정 보 보 호 학 회 지
제26권 제3호, 2016. 6

온라인 게임 보안을 위한 머신러닝 시스템 연동 방안

이 은 조*

요 약

온라인 게임에는 해킹이나 악성 코드와 같은 다른 분야에서도 널리 알려진 위협뿐만 아니라 계정 도용이나 자동 사냥 프로그램 사용과 같은 온라인 게임에서만 볼 수 있는 위협들이 존재한다. 온라인 게임은 게임 유저의 다양한 활동을 데이터로 기록하기 때문에 이런 풍부한 데이터를 활용한 머신 러닝 기반의 탐지 기법을 적용하기 적합한 분야이다. 그럼에도 불구하고 다른 보안 분야에 비해 상대적으로 연구가 많이 되지 않고 있으며 대부분의 연구가 탐지 모델링 단계에 집중되어 있다. 본 논문에서는 머신 러닝에 기반한 온라인 게임 보안 시스템을 효과적으로 구축하기 위한 연동 구조와 실전 적용 시 고려해야 할 점에 대해 소개한다.

Ⅰ. 서 론

온라인 게임에는 해킹이나 악성코드, 랜섬웨어와 같은 일반적으로 알려진 위협뿐만 아니라 게임 내 가상 세계를 통해 벌어지는 다양한 불법 행위 등이 존재한다. 여기에는 다른 게임 유저의 계정을 탈취하거나 신용 카드 도용을 통해 게임 아이템을 구매하는 등의 기존 온라인 서비스에서도 볼 수 있는 범죄 행위도 있지만 게임 약관 상에서 불법으로 규정하고 있는 자동 사냥 프로그램 이용이나 가상 아이템 거래를 이용한 돈 세탁과 사기 같은 온라인 게임 상에서만 볼 수 있는 유형도 있다. 이런 종류의 위협이나 범죄 행위는 기존에 연구되던 보안 대책과는 접근 방법이나 적용 방안에 있어서 다소 차이가 있으며 게임 생태계 및 사업 특성에 대한 충분한 이해가 기반 되지 않고서는 적절하게 실전에 적용할 수 없다.

과거 온라인 게임 보안은 주로 게임 클라이언트나 네트워크에서의 탐지 및 대응에 초점이 맞춰져 있었다. 그런데 최근 몇 년 사이에 빅데이터 인프라 및 머신 러닝에 대한 관심이 커지면서 서버 단에서 수집되는 각종 데이터를 이용한 탐지 기법이 많이 연구 되고 있다. 온라인 게임은 그 특성 상 게임 유저의 세세한 활동이 모두 기록되기 때문에 다른 온라인 서비스에 비해 다양한 데이터를 수집할 수 있다. 따라서 이런 대량의 데이터를 머신 러닝에 적용하여 위협을 탐지하고 대응하기에 적합한 분야이다. 그러나 데이터 기반의 보안과 머신 러닝 두 분야 모두 실전에 적용된 사례가 많지 않고 경험이 풍부한 전문가가 부족하기 때문에 어떻게 구축하고 적용해야 할지 어려워하는 것이 현재 업계의 상황이다.

한편 대부분의 머신 러닝 관련 연구들은 탐지 성능을 높이기 위한 학습 모델링을 어떻게 할 것인지에 대한 방법론에 집중하고 있다. 반면 실전에 탐지 모델을 적용할 때 학습 모델링을 통해 만든 탐지 모델을 서비스에 어떻게 연동할지에 대한 방법론이나 인프라 구조에 대해서는 공유되는 정보가 많이 부족하다.

본 논문에서는 온라인 게임에서 주로 발생하는 보안 문제가 어떤 것이 있는지 먼저 살펴본 후 이를 해결하기 위해 서버 단에서 수집하는 데이터를 머신 러닝에 적용하여 위협을 탐지하는 방법 및 시스템 구조를 소개하겠다. 아울러 머신 러닝 기반의 온라인 게임 보안 시스템을 실전에 적용할 때 고려해야 할 사항들에 대해서도 간략히 살펴보도록 하겠다.

Ⅱ. 온라인 게임 위협의 종류와 특징

온라인 게임에서 다루는 주요 보안 이슈는 계정 도용 및 결제 사기와 자동 사냥 프로그램 탐지가 있다. 최근에 문제가 되고 있는 대부분의 보안 위협들이 그러하듯

* 엔씨소프트 (gimmesilver@ncsoft.com)

온라인 게임 분야의 위협 행위 역시 금전적인 이득을 취하려는 목적 때문에 발생한다. 이런 금전적인 이득은 대부분 '아이템 현물 거래(Real Money Trading, 이하 RMT)'를 통해 이뤄진다. RMT는 게임 세계의 가상 재화나 물건, 게임 캐릭터를 현실 세계의 실물 화폐로 교환하는 행위를 말한다. 최초의 RMT는 1999년 '울티마 온라인'이라는 게임의 한 유저가 자신의 계정을 이베이에 경매로 올린 것으로 알려져 있다[1]. 이후 이런 RMT 수요가 점점 늘면서 2001년에 '아이템 베이'라는 게임 전문 거래 중개 사이트가 세계에서 최초로 우리나라에 생겼으며 이후 유사한 사이트들이 전 세계에 널리 퍼져 있다.

온라인 게임이 대중화되고 게임 세계에서의 높은 지위와 능력을 갖는 것을 갈망하는 사람들이 많아지면서 이 RMT 규모 역시 크게 성장하였다. 상당수의 RMT가 음성적으로 진행되기 때문에 정확한 규모를 추정하기는 어렵지만 우리나라의 경우 연간 최소 약 1조 5천억 원의 거래가 발생하는 것으로 추정하고 있다[2].

RMT는 온라인 게임이 인기를 끌기 시작한 2000년대 초반부터 지속적으로 규모가 성장했다. 특히, 2010년 1월에는 RMT를 합법적인 행위라고 인정한 대법원 판례가 나오면서 이에 대한 규제가 어려워 졌으며 이로 인해 앞서 언급한 계정 도용, 사기 결제, 자동 사냥 프로그램 사용과 같은 온라인 게임에서의 범죄 행위가 크게 증가하였다[3]. 그러나 2012년에 게임산업진흥에 관한 법률 시행령이 개정되면서 RMT를 업으로 하는 행위를 처벌 대상으로 규정하였으며 이에 따라 최근 대규모 RMT 전문 조직이 단속되기도 했다[4].

이제 앞서 언급한 온라인 게임 상의 범죄 유형들이 어떻게 RMT와 연결되고 이로 인해 어떤 피해가 발생할 수 있는지 좀 더 자세히 언급하면 다음과 같다.

2.1. 계정 도용

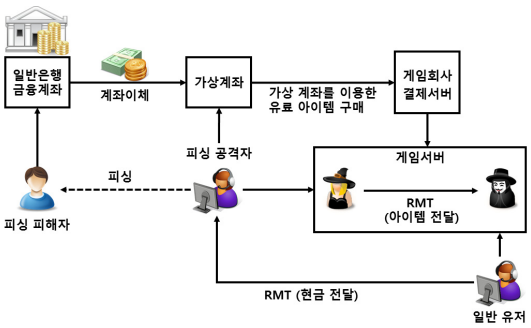
계정 도용이란 다른 사람의 계정 정보를 다양한 해킹 기법을 이용해 탈취한 후 허락을 받지 않은 상태에서 게임에 대신 접속하는 행위를 말한다. 대개 다른 온라인 서비스의 경우 이렇게 도용한 계정을 이용해 개인 정보를 빼돌리거나 해당 계정의 지인에게 접근하여 돈을 가로채는 피싱 등의 행위를 많이 하는데 온라인 게임에서는 해당 계정의 게임 캐릭터가 보유한 가상 재화를 RMT를 통해 제 3자에게 처분하는 경우가 가장 많다.

이런 경우 피해 계정 당사자가 차후 도용 사실을 안 후 신고를 하게 되면 조사 과정을 거쳐 피해 복구가 가능하다. 그러나 RMT를 통해 해당 아이템을 구매한 제 3자는 자신이 구매한 아이템이 도용 아이템이라는 사실을 모르고 구매하는 경우가 대부분이기 때문에 의도치 않은 손해가 발생하게 된다. 도용 피해자와 구매 유저 모두 손해를 보지 않도록 피해 아이템을 새로 게임 내에 생성하는 방법도 가능하지만 이런 식의 대응이 빈번해지면 게임 세계 내의 재화가 비정상적인 방법으로 증가하게 되기 때문에 게임 경제에 인플레이션을 유발할 수 있다. 따라서 계정 도용은 피해를 복구하는데 집중하기 보다는 사전에 도용 행위를 탐지하여 방지하거나 혹은 도용 이후 RMT를 통해 게임 재화가 처분되기 전에 빠르게 탐지하여 조치하는 것이 중요하다.

2.2. 결제 사기

결제 사기는 다른 사람의 카드나 계정을 도용하여 게임 회사에서 판매하는 유료 아이템을 구매한 후 이 아이템을 RMT 시장을 통해 처분하는 행위이다. 이 경우 역시 피해 당사자가 사후 신고를 할 경우 해당 비용은 환불 처리가 될 수 있지만 이로 인한 손실은 게임 회사가 지게 된다. 심지어 최근에는 은행 계좌 해킹이나 피싱을 통해 다른 사람의 계좌에 있는 현금을 가상 계좌로 이체한 후 이 가상 계좌를 통해 게임 아이템을 구매하고 다시 이 아이템을 RMT 를 통해 현금화함으로써 돈 세탁을 하는 신종 사기 범죄도 등장하였다(그림 1 참조).

RMT는 그 특성 상 게임 세계의 가상 재화와 현실 세계의 재화 간의 접점 역할을 할 수 있는 반면 기존 금



〔그림 1〕 금융 계좌 피싱과 RMT를 연계한 금융 사기

융 거래와 정보를 공유하는 채널이 아직 없기 때문에 수사 기관에서 자금 추적이 쉽지 않아 효과적인 돈 세탁 수단이 될 수 있다. 따라서 이와 같은 금융권의 해킹이나 피싱과 연계한 온라인 게임 금융 범죄는 앞으로 계속 증가할 것이라 예상되고 있다.

2.3. 자동 사냥 프로그램 사용

자동 사냥 프로그램은 사람을 대신해서 게임을 플레이하는 프로그램을 말한다. 보통 유저의 번거로운 조작을 줄이기 위한 보조 도구로 많이 사용한다. 그런데 RMT 시장의 규모가 커지면서, 자동 사냥 프로그램을 대규모로 운영하여 이를 통해 모은 가상 재화를 현금화하는 전문적인 사업체가 등장하게 되었다. 이를 '작업장(Gold Farming Group, GFG)'이라고 부른다. 일반 사용자가 편의성을 얻기 위해 자동 사냥 프로그램을 사용하는 것과 달리 이런 작업장은 수백 개 이상의 게임 캐릭터를 이용하여 대규모로 활동하기 때문에 게임 내 리소스를 독점함으로써 일반 사용자에게 심리적인 박탈감과 게임에 대한 흥미를 떨어뜨리게 만듦으로써 고객 이탈을 야기한다. 또한 게임 내 경제 시스템이나 컨텐츠 밸런스를 무너뜨릴 수 있기 때문에 대부분의 게임 회사에서는 약관상으로 자동 사냥 프로그램 사용을 불법으로 규정하여 제재하고 있다.

Ⅲ. 온라인 게임 보안 관련 기존 연구 사례

비록 다른 보안 분야에 비해 그 규모는 작지만 앞 장에서 소개한 온라인 게임 관련 보안 위협을 탐지하기 위한 방안을 제안하는 연구들은 2000년대부터 꾸준히 진행되었다. 먼저 Yan과 Randell[5]은 온라인 게임 분야에서 발생하는 다양한 범죄 행위에 대해 체계적인 분류 작업을 진행했다. 이 중 가장 많은 연구가 이뤄지는 분야는 자동 사냥 프로그램 탐지 분야이다. Ahmad 등[6]은 '에버퀘스트2'라는 게임을 대상으로 다양한 분류 알고리즘과 특질 데이터를 이용하여 자동 사냥 프로그램을 탐지하는 방법을 제안하였다. 이후 좀 더 전문화된 탐지 방법들이 연구되었는데 유저들의 다양한 활동량에 대한 통계를 이용한 연구로는 Thawonmas 등[7]이 있다. 또한 동일한 행동이나 패턴을 반복하는 특성을 이용한 연구도 있었는데 Kesteren 등[8]은 이동 경로 상의 반복 패턴을 이용하는 기법을 연구했으며 Lee 등[9]은 유저의 전반적인 행동을 자기 유사도라는 정량화 기법을 이용하여 탐지하는 기법을 제안하였다. 한편 온라인 게임은 여러 유저 간의 상호 작용 및 협업 활동이 중요한 역할을 하기 때문에 이런 협업 활동에서 나타나는 특성에 초점을 맞춘 연구도 있었다. Kang 등[10]이 제안한 파티 활동에 기반한 탐지 기법과 김하랑 등[11]이 연구한 길드 활동을 이용한 탐지 기법이 이에 해당한다. 마지막으로 작업장의 RMT 거래 네트워크를 분석한 연구도 있는데 Woo 등[12]과 Kwon 등[13]은 캐릭터들간의 거래 행위에 대해 네트워크 분석을 수행했으며 이를 통해 작업장 커뮤니티를 탐지하고 이들의 특징을 분석하였다. 한편 강성욱 등[14]은 네트워크 분석을 통해 RMT 구매자를 탐지하는 접근 방식을 제안하기도 했다. 마지막으로 Keegan 등[15]은 작업장의 거래 네트워크를 현실 세계의 마약 거래 네트워크와 비교하여 유사점을 분석한 흥미로운 연구를 진행하기로 했다.

반면 계정 도용이나 결제 사기에 대한 연구는 상대적으로 많이 진행되지 못했다. Chen 등[16]은 유저의 활동 정보 중 활동 시간 및 휴식 시간의 분포에 기반한 탐지 기법을 제안하였으며, Woo 등[17]은 유저의 활동 정보 뿐만 아니라 접속 IP와 같은 접속 관련 정보까지 포함한 다양한 특질을 이용한 머신 러닝 기법을 제안하였다. 김하나 등[18]은 게임 활동에 대한 시퀀스 분석을 통해 계정 도용을 탐지하는 기법을 제안하였는데 이렇게 탐지한 정보를 토대로 피해 계정으로부터 아이템이 이동하는 패턴에 대한 분석을 진행하기도 했다.

이렇듯 온라인 게임 보안 분야는 다양한 데이터를 이용한 탐지 기법들이 연구되었지만 이런 연구 결과가 실전에 활용된 사례는 상대적으로 드물다. 여러 가지 현실적인 이유가 있겠지만 그 중에서 한 가지 이유는 머신 러닝 탐지 모델을 실전 서비스에 적용하기 위한 구현 및 연동 방법에 대한 자료가 상대적으로 부족하기 때문이다. 따라서 다음 장에서는 이런 다양한 머신 러닝 탐지 기법을 서비스에 연동하기 위한 시스템 구조 및 방법에 대해서 소개하겠다.

Ⅳ. 머신 러닝 인프라 구축 방안

이번 장에서는 머신 러닝을 이용하여 온라인 게임 위협을 탐지하기 위한 시스템 구축 방안에 대해 소개하겠

다.

머신 러닝을 이용한 위협 탐지는 탐지 패턴을 찾는 분석 및 학습 모델링 단계와 생성한 탐지 모델을 적용하여 실제 위협을 탐지하는 서비스 단계로 나뉜다.

분석 및 학습 모델링 단계에서는 보안 전문가 혹은 데이터 분석가가 기존에 축적된 데이터와 보안 위협 사례를 분석하여 보안 위협 탐지를 위한 패턴을 찾거나 기계 학습을 위한 정답 집합을 구축한다. 이후 이런 패턴을 효과적으로 탐지하기 위한 모델을 만들기 위해 기계 학습 알고리즘을 이용해 모델을 학습한다.

서비스 단계에서는 위에서 생성한 학습 모델을 서비스 서버에 적용해 실제 이벤트 발생 시 이것이 정상 데이터인지 아니면 사기나 도용과 같은 보안 위협인지를 판단하고 적절한 대응을 한다.

대개의 경우 분석 및 모델링 단계와 서비스 단계는 개별 분야의 전문가에 의해 업무가 진행되며 개발 및 작업 환경 역시 상이한 경우가 많다. 예를 들어 엔씨소프트의 경우 분석 단계에서는 Hive와 R같은 데이터 분석용 도구를 사용해 통계 전문가나 보안 담당자가 작업을 수행하며, 서비스 단계는 게임 서버나 기타 인프라 서버 개발자가 C++를 이용해서 개발한다.

따라서 이런 이질적인 환경에서는 머신 러닝을 이용해 탐지 모델을 만들더라도 적절한 연동 방안을 마련하지 못한다면 실전에 적용되기까지 많은 시간과 비용이 필요하다.

4.1. 학습 모델과 서비스 연동

분석 및 학습 단계에서 만든 탐지 모델을 서비스에 적용하기 위해선 해당 규칙을 직접 구현하거나 혹은 '도메인 특화 언어(Domain Specific Language, 이하 DSL)'를 이용하는 방법이 가능하다.

서비스에서 직접 구현하는 방법은 매번 탐지 모델을 수정할 때마다 시스템 수정이 필요하기 때문에 구현 및 유지 보수 비용이 크다는 단점이 있다. 또한 '의사 결정 나무(Decision Tree)'처럼 사람이 이해하기 쉬운 방식으로 만들 수 있는 모델이 아니라면 탐지 모델을 구현하는 것은 매우 복잡하기 때문에, 이 방식을 이용할 때는 사용 가능한 머신 러닝 알고리즘에 제약이 생길 수 있다. 이런 제약을 피하려면 학습 단계에서 사용하는 분석 환경과 동일한 시스템 환경을 서비스에서도 이용하는 방법을 사용할 수 있다. 예를 들어 파이썬의 SciKit 라이브러리를 분석 환경에서 사용했다면 학습 모델을 적용하는 시스템에서도 탐지 모듈을 파이썬으로 구축하는 방식이다. 그러나 이렇게 분석 및 학습 단계와 서비스 단계에 의존성을 두게 되면 향후 유지 보수에 제약이 생길 수 있다.

결국 탐지 모델을 서비스에 적용할 때는 둘 사이에 구현 레벨에서의 의존성을 분리할 수 있는 DSL을 사용해야 한다. 이 때 DSL은 첫째, 다양한 머신 러닝 알고리즘의 모델을 명세할 수 있어야 하고, 둘째, 이 명세한 모델을 범용적으로 탐지에 활용할 수 있는 구현체가 있어야 하며, 셋째, 명세 결과를 디코딩하거나 인코딩하는 데 드는 비용이 적어야 한다.

위 요구 사항을 만족하는 대표적인 언어가 'Predictive Model Markup Language(이하 PMML)'이다[21]. PMML은 'Data Mining Group(이하 DMG)'에서 만든 통계 및 머신 러닝 모델 명세용 표준이다. XML을 기반으로 하며 여러 알고리즘의 모델 명세에 필요한 파라미터 정보를 표시할 수 있도록 규격화했는데 1998년에 최초로 규격이 공개되어 계속 확장 및 발전하였으며 2016년 5월 현재 4.2.1 버전까지 공개되어 있다. 자바, 파이썬, R 등 다양한 언어 및 머신 러닝 라이브러리에서 PMML을 지원하고 있으며 널리 알려진 학습 알고리즘의 대부분을 명세할 수 있기 때문에 적용

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<PMML xmlns="http://www.dmg.org/PMML-4_2" version="4.2">
    <Header copyright="Copyright (c) 2014 vfed" description="RPart Decision T
        <Extension extender="Rattle/PMML" name="user" value="vfed"/>
        <Application name="Rattle/PMML" version="1.4"/>
        <Timestamp>2014-07-06 23:51:05</Timestamp>
    </Header>
    <DataDictionary numberOfFields="5">
        <DataField name="Species" optype="categorical" dataType="string">
            <Value value="setosa"/>
            <Value value="versicolor"/>
            <Value value="virginica"/>
        </DataField>
        <DataField name="Sepal_Length" displayName="Sepal length in cm" optyp
            <Interval closure="closedClosed" leftMargin="4.3" rightMargin="7.
        </DataField>
        <DataField name="Sepal_Width" displayName="Sepal width in cm" optype=
            <Interval closure="closedClosed" leftMargin="2.0" rightMargin="4.
        </DataField>
        <DataField name="Petal_Length" displayName="Petal length in cm" optyp
            <Interval closure="closedClosed" leftMargin="1.0" rightMargin="6.
        </DataField>
        <DataField name="Petal_Width" displayName="Petal width in cm" optype=
            <Interval closure="closedClosed" leftMargin="0.1" rightMargin="2.
        </DataField>
    </DataDictionary>
    <TreeModel modelName="RPart_Model" functionName="classification" algorith
        <MiningSchema>
            <MiningField name="Species" usageType="predicted"/>
            <MiningField name="Sepal_Length" usageType="active"/>
            <MiningField name="Sepal_Width" usageType="active"/>
```

〔그림 2〕 Iris 데이터를 의사결정나무 알고리즘으로 분류한 모델에 대한 PMML 샘플

단계에서의 제약이 적고 이미 실전에서 검증된 de facto 표준이다 (그림 2 참조).

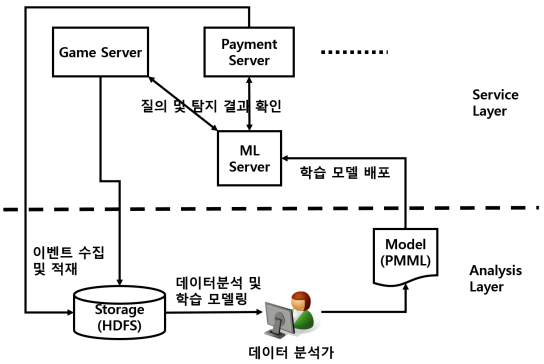
4.2. 머신 러닝 서버 구축 및 서비스 연동

앞 절에서 소개한 PMML을 이용하면 분석 및 학습 단계에서 만든 탐지 모델을 PMML 형태의 데이터로 변환하여 탐지 시스템에 전달할 수 있다. 그러면 탐지 시스템에서는 전달 받은 PMML 데이터를 디코딩하여 탐지 모델을 사용하게 된다. 그런데 2장에서 언급했듯이 온라인 게임 서비스에는 다양한 위협들이 존재하며 각 위협을 탐지하기 위한 서비스는 대개 독립적인 시스템으로 구축되어 있다. 가령, 계정 도용이나 자동 사냥 프로그램 탐지 및 제재는 게임 서버와 연동되어야 하지만 결제 사기 탐지는 결제 시스템과 연동되어야 한다. 따라서 만약 PMML 디코딩 및 탐지 모듈이 각 시스템마다 별도로 구현된다면 유사한 기능들이 중복 구현되기 때문에 비효율적이다.

이런 중복 구현에 따른 부담과 운영 및 유지보수 비용을 줄이려면 PMML 데이터를 디코딩하고 탐지 모델을 실행하는 기능은 별도의 서버로 구축하고 서비스 서버는 이 서버와 API를 통해 연동하는 것이 바람직하다.

예를 들어 분석 담당자는 분석 및 학습 단계에서 생성한 모델을 머신 러닝 서버에 배포하고, 서비스 서버에서는 이벤트 발생 시 이 이벤트가 계정 도용이나 결제 사기 같은 보안 위협인지 여부를 탐지하기 위해 머신 러닝 서버에 질의하여 그 결과에 따라 적절한 조치를 취하는 것이다. 이를 도식화하면 [그림 3]과 같다.

위에서 설명한 머신 러닝 서버를 구현한 사례 중 대표적인 것이 Openscoring이다[22]. Openscoring은 머신 러닝 서비스를 제공하는 REST API 기반의 오픈 소스 웹 애플리케이션이다. 모델 명세를 위해 PMML을 사용하고 있으며 모델의 배포와 예측 서비스를 HTTP 기반의 REST API 형태로 제공한다. 따라서 이를 이용하면 분석 및 학습 단계와 서비스 단계 사이의 의존성을 없앨 수 있으며 REST API를 통해 서비스를 중단하지 않고도 탐지 모델 변경 및 개선이 가능하다.



〔그림 3〕 머신 러닝 서버를 이용한 학습 모델과 서비스 서버 연동 구조

Ⅴ. 머신 러닝 적용 시 고려 사항

온라인 게임 보안에서 머신 러닝을 실전에 적용할 때는 아래와 같은 사항에 대해 고려해야 한다.

5.1. 정답 집합 구축

머신 러닝을 온라인 게임 보안 분야에 적용하기 위해선 먼저 탐지하고자 하는 위협에 대한 학습 및 평가를 위해 정답 집합이 준비되어야 한다. 비록 정답 집합 없이 학습을 수행하는 비지도 학습 알고리즘이 있지만 현재까지 비지도 학습 알고리즘은 지도 학습 알고리즘에 비해 성능이나 효율성 측면에서 미숙한 부분이 많다. 다음 절에서 언급하겠지만 보안 탐지 분야는 높은 정확도를 보장하지 못하면 실전에 적용하기 어려운 영역이기 때문에 비지도 학습 알고리즘을 사용하는 것은 아직 시기상조이다.

그러나 지도 학습 알고리즘을 적용함에 있어서도 한 가지 문제가 있는데 바로 지도 학습에 사용할 정답 집합을 어떻게 만들 것인가 하는 것이다. 머신 러닝 알고리즘은 주어진 정답 집합을 기반으로 탐지 모델을 만들기 때문에 만약 정답 집합이 실제 서비스에서 접하게 되는 여러 가지 보안 위협을 충분히 반영하지 못하면 분석 및 모델링 단계에서 아무리 높은 성능을 보이더라도 실전에서는 좋은 성능을 보장할 수 없다.

보통 정답 집합을 구축할 때 기존에 수집된 피해자나 일반 사용자의 신고 혹은 다른 수단을 통해 탐지된 데이터를 이용하는 경우가 많다. 이런 식으로 하면 비록 적은 비용으로 손쉽게 정답 집합을 구축할 수는 있지만 기존 탐지 내역에 편향된 탐지 모델이 만들어지기 때문에 실전에서 높은 성능을 기대하기 힘들다.

따라서 미탐(false negative) 데이터를 고려한 정답

집합 구축 방안을 고려해야 한다. 참고로 엔씨소프트에서는 자동 사냥 프로그램 탐지를 위한 정답 집합 구축을 위해 정기적으로 사용자에 대한 랜덤 샘플링 후 운영자 모니터링을 통해 정답 집합을 구축하고 있으며, 이때 운영자의 주관에 의한 편향이나 구축 과정에서의 실수를 방지하기 위한 체계를 갖추고 있다[9].

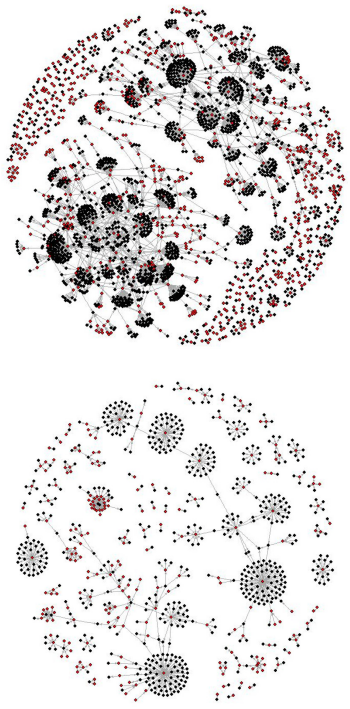
5.2. 오탐에 대한 대응

아무리 우수한 알고리즘과 풍부한 정답 집합을 이용해 학습한다고 해도 오탐이 전혀 없는 탐지 모델을 만드는 것은 현실적으로 불가능하다. 따라서 머신 러닝을 온라인 게임 보안에 적용할 때는 오탐이 발생할 경우 어떻게 처리할지를 미리 서비스 상에서 고려하거나 운영 정책을 만들어야 한다.

5.3. 탐지 모델 유지 보수

다른 머신 러닝 응용 분야에 비해 보안 분야는 탐지 모델에 대한 유지 보수 비용이 매우 크다. 왜냐하면 공격자는 자신들의 공격이 탐지당하지 않도록 적극적으로 우회 기법을 시도하기 때문에 이로 인해 과거 패턴에 기반한 탐지 모델이 쉽게 무력화될 수 있기 때문이다. 온라인 게임 분야 역시 마찬가지이다. 계정 도용이나 자동 사냥 프로그램에 대해서 제재를 효과적으로 할수록 공격자는 기존 행위와 다른 방식으로 공격 시도를 하면서 탐지 규칙을 파악하고 우회하려는 노력을 적극적으로 한다.

[그림 4]는 엔씨소프트에서 서비스하는 아이온이라는 MMORPG에서 캐릭터 간의 네트워크 분석 기법을 이용하여 작업장을 탐지하기 위해 사용한 패턴이 어떻게 우회되었는지를 보여주는 그림이다. 이 때 우리가 사용한 기법은 클라이언트 탐지 기법을 통해 탐지한 자동 사냥 캐릭터들과 빈번하게 거래를 주고받는 캐릭터들의 커뮤니티를 찾아 이들을 제재하는 방식이었다[19]. 첫 번째 그림이 당시 작업장 커뮤니티의 모습이다. 그림에서 보다시피 탐지된 자동 사냥 캐릭터(빨간색 노드)가 다른 캐릭터들과 대규모 커뮤니티를 형성하고 있다. 두 번째 그림은 이 탐지 모델이 시스템에 적용된 후 몇 개월이 지난 시점에서의 거래 네트워크 구조이다. 제재 전과 비교해 볼 때 캐릭터들 간의 거래 밀도가 매우 낮아



〔그림 4〕 네트워크 분석 기법을 이용한 자동 사냥 캐릭터 제재 전/후 거래 네트워크 비교 (위: 제재 전, 아래: 제재 진행 몇 개월 후)

거의 커뮤니티를 형성하지 않고 있다. 이것은 작업장이 여러 차례의 제재로 인해 탐지 패턴을 파악하고 이를 우회하기 위해 내부 캐릭터간의 거래를 중지했기 때문이다.

이렇게 온라인 게인 보안에서는 좋은 탐지 모델을 구축하더라도 시간이 지남에 따라 공격자의 패턴이 빠르게 바뀌기 때문에 그 변화를 계속 모니터링하고 그에 대응할 수 있게 모델 유지 보수를 계속 해야 한다. 그리고 이런 유지 보수 비용을 줄이기 위해서는 적절한 자동화 프로세스를 마련하는 것이 중요하다. 예를 들어 Lee 등[9]은 모델의 유지 보수 자동화를 위해 모델 성능에 대한 모니터링과 학습 자동화를 위한 프로세스를 제안하고 실전에 적용하였다.

Ⅵ. 결 론

온라인 게임 업계에서는 보안 시스템에서 머신 러닝을 적용하려는 시도와 관심이 최근 크게 증가하고 있으며 학계에서도 꾸준히 다양한 탐지 기법들이 연구되고

있다. 그러나 상대적으로 이를 실전에 적용한 사례는 많지 않다. 그 이유에 대해 다양한 원인 분석이 가능하겠지만 그 중 하나를 꼽자면, 이렇게 실험 환경에서 성공적으로 구축한 머신 러닝 모델을 실제 서비스에 어떻게 적용해야 할지에 대한 방법론이나 시스템 및 적용 사례에 대한 정보는 거의 공개되어 있지 않기 때문이다.

본 논문에서는 분석 단계에서 구축한 머신 러닝 모델을 라이브 서비스에 적용하기 위한 방법으로 머신 러닝 모델을 DSL로 명세하고 이 명세한 데이터를 API 기반의 머신 러닝 서버에 배포하여 서비스 서버와 연동하는 구조를 소개하였다. 그리고 이 방법은 PMML이라는 모델 명세용 규격과 Openscoring이라는 오픈 소스 웹 서비스를 이용하면 적은 비용으로 구축할 수 있다. 이 구조는 Airbnb와 같은 실전 서비스에서 실제 사용하고 있는 이미 검증된 방법이라는 점에서 충분히 업계에서 적용 가능한 방법이다[20].

한편 온라인 게임은 산업 규모와 사람들에게 미치는 영향이 갈수록 커지고 있으며 이에 따라 온라인 게임 서비스의 취약점을 이용한 범죄 역시 갈수록 다양해지고 있다. 그러나 현실에 미치는 영향에 비해 온라인 게임 보안에 대한 연구는 상대적으로 미약한 수준이다. 이것은 아마도 온라인 게임 업계의 폐쇄성에서 기인한 현상이라 생각한다. 따라서 이런 문제를 타계하기 위해서는 온라인 게임 업계에서 좀 더 열린 자세로 학계와 정보를 교류하고 협업하는 기회를 적극적으로 가져야 할 것이다.

아울러 학계에서 연구된 방법들이 실제 적용되어 효과를 발휘한 사례가 많이 늘어나 업계에서 협업에 대한 필요성을 직접적으로 느낄 수 있게 만드는 것도 중요하다. 이를 위해선 제한된 실험 환경을 통해 테스트가 진행되는 머신 러닝 기법에 대한 연구뿐만 아니라 실전 적용을 위한 시스템 구조나 구현 방안에 대한 부분에도 좀 더 많은 연구가 진행되어 학계의 연구 결과가 업계에 의미 있는 영향을 줄 수 있기를 바란다.

참 고 문 헌

[1] “Electronic Arts’ Ultima Online, The Best-Selling Internet-Only Game in History, Redefines the Meaning of Online Trading,” http://web.archive.org/web/20051124071021/http://retailsupport.ea.com/corporate/pressreleases/uo_ebay.html.

[2] 김연주, “1조 5천억 지하시장을 당당히 세상 밖에 세우다, ‘아이템베이’ 김치현 회장,” http://lecture.cfe.org/info/bbsDetail.php?cid=13113&idx=42943, 2016.

[3] 정우철, “대법원의 무죄판결, 현금거래 합법화인가?,” http://www.thisisgame.com/webzine/news/nboard/4/?n=14360, 2010.

[4] “대법원 판례 뒤집혔다... 중개 사이트 게임 리셀러 ‘적신호’,” http://lineage.playforum.net/bbs/view/1013?idx=38364, 2016.

[5] Jeff Jianxin Yan and Brian Randell, “An Investigation of Cheating in Online Games,” *IEEE Security and Privacy*, vol. 7, no. 3, pp. 37-44, 2009.

[6] Muhammad Aurangzeb Ahmad, Brian Keegan, Jaideep Srivastava, Dmitri Williams, and Noshir Contractor, “Mining for Gold Farmers: Automatic Detection of Deviant Players in MMOGS,” *in Computational Science and Engineering International Conference*, vol. 4, pp. 340-345, Aug, 2009.

[7] Ruck Thawonmas, Yoshitaka Kashifuji, and Kuan-Ta Chen, “Detection of MMORPG Bots based on Behavior Analysis,” *in Advances in Computer Entertainment Technology Conference*, pp. 91-94, Dec. 2008.

[8] Marlieke van Kesteren, Jurriaan Langevoort, and Franc Grootjen, “A step in the right direction: Botdetection in MMORPGs using movement analysis,” *Proceedings of the 21st Belgian-Dutch Conference on Artificial Intelligence*, 2009.

[9] Eunjo Lee, Jiyoung Woo, Hyoungshick Kim, Aziz Mohaisen, Huy Kang Kim, “You are a game bot!: uncovering game bots in MMORPGs via self-similarity in the wild,” *NDSS*, Feb, 2016.

[10] Ah Reum Kang, Jiyoung Woo, Juyong Park, and Huy Kang Kim, “Online game bot detection based on party-play log analysis,” *Computers & Mathematics with Applications*, vol. 65, no. 9,

pp. 1384-1395, 2013.

[11] 김하랑, 김휘강, "온라인 게임 봇 길드 탐지 방안 연구," *정보보호학회논문지* 제25권 제5호, pp. 1115-1122, 2015.

[12] Kyungmoon Woo, Hyukmin Kwon, Hyun-chul Kim, Chong-kwon Kim, Huy Kang Kim, "What can free money tell us on the virtual black market?," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 392-393, 2011

[13] Hyukmin Kwon, Kyungmoon Woo, Hyun-chul Kim, Chong-kwon Kim, Huy Kang Kim, "Surgical strike: A novel approach to minimize collateral damage to game BOT detection," *Proceedings of Annual Workshop on Network and Systems Support for Games*, IEEE Press, 2013

[14] 강성욱, 이진, 이재혁, 김휘강, "MMORPG에서 GFG 쇠퇴를 위한 현금거래 구매자 탐지방안에 관한 연구," *정보보호학회논문지*, 제25권, 제4호, pp. 849-861, 2015.

[15] Brian Keegan, Muhammad Aurangzeb Ahmad, Jaideep Srivastava, Dmitri Williams, and Noshir Contractor, "Dark Gold: Statistical Properties of Clandestine Networks in Massively Multiplayer Online Games," *in Social Computing (SocialCom), IEEE Second International Conference*, pp. 201-208, Aug, 2010.

[16] Kuan-Ta Chen and Li-Wen Hong, "User identification based on game-play activity patterns," *Proceedings of the 6th ACM SIGCOMM workshop on Network and system support for games*, ACM, pp. 7-12, 2007.

[17] Jiyoung Woo, Hwa Jae Choi, and Huy Kang Kim, "An automatic and proactive identity theft detection model in MMORPGs," *Appl. Math*, vol.6, no.1, pp. 291-302, 2012.

[18] 김하나, 곽병일, 김휘강, "MMORPG 게임 내 계정도용 탐지 모델에 관한 연구," *정보보호학회논문지* 제25권, 제3호, pp. 627-637, 2015.

[19] Eunjo Lee, Jina Lee, and Janghwan Kim, "Detecting the bank character in MMORPGs by analysis of a clustered network," *The 3rd International Conference on Internet*, 2011.

[20] Naseem Hakim and Aaron Keys, "Architecting a Machine Learning System for Risk," http://nerds.airbnb.com/architecting-machine-learning-system-risk/, 2014.

[21] http://dmg.org/pmml/v4-2-1/GeneralStructure.html

[22] http://openscoring.io/

〈저자소개〉

이 은 조 (Lee, Eun Jo)
정회원

2002년 2월 : 숭실대학교 검퓨터 학부 학사
2007년 2월 : 숭실대학교 정보통신대학원 석사
2015년 2월~현재 : 고려대학교 정보보호대학원 박사과정
2007년 5월~현재 : 엔씨소프트 데이터인텔리전스팀 팀장
<관심분야> 온라인 게임 보안, 데이터 마이닝