논문 21-46-11-12 The Journal of Korean Institute of Communications and Information Sciences '21-11 Vol.46 No.11
https://doi.org/10.7840/kics.2021.46.11.1862

대용량 인증 로그를 활용한 Lateral Movement 탐지 연구

심 신 우•, 김 상 수°, 임 선 영*, 구 성 모**, 조 병 모*, 김 광 수*, 김 태 규*

A Study of Detecting Lateral Movement Using Large-Scale Authentication Log

Shinwoo Shim•, Sang-soo Kim°, Sun-Young Im*, Sung-mo Koo**,
Byoungmo Cho*, Kwangsoo Kim*, Taekyu Kim*

요 약

측면이동(Lateral Movement)은 사이버 킬체인(Cyber Kill Chain) 단계 중 하나로 공격자가 초기 침투 이후 장악하는 대상을 늘려 나가는 과정이다. 공격자는 측면이동을 반복하여 점차적으로 활동 범위를 넓혀 나가 최종 목표물로 도달할 수 있기 때문에 측면이동을 탐지하여 공격자의 공격을 중간에 차단하면 공격자의 목적 달성을 제지할 수 있다. 하지만 공격자는 자신의 공격이 탐지되지 않기 위해 고도화된 공격 방법을 통해 은밀하게 공격을 수행하기 때문에 이러한 공격자의 행위를 탐지하기는 쉽지 않다. 본 연구에서는 대용량 인증 로그를 활용하여 공격자의 측면이동을 탐지하는 방안을 제시하며, 출발지 계정, 도착지 계정, 출발지 컴퓨터, 도착지 컴퓨터, 로그인 시각 정보를 활용하여 로그인 경로들의 평범도 지수를 산출하고 비정상적인 로그인을 식별하여 잠재적인 공격자의 측면이동을 탐지한다. 공개 데이터 셋인 로스 알라모스 연구소(Los Alamos National Laboratory) 데이터 셋을 대상으로 본 연구에서 제안하는 방법을 적용하여 실험하였으며, 1-홉(1-hop) 경로 분석의 경우 TPR은 98.66%, FPR은 0.74%, 정확도는 99.26%의 성능 수치를 기록하여 본 연구에서 제시하는 방법을 실제 데이터에 적용하였을 때 효과적이었음을 확인하였다.

Key Words : Information Security, Lateral Movement, Intrusion Detection, Anomaly Detection, Cyber Kill Chain

ABSTRACT

Lateral Movement is one of the stages of the Cyber Kill Chain, in which attackers increase the number of targets they control after initial compromise. Since the attacker can gradually expand the range of activity by repeating the lateral movement and reach the final target, if the defender detects the lateral movement and blocks the attacker's attack in the middle, the attacker's achievement of the objective can be stopped. However, it is not easy to detect the actions of such attackers because attackers secretly perform attacks through advanced attack methods to prevent their attacks from being detected. In this study, we propose a method to detect an attacker's lateral movement by using a large-scale authentication log, and calculate the normality index of the login paths. By comparing the normality index score, we can detect abnormal login to detect the lateral movement of potential attackers. The method proposed in this study was applied to the Los

◆ First Author : LIG Nex1 Co., shimshinwoo@lignex1.com, 정회원
° Corresponding Author : Agency for Defense Development, wisdory@naver.com, 정회원
* LIG Nex1 Co., {sunyoung.im, byoungmo.cho, kwangsoo.kim, taekyu.kim}@lignex1.com
** Agency for Defense Development, smkoo12@add.re.kr
논문번호 : 202106-134-B-RN, Received June 21, 2021; Revised August 11, 2021; Accepted August 17, 2021

Alamos National Laboratory public data set and tested, and in the case of 1-hop path analysis, the TPR was 98.66%, the FPR was 0.74%, and the accuracy was 99.26%. It was confirmed that the method was effective when applied to actual data.

Ⅰ. 서 론

사이버 킬체인(Cyber Kill Chain) 프레임워크는 록히드 마틴(Lockheed Martin)에서 제안한 개념으로 공격자가 목표를 이루기 위해 수행해야 하는 업무들을 단계별로 분석하여 공격자의 행동을 파악할 수 있는 프레임워크이다[1]. 사이버 킬체인은 사이버 공격의 기법을 분류하는 데 활용되기도 하며[2], 공격자의 행동을 분석하고 예측하여 공격자의 공격을 방어하는 데에도 활용 가능하다.

사이버 킬체인을 이루는 여러 단계 중 측면이동(Lateral Movement)은 공격자가 초기 침투 이후에 거점을 확보하고, 장악하는 대상을 늘려 나가는 단계로 볼 수 있다. 측면이동 사이클의 세부 내용은 그림 1에서 설명한다.

공격자는 공격 대상의 외부 환경을 조사(External Reconnaissance)하며, 초기 침입 대상을 선정하고 침투에 성공(Initial Compromise)한다. 이후 공격자는 공격 대상을 확대시키기 위해 내부 정찰(Internal Reconnaissance)을 수행하며, 공격에 필요한 자원에 대한 접근을 위해 권한 상승(Privilege Escalation)을 시도한다. 공격자는 상승된 권한을 활용하여 다른 디바이스에 접근할 수 있는 인증 정보를 획득(Harvest Credentials)하며, 획득한 인증 정보를 통해 다른 디바이스에 대한 제어권도 확보(Compromise Other Devices)한다. 공격자는 필요한 자원에 대한 접근 권한을 얻을 때까지 측면이동을 계속 수행하며 내부에서의 제어권을 확장시키며 데이터 유출, 시스템 파괴 등 목표에 대한 활동을 수행(Action on Objectives)하며 최종적으로 임무를 완수(Mission Complete) 시킨다. 공격자의 행동 패턴은 공격자의 공격 방법이나 공격 대상의 환경에 따라 상이할 수 있으며, 여기서는 일반적으로 많이 사용되는 공격자의 행위에 대해 설명하였다.

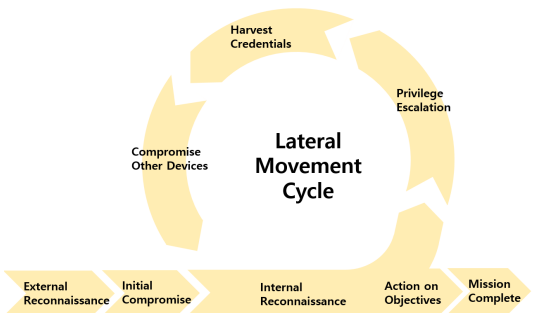

그림 1. 측면이동(Lateral Movement) 사이클
Fig. 1. Lateral Movement Cycle

이러한 측면이동은 고도화된 APT (Advanced Persistent Threat) 공격이나 표적공격 수행 시 자주 활용되며, 공격자는 오랜 기간 동안 은밀하게 공격을 수행하기 때문에 공격을 탐지하기가 쉽지 않다. 하지만 공격자는 다른 디바이스나 컴퓨터에 로그인할 때 공격자가 필요로 하는 정보를 가진 컴퓨터로 공격자가 획득 가능한 인증 정보를 활용하여 로그인하기 때문에 일반적인 사용자의 원격 로그인 패턴과는 다른, 평소에는 발생하지 않는 로그인 경로를 활용할 확률이 높다. 본 연구에서는 이러한 일반적인 사용자와는 차별되는 공격자의 특성을 고려하여 공격자의 행동을 탐지하는 방안을 제시한다. 로그인 정보를 분석하여 비정상적인 패턴을 탐지해 공격자의 행위를 찾아내는 연구는 기존에 연구되었지만, 본 연구에서는 비정상 탐지에 대한 새로운 방법을 제시하여 높은 성능을 보인다.

본 연구에서 기여한 바는 다음과 같다. a) 사용자와 컴퓨터 정보를 활용하여 로그인 경로를 식별하고, 로그인 경로의 평범도 평가 지수를 산출하는 방법을 제시하였다. b) 다중 홉 경로에서 비정상 경로의 미탐(False Negative)을 줄일 수 있는 다중 홉 경로의 평범도 평가 지수 산출 방법을 제시하였다. c) 공개된 대용량의 로그인 데이터 셋을 대상으로 본 연구에서 제시한 방법을 실험하였으며, 높은 성능 측정 결과를 얻어 본 연구에서 제시하는 방안이 실제 데이터에 대해 적용하였을 때 효과적임을 보였다.

본 장 이후의 논문 내용 구성은 다음과 같다. 2장에서는 측면이동 탐지 분석과 관련된 기존 연구에 대해 살펴보고, 3장에서는 인증 로그를 활용한 새로운 측면이동 탐지 방법에 대해 제시하고, 4장에서는 본 연구에서 제안하는 방법을 공개 데이터 셋에 적용한 실험 방법 및 실험 결과에 대해 설명하고, 5장에서는 결론과 향후 연구에 대해 기술한다.

Ⅱ. 관련 연구

측면이동 탐지에 대한 연구는 여러 가지 데이터 셋과 다양한 탐지 방법을 활용하여 지속적으로 이루어지고 있다.

Qingyun Liu 외 7명은 윈도우 이벤트 로그를 활용하여 시스템에서의 이상 행위를 탐지하는 연구를 수행하였다[3]. 이 연구에서는 침해된 컴퓨터나 계정이 발견되었을 때 공격과 관련된 다른 컴퓨터와 경로를 검사하는 포렌식 분석(Forensic Analysis)과 침해된 컴퓨터나 계정이 발견되지 않았을 때 알려지지 않은 노드 간의 측면이동을 탐지하는 일반 탐지(General Detection) 개념을 제시하였다. 기본적으로 컴퓨터 간의 원격 로그인 정보를 활용하여 공격자의 이상 행위를 탐지하였으며, 오탐을 줄이기 위해 로그온 이후 이루어지는 서비스 인스톨이나 태스크(task) 등록 및 이후에 이루어지는 프로세스 생성 이벤트 정보들을 조합하여 이상 행위를 탐지하였다. 이 연구에서는 원격 로그인 정보 활용 시 로그인하는 사용자의 정보는 배제하고 컴퓨터의 정보만을 활용했기 때문에 한 컴퓨터 내에 여러 사용자가 존재하는 환경에 적용하기에는 제한점이 있을 것으로 판단된다.

Rosa Maria Basagoiti 외 4명은 보안 이벤트 로그에서 추출할 수 있는 다양한 피처(feature)를 이용해 윈도우 서버에서 발생하는 이벤트들을 클러스터링하였다[4]. 이 실험에서 서로 다른 서버들은 서로 다른 이벤트 패턴을 보임을 확인함으로써 시스템에서 발생하는 이상 행위를 탐지할 수 있는 가능성을 보였다.

Hossein Siadati 외 1명은 정상적인 네트워크 로그인을 모델링하여 학습하고 정상 범위에서 벗어나는 악의적인 로그인을 식별하여 측면이동을 탐지하는 연구를 수행하였다[5]. 이 연구에서는 사용자, 출발지 컴퓨터(Source computer), 도착지 컴퓨터(Destination computer)의 속성으로부터 로그인 패턴을 추출하여 정상 로그인의 범주를 학습한 뒤 비정상 로그인을 분류하였으며, TPR(True Positive Rate)은 82%, FPR(False Positive Rate)은 0.3%의 성능 지표를 나타내었다.

Georgios Kaiafas 외 6명은 인증 이벤트로부터 기본적인 피처 외에 합성 피처들을 추출하고, 인증 로그를 분류하는 지도학습 기술에 대한 방식을 제시하였다[6]. 모델은 Random Forest, LogitBoost, Logistic Regression, 그리고 최종적으로 각 인증 이벤트에 대해 각 모델들의 예측 결과를 종합적으로 활용하는 다수결 원칙(Majority Voting)을 활용하여 악의적인 이벤트를 탐지하였다.

Mart Meijerink는 기업에서 수집된 운용 데이터를 수집하여 기업 IT 환경에서 발생하는 측면이동을 탐지하였다[7]. 이 연구에서는 이벤트 로그에서부터 피처들을 선정하고 클러스터링 알고리즘인 HDBSCAN (Hierarchical Density-Based Spatial Clustering for Applications with Noise)을 적용하여 클러스터에서 멀리 떨어진 아웃라이어를 탐지하여 비정상 행위 탐지를 시도하였다. 또한 차원 축소 방법인 주성분 분석(PCA, Principal Component Analysis)를 활용하여 고차원의 데이터를 데이터 간의 차이를 최대화할 수 있는 새로운 축을 가진 데이터로 변환시킨 후 통상적으로 표준편차의 3배의 차이가 나는 데이터를 outlier로 식별하여 악의적인 데이터로 분류하였다.

Harinder Pal Singh Bhasin 외 4명은 네트워크 트래픽에서 피처를 추출하고 두 집합의 교집합을 합집합으로 나누어 유사도 점수를 정량적으로 측정하는 Jaccard Similarity Coefficient 방식을 사용하여 네트워크 트래픽 간의 유사도를 측정한 후 측정된 유사도를 통해 비유사도를 정량적으로 도출하고 비정상적인 트래픽을 식별하였다[8].

Brian A. Powell은 로그인 활동 이력을 로그인 그래프로 나타내어, 컴퓨터 시스템을 노드로 표시하고 방향성 있는 에지를 로그인으로 표시하였다[9]. 이 그래프를 통해 시스템에 대한 악의적인 로그인 이벤트를 나타내는 특이한 노드를 식별하는 그래프 기반 비정상 탐지 방안을 제시하였다. 그래프 토폴로지로부터 피처들을 식별하여 NMF (Non-negative Matrix Factorization)와 PCA(Principal Component Analysis) 알고리즘을 활용해 데이터 압축 작업을 하여 변환하고 이를 다시 재변환(reconstruct)하였다. 재변환 시 발생하는 reconstruction error를 정량화하여 비정상 행위를 식별하기 위한 지표로 사용하였다.

Ⅲ. 인증 로그를 활용한 측면이동 (Lateral Movement) 탐지

측면이동을 탐지하기 위해 본 연구에서는 인증 로그를 활용하여 비정상적인 로그인을 탐지하는 방식을 사용한다. 공격자는 일반적인 사용자의 원격 로그인 패턴과는 달리 인증 정보를 획득한, 혹은 공격자가 필요로 하는 정보를 가진 컴퓨터로 평소에는 발생하지 않는 경로의 로그인을 시도하기 때문에 기존에는 발견되지 않은 로그인 패턴을 보인다. 따라서 시스템이 정상적인 상황에서 일반적인 사용자들이 시도하는 원

격 로그인 정보를 정상 데이터로 학습하고, 정상 원격 로그인 패턴을 벗어나는 리모트 로그인 행위를 비정상으로 탐지(anomaly detection)하여 공격자가 발생시키는 이상 원격 로그인을 탐지할 수 있다.

본 연구에서 제시하는 비정상 로그인 탐지 방식은 로그인 이력을 기반으로 로그인 경로의 출현 횟수를 집계하여 분석하는 빈도 기반 비정상 탐지 방식이다. 이와 비슷한 방식은 관련 연구[3]에서 제시되었지만 해당 연구에서는 사용자 정보를 배제하고 컴퓨터 노드 정보만을 사용하였다. 따라서 공격자가 인증 정보를 획득한 사용자 계정을 통해 일반 사용자들이 자주 사용하는 경로로 침투한다면 탐지를 회피할 수 있다. 본 연구에서는 컴퓨터 노드 정보에 사용자 정보까지 포함하여 경로를 식별하여 공격자가 탐지를 회피할 수 있는 가능성을 줄이며, 정확도를 높일 수 있는 방법을 제시한다. 또한 본 연구는 다중 홉 경로 비정상 탐지 방식에서도 차별성이 있는 방안을 제시하며 이는 다음 절에서 설명한다.

비정상 원격 로그인 탐지에 대한 기본 개념은 그림 2에서 설명한다. 큰 원으로 표시된 부분은 컴퓨터를 나타내며 원 안의 문자는 컴퓨터 이름을 나타낸다. 큰 원에 붙어 있는 작은 원들은 사용자 계정을 나타내며 원 안의 문자는 사용자 계정명을 나타낸다. 회색으로 표시된 화살표는 정상적인 시스템 운용 시 자주 일어나는 원격 로그인 경로를 나타내며, 빨간색으로 표시된 화살표는 정상적인 시스템 운용 시 발생하지 않는 원격 로그인이 발생한 경로를 나타낸다.

예를 들어, 그림 2에서 일반적인 운용 상황에서 컴퓨터 C1에서 U3 계정을 사용하는 사용자가 컴퓨터 C2로 U9 계정을 사용하여 원격 로그인하는 경우가 평소에 빈번하게 일어난다고 할 때 해당 원격 로그인은 정상적인 로그인 경로로 볼 수 있다. 반면, 컴퓨터 C4에서 U8 계정을 사용하는 사용자가 컴퓨터 C3으로 U6 계정을 사용하여 원격 로그인하는 경로가 일반적인 운용 시에는 발생하지 않다가 해당 원격 로그인이 어느 시점에 발생했을 때 해당 이벤트는 일반적인 상황에서 발생되지 않는 공격자의 행위로 의심할 수 있다.

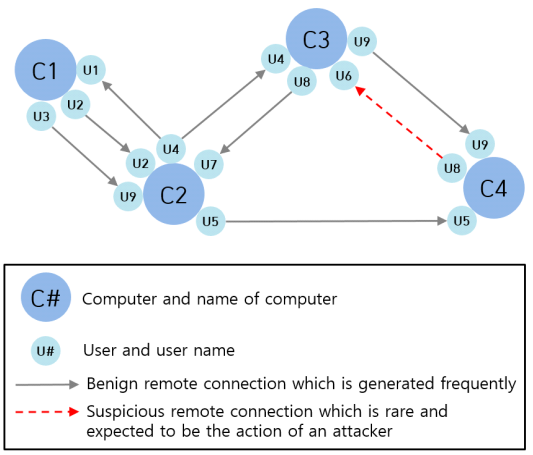

그림 2. 비정상 원격 로그인 탐지 개념 예시
Fig. 2. Concept of anomaly detection in remote logins

이를 분석하기 위해 인증 로그에서 출발지 계정(Source user), 출발지 컴퓨터(Source computer), 도착지 계정(Destination user), 도착지 컴퓨터(Destination computer), 로그인 시각(Time) 정보를 피처로 활용한다. 원격 로그인을 분석하는 방식은 크게 1-홉(1-hop) 로그인 분석과 다중 홉 로그인 분석으로 나뉘며, 다중 홉 로그인 분석은 원격 로그인한 컴퓨터에서 다시 다른 컴퓨터로 원격 로그인을 시도하는 2-홉(2-hop), 3-홉(3-hop) 등의 다중 홉 로그인 분석을 수행한다. 각 분석 방식에서 사용되는 피처(Feature)는 표 1과 같으며 분석 방법은 다음의 각 절에서 설명한다.

표 1. 비정상 탐지에 활용되는 피처(Feature)
Table 1. Features used for the anomaly detection

Features	1-hop analysis	multi-hop analysis
Source user	O	O
Source computer	O	O
Destination user	O	O
Destination computer	O	O
Time	X	O

3.1 1-홉 로그인에 대한 비정상 탐지

1-홉 로그인 분석은 사용자가 하나의 컴퓨터에서 다른 컴퓨터로 로그인하는 정보를 활용하여 분석을 수행한다. 여기서 사용자가 도착지 컴퓨터로 로그인하는 계정은 출발지 컴퓨터에서 사용하는 계정과 동일할 수도 있고 다를 수도 있다. 사용자가 컴퓨터 C_1에서 계정 U_1을 사용하다가 컴퓨터 C_2로 U_2 계정을 사용하여 로그인하는 경로의 평가 지수는 다음과 같이 계산된다.

$$N(U_1, C_1, U_2, C_2) = \frac{C(U_1, C_1, U_2, C_2)}{S} \quad (1)$$

여기서 $N(U_1, C_1, U_2, C_2)$는 U_1 계정이 컴퓨터 C_1

에서 컴퓨터 C_2로 U_2 계정을 사용하여 로그인하는 1-홉 경로의 평범도 평가 지수를 나타낸다. S는 전체 로그에서 1-홉 원격 로그인 총 횟수를 나타내며, $C(U_1, C_1, U_2, C_2)$는 전체 로그에서 컴퓨터 C_1의 계정 U_1이 컴퓨터 C_2로 U_2 계정을 사용하여 원격 로그인한 횟수를 나타낸다. 평범도 평가 지수의 산출은 정상적인 운영 환경에서 일반적인 사용자들이 사용하는 인증 로그를 통해 이루어져야 하며 정상적인 상태를 학습하는 과정으로 볼 수 있다.

평범도 평가 지수가 산출되면 이후에 발생하는 각각의 1-홉 원격 로그인이 비정상일 확률을 정량적으로 측정할 수 있다. 정상적인 운영 환경에서 사용자가 자주 사용하는 원격 로그인 경로에 대한 평범도 평가 지수는 다른 경로에 비해 상대적으로 높아질 것이며, 자주 사용되지 않는 원격 로그인 경로에 대한 평가 지수는 상대적으로 낮아질 것이다. 특히 일반적인 운영 환경에서 한 번도 출현하지 않은 원격 로그인 경로의 평범도 평가 지수는 0으로 해당 경로가 출현한다면 비정상일 확률이 매우 높다.

3.2 다중 홉 로그인에 대한 비정상 탐지

다중 홉 로그인 분석은 사용자가 출발지 컴퓨터에서 목적지 컴퓨터로 로그인하는데 중간에 1개 이상의 경유지 컴퓨터가 있는 경우에 수행한다. 사용자가 C_1 컴퓨터에서 U_1 계정으로 중간에 1개 이상의 경유지를 거쳐 최종적으로 C_k 컴퓨터에 U_k 계정으로 로그인하는 경우의 평범도 평가 지수는 다음과 같이 계산된다.

$$N(U_1, C_1, ..., U_k, C_k) = \frac{C(U_1, C_1, ..., U_k, C_k)}{S} \quad (2)$$

여기서 $N(U_1, C_1, ..., U_k, C_k)$는 U_1 계정이 컴퓨터 C_1에서 컴퓨터 C_2로 U_2 계정을 사용하여 경유하고, 최종적으로 C_k 컴퓨터로 U_k 계정을 사용하여 로그인하는 다중 홉 경로의 평범도 평가 지수를 나타낸다. 여기서 중간의 경유지는 출발지와 목적지를 제외한 k-2 개가 존재한다. S는 전체 로그에서 동일한 홉 수를 가지는 원격 로그인의 총 횟수를 나타내며, $C(U_1, C_1, ..., U_k, C_k)$는 전체 로그에서 U_1 계정이 컴퓨터 C_1에서 k-2개의 경유지를 거쳐 컴퓨터 C_k로 U_k 계정을 사용하여 원격 로그인한 횟수를 나타낸다.

다중 홉 로그인 경로에 대한 평범도 평가 지수의 산출도 1-홉 로그인 경로에 대한 평범도 평가 지수 산출과 마찬가지로 정상적인 운영 환경에서 일반적인 사용자들이 사용하는 인증 로그를 통해 이루어져야 하며, 이후 비정상 경로를 탐지하는 방식은 1-홉 로그인 경로에서 비정상적인 경로를 탐지하는 방식과 동일하다.

3.2.1 다중 홉 경로에서의 시간 제한

다중 홉 경로에서 계정 U_i가 컴퓨터 C_i에서 컴퓨터 C_{i+1}로 계정 U_{i+1}을 이용해 원격 로그인했을 때 이 경로를 E_i로, 로그인이 발생한 시각을 $T(E_i)$로 표기하고, 계정 U_{i+1}이 컴퓨터 C_{i+1}에서 컴퓨터 C_{i+2}로 계정 U_{i+2}를 이용해 원격 로그인했을 때 이 경로를 E_{i+1}로, 로그인이 발생한 시각을 $T(E_{i+1})$로 표기했을 때 $T(E_i) < T(E_{i+1})$의 관계가 성립한다.

홉 간의 시간 간격이 어느 범위에 있어야 사용자가 연속하여 원격 로그인한 경로로 볼 것인지도 시간 관련하여 고려하여야 할 사항이다. $T(E_{i+1}) - T(E_i)$이 너무 클 경우 이는 다중 홉 경로로 보기 어려우며, 분석 대상 경로 수가 너무 많아지게 되어 주어진 저장 공간과 분석 시간 내에서 처리하기가 어려워진다. 따라서 $T(E_{i+1}) - T(E_i)$가 특정 시간 이하일 때에만 다중 홉으로 간주하도록 한다. 정상적인 공격자라면 탐지를 피하기 위해 공격 대상의 네트워크에서 오래 머무는 행위를 지양할 것이며, 따라서 공격자의 공격 시간을 고려하여 홉 간의 제한 시간을 두는 것이 합리적이다. 홉 간의 제한 시간은 공격자의 공격 시간과 주어진 환경에서의 연산 시간 및 저장 공간을 고려하여 설정하여야 할 것이다. 만약 공격자가 이러한 제약 사항을 알고 고의로 홉 간의 이동 시간을 지연시킨다면 공격자는 공격 대상 네트워크에서 노출되는 시간이 늘어나게 되며, 이는 원격 로그인 세션 시간 등을 활용한 비정상 행위 탐지 등의 공격 탐지 방법 등을 함께 사용한다면 공격자의 행위를 탐지할 수 있을 것이다. 이러한 홉 간의 시간 제한 개념은 [3]에서도 소개되었고 [3]에서는 홉 간의 제한 시간을 2시간으로 설정하였으며, 본 연구에서의 실험에서는 홉 간의 제한 시간을 3시간으로 설정하였다.

3.2.2 다중 홉 경로 분석 방안 비교

다중 홉 경로의 출현 가능성을 정량적으로 측정하는 방안은 [3]에서도 제시되었다. 해당 연구에서는 각각의 1-홉 경로가 독립적이라 보고 다중 홉의 경로 출현 가능성은 다중 홉을 이루는 각각의 1-홉 경로에 대한 출현 가능성을 곱한 것으로 계산하였으며, (3)의 계산 식에 의해 계산된다.

$$p = w(v_1, e_1, v_2, e_2, ..., e_K, v_{K+1}) = \prod_{i=1}^{K} w(v_i, v_{i+1}) \quad (3)$$

$w(v_i, v_{i+1})$은 각각의 1-홉 경로의 출현 가능성을 의미한다. 이러한 방식으로 1-홉 경로 출현 가능성만을 이용하여 다중 홉에 대한 출현 가능성을 계산하면 다중 홉 경로에 대한 정보를 저장할 필요가 없어 저장 공간이 절약되고, 다중 홉 경로를 산출할 필요가 없어 연산 시간에서 유리할 수 있다. 하지만 이러한 방식은 비정상적인 다중 홉 경로를 정상적인 경로로 잘못 판단하는 경우가 발생할 수 있다.

그림 3의 다중 홉 경로 예시에서 1-홉 경로 (U1, C1, U2, C2)는 매주 월요일에 발생하고, 또 다른 1-홉 경로 (U2, C2, U3, C3)은 매주 금요일에 발생한다고 가정하면 두 가지 1-홉 경로 모두 정상적인 경로로 인식된다. 여기서 2-홉 경로 (U1, C1, U2, C2, U3, C3)의 발생 가능성을 계산한다고 할 때, 2-홉 경로를 이루는 1-홉 경로들의 발생 가능성을 곱하는 방식을 활용하면 각각의 1-홉 경로는 정상 경로이기 때문에 해당 2-홉 경로는 정상 경로로 인식된다. 하지만 해당 2-홉 경로는 정상적인 상황에서는 발생하기 힘든 경로이며, 공격자가 이러한 경로를 활용한다면 탐지가 어렵게 될 것이다. 본 연구에서 제시하는 2-홉 경로 분석 방식을 사용하면 2-홉경로 (U1, C1, U2, C2, U3, C3)는 3.2.1에서 언급한 시간 제한 때문에 사전에 인식되지 않는 경로이기 때문에 해당 경로가 발생했을 시 비정상 경로 탐지가 가능하다.

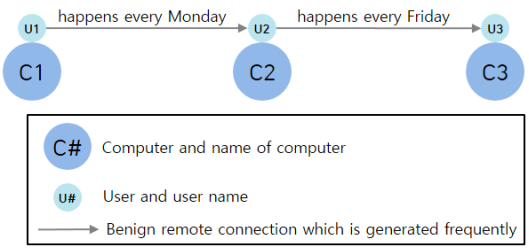

그림 3. 다중 홉 경로 예시
Fig. 3. Example of multi-hop path

Ⅳ. 실 험

실험은 윈도우10 운영체제를 사용하는 가상 데스크톱 PC를 활용하였으며, 프로세서는 인텔 Core i5-9400, 메모리는 12GB를 할당하여 실험을 수행하였다. 사용 언어는 파이썬(Python)을 사용하였으며, 대용량 데이터 처리를 위해 판다스(Pandas) 라이브러리[10]를 활용하였다.

4.1 데이터 셋

실험에서 사용한 데이터 셋은 공개 데이터 셋인 로스 알라모스 연구소(Los Alamos National Laboratory, LANL) 데이터 셋[11][12]을 활용하였으며, 다음 웹사이트에서 무료로 다운로드 가능하다. https://csr.lanl.gov/data/cyber1/

로스 알라모스 연구소 데이터 셋은 총 58일간 12,425 사용자가 17,684개의 컴퓨터에서 활동한 기록을 로그로 남긴 데이터 셋으로 본 연구에서는 인증 로그인 auth.txt와 공격자의 인증 기록을 나타낸 redteam.txt를 활용하였다. auth.txt에는 총 1,051,430,459개의 인증 기록으로 이루어져 있으며, 파일 크기는 압축 해제 시 68.3GB에 달한다. 인증 기록은 윈도우 데스크톱 컴퓨터, 서버, 액티브 디렉토리(Active Directory) 서버에서 수집된 인증 정보들을 나타내며, 이 로그에 기재된 정보는 표 2와 같다.

Time은 로그를 수집하기 시작한 시각부터 경과한 시간을 초 단위로 표시한 정보이며, 1부터 5011199의 값을 가진다. Source user@domain과 Source computer은 로그인을 시도하는 출발지 사용자명/도메인 이름과 출발지 컴퓨터를 나타내며, Destination user@domain과 Destination computer은 도착지 사용자명/도메인 이름과 도착지 컴퓨터를 나타낸다. Authentication type은 Kerberos 등 인증 방식을 나타내며, Logon type은 윈도우 로그온 유형을 나타낸다. Authentication orientation은 해당 기록이 로그온인지 로그오프인지를 나타내며, Success/failure는 성공/실패 여부를 나타낸다. 데이터 셋의 예시는 그림 4와 같다.

redteam.txt에는 총 749개의 레드팀의 활동, 즉 공격자의 로그인이 기록되어 있으며, 전체 58일 기간 중 18일에 걸쳐 일어난다. 공격자 로그인 비율은 전체 로

표 2. 로스 알라모스 인증 로그의 속성
Table 2. Attribute in Los Alamos authentication log

Attribute
Time
Source user@domain
Destination user@domain
Source computer
Destination computer
Authentication type
Logon type
Authentication orientation
Success/failure

```
1,C1241$@DOM1,SYSTEM@C1241,C1241,C1241,Negotiate,Service,LogOn,Success
1,C1250$@DOM1,C1250$@DOM1,C1250,C586,Kerberos,Network,LogOn,Success
1,C1314$@DOM1,C1314$@DOM1,C1314,C467,Kerberos,Network,LogOn,Success
1,C144$@DOM1,SYSTEM@C144,C144,C144,Negotiate,Service,LogOn,Success
1,C1444$@DOM1,C1444$@DOM1,C1444,C528,Kerberos,Network,LogOn,Success
```

그림 4. 로스 알라모스 연구소 인증 데이터 셋 예시
Fig. 4. Example of Los Alamos Laboratory authentication data set

그인 중 0.0000712%에 해당하여, 정상/공격 데이터 비율의 불균형이 심각하다. 이러한 정상/공격 데이터의 비율 차이를 줄이기 위해 데이터를 변형하는 것도 고려하였으나, 최대한 원본 데이터를 유지시키기 위해 노력하였으며, 1-홉 로그인 분석 실험은 데이터 셋 가공 없이 수행하였다.

공격자 로그인 기록에서 다중 홉을 거치는 공격자 로그인 기록은 발견되지 않았으며, 로스 알라모스 연구소 데이터 셋 생성 시 다중 홉을 이용한 공격은 수행하지 않은 것으로 보인다. 따라서, 다중 홉 로그인 분석을 위해 원본 데이터 셋의 공격자 로그인 기록에 가공의 공격자 로그인 기록을 추가하여 다중 홉 공격을 생성하여 실험하였다.

4.2 데이터 필터링

측면이동 공격을 성공적으로 수행했을 때 발생하는 원격 접속 로그는 출발지 컴퓨터와 도착지 컴퓨터가 상이하며 네트워크를 통해 성공적으로 로그인한 로그이다. 따라서 로스 알라모스 연구소 인증 데이터 셋 중 출발지 컴퓨터와 도착지 컴퓨터가 다르고, 로그온 타입(Logon type)의 속성은 네트워크를 통한 원격 접속을 의미하는 'Network'이고, 로그온/로그오프 구분(Authentication orientation) 속성은 'LogOn'이고, 성공 여부(Success/failure) 속성은 'Success'인 로그에 대해서 분석하였으며, 전체 1,051,430,459개의 로그인 기록 중 이러한 조건에 해당하는 로그인은 366,492,452개이다.

4.3 성능 측정 지표

실험 결과의 성능 측정 지표는 TPR(True Positive Rate), FPR(False Positive Rate), 정확도(Accuracy)를 사용하였다. TPR은 실제 공격자의 행위를 공격자의 행위로 예측한 비율이며 다음과 같이 계산한다.

$$TPR = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (4)$$

FPR은 정상적인 로그인을 공격자의 로그인으로 잘못 예측한 로그인의 비율이며 다음과 같이 계산한다.

$$FPR = \frac{False\ Positive}{False\ Positive + True\ Negative} \quad (5)$$

정확도는 테스트 대상 전체 로그인에서 악성은 악성으로, 정상은 정상으로 맞게 예측한 로그인 수의 비율을 뜻하며 다음과 같이 계산한다.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

여기서 TP는 True Positive, TN은 True Negative, FP는 False Positive, FN은 False Negative를 뜻한다.

4.4 1-홉 로그인 분석

4.2에서 언급한 바와 같이 분석 대상 원격 로그인 기록은 58일 간의 366,492,452개 데이터이며 공격이 발생한 날짜 중 6일간의 데이터(43,443,739개 데이터)를 테스트 셋으로 활용하고, 52일간의 데이터(323,048,713개 데이터)를 훈련 셋으로 사용하였다.

훈련 셋에서 레드 팀 로그인을 제외한 정상적인 원격 로그인 경로에 대해 출현 빈도를 집계하였으며, 이렇게 학습한 통계 데이터를 기반으로 테스트 셋에서 각각의 개별 로그인에 대해 정상/악성 여부를 판단하였다.

1-홉 원격 로그인을 악성으로 판단하는 임계값(Threshold)은 0으로 설정하였으며, 이는 훈련 셋에서 한 번도 나타나지 않았던 로그인을 공격자의 로그인 행위로 판단했음을 의미한다.

표 3에서는 테스트 셋에서 악성 로그인의 수, 정상 로그인의 수, 악성이라고 예측한 로그인 수, 정상이라고 예측한 로그인 수, 악성을 악성이라 판단한 수(TP), 정상을 악성이라 판단한 수(FP), 정상을 정상으로 판단한 수(TN), 악성을 정상이라 판단한 수(FN)를 나타낸 오차행렬(Confusion Matrix)를 나타내었다.

표 4에서는 본 연구에서 수행한 1-홉 로그인 분석 실험 결과 및 동일한 데이터 셋을 이용하여 로그인에 대해 비정상 탐지를 수행한 타 연구[13]에서 수행한 결

표 3. 1-홉 로그인 분석에 대한 실험 결과
Table 3. Experiment result for 1-hop login analysis

Category		Real state		Sum
		Malicious	Benign	
Prediction	Malicious	591 (TP)	319,564 (FP)	320,155
	Benign	8 (FN)	43,123,576 (TN)	43,123,584
Sum		599	43,443,140	43,443,739

표 4. 로스 알라모스 연구소 데이터셋에 대한 공격 탐지 실험 결과
Table 4. Experiment results of anomaly detection on Los Alamos National Laboratory dataset

		TPR(%)	FPR(%)
Anomaly Detection Results in [13]	UA	72	4.4
	FL	4	1.0
	LOF	12	9.6
	IF	9	16.9
	GL-LV	67	1.2
	GL-GV	85	0.9
1-hop login analysis in this research		98.66	0.74

과에 대해 성능 측정 지표인 TPR, FPR를 나타내었다.

[13]의 연구에서는 알려진 알고리즘과 해당 연구에서 제안한 기술을 사용하여 비정상 탐지를 수행하였다. 알려진 알고리즘은 UA(Unknown Authentication), FL(Failed Login), LOF(Local Outlier Factor), IF(Isolation Forest)를 활용하였으며, UA의 TPR은 72%, FPR은 4.4%였다. FL의 TPR은 4%이고 FPR은 1.0%였으며 LOF의 TPR은 12%이고 FPR은 9.6%였다. IF의 TPR은 9%이고 FPR은 16.9%의 성능을 보였다. 해당 연구에서 제안한 그래프 학습 기술인 Graph Learning with Local View (GL-LV)의 TPR은 67%이고 FPR은 1.2%였다. 또 다른 제안한 그래프 학습 기술인 Graph Learning with Global View (GL-GV)의 TPR은 85%이고 FPR은 0.9%였다.

본 연구에서는 악성 로그인 599개 중 591개를 악성으로 판단하여 TPR은 98.66%이며, 정상 로그인 43,443,140개 중 319,564개를 악성으로 판단하여 FPR은 0.74%였다. 전체 43,443,739개 데이터 중 정확하게 판단한 데이터 수는 43,124,167개로 정확도(Accuracy)는 99.26%였다.

4.5 2-홉 로그인 분석

2-홉 로그인 분석 실험에서도 1-홉 로그인 분석 실험에서와 동일하게 58일 간의 366,492,452개 데이터에서 공격이 발생한 날짜 중 6일간의 데이터(43,443,739개 데이터)를 테스트 셋으로 활용하고, 52일간의 데이터(323,048,713개 데이터)를 훈련 셋으로 사용하였다.

4.1에서 언급한 바와 같이 공격자 로그인 기록에서 다중 홉을 거치는 공격자 로그인 기록은 발견되지 않았으며, 따라서 다중 홉을 통한 공격 탐지에 대한 실험을 위해서 데이터 셋에 가공의 공격자 로그인 정보를 삽입하여 테스트 데이터 셋을 가공하였다. 테스트 셋에 있는 599개의 공격자 로그인 기록 중 약 5분의 1인 120개의 로그인에 대하여 2-홉 로그인 공격을 수행했다고 설정하여 기존에 존재하는 공격자 로그인의 목적지 컴퓨터를 출발지 컴퓨터로 하고 임의의 컴퓨터를 목적지 컴퓨터로 하는 공격자 로그인 120개를 테스트 셋에 추가하였다.

3.2.1에서 언급한 바와 같이 로그인과 로그인 사이의 시간 간격은 최대 3시간으로 설정하였으며, 목적지 사용자 계정과 목적지 컴퓨터를 출발지 사용자 계정과 출발지 컴퓨터로 하는 로그인이 3시간 내에 발생하면 2-홉 로그인으로 간주하였다. 이러한 방식으로 훈련 셋에서 정상적인 2-홉 원격 로그인 경로를 추출하고 각 경로에 대해 출현 빈도를 집계하였으며, 이렇게 학습한 통계 데이터를 기반으로 테스트 셋에서도 2-홉 원격 로그인 경로를 추출하고 각각의 경로에 대해 정상/악성 여부를 판단하였다.

테스트 셋에서 2-홉 로그인 경로 추출 결과 총 경로는 23,930,552개가 식별되었으며, 2-홉 원격 로그인을 악성으로 판단하는 임계값(Threshold)은 1-홉 분석에서와 마찬가지로 0으로 설정하였으며, 이는 훈련 셋에서 한 번도 발견되지 않았던 2-홉 로그인을 공격자의 로그인 행위로 판단했음을 의미한다.

표 5에서는 테스트 셋에서의 2-홉 악성 로그인의 수, 정상 로그인의 수, 악성이라고 예측한 로그인 수, 정상이라고 예측한 로그인 수, 악성을 악성이라 판단한 수(TP), 정상을 악성이라 판단한 수(FP), 정상을 정상으로 판단한 수(TN), 악성을 정상이라 판단한 수(FN)를 나타내었다.

표 6에서는 2-hop 로그인 분석 실험 결과에 대해 성능 측정 지표인 TPR, FPR, 정확도(Accuracy)를 나타내었다.

악성 2-홉 로그인 120개 중 120개를 악성으로 판단하여 TPR은 100%이며, 정상 2-홉 로그인 23,930,432개 중 86,381개를 악성으로 판단하여 FPR은 0.36%였다. 전체 23,930,552개 2-홉 로그인 중 정확하게 판단

표 5. 2-홉 로그인 분석에 대한 실험 결과
Table 5. Experiment result for 2-hop login analysis

Category		Real state		Sum
		Malicious	Benign	
Prediction	Malicious	120 (TP)	86,381 (FP)	86,501
	Benign	0 (FN)	23,844,051 (TN)	23,844,051
Sum		120	23,930,432	23,930,552

표 6. 2-홉 로그인 분석에 대한 성능 지표
Table 6. Evaluation metrics for 2-hop login analysis

TPR(True Positive Rate)	100.00 %
FPR(False Positive Rate)	0.36 %
Accuracy	99.64 %

한 데이터 수는 23,844,171개로 정확도는 99.64%였다.

V. 토의 및 결론

본 연구에서는 대용량 인증 로그를 대상으로 출발지 계정(Source user), 출발지 컴퓨터(Source computer), 도착지 계정(Destination user), 도착지 컴퓨터(Destination computer)의 피처를 활용하여 정상 상태의 로그인 경로를 학습하고 테스트 데이터에 대해 평범도 평가 지수를 산출하였다. 산출된 평범도 평가 지수를 활용하여 비정상 로그인 경로와 공격자의 행위를 구별하고 측면이동을 탐지하는 방안을 제시하였다. 또한, 공개 데이터 대상으로 본 연구에서 제안하는 방법을 적용하여 1-홉, 2-홉 경로에 대해 실험을 수행하였으며, 실험 결과 실제 상황에 적용 가능할 정도의 수준을 가진 높은 성능 지표를 보였다.

본 연구에서의 성능개선에 대한 분석평가를 위해 동일한 데이터 셋을 사용한 타 연구 결과와의 성능 지표 비교를 수행하였다. [13]의 연구에서는 해당 연구에서 제안한 그래프 학습 기술인 Graph Learning with Local View (GL-LV)와 Graph Learning with Global View (GL-GV) 및 알려진 알고리즘인 LOF(Local Outlier Factor), IF(Isolation Forest), UA(Unknown Authentication), FL(Failed Login)을 사용하여 로스 알라모스 연구소(Los Alamos National Laboratory) 데이터 셋에서 공격자 로그인을 식별하였다. [13]의 연구 결과에서 실험 결과 성능이 가장 좋은 알고리즘은 GL-GV였으며, 해당 실험 결과의 TPR은 85%, FPR은 0.9%였다. 본 연구에서는 [13]의 연구에서 활용한 데이터 셋과 동일한 로스 알라모스 연구소(Los Alamos National Laboratory) 데이터 셋 대상으로 TPR 98.66%, FPR 0.74%의 결과를 얻어 비정상 로그인을 분석하는 데 본 연구에서 제시하는 알고리즘의 성능이 우수함을 확인하였다.

본 연구에서 사용한 데이터 셋에서는 2-홉 이상의 로그인 경로를 가진 공격자 행위는 발견되지 않아 다중 홉 경로에 대한 실험은 인위적인 공격자 행위 데이터를 원본 데이터에 삽입하여 실험을 수행하였으며, 향후 다중 홉 공격자 경로를 가진 데이터 셋을 활용하여 성능을 측정해볼 필요가 있다. 또한 이번 연구에서는 동일한 수의 홉을 가진 경로들끼리 비교하여 평범도 평가 지수를 산출하였는데, 서로 상이한 홉을 가진 경로들끼리 비교하여 우선순위를 산출하거나, 1-홉 경로 분석과 다중 홉 경로 분석의 결과를 종합하여 정량적으로 지수를 산출하는 방안에 대한 연구 등도 이루어질 필요가 있다.

References

[1] https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

[2] G. M. Lee, et al., "The classification model of fileless cyber attacks," *J. KIISE*, vol. 47, no. 5, pp. 454-465, 2020.

[3] Q. Liu, et al., "Latte: Large-scale lateral movement detection," *IEEE MILCOM 2018*, Los Angeles, CA, USA, Oct. 2018.

[4] R. Basagoiti, et al., "Clustering of windows security events by means of frequent pattern mining," *Computat. Intell. Secur. Info. Syst.*, pp. 19-27, Springer, Berlin, Heidelberg, 2009.

[5] H. Siadati and N. Memon, "Detecting structurally anomalous logins within enterprise networks," in *Proc. CCS'17*, pp. 1273-1284, Dallas, TX, USA, Oct.-Nov. 2017.

[6] G. Kaiafas, et al., "Detecting malicious authentication events trustfully," *IEEE, NOMS 2018*, Taipei, Taiwan, 2018.

[7] M. M. J. Meijerink, "Anomaly-based detection of lateral movement in a Microsoft Windows environment," M.S. Thesis, University of Twente, 2019.

[8] H. P. S. Bhasin, et al., "Data center application security: Lateral movement detection of malware using behavioral models," *SMU Data Sci. Rev.*, vol. 1, no. 2, 2018.

[9] Brian A. Powell, "Detecting malicious logins as graph anomalies," *J. Info. Secur. and Appl.*, vol. 54, 2020.

[10] https://pandas.pydata.org

[11] A. D. Kent, "Comprehensive, multi-source

cybersecurity events," *Los Alamos National Laboratory,* 2015.

[12] A. D. Kent, "Cybersecurity data sources for dynamic network research," in *Dynamic Networks in Cybersecurity*, 2015.

[13] B. Bowman, et al., "Detecting lateral movement in enterprise computer networks with unsupervised graph AI," *23rd Int. Symp. RAID 2020,* 2020.

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 탐지, 사이버 위협 대응
[ORCID:0000-0003-0959-9200]

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]

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, 표적 공격 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

조 병 모 (Byoungmo Cho)

2001년 2월 : 인하대학교 컴퓨터공학과 졸업
2003년 2월 : 인하대학교 전자계산공학과 석사
2009년 9월~현재 : LIG넥스원 수석연구원

<관심분야> 사이버 보안, Modeling & Simulation
[ORCID:0000-0002-8068-6342]

김 광 수 (Kwangsoo Kim)

2009년 2월 : 아주대학교 정보 및 컴퓨터공학부 (공학사)
2017년 2월 : 아주대학교 대학원 컴퓨터공학과 (공학박사)
2017년 1월~현재 : LIG넥스원 수석연구원

<관심분야> 사이버전 훈련 기술, 네트워크 보안, 네트워크 M&S, 가상화 기술
[ORCID:0000-0003-0112-1464]

김 태 규 (Taekyu Kim)

2000년 2월 : 중앙대학교 컴퓨터공학 학사
2006년 5월 : the University of Arizona 컴퓨터공학 석사
2008년 5월 : the University of Arizona 컴퓨터공학 박사
2010년 2월~현재 : LIG넥스원 수석연구원

<관심분야> Cybersecurity Killchain and TTP (Tactics, Techniques, and procedures), 임베디드 시스템 보안, System Modeling and Simulation