

이벤트 네트워크 상관분석을 이용한 IoT 서비스에서의 침입탐지

박보석[†], 김상욱^{††}

Intrusion Detection on IoT Services using Event Network Correlation

Boseok Park[†], Sangwook Kim^{††}

ABSTRACT

As the number of internet-connected appliances and the variety of IoT services are rapidly increasing, it is hard to protect IT assets with traditional network security techniques. Most traditional network log analysis systems use rule based mechanisms to reduce the raw logs. But using predefined rules can't detect new attack patterns. So, there is a need for a mechanism to reduce congested raw logs and detect new attack patterns. This paper suggests enterprise security management for IoT services using graph and network measures. We model an event network based on a graph of interconnected logs between network devices and IoT gateways. And we suggest a network clustering algorithm that estimates the attack probability of log clusters and detects new attack patterns.

Key words: Intrusion Detection, Network Security, IoT Service, Event Correlation

1. 서 론

정보통신기술(ICT) 및 장치 기술의 급속한 발전을 통해 만물이 모바일과 인터넷을 통해 연결되어 서로 소통하는 초연결사회(Hyper Connected Society)로 진화되고 있다. 초연결사회를 구축하는 핵심 구성체가 사물통신(M2M : Machine to Machine), 사물인터넷(IoT : Internet of Things), 만물인터넷(IoE : Internet of Everything) 등이며, 이들이 ICT의 기술적 발전에 따라 인간과 사물을 둘러싼 소통의 요소들이 상호간 연결되어, 시공간의 제약을 극복하고 새로운 성장 기회와 가치 창출이 가능하게 되었다[1].

IoT가 다양한 분야에 활용됨에 따라서 IoT 서비스 보안 위협이 증가하고 있다. IoT 서비스 환경에서는 연결되는 사물의 수가 늘어나고 서비스 유형도 동적으로 변화하므로 보안 위협도 높아지고 공격 유형도 다양해진다. 대부분의 IoT 장치들은 저비용의 제약 조건 때문에 제한된 자원을 사용하여 자체적으로 많은 취약점을 가지고 있다. 장치의 인증 기능에 집중함에 따라 서비스의 유연성을 떨어뜨리는 문제가 있다.

IoT 장치에서 발생하는 로그와 기존의 네트워크 장비와 보안 장비에서 발생하는 로그 시퀀스를 분석하여야 침입에 대한 징후를 탐지할 수 있다. 이러한

※ Corresponding Author : Boseok Park, Address: (41566) Daehak-ro 80, Buk-gu, Daegu, Korea, TEL : +82-53-950-8684, FAX : +82-53-950-7452, E-mail : bouseok4u@knu.ac.kr

Receipt date : Sep. 8, 2019, Revision date : Dec. 16, 2019
Approval date : Dec. 23, 2019

[†] School of Computer Science and Engineering, Graduate School, Kyungpook National University

^{††} School of Computer Science and Engineering, Graduate School, Kyungpook National University
(E-mail : kimsw@knu.ac.kr)

※ This work was supported by the BK21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (21A20131600005)

다양한 장치로부터 로그를 수집하고 분석하기 위하여 통합 보안관제가 제안되었으며 활용되고 있다[2]. IoT 장치에서 발생하는 로그의 수가 많기 때문에 원시 로그를 축소하고 이상 징후를 탐지한다. 침입을 탐지하기 위한 상관 분석 기법에는 규칙 기반 추론(Rule Based Reasoning, RBR), 모델 기반(Model Based Reasoning, MBR), 상태 전이 그래프 기반 추론(State Transition Graph based reasoning, STG) 방법이 제시 되었다[3,4].

RBR은 작업 메모리(working memory), 규칙(rule), 추론 알고리즘(reasoning algorithm)으로 구성된다. 작업 메모리는 클러스터링을 위한 원시 로그의 집합이다. 규칙은 로그를 축약하고 이상 패턴을 탐지하기 위한 지식을 표현한 것으로 IF-THEN 형태를 가진다. 추론 알고리즘은 규칙을 작업 메모리와 매핑하기 위한 절차를 기술한다. 로그가 폭주하는 대규모의 네트워크에서 RBR은 작업 메모리와 추론 시간이 커지므로 적용하기 부적합하다.

MBR은 클러스터링 결과를 모델로 표현한다. 모델은 실제 개체인 네트워크/보안 장치이거나 논리적 개체인 네트워크 세션, 의심되는 프로세스로 구성된다. 하나의 모델은 속성, 다른 모델과의 관계, 액션으로 구성된다. 로그 클러스터링은 모델 간의 상호 협업에 의해 이루어진다. MBR은 모델 집합을 구성하기 어렵고 새로운 위협이 발생하는 현재의 IoT 서비스에서 동적인 모델을 생성하기 부적합한 단점이 있다.

STG는 토큰(token), 상태(state), 아크(arc)로 구성된다[4]. 공격 시나리오는 상태와 아크로 표현된다. 토큰이 로그에 의해 상태 전이가 발생하고 종말 상태로 토큰이 전이되면 공격이 탐지되었음을 의미한다. 공격 시나리오에 일치하는 로그를 클러스터링하여 공격을 탐지하는 기법이다. STG는 모든 로그를 분석하지 않아도 되는 장점이 있으나 공격 시나리오에 대한 STG를 구성해야 한다는 단점이 있다. MBR과 같이 새로운 취약점을 이용한 공격에 대한 탐지에 적합하지 않다.

본 논문에서는 네트워크 로그를 상관분석 하여 IoT 서비스의 침입을 탐지하는 알고리즘을 제안한다. 공격에 대한 규칙이나 상태를 정의하지 않고 이웃 노드의 평균 차수를 계산하여 동적으로 이벤트를 클러스터링 하는 알고리즘을 제안한다. 클러스터링된 장치를 연결하여 침입 경로를 추측하여 감염된

장치를 추측할 수 있다. 2장에서 IoT 서비스의 보안의 문제점과 네트워크 로그 수집을 위한 통합보안관제에 대하여 기술한다. 3장에서 네트워크 로그 수집과 IoT 서비스 보안 이벤트의 클러스터링 알고리즘에 대하여 기술한다. 4장에서 모델의 구현 및 평가를 하고 5장에서 결론을 맺고 향후 연구 방향을 기술한다.

2. IoT 서비스의 보안 관제

2.1 IoT 서비스의 보안 위협

IoT는 모든 사물을 네트워크상에서 연결하고 통합하는 기술이다. IoT 장치들의 종류가 많기 때문에 일반 IT 네트워크 장치들과 달리 표준화된 보안 솔루션을 적용하기 어렵다. IoT 장치들은 대량으로 보급하기 위하여 저비용으로 생산되어 보안 기능을 탑재하지 않는 경우가 많다. 자체 보안 기능을 탑재했다라도 사용 과정의 부주의로 비밀번호가 누설되거나 관리의 부주의로 보안 패치를 적용하지 않아 보안 침해 사고로 이어 질 가능성이 높다. 실제 세계 최대의 컴퓨터 보안 컨퍼런스이자 해킹 대회에서 온도 조절기를 해킹하여 방의 온도를 최고 온도로 한 뒤 비트코인을 요구하는 사건이 발생 하였다[6]. 다양한 IoT 서비스가 실생활에 보급되면 해킹과 랜섬웨어가 늘어날 가능성이 크다.

IoT 장치는 용도, 기능, 복잡도 그리고 운영체제에 따라 다양한 형태를 가지고 원시 로그의 크기와 복잡도가 높다. 원시 로그를 수집, 가공, 분석하여 이상 트래픽을 실시간으로 탐지하는 것이 어렵다. 또한 해킹 기법은 복합적인 기술이 적용되어 고도화, 지능화되고 있으므로 보안사고 사전에 대응체계를 수립하기 어렵다. IoT 장치에 대한 적극적인 보안 관리가 되지 않으면 취약점을 이용한 내부망 침입으로 전체 IT 자산에 위협이 된다. 따라서 다양한 유형의 유/무선 네트워크 장치, 플랫폼 및 IoT 장치가 연계되는 IoT 서비스에는 침입탐지 및 모니터링이 수행되어야 한다. 로그기록을 주기적으로 안전하게 저장/관리하여 침해사고 이후의 원인 분석이 가능해야 한다[7].

대부분의 IoT 장치는 저전력, 소형의 하드웨어 및 운영체제가 탑재되어 자체 로그 기록의 생성 및 보관이 불가능하고 자체 인증을 통한 권한 관리가 없다. 이러한 로깅 및 인증은 대부분 IoT 게이트웨이 층에서 관리하고, IoT 게이트웨이가 IoT 장치의 상태 정

치의 보안 문제를 해결하기 위하여 보안 정보를 클러스터링 엔진에 분산 저장하고 통합보안관계하는 프로세스이다. 각 모니터, 마스터, 슬레이브, 에이전트가 초기에 이벤트 네트워크를 구성하는 계층적 클러스터링 모델이다. 에이전트는 원시 로그를 수집하고 슬레이브에 전송한다. 슬레이브는 이벤트 네트워크를 구성하고 클러스터링 엔진의 축약된 이벤트를 마스터에 전송한다. 마스터는 슬레이브에서 전송된 이벤트를 다시 클러스터링하여 의심되는 네트워크를 모니터에 전송한다. ①단계에서 IoT 장치가 추가되면 IoT 게이트웨이를 통하여 등록 정보가 에이전트-슬레이브-마스터로 전송된다. ②단계에서 클러스터링 알고리즘으로 수집된 로그를 이벤트 네트워크로 구성하여 클러스터링할 것인지를 판단한다. ③단계에서 모든 네트워크 장치에 이벤트를 브로드캐스팅한다. ④단계에서 이벤트를 슬레이브들의 클러스터링 엔진에 로깅하고 클러스터링한다. ⑤단계에서 마스터는 모든 이벤트 네트워크를 업데이트하고 클러스터링한다. 마스터의 클러스터링 엔진은 의심 네트워크를 탐지하고 해당 IoT 장치에 대한 보안 정책 규칙을 갱신한다.

3. 이벤트 네트워크 모델 및 클러스터링

3.1 이벤트 네트워크 모델

수집된 원시 로그를 타임스탬프, 출발 장치, 목적 장치, 이벤트 ID로 정규화하여 이벤트 네트워크를 구성한다. Table 1은 내부 IPCam으로 스캐닝 공격이 발생하는 원시 로그를 정규화한 예이다.

이벤트 네트워크 모델에서 이벤트 e는 아래 정의(1)과 같이 출발 IP:Port를 가지는 s, 목적 IP:Port를 가지는 d, 가중치 r로 구성된다. r은 공격 가능성을 나타내는 [0,1] 사이의 값이며 초기 값은 일괄적으로 설정된다. 봇넷 위협이 많은 네트워크에서는 r 값을 1에 가깝게 초기화하고 가능성이 적은 구역은 0으로 초기화 한다. 이벤트 워드(word) w는 정의(2)와 같

이 이벤트 e의 나열이며 이벤트의 순서와 상관없다. 예를 들어 $e_1e_2e_3, e_2e_1e_3, e_3e_1e_2$ 는 동일한 이벤트 워드로 처리한다. k-이벤트 집합은 정의(3)과 같이 이벤트 워드의 크기가 k인 모든 워드의 조합이다. 예를 들어, $\Sigma = \{e_1, e_2, e_3\}$ 일 경우 2-이벤트 집합은 $\Sigma^2 = \{e_1e_2, e_1e_3, e_2e_3\}$ 이다.

event e as tuple $\langle s, d, r \rangle$ (1)

event word w is sequence of e (2)

k-event set, (3)

$\Sigma^k = \{w | w \text{ is an event word on } \Sigma, |w| = k\}$
event network k-EN, (4)

$$\Sigma^+ = \bigcup_{k=1}^{\infty} \Sigma^k = \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \dots \cup \Sigma^k$$

이벤트 네트워크 k-EN은 정의(4)와 같이 1-이벤트 집합에서 k-이벤트 집합까지의 합집합이다. 즉, 모든 네트워크/보안 장치 및 IoT 장치로부터 발생하는 이벤트의 집합이다. 봇넷 공격은 리포터 서버, 로더 서버, 다운로드 서버, 컨트롤러 서버 등과 IoT 장치에서 발생하는 로그 및 보안 장치로부터 발생하는 로그의 조합이 발생할 수 있으므로 k 값이 7 이상(서버 4 + IoT 장치 1 + 네트워크 장치 1 + 보안 장치 1)이다.

3.2 이벤트 클러스터링 알고리즘

이벤트 클러스터링 알고리즘은 이벤트 네트워크 k-EN에서 침입에 대한 네트워크를 탐지하기 위하여 봇넷과 같이 협업 공격이 발생하는 네트워크를 클러스터링한다. 네트워크를 클러스터링하기 위하여 인접 이벤트에 대한 평균 차수를 구하고 침입 확률에 반영한다. 정의 (5)는 임의의 이벤트 i의 이웃 노드 N개에 대한 인접 이벤트 평균 차수에 대한 식이다. k_i, k_j 는 각각 이벤트 i, j를 포함하는 k-EN의 k 값이며 A는 네트워크를 표현하는 인접 행렬이다. A_{paned} 는 이벤트 네트워크에서 관련 장치의 간접적인 활성화 정도를 나타낸다.

Table 1. The examples of normalized log

Timestamp	Event name	Src IP:Port	Tgt IP:Port	Dev_Type	Payloads
19-07-30 ...	accept	192.0.0.5:1355	10.10.10.10:80	CheckPoint	
19-07-30 ...	list 102 tcp	192.0.0.10:x	10.10.10.10:80	Cisco RT	
19-07-30 ...	IIS printer access	192.0.0.10:x	192.0.0.x:x	IPCam	id/pass

$$P_{aned}(k_i) = \frac{1}{k_i} \sum_{j=1}^N A_{ij} k_j$$

where k_i is the k value of k -EN and A is an adjacency matrix (5)

이벤트 네트워크에서 의심되는 네트워크를 탐지하는 알고리즘은 다음과 같다. 침입 확률 $P(k)$ 는 t 시간 동안 이웃 노드의 평균 차수 값으로 업데이트된다. 제시한 알고리즘은 이벤트 네트워크와 k 값, 탐지 시간 t , 위협 정도의 초기 값으로 사용되는 r_0 을 입력받아 각 이벤트의 침입 확률을 구한다. 이벤트 네트워크를 이웃 노드의 위협 값 r_0 로 초기화를 한다. 알고리즘의 2행~4행은 모든 이벤트의 위협 가중치 r 값을 이웃 노드의 위협 평균값으로 업데이트 한다. 5행에서 두 이벤트를 샘플링하고 6행에서 위협이 높은 값으로 업데이트한다. 이러한 과정을 시간 t 동안 반복함으로써 협업 공격으로 의심되는 위협 값을 가지는 $P(k)$ 로 업데이트된 이벤트 네트워크를 구할 수 있다.

ALGORITHM ESCA: Event Sampling & Clustering Algorithm

Input: An event network k -EN, expire time t and initial threat value r_0

Output: Probabilities P that have $P_{aned}(k_i)$ for all events i

- 1: For all events e , set $P(k_e)$ to r_0 value in event tuple
- 2: For the events i connected to EN do
- 3: $P(k_i) = \text{sum}(r_i)/k$
- 4: End
- 5: Sampling two links with probability $P(k)$ from the network: $e_1(s_1, d_1)$ and $e_2(s_2, d_2)$
- 6: Measure the degrees j_1, k_1, j_2, k_2 of nodes s_1, d_1, s_2, d_2 . Replace the two selected links with two new ones (s_1, s_2) and (d_1, d_2) with probability :
 If $(r_{j_1 j_2} \cdot r_{k_1 k_2} < r_{j_1 k_1} \cdot r_{k_2 j_2})$ $P_{e1e2}(k) = r_{j_1 k_1} \cdot r_{k_2 j_2}$
 Else $P_{e1e2}(k) = r_{j_1 j_2} \cdot r_{k_1 k_2}$
- 7: Repeat from step 2 until time t is not expired
- 8: Return the $P(k)$

4. 구현 및 평가

IoT 서비스만을 대상으로 정상 행위 300개와 비정

상 행위 30개를 데이터 셋으로 사용하였다. 정상 행위 데이터 셋은 ACCS(Australian Center for Cyber Security)의 Bot-IoT 데이터 셋[11]에서 샘플링하여 구성하였다. 비정상 행위 데이터 셋은 ACCS 데이터 셋에서 샘플링한 장치 이름을 사용하고 봇넷의 신종인 IoTroop 봇넷 공격 시나리오에서 발생하는 로그 메시지를 조합하여 구성하였다.

Fig. 3은 보안관제 시스템에서 공격 시나리오의 이벤트 네트워크를 클러스터링하고 침입 경로를 예측한 결과를 시각화한 것이다. Fig. 3 (a)는 이벤트 네트워크를 제안한 알고리즘으로 보안 장비와 IoT 장치에서 발생한 로그를 클러스터링한 예이다. Fig. 3 (b)는 공격 확률이 높은 클러스터를 연결하여 취약 경로를 시각화한 예이다.

클러스터링의 정확도는 정의(6)과 같이 정밀도(P, Precision), 재현율(R, Recall)을 사용한 F-measure 값을 구하여 평가한다. F-Measure는 분류의 정확성을 평가하기 위한 지표이다[10]. 제시한 봇넷 공격을 탐지하는 비율(F)과 오탐율(1-F)을 구하였다.

$$F = 2 \cdot \frac{P \cdot R}{P + R}, \text{ where } R = \frac{TP}{TP + FN}, P = \frac{TP}{TP + FP} \quad (6)$$

Table 2는 정상 흐름과 비정상 흐름을 샘플링할 때 각각의 데이터 셋의 다양성 정도를 정규 인자 값으로 조절하여 반복 실험한 결과이다. 로그의 다양성 정도가 작은 정규 인자 0에서 다양성이 높은 정규 인자 10일 때 평균 탐지율이 거의 균등하게 측정된다. 정규 인자(Regularization Parameter λ)가 1일 때 평균 정확성이 80%내외이므로 침입 흐름을 유효하게 탐지하였다.

클러스터링 알고리즘에 대한 클러스터링의 의미성 검증은 모듈값(modularity)을 사용한다. R의 igraph 라이브러리를 사용하여 이벤트 네트워크를 클러스터링(cluster_edge_betweenness(k EN))하고 modularity 함수로 모듈값(0.8359884)을 측정하였다.

Table 2. F-measure values on 300 normal & 30 abnormal traffic

Ensemble	Regularization Parameter λ			
	0	0.1	1	10
Mean	78.53	78.65	79.06	78.65
Min.	75.00	75.50	75.70	74.30
Max.	80.90	81.70	81.70	81.50

[6] K. Koh, S. Lee, and S. Ahn, "A Study on the Direction of Security Control of IoT Environment," *Journal of Korea Convergence Security*, Vol. 15, No. 5, pp. 53-59, 2015.

[7] D. Schnackengerg, H. Holliday, R. Smith, K. Djahandari, and D. Sterne, "Cooperative Intrusion Traceback and Response Architecture (CITRA)," *Proceeding of Defense Advanced Research Project Agency Information Survivability Conference and Exposition II*, pp. 56-68, 2001.

[8] B. Park, T. Lee, and J. Kwak, "Blockchain-Based IoT Device Authentication Scheme," *Journal of the Korea Institute of Information Security and Cryptology*, Vol. 27, No. 2, pp. 343-351, 2017.

[9] S. Sekharan and K. Kandasamy, "Profiling SIEM Tools and Correlation Engines for Security Analytics," *Proceeding of International Conference on Wireless Communications, Signal Processing and Networking*, pp. 717-721, 2017.

[10] D. Olson and D. Delen, *Advanced Data Mining Techniques*, Springer, New York, 2008.

[11] The BoT-IoT Dataset, https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/bot_iiot.php (accessed November 4, 2019).



박 보 석

1997년 경북대학교 컴퓨터과학과 학사
 1999년 경북대학교 컴퓨터과학과 석사
 2001년 경북대학교 컴퓨터과학과 박사과정 수료

2001년~2004년 어울림엘시스(주)책임연구원
 2004년~현재 경북대학교 소프트웨어교육센터 초빙교수
 관심분야 : 소셜네트워크 보안, 네트워크 보안, SIEM



김 상 욱

1979년 경북대학교 전자계산기공학과 학사
 1981년 서울대학교 컴퓨터과학 석사
 1989년 서울대학교 컴퓨터과학 박사

1982년~현재 경북대학교 IT대학 컴퓨터학부 교수
 관심분야 : 초연결 미디어, 소셜네트워크, 인간과 컴퓨터의 상호작용