Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.30, NO.4, Aug. 2020 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2020.30.4.693

FDANT-PCSV: Parallel Coordinates 및 Sankey 시각화를 이용한 신속한 이상 트래픽 탐지

한 기 훈,[1†] 김 휘 강[2‡]
[1,2]고려대학교(대학원생, 교수)

FDANT-PCSV: Fast Detection of Abnormal Network Traffic Using Parallel Coordinates and Sankey Visualization

Ki hun Han,[1†] Huy Kang Kim[2‡]
[1,2]Korea University(Graduate student, Professor)

요 약

기업의 네트워크가 대규모화되고 보안시스템 수가 많아짐에 따라 엄청난 양의 보안시스템 이벤트로부터 이상 트래픽을 신속하게 탐지하기란 쉽지 않다. 본 논문에서는 방화벽 등 정보보호 시스템의 보안이벤트를 실시간 탐지하고 분석할 수 있는 트래픽 시각화 분석시스템(FDANT-PCSV)를 제안한다. FDANT-PCSV는 보안이벤트 중 5가지 인자(출발지 IP, 목적지 IP, 목적지 포트, 패킷 길이, 처리상태)를 이용한 Parallel Coordinates 시각화와 4가지 인자(출발지 IP, 목적지 IP, 이벤트 수, 데이터 크기)를 이용한 Sankey 시각화로 구성된다. 또한, 빅데이터 기반 SIEM을 이용하기 때문에 인터넷 및 인트라넷에서 발생하는 네트워크 공격과 네트워크 장애 트래픽을 실시간 탐지할 수 있다. FDANT-PCSV는 사이버 보안 관제요원과 네트워크 관리자가 네트워크 이상 트래픽을 빠르고 쉽게 탐지하여 네트워크 위협에 신속히 대응할 수 있도록 해준다.

ABSTRACT

As a company's network structure is getting bigger and the number of security system is increasing, it is not easy to quickly detect abnormal traffic from huge amounts of security system events. In this paper, We propose traffic visualization analysis system(FDANT-PCSV) that can detect and analyze security events of information security systems such as firewalls in real time. FDANT-PCSV consists of Parallel Coordinates visualization using five factors(source IP, destination IP, destination port, packet length, processing status) and Sankey visualization using four factors(source IP, destination IP, number of events, data size) among security events. In addition, the use of big data-based SIEM enables real-time detection of network attacks and network failure traffic from the internet and intranet. FDANT-PCSV enables cyber security officers and network administrators to quickly and easily detect network abnormal traffic and respond quickly to network threats.

Keywords: abnormal network traffic, parallel coordinates, sankey, detection, visualization

I. 서 론

1.1 네트워크 트래픽 시각화의 필요성

과거와는 달리 현재의 기업 네트워크 환경은 점점 대규모화되고 복잡해지고 있다. 한 대의 보안장비에서 하루 발생하는 트래픽이 수십~수백 기가바이트에 이르기도 한다. 방화벽, IDS, IPS 등의 텍스트 로그를 일일이 분석하여 이상(비정상) 정보를 탐지하기는 쉽지 않다.

Received(04. 17. 2020), Modified(06. 03. 2020), Accepted(06. 19. 2020)

† 주저자, hbora10@korea.ac.kr
‡ 교신저자, cenda@korea.ac.kr(Corresponding author)

분석가의 신속하고 효율적인 네트워크 분석을 위해 분석시스템을 자동화해야 한다. 또한, 분석가는 주요 네트워크 시나리오를 완벽하게 인지하고 있어야 한다. 하지만 이를 위해서는 분석을 돕는 도구가 필수적이다. 이 도구는 보안장비의 방대한 이벤트로부터 이상 트래픽을 신속하게 탐지하도록 도와준다. 정보 시각화는 고차원 데이터에 대한 효율적인 분석 방법을 제시한다. 인간의 두뇌는 텍스트보다 이미지를 처리하는 능력이 더 우수하기 때문이다[1][2].

C.Wu 등[1]은 시각적 데이터 분석은 가장 복잡한 데이터로부터 패턴, 추세, 구조, 예외사항을 인식하는 데 도움을 준다고 했다. 빅데이터를 실시간으로 처리하는 SIEM을 이용하고 적절한 시각화 도구를 사용한다면 대규모 네트워크에서 발생하는 이상 트래픽을 효율적으로 탐지할 수 있다.

1.2 이상 트래픽(abnormal traffic) 분류

H. Choi 등[7]의 PCAV는 외부 인터넷 환경에서 DDoS, Worm, 스캔 공격 등과 같은 이상 트래픽을 탐지할 수 있다. V.T. Guimarães 등[8]은 이상 트래픽을 네트워크 관리(network management) 관점과 사이버 보안(cyber security) 관점으로 분류하였다. 본 논문에서는 V.T. Guimarães 등[8]의 분류체계를 사용하여 외부 인터넷을 통해 유입되는 공격성 트래픽뿐만 아니라 인트라넷(intranet)에서 발생하는 네트워크 장애 트래픽도 신속하게 탐지하는 방법을 제안한다.

1.2.1 네트워크 관리 관점의 이상 트래픽

인트라넷 접속이 매우 느려짐을 인지한 네트워크 관리자가 방화벽 로그를 분석해 보지만 엄청난 양의 로그로 인해 장애 원인을 신속하게 찾지 못한다. 분석한 지 몇 주 후에야, 회사 내 업무용 PC에서 불법 소프트웨어를 점검하는 서버로 매우 큰 패킷(초당 60개 이상)이 전송됨을 확인하였다. 회사 내 수천 대의 PC에서 발생한 트래픽이 네트워크 속도 저하의 원인이었다. 네트워크 장애를 일으키는 트래픽을 네트워크 관리 관점의 이상 트래픽(abnormal network traffic)으로 분류할 수 있다.

1.2.2 사이버 보안 관점의 이상 트래픽

사이버 보안 관제요원(cyber security operator)은 인터넷 공격을 인지하고 보안장비의 텍스트 이벤트로부터 공격 정보를 탐지하고자 한다. 하지만 이벤트 양이 너무 많아 공격에 해당하는 트래픽을 탐지하는 데 많은 시간 소요된다. 인터넷이나 인트라넷 상에서 사이버 공격을 통해 발생하는 네트워크 트래픽을 사이버 보안 관점의 이상 트래픽(abnormal security traffic)으로 분류할 수 있다.

1.3 연구의 방법과 구성

본 논문은 빅데이터 기반 SIEM을 이용하여 실시간 수집한 보안장비 이벤트로부터 이상 트래픽을 신속하게 탐지하여 분석하는 시스템을 제안한다. 2장에서는 정보 시각화 관련 연구를 살펴보겠다. 3장에서는 기존 네트워크(L4 스위치 환경) 트래픽 시각화에서 발생 가능한 오탐과 미탐을 줄이는 네트워크 트래픽 시각화 분석시스템(FDANT-PCSV)을 제시하겠다. 4장에서는 FDANT-PCSV를 이용한 신속한 이상 트래픽 탐지 사례를 살펴보겠다. 마지막으로 5장에서는 결론 및 향후 연구 방향을 제시하고자 한다.

II. 관련 연구

2.1 FRuVATS

김형석 등[3]은 충돌로 인한 복잡한 방화벽 정책(rule set)을 제어할 수 있도록 조건을 검색하는 시각화 모델(FRuVATS)을 제안했다. 방화벽 관리자는 정책을 설정(패킷의 허용/차단 상태 결정)할 때 여러 가지 조건을 고려한다. 하지만 시간이 흐를수록 방화벽 규칙은 복잡해지고 잘못된 정책을 설정하는 경우가 발생한다. 이로 인해 방화벽 관리 비용(시간 및 노력)이 증가하게 된다.

FRuVATS는 충돌을 분석하고 활성 영역(패킷이

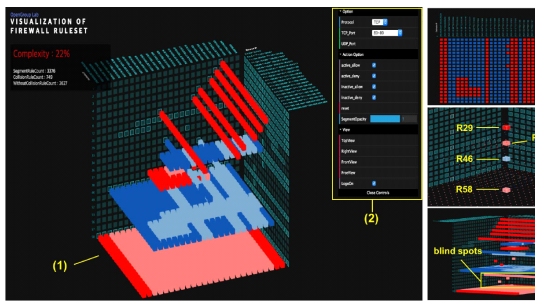

Fig. 1. 3D view of FRuVATS visualization[3]

정상 제어되는 영역) 및 비활성 영역(방화벽 정책이 충돌되어 무시되는 영역)을 자동으로 식별하여 방화벽 관리자가 정책을 정확하고 신속하게 설정하도록 돕는다.

2.2 Honeycomb

박재범 등[4]은 네트워크 보안 위협을 식별하는데 도움을 주는 벌집 구조 시각화를 제안하였다. 효과적인 위협상황 식별을 위해 정육각형을 이용한다. 허니컴 벌집 구조 시각화는 ① 선의 생략 ②목적지 중심의 시각화 ③ 표출정보 최소화 ④ 위협 별로 색 사용 등의 설계 특징이 있다. 모의 테스트(정보 유출 시도, FTP 연결, 원격 로그인 등)를 수행하여 보안 위협 식별·인지에 대한 효과성을 보여주었다.

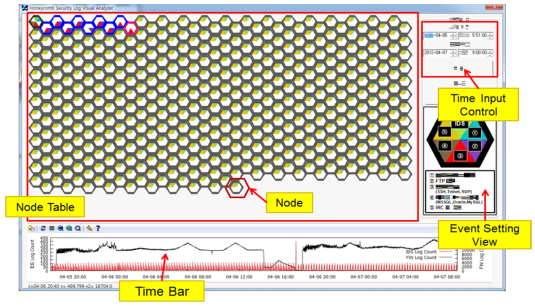

Fig. 2. A Main view of honeycomb structure visualization[4]

2.3 RGB Palette

이동건 등[5]은 RGB 색상 값으로 보안 위협상황을 직관적으로 파악할 수 있는 RGB Palette 시각화 시스템을 제안하였다. RGB Palette는 방화벽, IDS 로그 중 출발지 IP, 목적지 IP, 목적지 포트, 로그 발생량 정보를 이용한다. 대시보드에 표출되는 원은 RGB Palette의 핵심이다. 원의 가장자리는 3부분으로 나뉘며 각각 RGB 색상으로 표현된다. 중앙의 원은 로그 발생량이 많아지면 크기가 커진다. 색상은 출발지 IP, 목적지 IP, 목적지 포트 수에 따라 달라진다. VAST Challenge 2012 데이터 세트를 이용하여 포트 스캔, 호스트 스캔, DDoS 공격 등에 대한 효과성을 확인하는 테스트를 진행하였다.

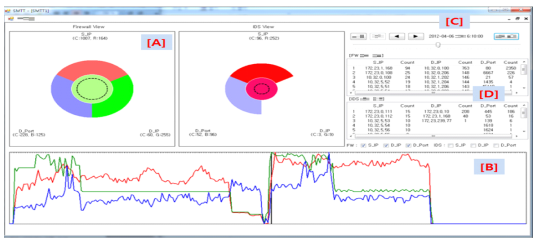

Fig. 3. RGB palette visualization dashboard[5]

2.4 Time-Tunnel

Y. Okada[6]는 3D 시각화 도구인 PCCT(Parallel Coordinates version of time-tunnel)를 제안하였다. PCCT는 수많은 다차원 데이터 레코드를 3D 공간에서 개별 차트로 시각화한다. Y. Okada[6]는 2Dto2D 시각화 기능과 Fig. 4.와 같은 SPC(Spline Parallel Coordinates)라는 새로운 PCCT 기능을 제시했다.

PCCT는 다차원 데이터 레코드를 중첩 라인으로 표시한다. 서로 다른 데이터 레코드에 대하여 인접한 인자들의 속성값이 같으면 레코드를 분류하기 어렵다. SPC를 사용할 경우 Fig. 4.와 같이 중첩현상을 막을 수 있다. PCCT는 네트워크 IP 패킷 데이터를 시각화하는 데 매우 유용하다.

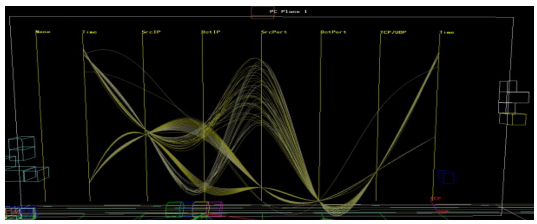

Fig. 4. Spline parallel coordinates representation as one of the new features of PCCT[6]

2.5 PCAV

H. Choi 등[7]은 Parallel Coordinates를 사용하여 인터넷 공격(DDoS, worm, 스캔 공격 등)을 실시간 탐지하는 시각화 시스템인 PCAV(parallel coordinates attack visualization)을 제시하였다. PCAV는 네트워크 스위치의 패킷 정보(출발지 IP, 목적지 IP, 목적지 포트, 평균 패킷 길이)를 이용하여 네트워크 트래픽을 공격유형에 따라 Fig. 5.과 같은 패턴으로 시각화한다.

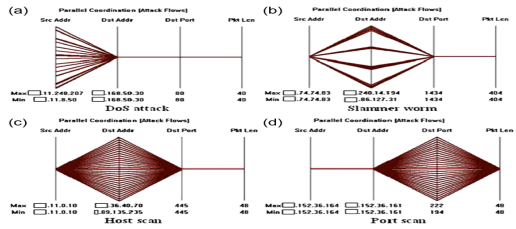

Fig. 5. Real-life attack graphs using PCAV[7]

III. 트래픽 시각화 분석시스템(FDANT-PCSV)

3.1 스위치 환경의 Netflow 시각화 시스템의 한계

본 논문에서 제안하는 FDANT-PCSV(Fast Detection of Abnormal Network Traffic using Parallel Coordinates and Sankey Visualization) 시각화 시스템은 H. Choi 등[7]의 PCAV 시각화 시스템을 기초로 설계되었다. PCAV 시각화 시스템은 스위치(switch) 환경의 Netflow 패킷 속성을 이용한다.

Table 1. Typical attributes of flow-based data such as Netflow using a network switch

Attribute	Description
Src IP	Source IP Address
Src Port	Source Port
Dest IP	Destination IP Address
Dest Port	Destination Port
Proto	Transport Protocol
Duration	Duration of the flow
Bytes	Number of transmitted bytes
Packets	Number of transmitted packets

스위치 환경의 Netflow 패킷은 Table 1.과 같은 흐름(flow) 속성 정보를 갖는다. 이 중 PCAV는 4가지 속성값(출발지 IP, 목적지 IP, 목적지 포트, 패킷 길이)을 이용하여 트래픽 시각화 패턴을 만든다. PCAV는 네트워크 공격 트래픽 탐지 시 오탐(false positive)과 미탐(false negative)이 발생할 수 있다.

3.1.1 오탐 발생

PCAV는 네트워크 스위치 환경에서 공격패턴 형성에 따라 이상 트래픽 여부를 판단한다. PCAV는 4가지 속성 정보(출발지 IP, 목적지 IP, 목적지 포트, 평균 패킷 길이)를 이용한다. 외부로부터 회사 내부로 유입되는 포트 스캔은 공격패턴을 형성하기 때문에 PCAV의 이상 트래픽 판단은 타당하다. 하지만 네트워크 보안 점검을 위해 네트워크 관리자에 의해 허용된 포트 스캔은 정상 트래픽으로 판단해야 한다. 스위치에 기록되는 이벤트에는 패킷에 대한 허용/차단 여부를 알려주는 상태 정보가 없어서 이러한 오탐이 발생한다.

정보보호 시스템(IDS, IPS, DDoS 등)은 사전에 정의된 탐지/차단 정책(rule)에 의해 패킷의 허용/차단 여부를 이벤트에 기록한다. 이상 트래픽 탐지 시 정보보호 시스템 이벤트를 이용하고 매개변수 action(패킷의 허용/차단 정보를 지시)을 추가할 경우 PCAV에서 발생 가능한 오탐을 줄일 수 있다.

3.1.2 미탐 발생

PCAV에서 정상 트래픽 형태를 보이지만, 나중에 이상 트래픽으로 확인된 경우에 미탐이 발생한다. 예를 들어, 특정 클라이언트(출발지 IP)가 특정 서버(목적지 IP)의 특정 목적지 포트로 매우 큰 다량의 패킷을 지속해서 보내는 경우, PCAV의 Parallel Coordinates 다이어그램에서는 공격패턴이 형성되지 않아 정상 트래픽으로 판단하게 된다.

하지만 이러한 과부하 트래픽은 네트워크 및 보안 장비의 성능을 크게 저하하는 네트워크 관리 관점의 이상 트래픽에 해당한다. Sankey 다이어그램 등 트래픽 흐름 정보 시각화를 이용하면 이러한 이상 트래픽에 대한 미탐을 줄일 수 있다.

3.2 시각화 시스템 설계 방향

FDANT-PCSV는 H. Choi 등[7]이 제안한 PCAV의 제한된 용도를 확장하고 발생 가능한 미탐 및 오탐을 개선할 목적으로 설계되었다.

FDANT-PCSV의 설계 방향은 다음과 같다. 첫째, 사이버 보안 관점의 외부 공격 트래픽뿐만 아니라, 내부 네트워크 보안을 위협하는 이상 트래픽을 탐지하는 것이다. 둘째, SIEM의 가용성을 극대화하는 것이다. 보통 상용 SIEM은 일일 허용 트래픽양을 라이선스로 사용한다. 트래픽 증가로 라이선스가 초과하면 SIEM의 실시간 이벤트 수집이 제한된다. 셋째, 시각화 결과의 오탐과 미탐을 줄이는 것이다.

3.3 시각화 시스템 구성

3.3.1 네트워크 트래픽 시각화 프로세스

Fig. 6.은 빅데이터 기반의 SIEM을 이용하여 네트워크 이상 트래픽을 탐지하는 FDANT-PCSV 프로세스를 도식화한 것이다. 정보보호 설비의 이벤트 로그는 시스로그(syslog) 형태 등으로 SIEM에 실시간 전송된다.

Fig. 7.에서 FDANT-PCSV 시각화 시스템은 입력 데이터로 방화벽, IPS, IDS의 이벤트를 사용한다. SIEM은 실시간 보안장비의 이벤트를 구문분석(parsing)하여 변환된 정보를 디스크에 저장한다. FDANT-PCSV는 구문분석 이벤트로부터 의미 있는 매개변수(출발지 IP, 목적지 IP, 목적지 포트, 패킷 길이, 허용 및 차단 여부, 이벤트 수)를 추출하고 실시간 시각화한다.

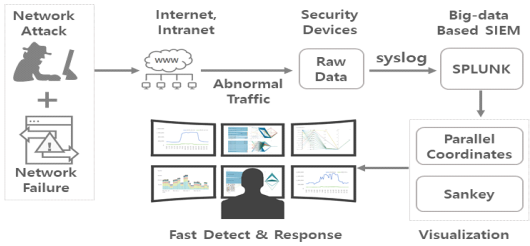

Fig. 6. Data processing and visualization using Big-data based SIEM(SPLUNK)

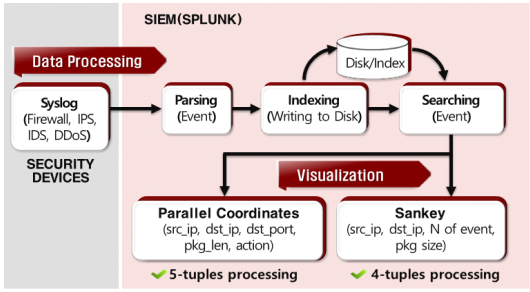

Fig. 7. Design structure of FDANT-PCSV visualization system

3.3.2 Parallel Coordinates 시각화

Parallel Coordinates 다이어그램은 다차원 수치 데이터 세트를 이해하는 데 매우 유용한 시각화 도구이다. n 차원 데이터의 특정한 기하학적 특성을 2D 패턴으로 쉽게 변환할 수 있다. 하지만 데이터 항목 수가 많아지게 되면 데이터의 특징을 선별하기 어려운 단점이 있다. 일반적으로 이 기술은 적당한 수의 차원과 데이터 레코드가 많은 경우 특정 패턴을 찾아 인식하고 상관관계를 신속하게 추정할 수 있다.

H. Choi 등[7]의 PCAV는 4-tuple(source IP, destination IP, destination port, packet length) 이용한다. 반면, FDANT-PCSV는 시각화 오탐을 줄이기 위해 Fig. 8.와 같이 4-tuple에 "action"을 추가한 5-tuple을 이용한다. 추가된 매개변수 "action"은 방화벽, IPS의 해당 패킷에 대한 사전 허용/차단 여부를 알려주는 지표이다.

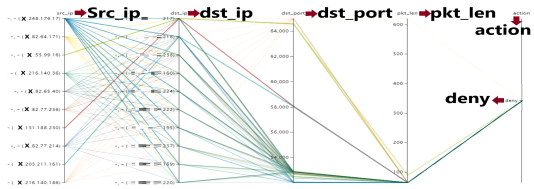

Fig. 8. Parallel coordinates security visualization of FDANT-PCSV

3.3.3 Sankey 시각화

Fig. 9.은 FDANT-PCSV 시각화 시스템에 사용된 Sankey 다이어그램이다. Sankey 다이어그램은 너비가 흐름양에 비례하는 흐름도이다. Sankey 다이어그램을 통해 네트워크에서 발생하는 이벤트 수와 트래픽 발생량을 파악할 수 있다. 이를 이용하여

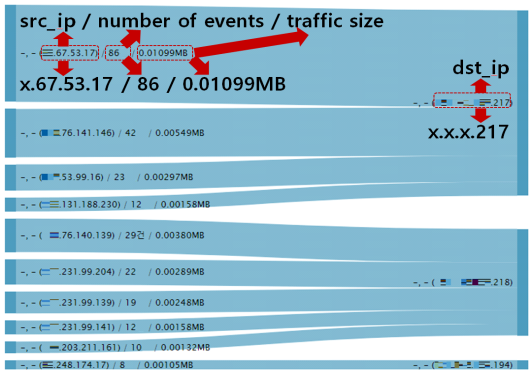

Fig. 9. Sankey visualization showing number of events and traffic size of firewall

네트워크 과부하를 초래하는 트래픽 정보 등을 신속하게 탐지하여 이상 트래픽 여부를 판단할 수 있다. H. Choi 등[7]의 PCAV는 Parallel Coordinates 다이어그램에 특정 패턴을 형성한 경우에만 공격으로 분류한다. 하지만 특정 패턴이 없어도 이벤트 수와 트래픽양이 과다하다면 이상 트래픽으로 보아야 한다. FDANT-PCSV는 Sankey 다이어그램을 사용함으로써 시각화에 따른 미탐을 줄인다. FDANT-PCSV는 Parallel Coordinates 다이어그램과 Sankey 다이어그램을 동시에 사용함으로써 이상 트래픽을 매우 효율적으로 탐지한다.

3.3.4 시각화 대시보드

Fig. 10.은 FDANT-PCSV 시각화 시스템의 대시보드이다. 대시보드는 빅데이터 기반의 플랫폼인 SPLUNK를 이용하여 구현하였다. 화면의 중앙은 Parallel Coordinates 다이어그램과 Sankey 다이어그램으로 양분된다. 사용자는 대시보드 상단의 대화식 인터페이스를 통하여 빠르고 편리하게 특정 트래픽을 검색할 수 있다. FDANT-PCSV에서 사용되는 상세 검색항목은 Table 2.와 같다.

Table 2. Information of FDANT-PCSV dashboard using big data based SIEM(SPLUNK)

Topic	Description
Time zone	Select dates, times and ranges
Index	Index name used by SPLUNK
Number of View Line	The number of flows to show
Source IP or Dest IP Range	outer: IP not used by the corporation, inner: IP used by the corporation, total: IP including both internal and external
5-tuples	action, source IP, destination IP, destination port range, packet length range, action
Include Source Port	Choose whether or not to show the source port

3.3.5 FDANT-PCSV의 상관관계 분석

FDANT-PCSV 시각화 시스템의 Parallel Coordinates 및 Sankey 다이어그램의 상관관계를 살펴보자. Fig. 11.의 Parallel Coordinates 다이어그램을 살펴보면 출발지 IP(a.b.f.167, a.b.f.141)에서 목적지 IP(a.b.x.99, a.b.x.100)로 포트 스캔 형태의 트래픽이 형성되어 있다.

Fig. 11.의 Sankey 다이어그램의 탐지 결과를 살펴보면 출발지 IP(a.b.f.167, a.b.f.141)에서 목적지 IP(a.b.x.99)로 24시간 동안 각각 약 1.75GB,

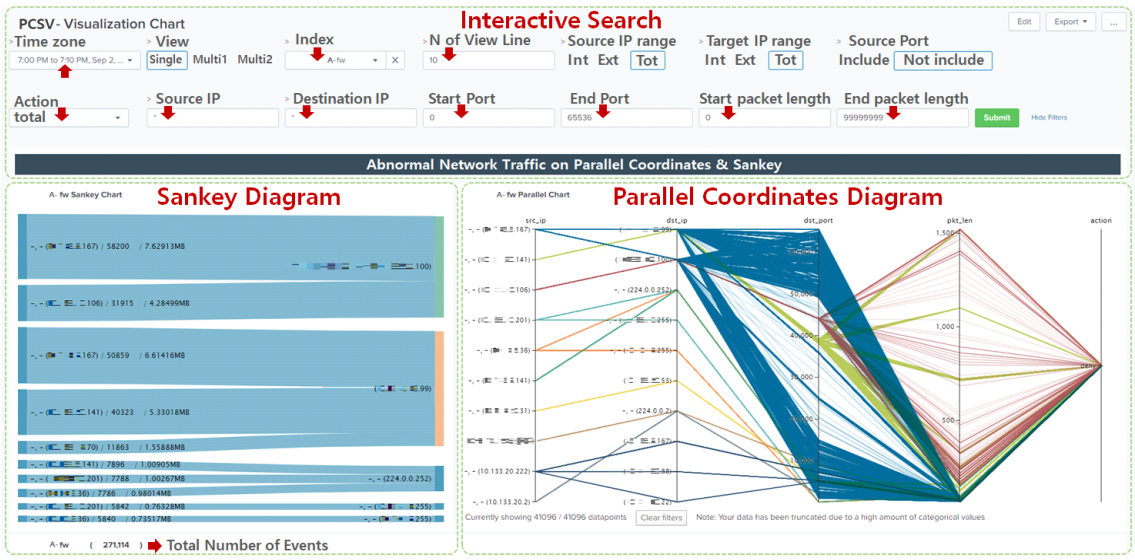

Fig. 10. Interactive dashboard of FDANT-PCSV

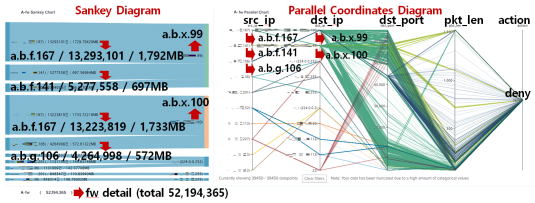

Fig. 11. FDANT-PCSV correlation

0.68GB의 트래픽양이 측정된다. 보통 포트 스캔은 하루 트래픽양이 수십 MB 정도만 측정된다는 점을 고려할 때, 이 트래픽은 네트워크에 과부하를 초래하고 있음을 확인할 수 있다.

FDANT-PCSV 시각화 시스템은 Parallel Coordinates 다이어그램과 Sankey 다이어그램을 동시에 분석함으로써 이상 트래픽에 대한 보다 정확한 결과 분석이 가능하다.

3.3.6 PCAV와 FDANT-PCSV 특성 비교

Table 3.은 본 논문의 관련 연구인 PCAV 시각화 시스템과 제안시스템 FDANT-PCSV의 특징을 비교한 것이다. PCAV의 경우 인터넷 환경의 네트워크 공격을 탐지하는 것이고 FDANT-PCSV는 인터넷 및 인트라넷 환경에서 네트워크 공격 및 장애를 탐지하는 것이다. PCAV는 대학 내 백본(L4) 스위치의 트래픽을 윈도우 PC를 이용하여 4가지 매개변수(4-factor parameter)를 사용한 Parallel Coordinates 다이어그램으로 시각화한다.

Table 3. network traffic visualization system comparison(PCAV vs. FDANT-PCSV)

Category	PCAV	FDANT-PCSV
detection target	network attack	network attack, network failure
event collection target	L4 switch	fire-wall, IPS, DDoS etc
data processing device	windows PC	big-data based SIEM(SPLUNK)
parallel coordinates visualization parameter	4-factor (src IP, dst IP, dest port, pkt length)	5-factor (src IP, dst IP, dest port, pkt length, action)
sankey visualization parameter	-	4-factor (src IP, dst IP, num of event, data size)

반면에 FDANT-PCSV는 스위치가 아닌 방화벽, IPS 등 정보보호 시스템의 이벤트 로그를 SIEM으로 실시간 수집하고 이를 5가지 매개변수를 사용한 Parallel Coordinates 다이어그램과 4가지 매개변수를 사용한 Sankey 다이어그램으로 시각화한다. 또한, 빅데이터 기반 SIEM을 이용함으로써 서로 다른 이기종 보안장비 이벤트에 대하여 매우 빠른 속도로 이상 트래픽을 탐지할 수 있다.

IV. FDANT-PCSV 기반 이상 트래픽 탐지

4.1 네트워크 관리 관점의 이상 트래픽 탐지

4.1.1 (사례 1) 라우팅 오류로 인한 이상 트래픽 탐지

Fig. 12.는 인트라넷에서 A-방화벽을 통과하는 트래픽을 FDANT-PCSV로 시각화 한 것이다. Fig. 12.(a)는 10분간(2019.9.2. 17:00~17:10) 탐지된 트래픽이며 Fig. 12.(b)와 Fig. 12.(c)는 2019.9.2.과 2019.10.10.에 발생한 24시간 트래픽이다. TOP10 모두 action 부분이 "deny"로 표시되는 차단 트래픽이다.

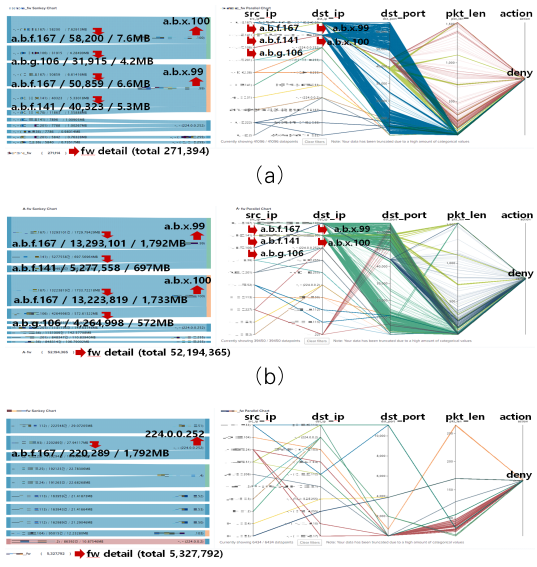

Fig. 12. FDANT-PCSV Diagram (a) 10-minutes diagram before abnormal traffic removal (b) 24-hour diagram before abnormal traffic removal (c) 24-hour diagram after abnormal traffic removal

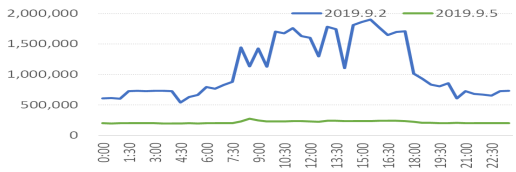

Fig. 13. Comparison of daily events in A firewall before and after abnormal traffic removal

Parallel Coordinates 다이어그램만 살펴보면 출발지 IP(a.b.f.167)에서 목적지 IP(a.b.x.100, a.b.x.99)로 포트 스캔을 수행하고 있는 것으로 판단된다. 하지만 Sankey 다이어그램으로 보면 발생하는 이벤트 수와 트래픽양이 상당히 많다. 인트라넷 포트 스캔은 다음과 같은 상황에서 발생할 수 있다. 첫째, 정보보안 담당자가 보안 점검을 위해 수행하는 경우와 둘째, 공격자가 사내 네트워크 정보를 수집하기 위한 경우이다.

네트워크 관리 담당자는 출발지 IP(a.b.f.167)과 목적지 IP(a.b.x.99, a.b.x.100) 간 트래픽이 A-방화벽에 탐지되면 안 된다고 했다. A-방화벽의 다른 트래픽도 마찬가지 상황이다. 이후 네트워크 담당자는 L3 스위치의 라우팅 설정을 변경하여 A-방화벽의 이상 트래픽 유입을 차단하였다.

Table 4.은 A-방화벽의 이상 트래픽 중 TOP10 현황을 나타낸 것이다. 목적지 IP가 224.0.0.252인 트래픽은 도메인 네임 시스템(DNS) 기반의 LLMNR(Link Local Multicast Name Resolution) 프로토콜을 사용한다. LLMNR 프로토콜은 호스트와 같은 로컬 링크 간에 네임 레졸루션을 수행하며 윈도우 비스타, 윈도우 2008 서버 등에 포함되어 있다. 문제는 224.0.0.0 대역에서 발생하는 트래픽이 멀티캐스팅 방식을 사용하고 있어 트래픽 발생량이 매우 많다는 사실이다. 네트워크 관리 담당자는 출발지 IP가 a.b.f.141, a.b.g.201, a.b.j.36인 윈도우 시스템의 LLMNR 프로토콜 사용이 불필요하다고 판단하고 이 서비스를 중지하였다.

Table 4.의 목적지 IP(a.b.g.255, g.g.k.255)는 브로드캐스트 IP로 불필요한 트래픽을 발생시킨다. 서버(a.b.g.201, g.g.j.36)에서 해당 브로드캐스트 트래픽이 발생하지 않도록 조치하였다. 이후 Fig. 12.(c)와 같이 이전에 발생하던 이상 트래픽이 A-방화벽에서 제거되었다.

Table 5.는 A-방화벽에서 발생하는 일일 이벤트 수와 트래픽양을 나타낸 것이다. 2019.8.28.부터 2019.9.3.까지 A-방화벽에서 발생한 전체 이벤트 수는 5천만 건에 이른다. 같은 기간 동안 이상 트래픽은 3천만 건을 넘어선다.

평일 기준, 4일간(2019.8.28.~2019.9.2.) 발생한 A-방화벽의 일일 총 이벤트 수는 206,905천 건이며 이상 트래픽의 총 이벤트 수는 140,582천 건이다. 이상 트래픽이 전체 트래픽양의 67.9%에 이를 정도로 매우 크다. 같은 기간, 전체 트래픽양은 26.41GB, 이상 트래픽양은 17.97GB이며 이상 트래픽양이 전체의 68.04%를 차지한다.

2019.9.3. 이상 트래픽을 제거한 후 2019.9.4.

Table 4. Abnormal traffic through the "A" firewall measured for 10 minutes

source IP	number of events	packet size(MB)	destination IP
a.b.f.167	58,200	7.76	a.b.x.100
a.b.g.106	31,915	4.28	
a.b.f.167	50,859	6.61	a.b.x.99
a.b.h.141	40,323	5.33	
a.b.i.70	11,863	1.55	
a.b.f.141	7,896	1.00	224.0.0.252
a.b.g.201	7,788	1.00	
a.b.j.36	7,786	0.98	
a.b.g.201	5,842	0.76	a.b.g.255
g.g.j.36	5,842	0.76	g.g.k.255

Table 5. Weekday "A" firewall total traffic vs. abnormal traffic using big data-based SIEM

date (`19)	"A" firewall daily total traffic		"A" firewall daily abnormal traffic	
	number of events	traffic (GB)	number of events	traffic (GB)
8.28	52,995k	6.76	36,364k	4.65
8.29	53,566k	6.84	37,034k	4.73
8.30	47,938k	6.12	31,782k	4.06
9.2	52,406k	6.69	35,402k	4.53
9.3	48,815k	6.23	33,048k	4.22
9.4	10,614k	1.33	0	0
9.5	10,435k	1.31	0	0
9.6	9,300k	1.17	0	0
9.9	8,145k	1.03	0	0
9.10	6,143k	0.78	0	0

부터 A-방화벽에서 문제를 유발하던 이상 트래픽은 탐지되지 않았다. 이후 계속해서 A-방화벽에 남아 있는 이상 트래픽을 FDANT-PCSV로 탐지하고 제거한 결과 2019.9.10. A-방화벽의 전체 이벤트 수는 6백만 건, 트래픽양은 0.78GB로 대폭 줄었다.

Fig. 13.은 A-방화벽에서 발생하던 이상 트래픽을 제거하기 전(2019.9.2.)과 후(2019.9.5.)의 발생 이벤트 수를 비교한 것이다. A-방화벽에서 발생하는 이벤트가 대폭 줄어든 것을 확인할 수 있다.

4.1.2 (사례 2) Server / Clients 이상 트래픽 탐지

Fig. 14.(c)는 B-방화벽 및 SIEM의 하루 동안 발생한 트래픽양을 보여준다. B-방화벽의 평일 트래픽양은 평균 7GB 정도 발생하다가 B-방화벽의 평일 트래픽양이 11.4GB(2019.12.5.), 13GB(2019.12.6.), 16.2GB(2019.12.9.)로 점점 증가했다.

사내 SIEM의 일일 허용 트래픽양은 32GB로 초과하게 되면, SIEM은 보안장비로부터 시스로그 수집이 중단된다. 2019.12.11. SIEM의 총 트래픽양이 30.8GB로 측정되었다. FDANT-PCSV를 이용하여 B-방화벽에서 발생하는 전체 트래픽을 분석하고 이상 트래픽 탐지를 시도하였다.

Fig. 14.(a)은 B-방화벽을 FDANT-PCSV로 1분간(2019.12.11. 09:30~09:31) 측정한 다이어그램이다. Parallel Coordinates 다이어그램 상 출발지 IP(a.b.x.99~100)에서 목적지 IP(a.b.f.167)로 다량의 트래픽이 발생한다.

또한, Sankey 다이어그램을 분석해 보면 출발지 IP(a.b.x.99, a.b.x.100)과 목적지 IP(a.b.f.167) 간 트래픽양이 매우 큰 것을 알 수 있다. 하나의 출발지 IP에서 목적지 IP(a.b.f.167)로 초당 전송되는 이벤트의 수가 56~57건에 이를 정도로 많은 트래픽이 발생한다.

IP가 a.b.f.167인 시스템은 회사 내 소프트웨어를 관리하는 시스템으로 확인되었다. 5분간(2019.12.11. 09:30~09:35) 목적지 IP(a.b.f.167)과 세션을 맺은 회사 내 PC 수는 1,778대로 그 양이 매우 많아 네트워크에 매우 큰 부하를 초래한다.

Fig. 14.(b)는 FDANT-PCSV로 B-방화벽의 발생 이벤트를 시각화한 것이고 Table 5.는 B-방화벽의 일일 이벤트 수와 트래픽양을 나타낸 것이다.

Table 6.를 보면 2019.12.11. B-방화벽 이벤트 수와 트래픽양이 최고치에 달한다. B-방화벽의 전체

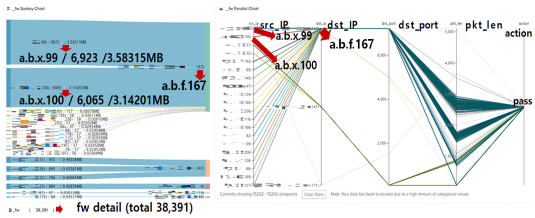

Fig. 14.(a) 1-minute traffic of SIEM(include B-firewall traffic)

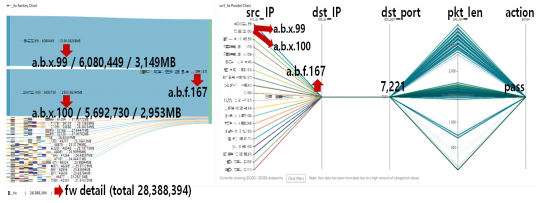

Fig. 14.(b) Daily traffic of SIEM(include B-firewall traffic)

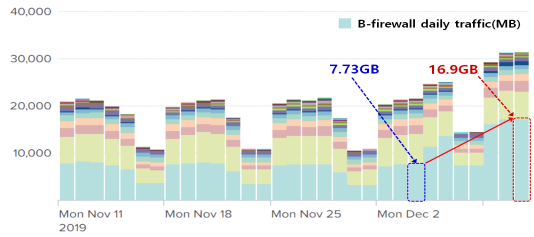

Fig. 14.(c) Daily traffic of SIEM(include abnormal traffic of B-firewall traffic)

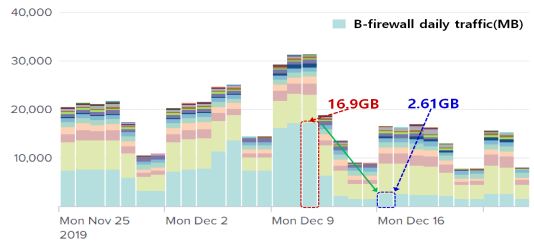

Fig. 14.(d) Daily traffic of SIEM(after removing abnormal B-firewall traffic)

발생 이벤트 수는 33,453천 건이고 트래픽양은 16.90GB이다. 이상 트래픽(목적지 IP: a.b.f.167, 목적지 포트: 7,221)의 이벤트 수는 28,388천 건(84.85%)이고 트래픽양은 14.38GB(85%)이다.

2012.12.12. 과도한 트래픽 발생 원인을 제거하자 2019.12.16.(평일 기준) B-방화벽의 전체 이벤

Table 6. Weekday "B" firewall total traffic vs. abnormal traffic using big data-based SIEM

date (`19)	"B" firewall daily total traffic		"B" firewall daily abnormal traffic (x.x.x.167)	
	number of events	traffic (GB)	number of events	traffic (GB)
12.2	30,949k	7.10	0	0
12.3	32,998k	7.57	0	0
12.4	33,718k	7.73	0	0
12.6	26,563k	13.41	22,211k	11.24
12.9	31,333k	15.83	26,247k	13.29
12.10	33,431k	16.89	28,253k	14.31
12.11	33,453k	16.90	28,396k	14.38
12.12	12,550k	6.32	7,566k	3.83
12.13	4,451k	2.22	0	0
12.16	5,235k	2.61	0	0
12.17	5,239k	2.62	0	0

트 수는 5,235천 건, 트래픽양은 2.61GB로 낮아졌다. FDANT-PCSV 시각화를 통해 Fig. 14.(d)와 같이 2019.12.13.부터 B-방화벽의 이상 트래픽이 제거되고 SIEM의 일일 트래픽양이 최적화되었다.

Fig. 15.는 B-방화벽의 이상 트래픽 제거 전(2019.12.11.)과 제거 후(2019.12.15.)의 일일 이벤트 수를 비교한 것이다. 이상 트래픽 제거 전에는 업무시간 대(9~18시) B-방화벽의 발생 이벤트가 매우 많았으나, 이상 트래픽 제거 후 B-방화벽의 일일 이벤트 수가 눈에 띄게 줄었다.

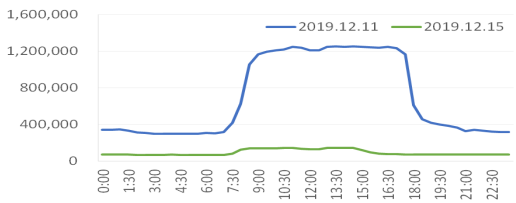

Fig. 15. Comparison of daily events in B firewall before and after abnormal traffic removal

4.2 사이버 보안 관점의 이상 트래픽 탐지

4.2.1 (사례 3) 포트 및 웹 스캔 탐지

Fig. 16.은 FDANT-PCSV를 이용하여 IPS의 보안이벤트 중 네트워크 포트 스캔과 웹 스캔 패턴을 탐지한 것이다. 공격자는 포트 스캔을 통해 취약한 포트를 탐색하고 연이어 웹 스캔을 시도한다. 공격자(x.244.44.36)는 2019.9.10. 16:36:58에 목적지 IP(x.x.x.133)에 대한 포트 스캔을 시작한다. 이후 16:42:03에 포트 스캔을 중지하고 16:43:58부터 80 포트를 이용한 웹 스캔을 시작했으나 웹 스캔 시도가 차단됨을 인지하고 17:01:23에 웹 스캔을 중지한다. FDANT-PCSV를 실시간 모니터링 화면에 띄워 놓음으로써 포트 스캔 및 웹 스캔 공격을 신속하게 탐지하고 대응할 수 있다.

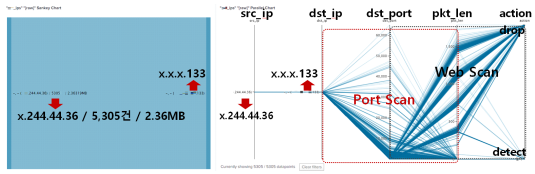

Fig. 16. Detection of web and port scans coming into the Internet using FDANT-PCSV

4.2.2 (사례 4) 느린 포트 스캔 탐지

M. Ring 등[9]은 느린 포트 스캔에 대한 탐지 방법을 제시했다. 일반적으로 포트 스캔은 짧은 시간 내에 대량의 스캔 트래픽을 발생시킨다. 포트 스캔은 출발지 IP의 요청 포트 수를 계산하는 등 간단한 메커니즘으로 쉽게 감지할 수 있다. 지능적인 공격자는 탐지를 우회하려고 느린 포트 스캔을 이용한다. 방대한 이벤트로부터 느린 포트 스캔을 탐지하기는 쉽지 않다. FDANT-PCSV 시각화를 통해 느린 포트 스캔을 효율적으로 탐지할 수 있다.

Fig. 17.(a)~(d)는 C-방화벽의 호스트 스캔(host scan) 및 포트 스캔(port scan) 공격을 이용한 느린 포트 스캔을 FDANT-PCSV로 탐지한 것이다. 공격자는 스푸핑 IP를 사용하여 목적지 IP 대역에 대해 미리 스캔 범위를 정해두고 포트 스캔을 시도하고 있다.

Fig. 17.(a)~(d)는 각각 1분, 5분, 1일, 8일에 해당하는 FDANT-PCSV 다이어그램이다. Fig. 17.(d)을 통해 알 수 있듯이 공격자는 장기간에 걸쳐 거의 모든 포트(1~65,536)에 대한 스캔을 시도하고 열려 있는 포트를 찾고 있다. Fig. 17.(a)을 보면 공격자는 스푸핑 IP 4개를 이용하여 1분 동안 전체 20개의 스캔 패킷을 보낸다. 느린 포트 스캔을

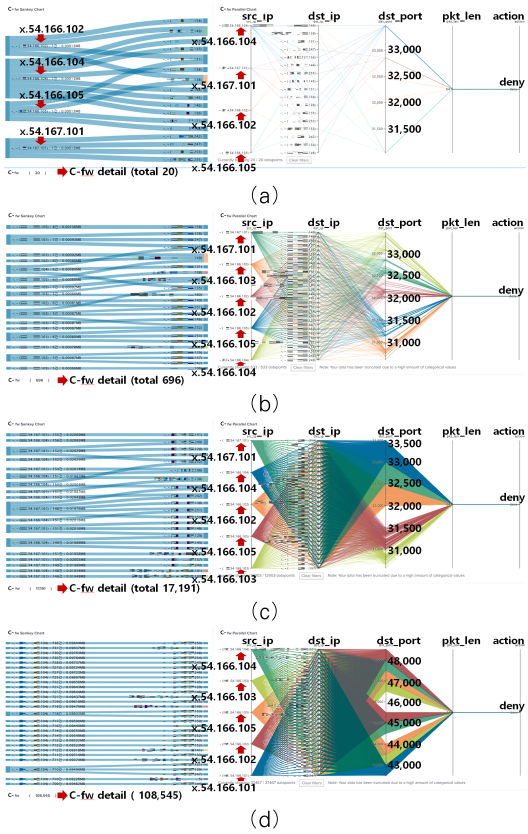

Fig. 17. Intelligent slow host scan and port scan traffic detection using FDANT-PCSV (a) 1-minute (b) 5-minute (c) 1-day (d) 8-day

시도하고 있다.

IPS나 방화벽의 방대한 이벤트를 일일이 분석하여 느린 포트 스캔을 탐지하기란 쉽지 않다. 규모가 작은 SOC(Security Operations Center) 또는 24시간 보안관제를 수행할 수 없는 회사의 경우, FDANT-PCSV를 이용하면 실시간 또는 장기간 스캔 트래픽을 신속하게 탐지할 수 있다.

4.2.3 (사례 5) 스캔 공격 자동 탐지 대시보드

Fig. 18.은 스캔 트래픽을 자동으로 탐지할 수 있도록 FDANT-PCSV를 이용한 대시보드를 구성한 것이다. 사이버 보안관제 인력이 한정된 작은 규모의 SOC에 많은 도움을 줄 수 있다. 자동화 스캔 FDANT-PCSV 대시보드는 호스트 스캔 및 포트 스캔, DDoS 패턴을 형성하는 트래픽을 집중적으로 찾을 수 있도록 최적화되어 있다. 10분간 30개 이상의 네트워크 스캐닝 플로우(network scanning flow)가 발생하면 해당 플로우가 FDANT-PCSV로 시각화된다.

자동화 스캔 FDANT-PCSV 대시보드는 3가지로 화면으로 구성되어 있다. 첫 번째, Fig. 18.의 상단에 있는 전체 트래픽 중 스캔 패턴만 FDANT-PCSV로 시각화한다. 두 번째, 스캔 패턴을 지닌 모든 IP 리스트를 대시보드 왼쪽에 보여준다. 세 번째, 해당 IP 클릭 시 Parallel coordinates 다이어그램에 스캔 트래픽을 오른쪽에 표출한다.

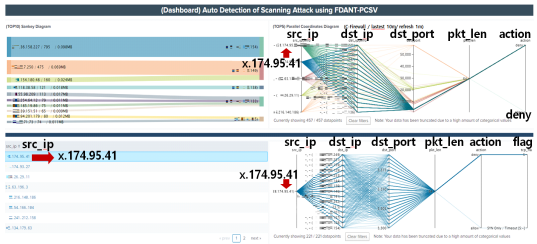

Fig. 18. Automatic detection of flow-based network scanning attacks using FDANT-PCSV

V. 결 론

사이버 공격과 위협이 증가하는 상황에서 정보보호 시스템의 발생 이벤트를 신속하고 효과적으로 분석하는 것이 가장 중요하다. 본 논문에서는 신속한 이상 트래픽을 탐지 시각화 시스템 FDANT-PCSV를 제안하였다.

FDANT-PCSV는 Sankey 및 Parallel Coordinates 다이어그램을 동시에 이용하여 네트워크 이상 트래픽을 신속하게 탐지하고 분석하는 시각화 시스템이다. FDANT-PCSV 시각화 시스템을 이용하면 두 가지 관점의 이상 트래픽을 탐지할 수 있다. 첫 번째는 DDoS 공격, 포트 스캔 등 외부 공격자와 직접 연관이 있는 이상 트래픽을 탐지하는 것이고, 두 번째는 네트워크 장비나 보안장비에 과도한 이벤트를 발생시키는 이상 트래픽 탐지하는 것이다.

본 논문에서는 FDANT-PCSV의 활용성 증대를 위해 스캔 공격 자동 탐지 대시보드를 제안하였다. 현재는 내부 네트워크에서 발생하는 이상 트래픽도 자동으로 탐지하고 경보를 발생하는 시스템을 구현중이다. 향후, FDANT-PCSV에 기계학습 기술을 적용하여 사람을 대신하여 이상 트래픽을 자동으로 탐지하는 시스템에 관한 연구를 계속 진행할 것이다.

References

[1] C. Wu, S. Sheng, and X. Dong, "Research on visualization systems for DDoS attack detection," 2018 IEEE Interna- tional Conference on Systems, Man, and Cybernetics, pp. 2986-2991 January. 2019.

[2] R. Marty, Applied Security Visualization, Pearson Education, Inc, Crawford- sville, 2008.

[3] Hyung Seok Kim, Suk Jun Ko, Dong Seong Kim, and Huy Kang Kim, "Firewall ruleset visualization analysis tool based on segmentation," 2017 IEEE Symposium on Visualization for Cyber Security(VizSec), Oct. 2017.

[4] Jae Beom Park, Huy Kang Kim, and Eun Jin Kim, "Design and implementation of the honeycomb structure visualization system for the effective security situ- ational awareness of large-scale networks," Journal of The Korea Institute of Information Security & Cryptology, 24(6), pp.1197-1213, Dec. 2014.

[5] Dong Gun Lee, Huy Kang Kim, and Eun Jin Kim, "Study on security log visualization and security threat detection using RGB Palette," Journal of The Korea Institute of Information Security & Cryptology, 25(1), pp. 61-73, Feb. 2015.

[6] Y. Okada, "Time-tunnel: 3D Visualization Tool and Its Aspects as 3D Parallel Coordinates," 2018 22nd International Con- ference Information Visualization, pp. 56-57, December. 2018.

[7] H. Choi, H. Lee, and H. Kim, "Fast detection and visualization of network attacks on parallel coordinates," Computers and Security, vol. 28, no. 5, pp. 276-288, July. 2009.

[8] V.T. Guimarães, C.M.D.S. Freitas, R. Sadre, L.M.R. Tarouco, and L.Z. Gran- ville, "A Survey on Information Visualiza- tion for Network and Service Manage- ment," IEEE Communication surveys & tutorials, vol. 18, no. 1, 2016.

[9] M. Ring, D. Landes, and A. Hotho, "Detection of slow port scans in flow-based network traffic," PLOS ONE, vol. 13, no. 9, , Article number. e0204507, September. 2018.

〈저자소개〉

한 기 훈 (Ki Hun Han) 정회원
1995년 2월: 전북대학교 전기전자공학과 학사
2002년 2월: 전북대학교 전자공학과 석사
2018년 9월~현재: 고려대학교 정보보호대학원 석사과정
2015년 12월~2019.12월: 한국서부발전 정보보안부 근무
〈관심분야〉 네트워크 보안, 보안 시각화, 빅데이터 보안관제, 사이버 침해 대응

김 휘 강 (Huy Kang Kim) 종신회원
1998년 2월: KAIST 산업경영학과 학사
2000년 2월: KAIST 산업경영학과 석사
2009년 2월: KAIST 산업 및 시스템공학과 박사
2004년 5월~2010년 2월: 엔씨소프트 정보보안실장, Technical Director
2010년 3월~현재: 고려대학교 정보보호대학원 교수
〈관심분야〉 온라인게임 보안, 네트워크 보안, 네트워크 포렌식, 침입탐지시스템, 봇넷탐지