Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.30, NO.1, Feb. 2020 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2020.30.1.157

보안로그 빅데이터 분석 효율성 향상을 위한 방화벽 로그 데이터 표준 포맷 제안

배 춘 석,[1†] 고 승 철[2‡]
[1,2]수원대학교(대학원생, 교수)

For Improving Security Log Big Data Analysis Efficiency, A Firewall Log Data Standard Format Proposed

Chun-sock Bae,[1†] Sung-cheol Goh[2‡]
[1,2]The University Of Suwon(Graduate Student, Professor)

요 약

최근 4차 산업혁명 도래의 기반을 제공한 빅데이터와 인공지능 기술은 산업 전반의 혁신을 견인하는 주요 동력이 되고 있다. 정보보안 영역에서도 그동안 효과적인 활용방안을 찾기 어려웠던 대규모 로그 데이터에 이러한 기술들을 적용하여 지능형 보안 체계를 개발 및 발전시키고자 노력하고 있다. 보안 인공지능 학습의 기반이 되는 보안로그 빅데이터의 품질은 곧 지능형 보안 체계의 성능을 결정짓는 중요한 입력 요소라고 할 수 있다. 하지만 다양한 제품 공급자에 따른 로그 데이터의 상이성과 복잡성은 빅데이터 전처리 과정에서 과도한 시간과 노력을 요하고 품질저하를 초래하는 문제가 있다. 본 연구에서는 다양한 방화벽 로그 데이터 포맷 관련 사례와 국내외 표준 조사를 바탕으로 데이터 수집 포맷 표준안을 제시하여 보안 로그 빅데이터를 기반으로 하는 지능형 보안 체계 발전에 기여하고자 한다.

ABSTRACT

The big data and artificial intelligence technology, which has provided the foundation for the recent 4th industrial revolution, has become a major driving force in business innovation across industries. In the field of information security, we are trying to develop and improve an intelligent security system by applying these techniques to large-scale log data, which has been difficult to find effective utilization methods before. The quality of security log big data, which is the basis of information security AI learning, is an important input factor that determines the performance of intelligent security system. However, the difference and complexity of log data by various product has a problem that requires excessive time and effort in preprocessing big data with poor data quality. In this study, we research and analyze the cases related to log data collection of various firewall. By proposing firewall log data collection format standard, we hope to contribute to the development of intelligent security systems based on security log big data.

Keywords: Logdata Standard, Firewall Logdata, Security Bigdata Analysis

Received(12. 10. 2019), Modified(02. 12. 2020), Accepted(02. 12. 2020)

† 주저자, csbae01@suwon.ac.kr
‡ 교신저자, goh5703@suwon.ac.kr(Corresponding author)

I. 서 론

1.1 연구의 배경 및 목적

빅데이터와 인공지능 기술을 활용한 비즈니스와 정보기술 서비스 전반의 최근의 혁신 성공사례들로 인해 이러한 기술들은 미래의 성장 동력으로 더욱 주목받고 있다[1]. 정보보안 영역에서도 이를 활용하여 능동적이고 지능적인 보안 대응 체계를 만들어 내기 위한 다양한 시도가 이루어지고 있다[2].

이러한 과정에서 보안 인공지능 학습의 기반이 되는 빅데이터 특성을 갖는 보안 로그 데이터의 품질은 곧 지능형 보안 체계의 성능을 결정짓는 매우 중요한 요소라고 할 수 있다. 반면에 다양한 공급자의 장비별 발생된 대용량의 로그를 통합 및 분석하는 과정에서, 수집된 로그 데이터의 상이성은 빅데이터 전처리 과정에서 과도한 시간과 노력을 요하는 문제가 있다. 전처리 과정이 미흡해 정규화 수준이 낮은 빅데이터는 머신러닝 등을 통한 보안 인공지능 학습의 신뢰 수준을 상당히 저하시키는 품질저하 원인이 된다.

본 연구에서는 방화벽 장비의 로그 수집 데이터 포맷 사례분석과 국내외 표준 조사를 통해 해당 로그 데이터 표준 포맷안을 제시하여 지능형 보안 체계 발전에 기여하고자 한다.

1.2 연구의 범위 및 방법, 선행 연구조사

본 연구의 범위는 방화벽 로그 데이터로 한다. 방화벽 로그 전송 포맷 사례와 로그 통합관리 솔루션의 로그 수집 포맷 분석을 통해 데이터 포맷 표준화 안을 제시한다.

관련 선행 연구 결과를 보면, 조석상은 '방화벽 로그를 이용한 침입탐지정보 추출방법'에서 방화벽 로그 필드별 침입탐지 가설을 수립하고 시험환경에서 공격을 수행한 후 가설에 따른 스캔공격(3가지 유형), 디도스공격(6개 유형)을 탐지할 수 있음을 제시하였으며[3] 로그 데이터의 보안 유용성만을 제시하는데 그쳤다.

박정국은 '시스템프로파일 기반 IT 위협관리 방안에 관한 연구 : IDS(Intrusion Detection System)가 발생시키는 알람이벤트를 중심으로'에서 금융결제원의 정보공유분석센터시스템 사례를 바탕으로 침입탐지 및 대응을 위한 보호대상 IT자원의 프로파일, 취약성, 위협의 상관관계분석을 핵심으로 하는 TPS(Threat Prioritization System)을 제안하였다. 또한 다양한 보안 제품들의 상호운영을 위해 로그 표준화와 전송관련 프로토콜의 제정의 시급성을 언급하는 데에만 머물렀다[4].

전병진 외는 '로그 데이터를 이용한 통합모니터링 시스템'에서 개별 보안시스템의 대용량 로그 데이터를 전송하는 모듈 개발과 수치화 방법을 논하였다[5]. 소우영은 '통합 보안 관리 시스템 표준화에 대한 연구'에서 국내 민간 보안업계 표준화 포럼인 ISTF(Internet Security Technology Forum)에서 제안한 침입차단시스템과 침입탐지시스템의 로그 형식포맷 표준화뿐만 아니라 제어 메시지 표준화를 통한 이벤트 통합보안관리 방법을 논하였다[6].

김종현 외는 '빅데이터를 활용한 사이버 보안 기술 동향'에서 진화하는 사이버 표적공격 위협에 대응하여 최근 주목받는 빅데이터 처리기술을 기반으로 다중 소스 데이터 수집 및 분석을 통한 지능형 보안 기술에 대해 논하였으며 현저히 낮은 국내 기술경쟁력을 높이기 위한 연구개발의 필요성을 주장하였다[7].

김도근 외는 '빅데이터 분석기술을 기반으로 하는 APT공격 대응 기술에 관한 방안 연구'에서 APT공격에 대응하기 위해 빅데이터 기반의 보안시스템 필요성을 제안하였으며, 특히 빅데이터 처리 과정에서 속도가 느리다는 단점에 대해 지적하였다[8]. 기타 다수 논문들에서도 빅데이터 기술을 활용한 보안 향상, 인공지능을 적용한 보안 대응 방법을 연구 제시하였다[9-15].

한편, 이준석은 '효율적인 데이터베이스 마케팅을 위한 데이터마이닝 전처리도구에 관한 연구'에서 대용량 데이터의 전처리 과정이 매우 높은 비중을 차지하며[16], 조준모는 '빅데이터의 정규화 전처리 과정이 기계학습의 성능에 미치는 영향'에서 빅데이터 분야의 양적 팽창에 따른 정규화 전처리가 기계학습의 성능에 역시 큰 영향을 준다[17]는 연구결과를 제시하였다. 이외에도 빅데이터 전처리의 중요성에 대한 다수 논문들의 연구결과가 발표되었다[18-19].

이와 같이 침입탐지 및 대응의 핵심 기반인 보안 로그의 빅데이터 분석을 통한 획기적인 보안 향상, 지능화된 대응체계로 나아가기 위해서는 다시 기본으로 되돌아가 우선 보안 로그 전처리를 효과적으로 할 수 있도록 방화벽 로그 데이터 포맷 표준 정립을 위한 방안 연구가 필요하다.

II. 방화벽 로그 데이터 수집 포맷 사례 분석

2.1 국내 주요 제품 로그 전송 데이터 분석

특정 데이터센터에 도입된 국내 4개사의 방화벽 솔루션의 로그 데이터 전송 양식에 대한 분석을 한 결과 메시지 항목 및 수, 속성 정보들이 각각 상이하여 로그 데이터의 효과적인 수집 및 분석에 저해 요소로 작용하고 있음을 확인 하였다. Table 1 ~ Table 4는 각 사별 다양한 로그 전송 내역을 보여주고 있다. 분석 및 활용을 잘 하기 위해 일부 필수적으로 여겨지는 정보의 추가 필요, 핵심 데이터 정의 및 전송방안 제공 등 보완점이 도출 되었다.

Table 1. Company A's Log Data Transfer Format

Item	Description
Number Count	20EA
Message Items	Timestamp, Type, protocol, Sip, Dip, Dport, In nic, Out nic, Nat type, Nated ip, Nated port, Sent data, sent pkt, rcvd data, rcvd pkt, Duration, State, Reason
Status	▪ Not included forwarding firewall information (IP) ▪ Difficulty in checking traffic direction without NIC number information
Improvement opportunities	▪ Need to add data elements for analysis / utilization, such as adding firewall individual identification information ▪ Define core data / add transmission method

Table 2. Company B's Log Data Transfer Format

Item	Description
Number Count	Allow Type Log 11EA, Deny Type Log 10EA
Message Items	Allow: Time, Rule ID, Src IP, Protocol, Dest IP, Dest Port, Send Byte, Rev Bytes, Duration, Interface
	Deny: Time, Rule ID, Src IP, Dest IP, Protocol, Src Port, Denied_Action, Count, Interface
Status	▪ Dependent on the response type (allowed, blocked) ▪ Not included forwarding firewall information (IP) ▪ Manage TCP connection/termination status information
Improvement opportunities	▪ Need to reduce dependency according to response type ▪ Need to add data elements for analysis / utilization, such as adding firewall individual identification information ▪ Add core data definition / transfer scheme

Table 3. Company C's Log Data Transfer Format

Item	Description
Number Count	6EA
Message Items	Date / Time, Protocol, Service, User, Path, Details
Status	▪ Difficult to utilize / analyze data by including multiple attribute information within a few message items ▪ Not included forwarding firewall information (IP) ▪ Absence of traffic confirmation information ▪ Lack of IP and Port Translation (NAT) information
Improvement opportunities	▪ Need for structural improvement for data analysis / utilization ▪ Need to add data elements for analysis / utilization, such as adding firewall individual identification information ▪ Add core data definition / transfer scheme

Table 4. Company D's Log Data Transfer Format

Item	Description
Number Count	20EA
Message Items	Type, DateTime, Origin IP, Source IP/NAT SIP, Source Port/NAT Sport, Destination IP/NAT DIP, Destination Port/NAT Dport, Protocol, Log Type, Action, Level, Product, Note, Count,

Item	Description
	Authentication or Category, Rule, Log NO, Usage, User, Interface
Status	▪ Not delivering correspondence information (Allow, Deny, etc.) ▪ Difficulty in checking TCP connection status information ▪ Undeliver session duration information ▪ Sending / receiving data amount not transmitted
Improvement opportunities	▪ Need to add data elements for analysis / utilization, such as adding firewall individual identification information ▪ Add core data definition / transfer scheme

2.2 통합 로그관리 시스템 수집 포맷 분석

통합 로그관리 시스템은 일반적으로 소스영역, 수집영역, 통합영역, 응용영역으로 구성되어 있으며 소스영역은 방화벽 장비로서 로그 수집요구에 대한 전송 서버 역할도 함께 가지고 있다.

본 사례의 특정 데이터센터에서 사용 중인 통합 로그관리 시스템의 수집 양식에 대한 현황은 Table 5와 같다. 수집을 위한 컬럼을 41개 사용 중이나 방화벽 로그 수집 데이터 비표준화로 컬럼과 속성 불일치, 데이터 누락 등이 발생하여 분석 및 활용에 적합한 데이터가 축적 되고 있지 않는 문제점이 있다.

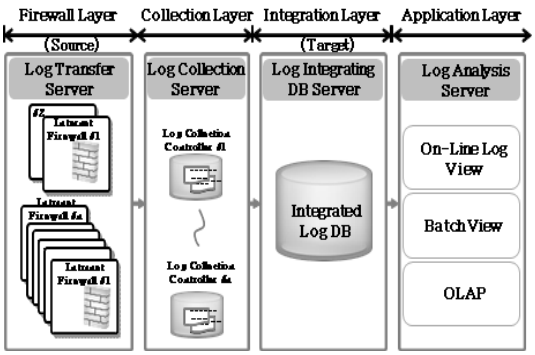

Fig. 1. Integrated Log Management System Architecture

Table 5. Collection Log Format of Integrated Log Management System

Item	Description
Number Count	41EA
Message Items	Device Address, Device Host Name, Date/Time, MessageID, Event Time, Duration, Network Port, Action, Direction, Interface, Protocol, Protocol Detail, Source Address, Source Port, Destination Address, Destination Port, Translated Source Address, Translated Source Port, Translated Destination Address, Translated Destination Port, ICMP Code, ICMP Type, Result Code, Bytes, Received Bytes, Sent Bytes, Packets, Connection ID, Policy ID, Risk Number, Filename Size, Group ID, Process ID, Counter1, Counter2, Inbound-Outbound, Level, Event Category1, Event Category2, Name Type, WIFI Channel
Status	160 total collection columns, 119 spare columns in addition to 41 items Collecting logs from non-standardized source firewalls causes column and attribute value mismatches Missing received data value
Improvement opportunities	Send / receive log transmission data standard required Need to collect data in absence of source firewall

2.3 분석 결과 시사점

로그 통합관리 시스템의 여러 사용자 그룹은 방화벽장비 자체의 장애 식별 및 조치, 비정상 트래픽 분석, 불법 IP 접근 또는 유해사이트 접근 확인, 서비스 동작 여부 확인 등 다양한 활용을 목적으로 하고 있다. 이와 대비하여 데이터 컬럼 항목의 불일치, 데이터 누락 등 결함의 원인을 제공하고 목적 달성을 저해하는 다양한 방화벽 로그 수집 데이터 포맷에 대한 표준화가 절실하다.

III. 보안 로그 관련 표준 현황 분석

3.1 해외 표준 현황

해외 표준은 국제인터넷표준화기구(IETF), 미국 표준연구소(NIST)의 보안 로그 관련 내용을 참조하였으며 현황은 Table 6과 같다.

국제인터넷표준화기구는 보안 시스템의 자체 시스템 로그 표준, 로그 송·수신 개체간의 세션 인증 프로토콜(IDXP)을 정의하였고, 미국표준연구소는 보안 로그 관리 가이드, 방화벽 관련 정책, 도입 가이드를 제공하고 있다. 반면 보안 로그 데이터 항목에 대하여는 아직까지는 글로벌하게 표준화된 내용이 없는 것으로 파악된다.

Table 6. Major Overseas Standards for Security Logs

Reference Code	Title of Standard	Description
IETF RFC 5424	The Syslog Protocol[20]	Provide standard for Syslog hierarchy and message format for event notification message delivery
IETF RFC 4767	The Intrusion Detection Exchange Protocol (IDXP)[21]	IDXP profile definition and data format definition related to authentication for session establishment as an application level protocol for data exchange between ingress and detection objects.
NIST SP800-92	Guide to Computer Security Log Management [22]	Log Management Infrastructure, Management Plan, Procedure Guide
NIST SP800-41	Guidelines on Firewalls and Firewall Policy[23]	Guide to Considera-tions for Firewall Types, Firewall Policy Enforcement, and Adoption

3.2 국내 표준 현황

국내의 경우에는 한국정보통신기술협회(TTA)의 관련 표준 내역 중 세션 정보 메시지 교환 포맷 표준에서 방화벽(침입차단시스템)등의 보안로그 교환 포맷 정의가 통합모델링언어(UML)를 활용한 클래스 다이어그램으로 상세하게 제시되어 있다. 관련 현황은 Table 7과 같다. 개발자 관점에서 보안 솔루션 개발 시 고려해야 할 모든 로그 클래스와 속성정보를 정의하다 보니, 로그 통합관리 실무에서 필요한 로그 수집 수준보다 지나치게 많고, 수집된 보안 로그 사용자의 활용 용도에 맞지 않아 적용이 어려운 문제가 있다.

Table 7. Major Domestic Standards for Security Logs

Reference Code	Title of Standard	Description
TTA K.KO-12.0279	SIMEP, Security Information Exchange Pro-tocol[24]	Domestic Security Information Message Exchange Protocol Created by TTA with Reference to RFC 4767 IDXP
TTA K.KO-12.0242	SIMEF, Session Information Message Exchange Format [25]	Define a message exchange format for sharing session information detected by heterogeneous security systems, such as intrusion prevention systems. SIMEF-Messages Consists of two subclasses, Connect, which contains session log information for NW connections under the top class, and Heartbeat, which contains the operating status of the system.
TTA K.KO-12.0256	SLMEF, System Log Message Exchange Format [26]	Collective of all kinds of exchange messages related to the system log in a normalized form of messages for status information related logs that may

Reference Code	Title of Standard	Description
		occur in this type of security system. It consists of two subclasses, Log, which contains the system information under the SLMEF top class, and Heartbeat, which contains the operating status information of the system.
TTAK.KO-12.0229	IDMEF, Intrusion Detection Message Exchange Format [27]	Defines the normalized format of detection messages for intrusions that can occur in heterogeneous security systems. The top class, IDMEF, is a generic term for all kinds of exchange messages. It consists of two classes, Heartbeat, which contains information on the operation status of the system and Alert containing detection information about network attacks.
TTAS.KO-12.0003/R1	Guidelines for Selecting Intrusion Prevention Systems for Network Operators [28]	Guide created by the TTA with reference to the NIST SP800 standard. Divides the information network risk level into high, medium, and low and provides a guide for selecting a firewall configuration type accordingly.

3.3 표준 분석 결과 시사점

보안 로그들과 관련한 국내외 표준들에서는 로그 수집을 목적으로 한 로그데이터 요건에 대해 표준화가 미진하며, 비 표준화된 보안 로그 데이터 환경으로 인해 관리상의 이슈, 기술상의 이슈들을 초래하고 보안투자의 축소, 빅데이터 및 인공지능 적용의 장애물로 작용하고 있는 실정이다. 보안 로그 수집의 기반이 되는 방화벽 로그 데이터에 대한 포맷 표준화가 우선적으로 시급하게 진행될 필요가 있다.

Table 8. Issues on Non-Standardized Log Data

Item	Issues	Negative Impacts
Management Aspects	▪ Difficult to consolidate log management ▪ Reduced productivity due to increased refining efforts ▪ Lack of utilization leads to budget waste	Reduced security investment
Technical Aspects	▪ Degradation of collected log data ▪ Difficult to extract patterns ▪ Difficult to create training / validation data sets ▪ Difficulty in automating detection, analysis and response systems	Obstacles to Applying Big Data and AI

IV. 방화벽 로그 데이터 표준 포맷안 제안

현장에서의 방화벽 운영관리 경험, 통합 로그관리 시스템 구축 및 운영 경험을 바탕으로 표준화 기본원칙(안), 핵심 데이터 요소(안), 방화벽 로그 데이터 표준 포맷(안)을 제시 한다. 1)로그포맷현황 분석, 2)표준(안) 정의, 3)표준(안) 적용방안 수립의 3단계로 연구를 진행하였으며, 표준화 기본원칙(안), 핵심 데이터 요소(안), 표준 포맷(안)은 2단계 세부연구 활동에서 도출 되었다. 제조사 엔지니어, 도입기관의 보안 전문가, 운영 및 유지보수 담당자 등 다수 전문가를 통한 의견수렴 및 결과 검토를 수행하였다.

4.1 표준화 기본 원칙(안)

로그 전송 데이터 표준화를 위해 우선적으로 Table 9와 같이 1)최우선적으로 핵심(필요) 데이터 요소 선정 및 검토가 선행되어야 한다. 2) 선정된 핵심 데이터요소는 컬럼으로 분류하여 로그포맷에 반영한다, 3) 기타 환경요소로부터 독립적인 로그포맷을 갖추어야 한다, 4) 로그포맷 표준화의 대상은 구조 및 용어이다, 5) 기관에서 관용화 된 용어는 우선하여 사용한다, 6) 사용빈도수와 분석가시성을 고

려한다, 7) 연계성을 고려한다, 8) 영문명 사용을 원칙으로 하며, 9) 영문명 전환 시, 발음식은 최대한 지양하고, 10) 기관명은 해당 기관에서 사용하는 약어(영문)를 따른다, 11) 중간에 띄어쓰기를 포함하지 않는다, 12) 특수문자 사용과 띄어쓰기는 하지 않는다("_"는 예외)와 같은 기본 원칙 12개를 정의하였다.

Table 9. Basic Principles of Data Standardization for Log Transfer

#	12 Basic Principles
1	First and foremost, the selection and review of key (required) data elements should be preceded.
2	The selected key data elements are classified into columns and reflected in the log format.
3	It should have log format independent from other environmental factors.
4	The object of log format standardization is structure and terminology.
5	The idiomatic terminology used in institutions already is used first.
6	In order to standardize the data structure, consider the frequency of use and the visibility for easy analysis.
7	If there is data exchange with other interlocking systems, consider the connectivity.
8	In principle, the Alphabet name should be used. Hangul names are not used.
9	When switching English names, avoid phonetic expressions as much as possible and use normal English(ex. CHBGY(X), Src(Source)(O).
10	The name of the agency entrusted with the security management follows the abbreviation used by the organization.
11	Do not include spaces in the middle of data. (Use "_" if you need a clear distinction)
12	Special characters and spaces are not used when assigning English names. (Except the special character "_")

4.2 핵심 로그 데이터 요소(안)

핵심 로그 데이터 요소는 총 6개로 정의하였다. 탐지 이벤트 정보, 방화벽 식별 정보, 송신지 정보, 수신지 정보, 통신정보, NAT 변환정보이며 세부 요소는 Table 10에 설명된 내용과 같다.

Table 10. Key Data Elements for Firewall Log Transfer

Key Data element	Description	Detail element
1. Detection Event Information	Traffic (packet) information detected by the firewall	▪ Event Detection Time ▪ Event response form ▪ Detection Rule Identification Code
2. Firewall Identification Information	Firewall identification information that raised the event	▪ Event Detection Firewall IP ▪ Event Detection Firewall Name
3. Sender Information	Sender information of detected packets	▪ Packet Sender IP ▪ Packet Send Port
4. Destination Information	Destination Information for Detected Packets	▪ Packet Destination IP ▪ Packet destination port
5. Communication Information	Communication information of packet accessing firewall	▪ Traffic direction ▪ protocol ▪ Session Duration ▪ Session Termination Status ▪ Sending / Receiving data amount

Key Data element	Description	Detail element
6. NAT translation information	NAT conversion information of source and destination	▪ The translated IP address of the source ▪ The translated port of the source ▪ The translated IP address of the destination ▪ The translated port of the destination

4.3 방화벽 로그 데이터 표준 포맷(안)

방화벽 로그 데이터 표준 포맷은 Table 11에 제시한 24개 데이터 요소로 정의 하였다. 해당 데이터 요소별 데이터 값의 예시 및 범위를 함께 제시하여 데이터 요건을 명확화 하였다.

Table 11. Standard Data Log Format for Firewall Log Transfer

order #	Data Element Name	Data value (example / range)
1	Event_Time	YYYY-MM-DD.HH:MM:SS.m
2	FW_IP_v4	0.0.0.0 - 255.255.255.255
3	FW_IP_v6	1111111111111111:::::::1111111111111111 - FFFF:::::::FFFF(:: is equal to the first 16 bits. Omitted for convenience of notation)
4	FW_Name	Name by unique notation for each institution
5	Action	Allow, Deny, Expire
6	Rule_ID	0 - 2147483647
7	Src_IP_v4	0.0.0.0 - 255.255.255.255
8	Src_IP_v6	1111111111111111:::::::1111111111111111 - FFFF:::::::FFFF(:: is equal to the first 16 bits. Omitted for convenience of notation)
9	Src_Port	1 - 65534
10	Dst_IP_v4	0.0.0.0 - 255.255.255.255
11	Dst_IP_v6	1111111111111111:::::::1111111111111111 - FFFF:::::::FFFF(:: is equal to the first 16 bits. Omitted for convenience of notation)
12	Dst_Port	1 - 65534
13	Protocol	ICMP, TCP, UDP, ARP, IP
14	Duration	0 - 2147483647
15	Sent_Data	0 - 2147483647
16	Rcvd_Data	0 - 2147483647
17	Entry_Nic	Internal, External, DMZ
18	Session State	Fin_Syn_SynAck_Ack, Timeout_Syn, Timeout_Syn_SynAck, Timeout_Syn_SynAck_Ack, Rst_Syn, Rst_Syn_SynAck, Rst_Syn_SynAck_Ack, Invaild_TCP, Null
19	SNAT_IP_v4	0.0.0.0 - 255.255.255.255
20	SNAT_IP_v6	1111111111111111:::::::1111111111111111 - FFFF:::::::FFFF(:: is equal to the first 16 bits. Omitted for convenience of notation)
21	SNAT_Port	1 - 65534
22	DNAT_IP_v4	0.0.0.0 - 255.255.255.255
23	DNAT_IP_v6	1111111111111111:::::::1111111111111111 - FFFF:::::::FFFF(:: is equal to the first 16 bits. Omitted for convenience of notation)
24	DNAT_Port	1 - 65534

V. 결 론

본 방화벽 로그 데이터 포맷 표준화를 통하여 다음과 같은 효과가 기대 된다.

- 방화벽 로그 수집 및 저장, 분석 효율화
- 방화벽 로그 빅데이터 분석을 위한 전처리 과정 시간 단축 및 데이터 품질 향상
- 지능형 보안대응 솔루션 개발을 위한 고품질의 학습 및 검증 데이터 세트 제공

예시로 아래 Table 12에서 이벤트 발생 시간을 제시하였다. 현행 4개사의 상이한 데이터 포맷은 제시된 표준 포맷으로 통합 및 고품질 데이터로 제공될 수 있다.

본 연구결과를 바탕으로 국내 보안 업계의 표준으로 정의할 필요가 있으며, TTA와 같은 표준화 기관의 주관으로 표준화를 위한 과제 선정, 관련 전문가 구성, 본 논문에서 제안 된 표준안 검토 및 보완, 수립된 표준에 대한 방화벽 제조업체들의 수용성 조사 및 합의, 표준 변경사항 발생 시 관리 절차 수립 등을 추가로 수행할 필요가 있다. 향후에는 각 방화벽 제조사의 로그 데이터를 표준 모델로 변환하는 시스템 설계방안을 연구하고자 한다. 또한 방화벽 외의 다른 보안장비를 대상으로 로그 데이터 관련 표준 조사 및 연구를 지속 수행하고자 한다.

Table 12. Improved Example for Standard Data Log Format(Event_Time)

Item	Description
Current (As-Is)	(Company)Log Data Sample (A)20120504`15:13:40 (B)2012-05-04 15:13:40 (C)May 04 15:13:40 2012 (D)20120504 151340
Common Elements	Year, Month, Day, Hour, Minute, Second
Proposed Format (To-Be)	YYYY-MM-DD.HH:MM:SS.m (2012-05-04.15:13:40.1)

References

〔1〕 Young-Im Cho, "Big Data Technology and Major Issues in the Smart Age," Journal of Control, Robotics and Systems, Vol. 18, No. 4, pp. 23-33, 2012

〔2〕 Jong-hyun Kim, Sun-hee Lim, Ik-kyun Kim, Hyun-sook Cho, Byung-kyu Roh, "The Trend of Cyber Security Technology Using Big Data," Electronic Communication Trend Analysis, Vol. 28, No. 3, pp. 19-29, 2013

〔3〕 Seok-Sang Cho, "How to Extract Intrusion Detection Information Using Firewall Log," Master's Thesis, Graduate School of Chungnam National University, 2015

〔4〕 Jung-kook Park, "A Study on System Profile-based IT Threat Management: Focused on Alarm Events Generated by Intrusion Detection System," Master's Thesis, Dongguk University, Graduate School of International Information, 2002

〔5〕 Byung-Jin Jeon, Deok-Byung Yun , Shin-Sung Shin, "Integrated Monitoring System Using Log Data," Journal of Convergence Information, Vol. 7, No. 1, pp. 35-42, 2017

〔6〕 Woo-young So, "A Study on Standardization of Integrated Security Management System," Journal of Convergence Security, Vol. 2, No. 2, pp. 109-121, 2002

〔7〕 Jong-Hyun Kim, Sun-Hee Lim, Ik-Kyun Kim, Hyun-Sook Cho, Byung-Gyu Roh, "The Trend of Cyber Security Technology Using Big Data," Electronic Communication Trend Analysis, Vol. 28, No. 3, pp. 19-29, 2013

〔8〕 Kim Do-Geun, Sung-Bin Pyo, Chang-Hee Kim, "A Study on the APT Attack Response Technology Based on Big Data Analysis Technology," Journal of Convergence and Knowledge Association, Vol. 4, No. 1,

pp. 29-34, 2016
〔9〕 Duk-jo Chun, Dong-kyu Park, "A Cyber Threat Prediction Analysis Model Using Big Data Technology," Journal of the Korea Information Technology Society, Vol. 12, No. 5, pp. 81-100, 2014
〔10〕 Chan-young Choi, Dae-woo Park, "Analysis of APT Attack Prediction through Big Data Analysis," Journal of The Korea Institute of Information and Communication Sciences, Vol. 20, No. 6, pp. 1129-1135, 2016
〔11〕 Bo-Min Choi, Jong-hwan Kong, Sung-sam Hong, Myeong-mook Han, "Firewall Log Analysis Using NoSQL-based MapReduce," Journal of the Korea Institute of Information Security and Cryptology, Vol. 23, No. 4, pp. 667-677, 2013
〔12〕 Hye-Geun Lee, Young-Woon Kim, Ki-Young Kim, Jong-Seok Choi, "Design of Big Data Analysis System for Harmful Information Detection Using Splunk Platform," Journal of the Institute of Information, Electronics, and Communication Technology, Vol. 11, No. 1, pp. 76-81, 2018
〔13〕 Young-Taek Oh, In-Joon Cho, "Development of Integrated Security Control Service Model based on Artificial Intelligence Technology," Journal of the Korea Contents Association, Vol.19, No.1, pp. 108-116, 2019
〔14〕 Jung-bin Yoo, Min-sik Shin, Tae-kyung Kwon, "An Analysis of Trends in Malicious Code Identification Research Using Machine Learning," Journal of the Korea Institute of Information Security and Cryptology, Vol. 12-19, 2017
〔15〕 Mun-gu Lee, Chun-seok Bae, "Next-Generation Convergence Security Framework for Intelligent Sustainable Threat," Journal of the Institute of Electronics Engineers of Korea, Vol. 50, No. 9, pp. 92-99, 2013
〔16〕 Jun-Seok Lee, "A Study on Data Mining Preprocessing Tool for Efficient Database Marketing," Journal of Digital Convergence, Vol. 12, No. 11, pp. 257-264, 2014
〔17〕 Joon-Mo Cho, "The Effect of Normalized Preprocessing of Big Data on the Performance of Machine Learning," Journal of the KIECS, Vol. 14, No. 3, pp. 547-552, 2019
〔18〕 Seong-Hae Jeon, "Big Data Preprocessing Using Statistical Text Mining," Journal of Korean Institute of Intelligent Systems, Vol. 25, No. 5, pp. 470-476, 2015
〔19〕 Dong-hyun Kim, Seung-eon Yoo, Byung-jun Lee, Kyung-tae Kim, Hee-yong Yun, "Data Preprocessing for Efficient Machine Learning," Proceedings of the Korean Society of Computer Information, Winter Conference, Vol. 27, No. 1, pp. 48-50, 2019
〔20〕 R. Gerhards, "RFC 5424 The Syslog Protocol," IETF Request For Comments, 2009
〔21〕 B. Feinstein, G. Matthews, "RFC 4767 The Intrusion Detection Exchange Protocol (IDXP)," IETF Request For Comments, 2007
〔22〕 Karen Kent, Murugiah Souppaya, "NIST 800-92 Guide to Computer Security Log Management," National Institute of Standards and Technology Special Publication 800, 2006
〔23〕 Karen Scarfone, Paul Hoffman, "NIST 800-41 Guidelines on Firewalls and Firewall Policy," National Institute of Standards and Technology Special

Publication 800, 2009
〔24〕 Telecommunications Technology Association, "TTAK.KO-12.0242 Session Information Message Exchange Format," Information and Communication Organization Standard (Korean Standard), 2014
〔25〕 Telecommunications Technology Association, "TTAK.KO-12.0279 Security Information Message Exchange Protocol," Korea Communications Standards, 2015
〔26〕 Telecommunications Technology Association, "TTAK.KO-12.0256 System Information Message Exchange Format for Security Control," Information and Communication Organization Standard (Korean Standard), 2014
〔27〕 Telecommunications Technology Association, "TTAK.KO-12.0229 Extended Intrusion Detection Message Exchange Format," Korea Communications Commission, 2013
〔28〕 Telecommunications Technology Association, "TTAK.KO-12.0003 / R1 Guidelines for Selecting Intrusion Prevention System for Network Operators," Information and Communication Organization Standard (Korean Standard), 2006

〈저자소개〉

배 춘 석 (Chun-sock Bae) 정회원
1993년 2월: 전남대학교 경영학과 학사
2017년 2월: 건국대학교 정보보안학과 석사
2018년 3월~현재: 수원대학교 컴퓨터학과 박사과정,
1993년 4월~현재: ㈜LG CNS 클라우드아키텍처팀 재직, 정보관리기술사(2008)
〈관심분야〉 정보보호, 데이터센터 구축 및 운영, 클라우드컴퓨팅

고 승 철 (Sung-cheol Goh) 종신회원
1981년 2월: 연세대학교 수학과 학사
1983년 2월: 연세대학교 수학과 석사
1992년 8월: 포항공과대학교 수학과 박사
2011년 9월~현재: 수원대학교 정보보호학과 교수
〈관심분야〉 정보보호, 국방사이버보안, 암호학, 클라우드컴퓨팅