를 능동적으로 탐지하고 제거하는

방식을 묘사할 때 쓰이던 명칭 입니다.

모든 사이버 침입을 방어할 수 없다는 현실도 전제로 하고 있어 방어를 뚫고 들어온 공격을

완화시키는 활동을 통칭하며, 해당 데이터의 질과 분석 능력이 중요하고, 해당 툴을 다룰

줄 아는 사람이 필요합니다.

즉 하나의 대세 기술이 있는 것 보다는 각자가 여러 가지 기술과 경험을 필요로 합니다.

- 머신러닝(Text mining)

- Apriori algorithm

빈발 패턴 마이닝은 주어진 데이터 집합에서 재귀 관계(recurring relationships)를

찾습니다.

빈발 항목집합 마이닝의 전형적인 예는 장바구니분석(market basket analysis)입니다.

예를 들어, "고객이 우유를 샀다고 한다면, 그 쇼핑에서 빵을 함께 살 가능성은 얼마나

되겠는가?" 이 질문에 답하기 위해 상점의 고객 트랜잭션에 관한 데이터에 대해 장바구니

분석을 수행해서 상점에서 제공되는 상품의 집합으로 이루어진 세계가 있다 가정하고,

각 상품들은 상품의 존재 유무를 표현하는 이진(boolean/binary)변수를 갖는다 가정

합니다. 이 변수들에 할당된 값으로 구성된 이진벡터(boolean/binary vector)로 표현할

수 있습니다.

지지도(support)와 신뢰도(confidence)는 규칙의 흥미도를 측정하는 두가지 기준입니다.

지지도가 낮은 연관규칙은 여러 트랜잭션에 적용되지 않는 규칙이어서 우연히 발생할 수

있는 규칙 또는 흥미가 없는 규칙일 가능성이 많습니다.

따라서 최소 지지도 임계값(minimum support threshold)을 정하여 그 이하의 규칙은 버립

니다. 즉 지지도는 좋은 규칙을 찾거나 흥미 없는 규칙을 버릴 때의 기준으로 사용됩니다.

반면에 신뢰도는 연관규칙의 신뢰성에 대한 측도 입니다.

주어진 연관규칙 X⇒Y가 있을 때, 신뢰도가 높을수록 항목 X를 포함하는 트랜잭션은 항목

Y도 포함할 가능성이 많게 됩니다.

최소지지도를 min\_sup, 최소신뢰도를 min\_conf라 할때, 연관규칙을 탐색하는 일반적인

원칙은 최소지지도보다 크면서 동시에 최소신뢰도보다 큰 연관규칙, 즉 강한(strong)규칙

을 찾는 것입니다.

- TF-IDF

TF-IDF(Term Frequency - Inverse Document Frequency)는 정보 검색과 텍스트 마이닝에

서 이용 하는 가중치로, 여러 문서로 이루어진 문서 군이 있을 때 어떤 단어가 특정 문서

내에서 얼마나 중요한 것인지를 나타내는 통계적 수치입니다. 문서의 핵심어를 추출

하거나, 검색 엔진에서 검색 결과의 순위를 결정하거나, 문서들 사이의 비슷한 정도를

구하는 등의 용도로 사용할 수 있습니다.

TF(단어 빈도, term frequency)는 특정한 단어가 문서 내에 얼마나 자주 등장하는지를 나

타내는 값으로, 이 값이 높을수록 문서에서 중요하다고 생각할 수 있습니다. 하지만 단어

자체가 문서군 내에서 자주 사용되는 경우, 이것은 그 단어가 흔하게 등장한다는 것을

의미 합니다. 이것을 DF(문서 빈도, document frequency)라고 하며, 이 값의 역수를

IDF(역문서 빈도, inverse document frequency)라고 합니다. TF-IDF는 TF와 IDF를 곱한

값 입니다. (TF-IDF = TF \* IDF)

IDF 값은 문서군의 성격에 따라 결정 됩니다. 예를 들어 '원자'라는 낱말은 일반적인

문서들 사이에서는 잘 나오지 않기 때문에 IDF 값이 높아지고 문서의 핵심어가 될 수

있지만, 원자에 대한 문서를 모아놓은 문서군의 경우 이 낱말은 상투어가 되어 각

문서들을 세분화하여 구분 할 수 있는 다른 낱말들이 높은 가중치를 얻게 됩니다.

**3. 데이터**

- 수집 및 전처리

수집된 데이터는 최초에 각 알고리즘에 입력되는 변수들을 추출하는 과정이 실제로 전체 과

정 (수집, 데이터준비, 알고리즘적용, 결과추출, 결과분석) 에서 70% 이상의 시간과 노력이

소모됩니다.

본 실험에서는 내부 데이터의 워너크라이 관련 IOC 값을 기반으로, 외부 테이터 상에서

내부 데이터 IOC 값 검색 결과를 기반으로 데이터를 수집하여, 기계 학습에 맞는 형태로

가공 한 후 실험 데이터를 생성 하였습니다.

- 내부 + 외부 데이터

C-TAS(내부 데이터) 에서 워너크라이 관련 IOC 를 csv 형태로 생성 하였습니다.

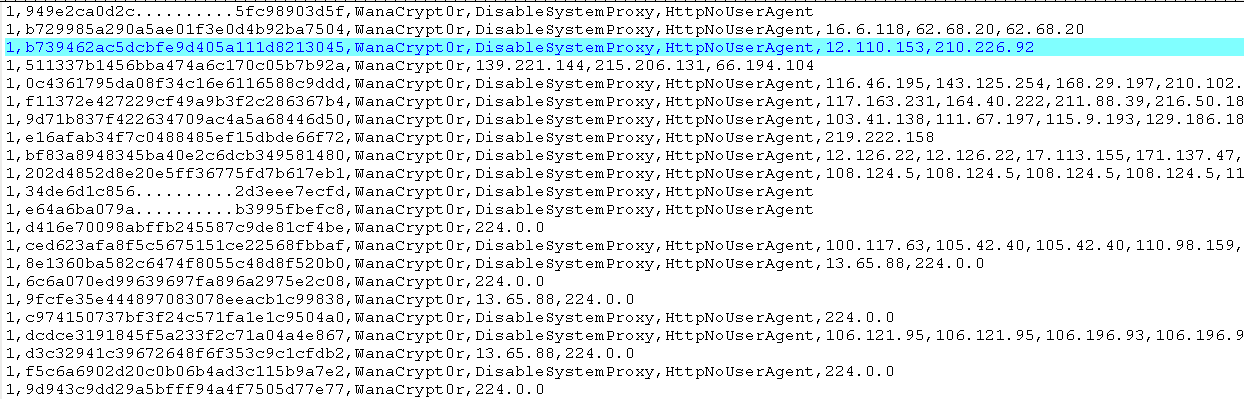
또한 상용 인텔리전스인 Aloalto Autofocuss(외부 데이터)통해 워너크라이 관련 IOC 를 수집

하고 그와 관련된 나머지 데이터를 export 하였으며, 로컬로 내려진 두 파일을 merge

합니다.

본 실험에서는 수동으로 200건의 데이터를 생성하여 실험을 진행 하였고, 향후 이 활동은

API 연동 등 자동화가 이루어져야 합니다.

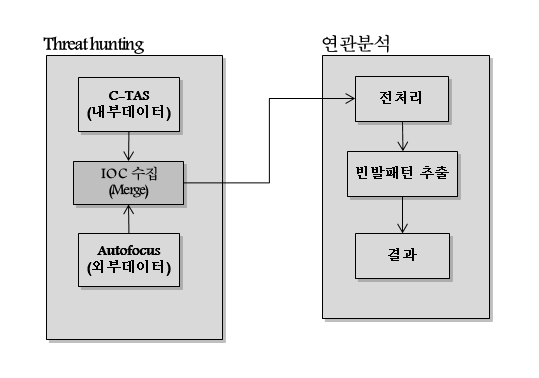


**4. 실험**

실험 데이터에서 해당 악성코드와 관련 있는 Autofoucs 데이터 중 Tag 값과 IP값을 대상으로

IP를 C class 대역으로 전처리 하고 IP들간 빈발 패턴을 찾기 위해 연관분석을 하였고, Apriori

알고리즘을 적용한 코드를 파이썬으로 구현하여 실행 했습니다.



예를 들어 특정 코드의 해시 값이 192.168.100.10, 192.168.100.11 과 같은 IP에 어떤 이벤트가

있었다 가정하면, 해당 해시와 IP가 빈발 패턴으로 잡힐 것입니다. 즉 192.168.100.xxx 대의

나머지 IP들도 의심해 볼 수 있으며, 이를 새로운 비즈니스로 활용 가능할 것입니다.

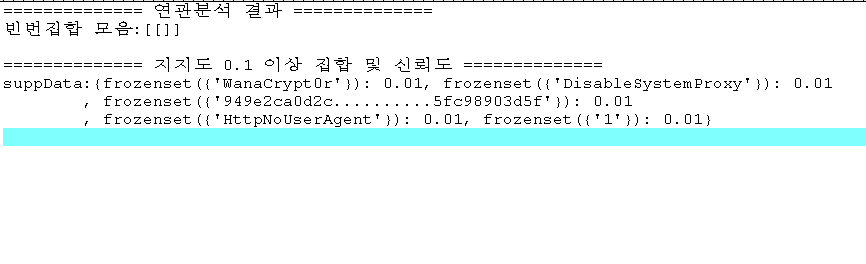
또한 이런 정보는 Autofoucs 상에서는 알려주지 않는 내용입니다

- 실험 결과

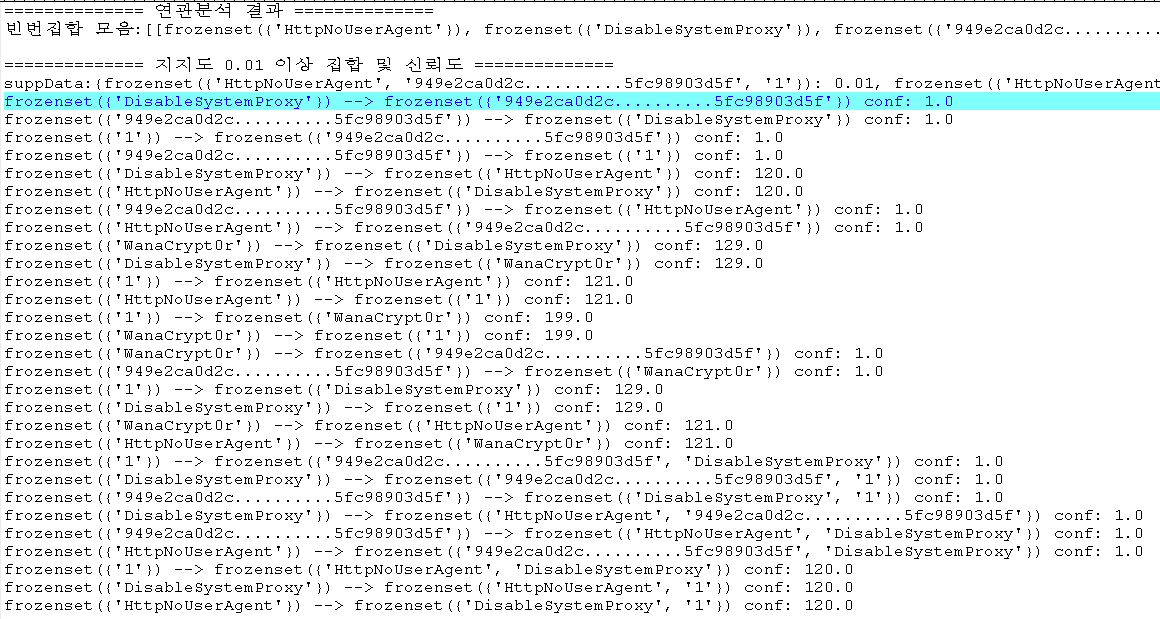
실험 데이터 200건의 해시, Tags, IPs 을 대상으로 한 연관분석에서 악성코드 해시 값과 IP들

간 빈발 패턴을 추출 하였습니다.

지지도 0.1 이상 신뢰도 0.1 이상인 값으로 설정하고 실행한 결과는 다음과 같습니다.



지지도 0.01 이상 신뢰도 0.1 이상인 값으로 설정하고 실행한 결과는 다음과 같습니다.



지지도 0.1 이상일 때의 빈발 집합은 하나도 잡히지 않습니다(공집합). 반면 지지도 0.01

이상으로 실행 시는 빈발 집합이 잡히는 결과가 나왔습니다. 추출되어진 빈발 항목간

신뢰도는 대체적으로 정확한 수준임을 확인 할 수 있습니다. 그러나 지지도 0.1 인 항목이

잡히지 않는 부분은 논리적으로 이해 할 수 없는 부분이며 그 원인은 크게 다음과 같이

예상해 볼 수 있습니다.

1. 데이터 전처리

데이터 취합 및 생성을 수작업으로 하였기 때문에 그 과정에서 결과물이 잘못 생성될

확률이 존재합니다. 또한 최초 데이터 분석 시 변수들간 관계를 아예 잘못 설정 했을

수 있습니다. 데이터 분석가가 업무 전문성이 떨어졌을 때 나타날 수 있는 현상 입니다.

2. 코드 오류

연관분석 코드는 비교적 짧은 시간에 작성되어진 것으로, 체계적인 알고리즘 분석,

설계 및 검증(Inspection) 이 이루어지지 않았습니다.

3. 데이터

실행 코드가 완전 무결하다는 전재하에 실제 빈발 집합이 없어서 그런 결과가 나올 수

있습니다. 그러나 실제 검증(Eye check) 결과 그럴 가능성은 희박합니다.

그리고 Input 데이터 건수가 2백 건 뿐 이었다는 것도 결과에 영향을 미칠 수 있는

요소라 할 수 있습니다.

**5. 향후 연구**

텍스트 마이닝 기법 중 하나인 연관 분석을 사용하여 IOC 간 빈발 항목 추출 및 빈발 항목

간 어떤 연관도가 있는지 알아보았습니다. 이 결과를 이용하여 변수들 간 기존에 알려지지 않은

패턴을 발견하여 업무에 적용할 수 있습니다.

향후 내부 데이터와 외부 데이터 지표들 간 상관관계 분석 고도화가 이루어지고 연관 분석 외

다른 알고리즘 적용으로 변수들간 새로운 패턴 발견을 시도해 볼 수 있으며, 그 결과를 새로운

기계학습(Supervised learning or Unsupervised learning)의 Input으로 사용할 수 있습니다.

**6. 결론**

본 실험에서 Threat hunting 을 통해 내부 데이터 IOC와 및 외부 데이터 IOC를 수집하고,

기계학습을 통한 연관분석 또는 TF-IDF 수행으로 기존의 통계기반 분석 패턴과 다른 의미

있는 결과 도출 가능성을 확인하는 실험을 했습니다.

향후 새로운 ICO 연구(변수간 상관관계), 상용 인텔리전스 활용 범위 확대로 다른 변수들 간

빈발 패턴을 탐색한다면 좀 더 의미 있는 결과를 도출 할 수 있습니다.

또한 이 결과를 기반으로 기계학습(Supervised learning or Unsupervised learning)을 시도

하여 새로운 통찰을 기대 해 볼 수 있을 것입니다.