논문 21-46-11-11 The Journal of Korean Institute of Communications and Information Sciences '21-11 Vol.46 No.11
https://doi.org/10.7840/kics.2021.46.11.1853

사이버 위협 헌팅을 위한 사용자 행위 정보 기반 위협 우선순위 산정 기법 연구

김 상 수◆, 심 신 우°, 임 선 영*, 구 성 모**

A Threat Prioritization Method Using User Behavior Data for Cyber Threat Hunting

Sang-soo Kim◆, Shinwoo Shim°, Sun-Young Im*, Sung-mo Koo**

요 약

사이버 위협 헌팅은 활발한 사이버 방어 활동으로, 기존의 보안 솔루션을 회피하는 지능형 위협을 탐지하고 격리하기 위해 네트워크를 통해 사전 및 반복적으로 검색하는 프로세스이다. 헌팅을 위해 수집된 다양한 행위 정보들 속에서 복수 개의 위협이 탐지되었을 때 어떤 위협을 먼저 대응해야 할지를 결정하는 것은 무척 중요한 일이다. 본 논문은 사이버 위협 헌팅 환경에서 사용자 행위 정보를 이용하여 위협의 우선순위를 산정하는 기법을 새롭게 제안하였다. 제안된 알고리즘은 노드 가중치, 에지 가중치, 룰 위험도를 이용하였고, 사례 연구를 통해 테스트 망의 모의 공격 시나리오에 대해 실적용 가능성을 제시하였다.

Key Words : Threat Prioritization, Cyber Threat Hunting, Information Security, Threat Assessment, Risk Assessment

ABSTRACT

Cyber threat hunting is an active cyber defence activity which is the process of proactively and iteratively searching through networks to detect and isolate advanced threats that evade existing security solutions. It is very important to decide which threat to respond first when multiple threats are detected among various behavioral information collected for hunting. This paper proposes a new technique for calculating the priority of threats using user behavior information in a cyber threat hunting environment. The proposed algorithm used node weight, edge weight, and rule risk, and the possibility of practical application to the simulated attack scenario of the test network was presented through a case study.

Ⅰ. 서 론

최근 발생하고 있는 보안 사고들은 장기간에 걸쳐 은밀하게 침투하는 표적공격을 통해 발생되고 있으며, 공격자의 공격이 진행되는 기간 동안 수개월 이상 탐지되지 않아 조직에 치명적인 피해를 주고 있다.

하지만, SIEM(Security Information and Event Management) 및 IDS(Intrusion Detection System)와 같은 기존의 보안체계는 주로 시그니처 기반 탐지나

◆ First Author : Agency for Defense Development, wisdory@naver.com, 정회원
° Corresponding Author : LIG Nex1 Co., shimshinwoo@lignex1.com, 정회원
* LIG Nex1 Co., sunyoung.im@lignex1.com, 정회원
** Agency for Defense Development, smkoo12@add.re.kr
논문번호 : 202106-143-B-RN, Received June 23, 2021; Revised August 13, 2021; Accepted August 17, 2021

기 분석된 위협모델 기반 탐지에 의존하고 있어 공격자에 의한 우회가 용이하여 최신의 APT(Advanced Persistent Threat)나 랜섬웨어 공격에 매우 취약하다.

더욱이 대부분의 공격자는 VirusTotal[1] 등을 통하여 사전에 다양한 회사의 AV(Anti-Virus) 프로그램에 대한 시험을 거치는 것으로 알려져 있으며 최근에는 파워쉘 스크립트, WMI(Windows Management Instrumentation) 등의 LotL(Living off the Land) 기법을 사용하므로 보안 시스템에서 탐지하기가 무척 힘들다.

이러한 문제를 해결하기 위해서 등장한 것이 EDR(Endpoint Detection and Response) 솔루션이며 이를 이용하여 PC, 서버 등의 단말에서 발생하는 위협 행위를 수집 및 탐지하고 대응하고 있다.

EDR과 유사한 솔루션인 사이버 위협 헌팅은 모든 위협은 사전에 방어할 수 없다는 가정 하에 조직 내 숨어있는 보안 위협을 찾아서 대응하는 모델이다.

사이버 위협 헌팅 모델은 기존의 보안 솔루션을 회피하는 지능형 위협을 탐지하고 대응하기 위해 네트워크를 통해 사전에 반복적으로 위협 행위를 검색하는 프로세스이다.

사이버 위협 헌팅을 위해서는 조직 내의 클라이언트 수준에서 네트워크 수준까지 가능한 모든 로그를 수집하고 로그에서 사이버 위협을 탐지하기 위해 다양한 분석 방법을 이용해야 한다.

헌팅을 통해 다수의 위협 행위를 발견하였을 때 어떤 위협에 대해 먼저 대응할 것인가를 결정하는 것은 무척 중요한 일이다. 왜냐하면, 이러한 위협 우선순위에 따라서 시스템 환경, 프로세스 상태, 악성코드 사용 여부 등의 추가 피해 여부를 조사하기 때문이다. 만약 이러한 위협 우선순위가 없다면 조직 내에서 중요한 자산에서 발생한 심각한 보안 위협에 대해 늦게 조사하거나 대응하게 될 수도 있다.

본 논문은 사이버 위협 헌팅에서 사용자 행위 정보를 이용하여 위협 우선순위를 산정하는 기법을 제안하고자 한다. 본 논문 구성은 다음과 같다. 다음 2장에서는 관련 연구 및 배경지식을 기술하고, 3장에서는 우선순위 산정 기법을 제안하고 4장에서는 사례연구를 통해 제안된 기법의 실제 적용가능성을 확인하며, 마지막 5장에서 결론을 맺는다.

Ⅱ. 관련 연구 및 배경 지식

2.1 관련 연구

기존의 연구는 주로 침해지표에 의한 위협 우선순위와 SIEM 이벤트에 대한 위협 우선순위를 결정하는 연구가 이루어졌으나, 수집된 사용자 행위 정보에서 위협 우선순위를 정하는 연구는 아직까지 이루어지지 않았다.

침해지표를 이용하여 위협의 우선순위를 결정하는 연구는 S. Kim 등[2]에 의해 이루어졌는데, 침해지표의 위험도 수준과 자산 중요도를 이용하였다. 침해지표의 위험도 수준은 MISP(Malware Information Sharing Platform)[3]의 점수를 이용하였고 자산 중요도는 특정 침해지표의 확산을 고려하여 산정하였다.

SIEM에서 발생하는 다양한 이벤트 중에서 먼저 대응해야 하는 위협에 대한 우선순위를 선정하는 연구는 A. Kim 등[4]에 의해 이루어졌다. 그 연구에서 연결성을 고려한 자산 중요도와 이벤트 영향도를 이용하여 전체 이벤트 효과도를 계산하고, 이벤트 전파도, 이벤트 클러스터링, 이벤트 상관분석을 고려하여 최종적으로 이벤트 우선순위를 계산한다.

본 연구는 조직 내의 사이버 자산에 대한 중요도, 자산과 자산 사이의 에지에 대한 중요도, 단말에서 수집된 사용자 행위에 대하여 사전에 정의된 위협 룰의 위험도를 이용하여 복수 개의 위협이 발견되었을 때 어떤 위협을 먼저 대응해야 하는가에 대한 위협 우선순위 산정 기법을 새롭게 제안한다.

2.2 배경 지식

본 연구에서 사용되는 공개 소프트웨어 도구 및 프레임워크는 SigmaHQ/sigma[5], ATT&CK(Adversarial Tactics, Techniques, and Common Knowledge)[6], CAPEC(ommon Attack Pattern Enumeration and Classification)[7]이며, 주요 특징은 다음과 같다.

2.2.1 SigmaHQ/sigma

SigmaHQ/sigma 프로젝트의 주요 목적은 연구자들 또는 분석가들이 개발된 탐지 기법을 한번만 기술하고 그것을 다른 이들과 공유가 가능한 구조화된 형식을 제공하는 것이다.

SigmaHQ/sigma는 직관적인 방식으로 관련 로그 이벤트를 기술할 수 있는 일반적이고 공개된 시그니처 포맷이다.

룰 포맷은 아주 유연하며, 작성하기 쉽고 어떤 종류의 로그 파일에도 적용 가능하며, SigmaHQ /sigma 포맷으로 기술된 룰을 sigma 컨버터를 이용하여 elasticsearch[8], splunk[9] 등 다양한 SIEM 벤더 고유의 쿼리로 변경한다. SigmaHQ/sigma 룰은 MISP나 STIX(Structured Threat Information Expression)[10]

등의 사이버 위협 인텔리전스 서비스 등을 통해 서로 공유가 가능하다. SigmaHQ/sigma 룰은 지속적으로 업데이트 되고 있으며, 2021년 1월 현재 SigmaHQ/sigma 룰은 670개이다. SigmaHQ/sigma는 Sigma로 표기하기도 하며, 본 장 이후에 SigmaHQ/sigma은 Sigma로 표기한다.

2.2.2 ATT&CK

MITRE의 ATT&CK는 시스템 및 네트워크에 대하여 보안 관점의 약점을 분석해 위협적인 전술과 기술을 체계화하여, 공격자의 공격 행동 패턴을 기반으로 전술(tactics), 기술(techniques), 절차(procedure)로 매핑한 프레임워크이다. ATT&CK 프레임워크는 엔터프라이즈, 모바일, ATT&CK for ICS 3개 영역에 대한 지식 베이스를 제공한다.

ATT&CK는 기 발생한 사이버 위협을 기반으로 공격자의 전술과 기술에 대해 공개하여 누구나 접근 가능한 지식 베이스이다.

전술은 공격자의 공격 목표에 따른 행동을 나타내고 상황에 따른 각 기술의 범주이다. 공격자의 목적에 따라 정찰, 자원 개발, 초기 탐색, 실행 등 엔터프라이즈에는 14개의 범주로 구성되어 있다.

기술은 공격자가 공격 목표에 대한 전술을 달성하는 방법이다. 사이버 공격을 통해 발생하는 결과(또는 피해)를 명시하고, 공격 유형에 따라 앞서 분류된 전술에 다양한 기술을 포함할 수 있다. 2021년 3월 기준으로 엔터프라이즈에는 기술 178개, 하부 기술 352개가 존재한다.

2.2.3 CAPEC

CAPEC은 공격자가 어플리케이션과 그 외 사이버 지원 기능이 가지고 있는 약점에 대해 익스플로잇(exploit)하는 방법을 사용자에게 제공하는 공개된 공통 공격 패턴 목록이다.

공격 패턴은 공격자가 사이버 지원 기능에서 알려진 약점을 익스플로잇 하는데 사용되는 공통 속성과 기법을 말한다. 각각의 공격 패턴은 공격의 구체적 부분이 어떻게 설계되고 실행되는지에 대한 지식을 포함하고 공격의 효과를 완화할 수 있는 지침을 제공한다.

공격 패턴은 어플리케이션을 개발하거나 사이버 지원 기능을 관리하는 사람들에게 구체적 공격 요소에 대해 파악하여 공격 성공을 막는 방법을 제공한다.

Ⅲ. 사이버 위협 헌팅을 위한 사용자 행위 정보 기반 위협 우선순위 산정 기법

그림 1은 사용자 행위 정보 기반 위협 우선순위 산정을 위한 전체 아키텍처이다. 입력으로 윈도우, 리눅스, Solaris의 감사 로그(audit log)를 사용하며 logstash[11]와 elasticsearch를 이용하여 데이터를 수집 및 저장한다.

수집된 데이터는 미리 정의된 sigma 룰 및 elasticsearch 쿼리 변환을 통해 룰에 매칭되는 사용자 행위를 검색한다. 검색된 결과는 타임스탬프, 출발지 IP, 목적지 IP, 적용된 룰로 구성될 수 있으며 최종적으로 각 결과에 대한 우선순위를 산정한다. 우선순위 산정에 필요한 노드 가중치 계산을 위하여 노드 정보를 별도로 수집하여 elasticsearch에 저장한다.

위협 우선순위 계산에 대한 전체적 그림은 그림 2와 같다. 데이터 저장소에 미리 정의된 다양한 속성값을 이용하여 먼저 노드 가중치를 계산하고, 필요시 에지 가중치를 계산한다. 노드 및 에지 가중치 계산과는 별도로 사전에 정의된 sigma 룰에 대한 위험도를 계산한다. 마지막으로 노드 가중치, 에지 가중치, 룰 위험도를 이용하여 전체 위협 우선순위를 계산한다.

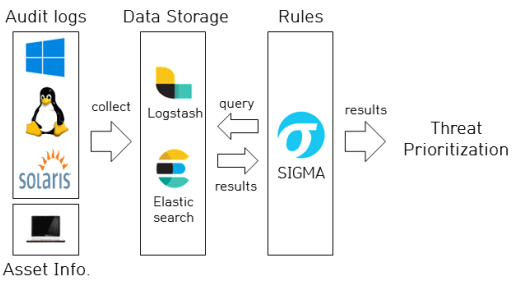

그림 1. 위협 우선순위 산정을 위한 전체 흐름
Fig. 1. The Overview of Workflow for Threat Prioritization Computation

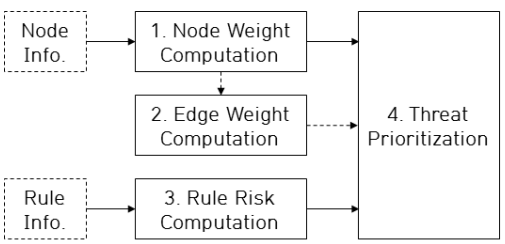

그림 2. 위협 우선순위 산정 아키텍처
Fig. 2. The Architecture of Threat Prioritization Computation

3.1 노드 가중치 계산

사이버 상의 노드는 PC, 서버, 보안 장비 등이 있으며 이들은 사이버 방어 관점에서 중요도가 다르다. 동일한 위험도를 가진 룰이 PC에 발생한 경우와 서버에 발생한 경우는 그 피해 정도가 다르기 때문에 노드 가중치를 결정하는 것이 무척 중요하다.

사이버 상에서 사용하는 자산에 대한 중요도는 그 자산을 운용하는 운용자가 미리 알고 있는 경우도 있지만, 대규모 자산으로 구성되어 있는 경우 이에 대한 가중치를 자동으로 계산하여야 한다.

하나의 자산을 구성하는 속성 값은 표 1과 같다. 첫 번째 속성 값은 네트워크의 종류이다. 예를 들어, 군의 네트워크를 크게 나누면 인터넷망, 국방망, 전장망으로 나눌 수 있으며 각각의 망에 대한 중요도가 다르다. 동일한 망에서 다양한 영역이 존재 할 수 있으므로 이를 PC 영역, 외부 사용자를 위한 서비스 영역, 중요 데이터베이스를 가지고 있는 서버 영역으로 나눌 수 있다. 다음으로 자산을 사용하는 부대의 중요도에 따라서 전술급, 작전급, 전략급으로 나눈다. 마지막으로 자산에 운용되는 OS의 종류는 윈도우, 리눅스, 유닉스로 나눈다.

네트워크의 경우 인터넷망 값은 국방망이나 전장망에 비해 낮고, 영역의 경우 PC 영역이 서비스 및 서버 영역보다 그 값이 낮다. 부대 및 OS의 경우도 전략급 및 유닉스인 경우의 값이 더 높다.

다양한 속성을 가진 자산에 대한 가중치를 계산하기 위하여, 우리는 TOPSIS(Technique for Order of Preference by Similarity to Ideal Solution)[12] 알고리즘을 사용하였다. TOPSIS 알고리즘은 가상 양의 이상치(최상점)와 가상 음의 이상치(최하점)와의 근접도에 따라서 대안을 정렬시키는 기법이다. 이 기법을 사용하여 솔루션(예를 들어, 가장 선호되는 대안)은 최저점에서 가장 멀고 동시에 최상점에서 가장 가까운 대안인데, 여기서 거리는 유클리디언 거리로 계산한다.

표 1. 자산을 구성하는 속성 값
Table 1. The Attributes of an Asset

Attribute	Value	
Network	Internet	1
	Military Network	3
	Combat Network	5
Region	PC	1
	Service	3
	Server	5
Command	Tactical	1
	Operational	3
	Strategic	5
Operating System	Windows	1
	Linux	3
	Unix	5

3.2 에지 가중치 계산

sigma로 정의된 룰은 크게 노드의 행위와 관련된 룰과 노드와 노드 사이의 에지와 관련된 룰로 나눌 수 있다. 앞서 계산한 노드 가중치는 노드 행위와 관련된 룰에 적용된다.

표 2는 파워쉘을 이용한 리모트 연결 행위를 탐지하는 룰의 예이다. 룰에서 detection 부분에 DestinationPort가 5985 또는 5986인 행위들 중에서 User: 'NT AUTHORITY\NETWORK SERVICE가 아닌 것을 검색하게 된다.

표 2와 같은 룰은 에지에 적용되므로 이러한 경우 에지의 가중치를 계산해야 한다. 검색 결과 Remote powershell session에 해당하는 행위가 검색된 경우에 에지에 대한 가중치가 필요하며 이는 앞서 계산한 노드 가중치를 이용하여 동적으로 계산해야 한다.

표 2. Sigma 룰 예시
Table 2. The Example of Sigma Rule

```
title: Remote PowerShell Session
id: c539afac-c12a-46ed-b1bd-5a5567c9f045
description: Detects remote PowerShell connections by monitoring network outbound connections to ports 5985 or 5986 from not network service account
status: experimental
date: 2019/09/12
modified: 2020/08/24
author: Roberto Rodriguez @Cyb3rWard0g
references:
    -
https://github.com/Cyb3rWard0g/ThreatHunter-Playbook/tree/master/playbooks/windows/02_execution/
        T1086_powershell/powershell_remote_session.md
tags:
    - attack.execution
    - attack.t1059.001
    - attack.t1086   # an old one
    - attack.lateral_movement
    - attack.t1021.006
    - attack.t1028   # an old one
logsource:
    category: network_connection
    product: windows
detection:
    selection:
        DestinationPort:
            - 5985
            - 5986
    filter:
        User: 'NT AUTHORITY\NETWORK SERVICE'
    condition: selection and not filter
falsepositives:
    - Leigitmate usage of remote PowerShell, e.g. remote administration and monitoring.
level: high
```

그림 3과 같이 두 개의 노드 i와 j가 있고, 노드 i의 가중치와 노드 j의 가중치가 있을 때 에지 가중치를 구한다.

노드 가중치 W_{n_i}와 W_{n_j}를 이용하여 에지 가중치 $W_{e_{ij}}$를 구하는 방법은 크게 3가지 방법으로 생각해 볼 수 있다. 첫 번째는 두 노드 가중치의 평균(mean)을 이용하는 방법, 두 번째는 두 노드의 가중치 중 최대값(max)을 에지 가중치로 이용하는 방법, 마지막으로 목적지 노드의 가중치를 그대로 이용하는 방법이 있으며 에지 가중치는 3가지 중의 하나를 선택하여 계산한다.

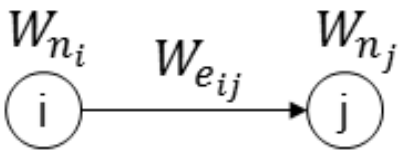

그림 3. 노드 가중치를 이용한 에지 가중치 계산
Fig. 3. The Edge Weight Computation using Node Weight

3.3 룰 위험도 계산

모든 sigma 룰은 각 룰마다 위험 등급(level)을 가지고 있으며 low, medium, high, critical 4개로 구성되어 있다. 이러한 등급은 룰을 작성하는 사용자에 의해서 결정되는데, 사용자에 의한 등급이 효과적일 수 있지만 등급 결정 사유가 모호하고 새로운 룰 생성 시 객관적 기준이 없어 상황에 따라 위험 등급이 맞지 않을 수 있다.

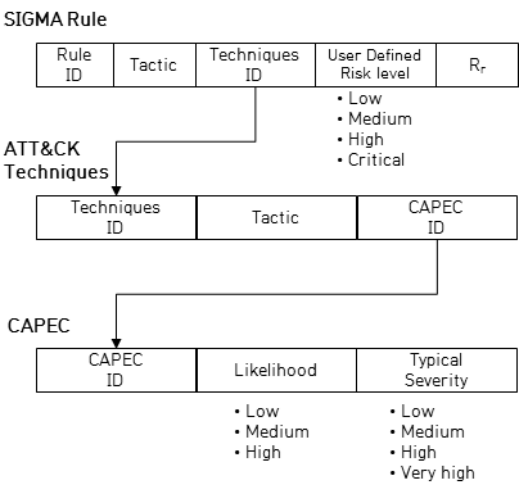

그림 4. Sigma 룰과 ATT&CK 기술 및 CAPEC 과의 관계
Fig. 4. The Relation of Sigma Rule, ATT&CK Techniques ID and CAPEC ID

이러한 문제점을 해결하기 위하여 사용자 정의 등급과 별도로 sigma 룰에서 가지고 있는 태그(tags) 정보를 이용하여 등급을 계산하는 방법을 제안한다. 그림 4와 같이 sigma 룰에서 가지고 있는 태그 중 MITRE의 ATT&CK 프레임워크의 기술(techniques) ID를 연결할 수 있고, 기술 번호는 MITRE의 CAPEC ID와 연결할 수 있다.

ATT&CK의 기술 ID가 모든 sigma 룰의 태그 정보에 존재하는 것은 아니므로 이 경우에는 사용자 정의 위험 등급을 이용한다. 태그 정보 중 기술 ID가 존재하는 경우는 그에 해당하는 기술 ID를 사용하여 CAPEC의 발생 가능성(likelihood)과 심각도(typical severity) 정보를 이용하여 위험도를 계산한다. CAPEC 정보 중 발생 가능성 및 심각도가 모든 CAPEC 항목에 존재하는 것은 아니므로 이를 고려하여 발생 가능성만 있는 경우는 sigma 룰의 사용자 정의 위험 등급을 이용하고 심각도가 있는 경우는 이를 이용한다.

CAPEC 정보 중 발생 가능성 및 심각도가 모두 존재하는 경우는 다음의 수식을 이용하여 임시 위험도(t_Risk)를 식 (1)을 이용하여 계산한다.

$$t\_Risk = Likelihood \times \alpha + Typical\ Severity \times (1-\alpha) \quad (1)$$

위 식에서 α는 발생 가능성과 심각도에 대한 반영 비율이며, 0과 1 사이의 값이다.

발생 가능성은 존재하지 않고 심각도만 존재하는 경우는 심각도를 임시 위험도로 사용하고 그 식은 다음과 같다.

$$t\_Risk = Typical\ Severity \quad (2)$$

사용자 정의 위험도와 CAPEC의 발생 가능성 및 심각도를 고려한 전체 위험도는 식 (3)과 같으며, β는 0과 1 사이의 값이다.

$$R_r = t\_Risk \times \beta + UserDefinedRisk \times (1-\beta) \quad (3)$$

sigma 룰에서 기술 ID가 존재하지 않는 경우, 기술 ID가 존재하더라도 이에 대응되는 CAPEC ID가 존재하지 않는 경우, 기술 ID와 CAPEC ID 모두 존재하더라도 심각도 필드가 존재하지 않는 경우는 사용자 정의 위험도를 룰 위험도로 사용하며, 전체 룰 위험도 계산 알고리즘은 그림 5와 같다.

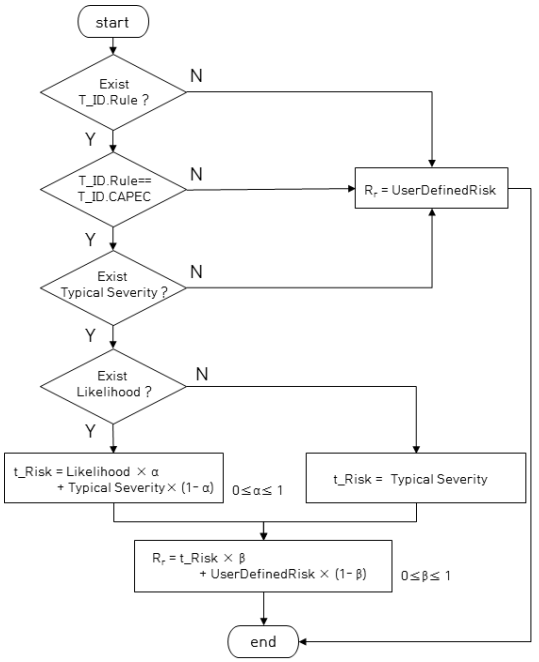

그림 5. 룰 위험도 계산 흐름도
Fig. 5. The flowchart of Rule Risk

3.4 위협 우선순위 계산

위협 우선순위 계산은 노드와 에지를 구분하여 계산한다. 노드에 대한 위협 우선순위는 식 (4)와 같이 노드 가중치(W_n)와 룰 위험도(R_r)를 계산한다. 식에서 γ는 0과 1 사이의 값이다.

$$P = W_n \times \gamma + R_r \times (1-\gamma) \quad (4)$$

에지에 대한 위협 우선순위는 식 (5)와 같이 에지 가중치(W_e)와 룰 위험도(R_r)를 고려하며, γ값은 식 (4)와 동일하게 0과 1 사이의 값이다.

$$P = W_e \times \gamma + R_r \times (1-\gamma) \quad (5)$$

Ⅳ. 사례 연구

앞서 제안한 사용자 행위 정보 기반 위협 우선순위 계산의 적용 타당성을 검증하기 위하여 아래 그림 6과 같은 가상의 테스트망을 구성하였다. 테스트망은 인터넷망, 국방망, 전장망 3개로 구성[13]하고 각 망 별로 각각 9개의 노드를 구성하였다.

각 망별로 위협 행위는 다음과 같다. 먼저 인터넷망에서는 스피어 피싱 첨부파일 공격을 수행하여 노드 N8에 침투하고, 사용자 계정 통제를 우회하고

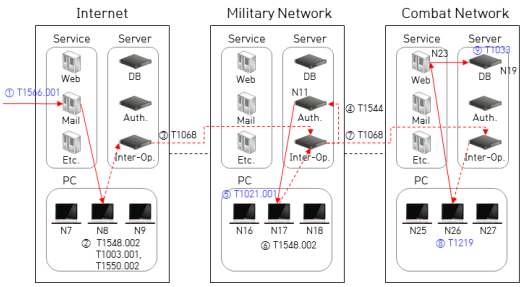

그림 6. 테스트 망 구성도
Fig. 6. The Configuration of Testbed

LSASS(Local Security Authority Subsystem Service) 메모리를 모니터링하여 패스워드를 획득하여 연동서버를 장악한다. 연동서버를 통하여 국방망 노드에 침투하고 미리 획득한 credential을 이용하여 원격 서비스 공격을 노드 N17에 수행한다. 국방망을 통한 합법적인 원격 접근 도구인 VNC(Virtual Network Computing)를 이용하여 전장망의 노드 N23에 침투하고 마지막으로 시스템의 소유자 및 사용자 검색 행위를 N19에 수행한다. 전체 가상 모의 공격 시나리오 및 각 단계별 사용된 ATT&CK 기술 ID는 표 3과 같다.

아래 표에서 위협 헌팅을 위한 시그마 룰에서 탐지가 가능한 기술 ID 및 기술명은 음영 처리가 되어있다.

표 3. 모의공격 시나리오
Table 3. The Scenario of Attack

No.	Techniques ID	Techniques Name
①	T1566.001	Phishing : Spearphishing Attachment
②	T1548.002	Abuse Elevation Control Mechanism : Bypass User Account Control
	T1003.001	OS Credential Dumping : LSASS memory
	T1550.002	Use Alternative Authentication Material : Pass the Hash
③	T1068	Exploitation for Privilege Escalation (CVE-2017-0263)
④	T1544	Remote File Copy
⑤	T1021.001	Remote Services : Remote Desktop Protocol
⑥	T1548.002	Abuse Elevation Control Mechanism : Bypass User Account Control
⑦	T1068	Exploitation for Privilege Escalation (CVE-2014-4113)
⑧	T1219	Remote Access Software (VNC)
⑨	T1033	System Owner/User Discovery

4.1 노드 가중치 계산

테스트망의 노드 가중치를 계산하기 위한 전체 노드의 속성 값은 표 4와 같다. 모의공격 시나리오와 관련된 노드는 N8, N11, N17, N19, N23이다.

공격과 관련된 노드의 속성을 수치화하면 표 5와

표 4. 노드 속성
Table 4. The Attributes of Nodes

Node	Network	Region	Command	Operating System
N1	Internet	Server	Operational	Unix
N2	Internet	Server	Operational	Unix
N3	Internet	Server	Operational	Windows
N4	Internet	Service	Operational	Linux
N5	Internet	Service	Operational	Unix
N6	Internet	Service	Operational	Linux
N7	Internet	PC	Operational	Windows
N8	Internet	PC	Operational	Windows
N9	Internet	PC	Operational	Windows
N10	Military Network	Server	Operational	Linux
N11	Military Network	Server	Operational	Unix
N12	Military Network	Server	Operational	Linux
N13	Military Network	Service	Operational	Linux
N14	Military Network	Service	Operational	Unix
N15	Military Network	Service	Operational	Unix
N16	Military Network	PC	Operational	Windows
N17	Military Network	PC	Operational	Windows
N18	Military Network	PC	Operational	Windows
N19	Combat Network	Server	Strategic	Windows
N20	Combat Network	Server	Strategic	Unix
N21	Combat Network	Server	Strategic	Windows
N22	Combat Network	Service	Strategic	Linux
N23	Combat Network	Service	Strategic	Windows
N24	Combat Network	Service	Strategic	Unix
N25	Combat Network	PC	Strategic	Windows
N26	Combat Network	PC	Strategic	Windows
N27	Combat Network	PC	Strategic	Windows

같다. 표 4에서의 각 노드의 속성과 표 1의 자산을 구성하는 속성 값을 조합하여 산출 가능하다. 예를 들어 N8의 Network 속성은 internet으로 표 1에서 internet에 해당하는 수치는 1이고, Region 속성은 PC로 수치는 1, Command는 Operational로 수치는 3이며, Operating System은 Windows로 수치가 1로 표현된다.

TOPSIS 알고리즘에서는 각 속성 간의 가중치를 설정할 수 있으며, 본 연구에서는 Network의 가중치를 0.25, Region의 가중치를 0.1, Command의 가중치를 0.4, Operating System의 가중치를 0.25로 설정하였다.

각 속성에서 발생하는 가장 높은 값에 속성 가중치를 곱하면 최상점(Zenith) 속성 값을 구할 수 있고, 각 속성에서 발생하는 가장 낮은 값에 속성 가중치를 곱하면 최하점(Nadir) 속성 값을 구할 수 있다. 예를 들어 표 5의 5개 노드에서 가장 낮은 Network 속성 값은 1로 여기에 Network 속성 가중치인 0.25를 곱하면 최하점의 network 값은 0.25가 된다. 비슷한 방식으로 최상점의 network 값은 표 5의 network 수치에서 가장 높은 수치인 5에 network 속성 가중치인 0.25를 곱하면 1.25가 된다. 이러한 방식으로 최상점과 최하점의 속성을 구하면 표 6과 같다.

각 노드의 속성에 속성 가중치를 곱한 후 구해진 최상점과 최하점과의 유클리드 거리를 구하면 각 노드의 최상점까지의 거리(Distance to Zenith) 및 최하점까지의 거리(Distance to Nadir)를 구할 수 있으며, 이를 이용해 최종적으로 노드 가중치를 계산할 수 있다. TOPSIS 값인 노드 가중치는 식 (6)과 같이 최하점까지의 거리와 최상점까지의 거리를 더한 값으로 최하점까지의 거리를 나누어서 계산할 수 있다.

$$Weight = \frac{Distance\ to\ Nadir}{Distance\ to\ Nadir + Distance\ to\ Zenith} \quad (6)$$

표 5. 노드 속성 값
Table 5. The Node Attribute Value

Node	Network	Region	Command	Operating System
N8	1	1	3	1
N11	3	5	3	5
N17	3	1	3	1
N19	5	5	5	1
N23	5	3	5	1

표 6. 최상점과 최하점
Table 6. Zenith and Nadir

Node	Network	Region	Command	Operating System
Zenith	1.25	0.50	2.00	1.25
Nadir	0.25	0.10	1.20	0.25

표 7. 노드 가중치 계산
Table 7. The Node Weight Computation

Node	Distance to Zenith	Distance to Nadir	Node Weight (TOPSIS value)
N8	1.673320	0.000000	0.000000
N11	0.943398	1.187434	0.557263
N17	1.431782	0.500000	0.258828
N19	1.000000	1.341641	0.572949
N23	1.019804	1.296148	0.559661

이러한 방식으로 공격과 관련된 노드에 대한 TOPSIS 계산은 표 7과 같다. 노드 가중치가 가장 높은 것은 N19 노드이며 가장 낮은 것은 N8이다.

4.2 에지 가중치 계산

에지 가중치 계산을 위해 노드 N11 및 노드 N17의 가중치를 이용하여 노드 가중치의 평균을 이용하였다. N11의 가중치는 0.557263, N17의 가중치는 0.258828로 두 값의 평균은 0.408046으로, 두 노드 간 에지 가중치는 0.408046이다.

4.3 룰 위험도 계산

Sigma 룰은 각 룰마다 yml 확장자를 가진 파일이 존재하며, 해당 파일 안에는 해당 룰이 탐지하고자 하는 공격에 대한 설명, 탐지 방법, 심각도 레벨 등이 기재되어 있다. 본 연구에서는 전체 Sigma 룰 670개 중 4개의 룰을 활용하였으며 룰 위험도 계산은 수식 (1)~(3)을 이용하여 계산하였고, 그 결과는 표 8과 같다. 룰 위험도가 가장 높은 것은 두 번째 및 세 번째 룰이며 가장 마지막 룰의 위험도가 가장 낮다.

룰 위험도 계산 시 발생 가능성 및 심각도에 대한 반영 비율인 α는 0.5을 사용하여, 발생 가능성 및 심각도를 동일한 비율로 적용하였다. 사용자 정의 위험도와 CAPEC의 발생 가능성 및 심각도의 반영 비율인 β는 0.6을 사용하여 발생 가능성을 60%, 심각도를 40%로 계산하였다.

표 8. 룰 위험도 계산
Table 8. The Rule Risk Computation

Sigma Rule	ATT&CK Techniques ID	CAPEC ID	Likelihood	Severity	User Defined Risk	Rule Risk
win_hwp_exploit.yml	T1566.001	163 (Spear Phishing)	High	High	High	0.75
win_protected_storage_service_access.yml	T1020.001	555(Remote Service with Stolen Credentials)	None	Very High	Critical	1.0
av_exploiting.ymal	T1219	None	None	None	Critical	1.0
win_susp_whoami.yml	T1033	577(Owner Footprinting)	Low	Low	High	0.6

4.4 위협 우선순위 계산

최종 위협 판단지수는 식 (4)와 (5)를 이용하여 계산하였으며 그 결과는 표 9와 같다. 노드 및 에지 가중치와 룰 위험도의 반영 비율인 γ는 0.3을 사용하여 노드 및 에지 가중치를 30%, 룰 위험도를 70%로 적용하였다.

표 9. 위협 우선순위 계산
Table 9. The Threat Prioritization Computation

From	To	Techniques ID	Node/Edge Weight	Rule Risk	Threat Prioritization Index
N8	-	T1566.001	0.000000	0.75	0.357500
N11	N17	T1021.001	0.408046	1.00	0.704023
N23	-	T1219	0.559661	1.00	0.779831
N19	-	T1033	0.572949	0.60	0.586475

최종 위협 판단 지수 중 가장 시급한 위협은 N23에서 발생한 위협으로 판단 지수가 0.779831이며, 다음은 N11과 N17 사이의 에지에 발생한 위협으로 위협 판단 지수가 0.704023이다.

4.5 노드 수에 다른 성능 예측

본 연구의 사례 연구에서는 총 3개의 망에 각 망별로 9개의 노드로 구성되어 총 27개의 노드가 존재하는 비교적 소규모의 환경으로 설정하여 실험을 수행하였다. 본 연구에서 제시하는 방안은 망에 존재하는 전체 노드를 대상으로 연산을 수행하지 않고, 위협이 발생한 노드만을 대상으로 연산을 수행하기 때문에 노드 수가 증가하여도 발생한 위협의 수가 일정하다면 우선순위를 산정하는 계산량은 증가하지 않을 것이며, 계산량은 발생한 위협 수와 위협이 발생한 노드 수에 비례하여 증가할 것으로 예상된다.

Ⅴ. 결론 및 향후 연구

본 연구에서는 사이버 위협 헌팅에서 사용하는 사용자 행위 정보를 이용하여 위협의 우선순위를 정하는 기법을 제안하였다. 이를 위해서 자산을 구성하는 속성 값에 따라서 노드의 자산 중요도를 계산하였고 룰이 노드와 노드 사이의 에지에 적용되는 경우에는 에지 중요도를 계산하였다. 다음으로 룰 위험도 계산을 위해 사용자 정의 필드뿐만 아니라 ATT&CK 기술 ID를 이용하여 CAPEC의 발생 가능성 및 심각도 필드도 함께 이용하였다.

이러한 방식은 기존의 사용자 정의 정보만을 이용하여 룰 위험도를 정의할 때 위험도에 대한 설명이 부족했던 단점을 보완할 수 있으며 ATT&CK와 CAPEC이 업데이트될 때마다 룰 위험도 함께 갱신할 수 있는 장점이 있다.

sigma 룰의 위험도를 ATT&CK 기술 ID를 이용한

CAPEC 정보 조회를 통해 위협 우선순위를 계산하였으나, 룰에 따라서 복수 개의 ATT&CK ID가 존재하는 것도 있는데 이러한 경우에 대해 보완이 필요하다. 아울러, CAPEC에서 제공하는 공격 패턴 527개 중 108개만이 ATT&CK 기술 ID를 보유하고 있어 sigma 룰과 CAPEC의 매핑 방안도 향후에 더 연구해야 할 분야이다.

References

[1] https://www.virustotal.com

[2] S. Kim, S. Koo, L. Kim, and S. Shim, "A study on threat prioritization considering diffusion of IOC in enterprise network," in *Proc. Korea Inst. Military Sci. Technol. Conf. 2020*, pp. 1111-1112, Online, Nov. 2020.

[3] S. Mokaddem, G. Wagner, C. Wagner, A. Dulaunoy, and A Iklody, "Taxonomy driven indicator scoring in MISP threat," arXiv preprint arXiv:1902.03914, 2019.

[4] A. Kim, M. H. Kang, J. Z. Luo, and A. Velazquez, "*A framework for event prioritization in cyber network defense*," NRL/MR/5540－14-9541, 2014.

[5] https://github.com/SigmaHQ/sigma

[6] https://attack.mitre.org

[7] https://capec.mitre.org

[8] https://www.elastic.co

[9] https://www.splunk.com

[10] https://stixproject.github.io

[11] https://www.elastic.co

[12] Y. J. Lai, T. Y. Liu, and C. L. Hwang, "TOPSIS for MODM," *Eur. J. Operational Res.*, vol. 76, no. 3, pp. 486-500, Aug. 1994.

[13] D. H. Kim and H. J. Park, "A study on the hacking countermeasures in military security," *J. Convergence Secur.*, vol. 17, no. 5, pp. 134-142, Dec. 2017.

김 상 수 (Sang-soo Kim)

1997년 7월 : 경북대학교 전자공학과 졸업

2003년 7월 : 경북대학교 컴퓨터공학과 석사

2003년 8월~현재 : 국방과학연구소 연구원

<관심분야> 사이버보안, 사이버 상황인식, 인공지능

[ORCID: 0000-0001-7975-673X]

심 신 우 (Shinwoo Shim)

2007년 2월 : 포항공과대학교 컴퓨터공학 학사

2019년 2월 : 고려대학교 정보보호학 석사

2007년 1월~현재 : LIG넥스원 수석연구원

<관심분야> 사이버 지휘통제, 임무영향평가, 사이버 위협 대응

[ORCID:0000-0003-0959-9200]

임 선 영 (Sun-Young Im)

2015년 2월 : 아주대학교 컴퓨터공학 학사

2017년 2월 : 아주대학교 컴퓨터공학 석사

2017년 1월~현재 : LIG넥스원 선임연구원

<관심분야> 사이버전, 사이버 위협 피해평가, 표적공격 분석

[ORCID:0000-0003-4385-173X]

구 성 모 (Sung-mo Koo)

1994년 2월 : 홍익대학교 전자계산학과 졸업

1996년 2월 : 홍익대학교 컴퓨터공학과 석사

1996년 3월~현재 : 국방과학연구소 연구원

<관심분야> 사이버보안, 사이버 상황인식, 인공지능