Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.25, NO.1, Feb. 2015 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2015.25.1.61

RGB Palette를 이용한 보안 로그 시각화 및 보안 위협 인식*

이 동 건,[1†] 김 휘 강,[1] 김 은 진[2‡]
[1]고려대학교, [2]경기대학교

Study on security log visualization and security threat detection using RGB Palette*

Dong-gun Lee,[1†] Huy Kang Kim,[1] Eunjin Kim[2‡]
[1]Korea University, [2]Kyonggi University

요 약

빠르게 증대되고 있는 다양한 보안 위협에 신속하게 대응하기 위해 기업들은 보안관제 담당자에게 방화벽, IDS 등의 보안장비에서 생성되는 방대한 양의 로그 등을 신속하게 분석하여 보안 위협을 파악하도록 요구하고 있다. 그러나 인간의 정보처리 능력의 한계로 인하여 방대한 양의 보안 관련 로그를 분석하는 데 있어 보안 관제 담당자는 많은 시간을 필요로 하게 되고 결과적으로 보안 위협 탐지와 대응이 늦어진다는 문제점이 존재해 왔다. 시각화 기법은 이러한 문제를 해결할 수 있는 효과적인 방법으로 본 논문은 RGB palette를 이용하여 보안 로그를 시각화, 보안 위협 상황 발생 여부를 보다 빠르고 효과적으로 인지할 수 있는 방법을 제안하고 이를 VAST Challenge 2012 데이터 세트에 실증적으로 적용하였다.

ABSTRACT

In order to respond quickly to security threats that are increasing fast and variously, security control personnel needs to understand the threat of a massive amount of logs generated from security devices such as firewalls and IDS. However, due to the limitations of the information processing capability of humans, it takes a lot of time to analyze the vast amount of security logs. As a result, there is problem that the detection and response of security threats are delayed. Visualization technique is an effective way to solve this problem. This paper visualizes the security log using the RGB Palette, offering a quick and effective way to know whether the security threat is occurred. And it was applied empirically in VAST Challenge 2012 dataset.

Keywords: security, visualization, RGB Palette, log, VAST Challenge

I. 서 론

1.1 연구배경 및 연구 목적

2013년 3월 20일 방송국 및 금융권을 공격한 3.20 사이버테러, 2013년 6월 25일 청와대 등 정부기관을 공격한 6.25 사이버 공격, 2011년 4월 농협 전산망을 마비시킨 사이버 공격 등 이제 보안 사고는 영화나 책에서만 발생하는 이야기가 아니라 실제 우리 주변에서도 발생하는 현실이 되었다. 실제로 KISA

접수일(2014년 9월 30일), 수정일(2014년 12월 16일), 게재확정일(2014년 12월 23일)
* 본 논문은 미래창조과학부 및 정보통신산업진흥원의 대학 IT연구센터육성 지원사업의 연구결과로 수행되었음 (NIPA-2014-H0301-14-1004)
† 주저자, eastgun@hanmail.net
‡ 교신저자, ejkim777@kyonggi.ac.kr(Corresponding author)

인터넷 침해사고 대응 통계를 보면 2013년 한 해 동안 접수된 해킹사고 건수는 10,600건이나 된다. 3.20, 6.25와 같은 대형 보안사고 뿐만 아니라 우리 주변에서 크고 작은 수많은 보안 사고들이 발생하고 있는 것이다.

최근 스마트폰, IoT 등 IT 기기들의 발달로 우리의 생활은 더욱 더 IT 기기들에 의존하는 형태로 변화하고 있다. 이에 따라 각종 기기들로부터 발생하는 트래픽 양 또한 늘어나고 있으며, 이에 따라 각종 보안 로그의 양 또한 증가하고 있다. 시스코의 'Global Mobile Data Traffic Forecast Update, 2013-2018'에 따르면 전 세계 모바일 데이터 트래픽 발생량은 2018년까지 11배 증가하여 매월 15.9 엑사바이트에 달할 것이라고 한다. 이러한 트래픽량의 증가는 네트워크 보안 장비의 로그 발생량도 기하급수적으로 증가함을 의미한다. 기하급수적으로 증가하는 로그 속에서 보안 위협 정보를 정확하게 인지하는 것은 결코 쉽지 않은 일이다. 이는 더 나아가 보안 사고가 발생하여도 우리가 발생 여부조차 인지하기 쉽지 않음을 의미한다.

방대한 양의 로그는 이를 살펴보는 것에 제약이 따르기 때문에 시각화의 기술적인 요소와 더불어 데이터를 요약하고, 한 눈에 살펴볼 수 있도록 돕는 시각화 방법론적 요소의 중요성이 커지고 있다[1]. 데이터 시각화란 데이터 생산자, 수급자, 소비자 등 데이터에 연결된 주체들과 다양한 종류의 데이터세트를 '정확하고, 유용하며, 의미있게' 연결해주는 시스템 디자인이다[2].

본 연구에서는 방대한 데이터세트 표현에 효과적인 색상을 통한 시각화 방법을 제안하고, 보안 관제에 대한 깊은 지식이 없는 사람들도 보안 위협 요소가 발생하고 있다는 것을 손쉽게 인지할 수 있는 보안 시각화 도구를 구현하였다.

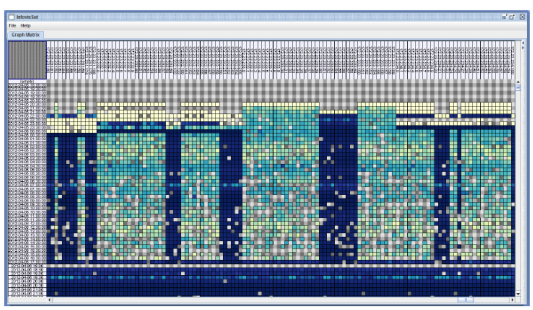


Fig. 1. example of visualization[3]

1.2 연구의 방법과 구성

2장에서는 보안 분야에서 적용되고 있는 시각화 방법과 그 한계점을 살펴보고 3장에서 본 연구의 시각화 방법과 이를 구현한 시스템을 소개하고 인지 가능한 보안 위협 패턴을 설명한다. 4장에서는 VAST 2012 데이터를 이용하여 수행한 모의 테스트 결과와 분석 내용을 소개한다. 마지막으로 5장에서는 본 연구의 결론과 한계점 및 향후 연구 과제를 논한다.

II. 선행 연구 검토

보안 관련 데이터 분석의 효율성과 직관성 향상은 급격히 증가하고 있는 보안 위협과 함께 그 중요성이 더욱 커지고 있다. 시각화는 효율성과 직관성 향상의 대표적인 방법으로 최근 보안 분야에서 활발히 연구되고 있다. VAST(Visual Analytics Science and Technology) challenge는 보안 분야 시각화 연구와 관련한 대표적인 컨퍼런스로 이 컨퍼런스를 통해 이 분야의 여러 우수 연구가 소개되고 있다.

VAST challenge는 방화벽, IDS 로그 등의 정해진 데이터세트를 제공하고, 컨퍼런스의 참가자들로 하여금 자신들만의 시각적 분석 기법으로 해당 데이터세트를 분석하게 함으로써 그 우수성을 평가, 우수한 시각화 방법들을 발굴해 내고 있다.

VAST challenge 2012의 우수 연구들에서 제시한 시각화 방법들은 Table. 1.과 같다.

Shi[4]는 "Investigating Network Traffic Through Compressed Graph Visualization"에서 네트워크 노드들 간의 트래픽 정보를 연결하여 데이터 시각화를 시도하고 있다. 하지만 모든 데이터 전체를 표시할 경우 시각적으로 유의미한 정보를 얻을 수 없어 압축을 통해서 정보를 간략화 하여 표현함으로써 유해 정보를 탐지하는 방법을 제안하고 있다.

Gilbson[5]은 "Network Infrastructure Visualization Using High-Dimensional Node-Attribute Data"에서 IDS와 방화벽 로그에서 각각의 IP주소 발생 빈도를 통해 노드를 결정하고, 출발지와 목적지 간의 연결을 선으로 표시하였다. 그리고 사용자가 각 포인트를 조작하여 이상 현상들을 추적할 수 있는 모델을 제안하고 있다.

Zhao[6]는 "NetSecRadar: A Real-Time Visualization System for Network Security"에서 원형 그래프 중앙에 서버와 호스트들을 배치하고,

Table 1. Visualization strategies on VAST challenge 2012 papers

Category	Subject	Key strategy and tool
network graph	Investigating Network Traffic Through Compressed Graph Visualization[4]	compressed network traffic graph
	Network Infrastructure Visualization Using High-Dimensional Node-Attribute Data[5]	high dimensional data exploration tool adapted for producing graph layouts using node-attributes
radial graph	A Real-Time Visualization System for Network Security[6]	radial graph(hosts, attack types, timeline and histogram, attack correlation and interaction)
parallel coordinates	Dynamic Analysis of Large Datasets with Animated and Correlated Views[7]	GPU-accelerated Tool, histogram view, parallel coordinate view, dynamic view(shows the count change of certain filtered events)
two-axis graph	Combining traditional and high-density visualizations in a dashboard to network health monitoring[8]	icon-based high-density visualization
	situ: Situational Understanding and Discovery for Cyber Attacks[9]	streaming visual analytic system consists of two scores for each event: anomalousness and maliciousness
	Chart- and Matrix-based Approach to Network Operations Forensics[10]	KNIME(Konstanz Information Miner System), Tableau and MySQL
heatmap	Visual Analytics for Network Security[3]	InfoVis Toolkit(IVTK), heatmap, parallel coordinates
geograph	Federating Geovisual Analytic Tools for Cyber Security Analysis[11]	combination of geographic visualization tools(ArcGIS, GeoDa and GeoViz Toolkit)

IDS 로그를 종류별로 밴드 형태로 만들어 원형 테두리에 배치한 다음 각 시점에서 상관관계를 연결하여 네트워크 전체의 상태를 모니터링 할 수 있는 방법을 제시하고 있다.

Cao[7]는 "Dynamic Analysis of Large Datasets with Animated and Correlated Views"에서 GPU 가속에 기반하여 대량의 데이터세트에 대한 실시간 처리 방안을 제시하고 있다. 시각화 표현 방법으로는 평행좌표계, 히스토그램 그리고 이벤트 로그 발생 수치의 변화를 보여주는 다이나믹뷰를 사용하고 있다.

Barcelos[8]는 "Combining traditional and high-density visualizations in a dashboard to network health monitoring"에서 특정 시점에서 어떠한 정책이나 행위의 존재 유무 및 비율을 아이콘화 하여 표현하는 방법을 제시하고, 이차원 배열 위에 아이콘화된 이미지를 시간별로 배치하여 전체 분석 구간의 유해 정보를 분석할 수 있는 방법을 보여주고 있다.

Harrson[9]은 "situ: Situational Understanding and Discovery for Cyber Attacks"에서 특정 시점에서 발생한 로그의 변칙성과 유해성을 점수화 한 뒤 전체 구간을 스트리밍화 하여 표현하는 방법을 제안하고 있다. 변칙성은 평소 발생 빈도에 근거하여 점수화하고, 유해성은 IDS 정책 위반에 근거하여 점수화 하는 방법을 사용하고 있다.

Hildenbrand[10]는 "VAST 2012 Mini-Challenge 2: Chart- and Matrix-based Approach to Network Operations Forensics"에서 새로운 시각화 도구에 대한 제안 보다는 기존 시각화 도구를 활용한 데이터 분석에 집중하고 있다. 전처리 과정을 거쳐 로그를 데이터베이스화 한 후에 Tableau를 이용하여 선-차트에 기반한 분석을 실시하였으며, 분석 결과를 표현하기 위하여 이차원 배열에 기반한 표현 도구를 구현하고 있다.

Shurkhovetskyy[3]는 "Visual Analytics for Network Security"에서 Python을 이용한 데이터 전처리 작업을 거친 뒤 IVTK(InfoVis ToolKit)를 이용한 시각화 분석을 제시하고 있으며 히트맵, 시계열, 평행좌표계 형태를 이용한 표현을 보여주고 있다.

Zhao[11]는 "Federating Geovisual Analytic Tools for Cyber Security Analysis"에서 분석 대상 노드들의 위치 정보를 GeoViz Toolkit을 이용하여 표현하는 방법을 보여주고 있다. 그리고 동일 노드가 동일 위치에 반복적으로 표현되는 것을 시각적으로 구분하여 표시하기 위해 난수를 이용하여 주변으로 흩어져 표현되도록 하는 기법을 소개하고 있다.

지금까지 살펴본 VAST challenge 2012 우수 논문들 외에도 보안과 관련한 여러 분야에서 다양한 시각화 관련 연구들이 Table. 2.와 같이 진행되어 왔다. 그 내용들을 살펴보면 다음과 같다.

Choi[12]는 "Fast detection and visualization of network attacks on parallel coordinates"를 통해서 평행좌표계를 이용한 시각화를 제안하고 있다. 평행좌표계의 각 축을 출발지IP, 목적지IP, 목적

Table 2. Visualization strategies on other papers

Category	Subject	Key strategy or tool
parallel coordinates	Fast detection and visualization of network attacks on parallel coordinates[12]	recognition of network attack using parallel coordinates
geograph	An Efficient Method for Analyzing Network Security Situation Using Visualization[13]	Connected lines of Source Country / Organization / IP / Port, Destination Port / IP / Organization / Country Information
	Cover-VT: Converged Security Visualization Tool[14]	Security threat visualization using GIS & GPS technology
treemap	A Visual Analytic Framework for Exploring Relationships in Textual Contents of Digital Forensics Evidence[15]	frequency of relevant terms and their location visualization
hierachical diagram	Closing-the-Loop: Discovery and Search in Security Visualizations[16]	recognition of security logs pattern & pattern tree search
	A Visualization and Modeling Tool for Security Metrics and Measurements Management[17]	Visualization tool using hierarchical structure
two-axis graph	A Visualization Methodology for Characterization of Network Scans[18]	scan fingerprint comparison using two-dimension array
	Accommodating IPv6 addresses in security visualization tools[19]	visualization methodology for IPv6
	Countering Security Information Overload through Alert and Packet Visualization[20]	display of hole host ids alert on two-dimension plane
	Designing Visualization Capabilities for IDS Challenges[21]	network activity visualization of relations between remote IP, port and local IP
	IDGraphs: Intrusion Detection and Analysis Using Histographs[22]	security log visualization using histograph technique
	Improving Security Visualization with Exposure Map Filtering[23]	traffic volume downsize and visibility improvement by filtering technique

지 포트 등으로 표현하고 공격 유형에 따라 평행좌표계 상에서 특정한 형태를 나타냄을 보여주고 있다.

Jeong[13]는 "An Efficient Method for Analyzing Network Security Situation Using Visualization"을 통해서 5-tuple(출발지IP, 출발지포트, 프로토콜, 목적지IP, 목적지포트)와 지리 정보를 조합하여 시각화하는 방법을 제안하고 있다. 각각의 정보를 하나의 평면으로 표시하고 각 평면들 간의 연결을 입체적으로 표시하여 보안 상황을 시각적으로 표현하고 있다.

Urbanski[14]는 "Cover-VT: Converged Security Visualization Tool"을 통해서 보안 위협 분석을 위하여 지리정보를 이용한 고차원의 네트워크 위협 인지 시스템을 제안하고 있다. 방화벽이나 IDS 상의 정보를 지리적으로 표시함으로써 보안 위협 분석가들이 특정 지역의 집중적인 문제 발생이나 주요 노드 주변의 이상 발생 등을 직관적으로 인식할 수 있도록 하고 있다.

Jankun-Kelly[15]는 "A Visual Analytic Framework for Exploring Relationships in Textual Contents of Digital Forensics Evidence"를 통해서 텍스트 기반의 디지털 포렌식 증거를 분석할 수 있는 방법을 제시하면서, 시각화 표현 방법으로 트리맵 형태를 사용하고 있다. 분석 대상 단어들의 출현 빈도를 기반으로 사각형 형태의 노드 사이즈를 결정하여 트리맵으로 표현하고 있다. 빈도를 노드의 크기로 표현함으로써 직관적으로 분석 대상들의 상대적인 출현 빈도를 가늠할 수 있게 한다.

Lakkaraju[16]는 "Closing-the-Loop: Discovery and Search in Security Visualizations"를 통해서 보안 로그에서 의미 있는 정보를 '발견'하는 것과 그 발견된 정보를 '탐색'하는 것을 구분하여 제시하고 있다. 그리고 탐색을 위하여 정보를 계층화된 트리 구조로 시각화하여 사용자가 단계별로 접근할 수 있는 방법을 제안하고 있다.

Savola[17]는 "A Visualization and Modeling Tool for Security Metrics and Measurements Management"를 통해서 효율적인 보안 지표 및 측정을 관리하기 위해서 시각화가 필요함을 언급하면서, 계층화된 트리 구조로 보안 지표들을 시각화 하는 방법을 제안하고 있다.

Muelder[18]는 "A Visualization Methodology for Characterization of Network Scans"을 통해서 fingerprint라는 시각화 방법을 제시하면서 네트워크 스캐닝 별로 구분된 형태를 나타냄을 보여주고 있다. fingerprint는 256x256 이차원 배열 위에 IP 주소의 세 번째와 네 번째 자리 값을 표현하는 방법을 사용하고 있다.

Barrera[19]는 "Accommodating IPv6 addresses in security visualization tools"를 통해서 현재 많은 보안 시각화 도구들이 IPv4에 맞추어져 있음을 지적하고, IPv6 시대를 대비하여 IPv6에 적용할 수 있는 시각화 방법을 제안하고 있다. IPv6는 IPv4와

비교하여 표현해야 하는 범위가 더 넓기 때문에 이차원 형태의 그래프에서 좌표축에 데이터가 없는 빈 공간을 제거하는 축약된 형태의 시각화를 제안하고 있다.

Conti[20]는 "Countering Security Information Overload through Alert and Packet Visualization"을 통해서 IDS에서 발생하는 전체 경고를 이차원 형태의 그래프에 표시함으로써 보안 관제 담당자들에게 시스템 전체에 대한 통찰력을 키워주고자 한다. 이차원 평면을 8개의 세로축으로 나누어 전체 IP대역을 나누어 배분하고, 각 구간 마다 가로축에 0시부터 24시까지의 경고를 표시하여 전체 IDS 경고를 한 눈에 보여줄 수 있는 방법을 제안하고 있다.

Erbacher[21]는 "Designing Visualization Capabilities for IDS Challenges"를 통해서 IDS 로그를 효과적으로 분석하기 위한 시각화 방법을 제시하고 있다. 그 방법으로 local IP 주소를 중앙에 원형으로 배치한 다음 그 주위를 둘러싼 사각형에서 가로 테두리에는 remote IP 주소를 배치하고 세로 테두리에는 서비스 포트들을 배치한다. 그리고 이 세 가지 요소를 선으로 연결하여 표시한 다음 일정 시간 간격마다 이전에 표시한 원의 크기는 줄이고 현재 시점의 원을 다시 그려 시간의 흐름에 따른 IDS 로그의 변화를 표현하는 방법을 제안하고 있다.

Ren[22]은 "IDGraphs: Intrusion Detection and Analysis Using Histographs"를 통해서 네트워크 트래픽에서 이상 트래픽 발생과 공격을 탐지하기 위하여 시간을 가로축으로 하고, 비정상적인 connection의 수를 세로축으로 하는 이차원 그래프를 제안하고 있다. 그리고 각 픽셀의 빈도를 표현하기 위하여 블루어링 (흐려짐)기법을 사용한다.

Alsaleh[23]는 "Improving Security Visualization with Exposure Map Filtering"을 통해서 많은 양의 네트워크 정보를 효율적으로 시각화하기 위해서 필터링을 통해 네트워크 트래픽의 볼륨을 줄이는 방법을 제안하고 있다. 네트워크의 볼륨을 줄이기 위해서 네트워크에서 오픈된 호스트와 서비스를 정리하여 NEM(Network Exposure Map)을 정의하고 네트워크 접근이 이 NEM의 어떠한 영역에 접근하는지에 따라 '위험', '의심' 등으로 필터링 한다. 그리고 '위험' 으로 필터링 된 정보만 이차원 그래프 상에 표현함으로써 시각적으로 인식이 쉽도록 제안하고 있다.

앞에서 살펴본 바와 같이 기존 연구들을 살펴보면 위치정보, parallel coordinates 등 다양한 시각화 전략 또는 도구들이 사용되었던 것을 확인할 수 있다. 하지만 많은 연구들이 다양한 시각화 요소 중 형태적인 파악을 통한 시각화에 많이 집중하고 있다는 것을 알 수 있다. 본 논문은 색상을 통해서 보안 요소들을 표현하고, 위협 요소들을 인지하기 위한 시도를 하였다는 점에서 다른 논문들과 차별성이 있다. 색상은 데이터세트가 방대할 때 매우 효과적인 표현 방법으로 다음 장들에서는 본 논문의 RPG Palette를 이용한 시각화 방법을 제시하고 이를 VAST challenge 2012 dataset을 적용해 봄으로써 그 유용성을 검증해보고자 한다.

III. 시스템 구현

이 연구의 목적은 방화벽, IDS 로그 분석을 통하여 관제 요원이 보안 위협 상황의 발생을 쉽게 인지할 수 있는 RGB 속성을 이용한 도구를 설계 및 구현하는 것이다. 이 장에서는 효과적인 시각화를 위해 우리가 선택한 시각화 전략을 소개하고, 그 시각화 전략을 이용하여 구현한 시각화 도구를 설명할 것이다. 그리고 이 도구를 이용하여 인지할 수 있는 보안 위협의 유형을 알아볼 것이다.

3.1 시각화 전략

본 시각화 전략을 수립함에 있어서 첫 번째 목표는 단순성이다. 보안 위협 상황을 나타내기 위해서 방화벽이나 IDS 로그에서 보여주어야 할 다음 4 가지 요소를 선정하였다.

1) 출발지 IP
2) 목적지 IP
3) 목적지 포트
4) 로그 발생량

출발지 IP는 공격의 근원지가 어디인지를 나타내는 정보라는 점에서 중요한 정보이다. 목적지 IP와 포트는 공격의 대상이 어디인지를 나타내는 정보라는 점에서 또한 중요한 정보이다. 마지막으로 로그 발생량은 유해 정보의 발생 정도라는 측면에서 의미 있는 정보라 할 수 있다. 이렇게 선택된 4 가지 정보를 시각적으로 효과적으로 표현하기 위한 방법으로 사람의 눈에서 쉽게 인지할 수 있는 색상과 크기라는 속성을 선택하였다. 크기는 가장 자주 사용되는 시각 표현이다. 사람은 두 가지 물체를 구별할 때 그 크기 차이를 매우 빠르게 구분할 수 있기 때문이다[24]. 그리고 색상은 데이터세트가 방대할 때 매우 훌륭한 표현 방법이

다. 사람은 색상의 다양한 색조나 음영의 차이를 높은 해상도로 구별할 수 있다[24]. 컴퓨터에서는 색상을 표현하기 위하여 Red, Green, Blue라는 세 가지 속성을 활용한다. 그리고 각각의 속성은 0 ~ 255 사이의 고유한 값을 표현한다. 따라서 본 연구에서는 Red, Green, Blue 라는 세 가지 속성에 앞에서 선택한 보안 로그에서 보여지는 요소들을 대입하여 표현하였다. 그 결과 특정한 보안 위협 상황은 고유의 RGB 값의 조합을 통해 보안 상황마다 고유한 색상을 표현한다는 것을 발견할 수 있었다.

3.1.1 Red

Red의 색상 속성에 우리는 방화벽과 IDS 로그의 Source IP를 표현하도록 설정하였다. 발생한 Source IP의 수가 많아질수록 붉은 색이 짙어지도록 하였다. 분석 구간을 10분 단위로 나누어 해당 구간에서 발생하는 Source IP의 수를 분석하여 가장 많이 발생한 구간의 값을 255로 설정하였다. 그리고 최소치를 0으로 설정하여 특정 시점의 Source IP의 수를 0 ~ 255 사이의 값으로 표현하였다.

3.1.2 Green

Green의 색상 속성에 우리는 방화벽과 IDS 로그의 Destination IP를 표현하도록 설정하였다. 발생한 Destination IP의 수가 많아질수록 녹색이 짙어지도록 하였다. 분석 구간을 10분 단위로 나누어 해당 구간에서 발생하는 Destination IP의 수를 분석하여 가장 많이 발생한 구간의 값을 255로 설정하였다. 그리고 최소치를 0으로 설정하여 특정 시점의 Destination IP의 수를 0 ~ 255 사이의 값으로 표현하였다.

3.1.3 Blue

Blue의 색상 속성에 우리는 분석 로그의 Destination Port 수를 표현하도록 설정하였다. 발생한 Destination Port의 수가 많아질수록 푸른색이 짙어지도록 하였다. 분석 구간을 10분 단위로 나누어 해당 구간에서 발생하는 Destination Port의 수를 분석하여 가장 많이 발생한 구간의 값을 255로 설정하였다. 그리고 최소치를 0으로 설정하여 특정 시점의 Destination Port의 수를 0 ~ 255 사이의 값으로 표현하였다.

3.1.4 크기

앞에서 결정된 RGB 값으로 고유한 색상의 원을 그리도록 하였다. 이 때 이 원의 크기라는 속성에 보안 로그의 발생량을 표현하도록 설정하였다. 유해한 보안 로그의 발생량이 많아질수록 원의 크기는 커지게 되며, 발생량이 줄어들수록 원의 크기는 작아지게 된다. 그리고 효과적인 인지를 위해서 로그 발생량의 평균치를 점선 원으로 표시하여 보는 사람들이 쉽게 인지할 수 있도록 하였다.

3.1.5 색의 보정

특정 보안 상황에서 결정된 RGB 값과 로그 발생량으로 크기가 결정된 원을 그렸고, RGB 값에 대한 사용자의 이해를 돕기 위하여 원 주변으로 각각 Red, Green, Blue 3가지의 파이를 그렸다. 하지만 파이 부분은 Red, Green, Blue의 색상을 가장 직관적으로 잘 표현하기 위하여 색의 보정 작업을 거쳤다. Red 색상을 예로 들면, 우리가 흔히 생각하는 가장 선명한 붉은 색은 RGB 값이 (255, 0, 0)인 경우이다.

Fig. 2.를 보면 붉은 색이 옅어질 때에는 Green과 Blue의 값이 커지는 것을 알 수 있다. 그리고 Green과 Blue의 값이 0으로 고정된 상태에서 Red의 값이 작아지면 그 색이 너무나 짙어져서 붉은색의 선명도가 오히려 떨어진다는 것을 알 수 있었다. 그래서 우리는 중앙 원의 색상 RGB 값에서 Red의 값이 얼마나 커지고 있냐는 것을 직관적으로 표현하기 위하여 원 주변의 Red 색상 파이의 RGB 값을 (255,

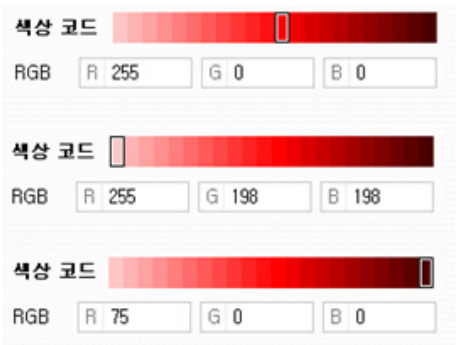


Fig. 2. Change of Red color according to the RGB values

255-red, 255-red) 값으로 보정하는 작업을 진행하였다. 그리고 Green 색상 파이의 RGB 값 역시 (255-green, 255, 255-green)으로, Blue 색상 파이의 RGB값 역시 (255-blue, 255-blue, 255)로 보정하였다.

3.2 System 소개

우리는 구현된 시각화 도구를 편리하게 지칭하기 위하여 'RGB Palette' 이라는 이름을 부여하였다. 이 절에서는 'RGB Palette'의 구성에 대하여 소개하고자 한다.

Fig. 3.에서 'A' 부분은 'RGB Palette'의 핵심인 부분으로, 방화벽과 IDS 로그를 RGB 색상과 원의 크기로 표현하는 부분이다. 원의 중앙부분은 가장자리의 3가지 Red, Green, Blue의 값의 조합으로 색상이 결정된다. 그리고 로그의 발생량이 많아질수록 중앙 원의 크기가 커지게 된다. 점선으로 그려진 원은 로그 발생량의 평균을 표시한다. 중앙원의 색상은 Source IP의 수, Destination IP의 수, Destination Port의 수에 따라서 각 상황마다 고유의 색을 띄게 된다.

'B' 부분은 방화벽과 IDS 로그의 RGB 값들이 시간 별로 얼마였는지를 보여준다. 그래서 시간이 흘러감에 따라 색의 변화가 어떠했는지를 알 수 있는 보조적인 수단을 제공한다. 방화벽 로그의 RGB 값 3 가지와 IDS 로그의 RGB 값 3 가지를 더하여 총 6가지 로그의 변화량을 선택 옵션에 따라 표시할 수 있도록 하였다.

'C' 부분은 로그 분석 구간의 시간을 변화시키기 위한 조작을 할 수 있는 버튼과 원하는 시간을 선택할 수 있는 프로그레스바를 위치시켰다.

'D' 부분에서는 특정 로그 분석 시점에서의 Source IP top 10 값, Destination IP top 10 값, Destination Port top 10값을 보여주는 부분이다. 각 부분의 실제 상위 값을 보여주어 사용자가 로그 분석에 참고 값을 얻을 수 있도록 하였다.

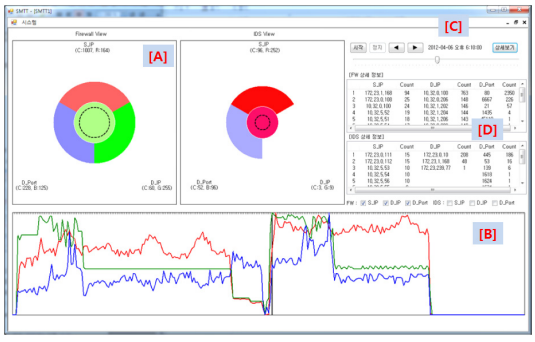


Fig. 3. RGB Pallete

3.3 패턴 인식

이 절에서는 'RGB Palette'를 이용하여 인식할 수 있는 몇몇 공격 패턴에 대하여 알아보고자 한다.

3.3.1 Port Scanning

포트 스캐닝 공격이 발생하면 방화벽 로그 상에서 Source IP와 Destination IP 숫자의 변화에 비하여 Destination Port의 숫자가 증가하게 될 것이다. 즉 RGB Palette에서 Fig. 4.와 같이 Destination Port의 숫자를 의미하는 Blue 색상이 짙어짐을 의미한다.

로그의 발생량이 적어서 원의 크기가 작아진 상태에서 Blue의 색상이 짙어진다면 Port Scan 공격의 가능성은 더욱 높아질 것이다.

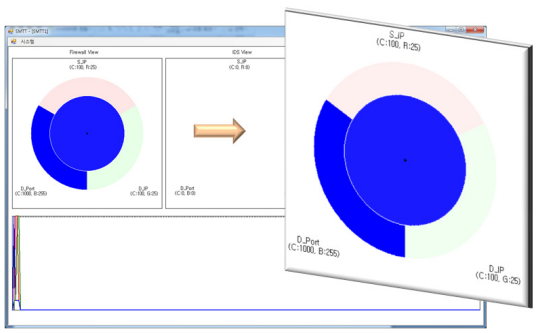


Fig. 4. Recognition of port scanning

3.3.2 DDos Attack

DDos 공격이 발생하게 되면 다양한 출발지에서 공격 목적지로 트래픽이 발생할 것이다. 다수의 Source IP가 감지될 것이며, Destination IP의 숫자와 Destination Port의 숫자는 소수가 될 것이다. 그리고 DDos 공격의 특성상 로그 발생량이 많아서 평균치 보다 큰 크기의 원이 그려지게 될 것이다.

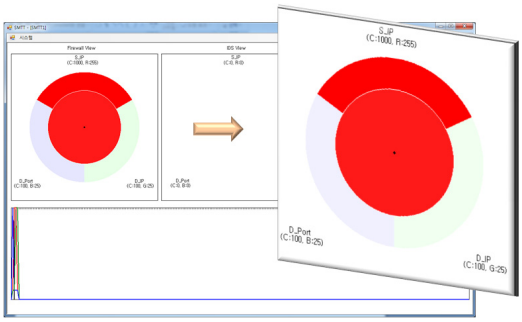


Fig. 5. Recognition of DDos attack

이러한 DDos Attack의 특성은 RGB Palette에서 Fig. 5.와 같이 Red 색상이 짙어진 형태로 인식될 것이다.

3.3.3 Host Scanning

호스트 스캐닝 공격이 발생한다면, Source IP와 Destination Port 수의 변화 보다는 Destination IP의 수가 증가할 것이다. 이것은 RGB Palette에서 Fig. 6. 과 같이 Destination IP의 수를 의미하는 Green 색상이 진하게 변하는 것을 의미한다.

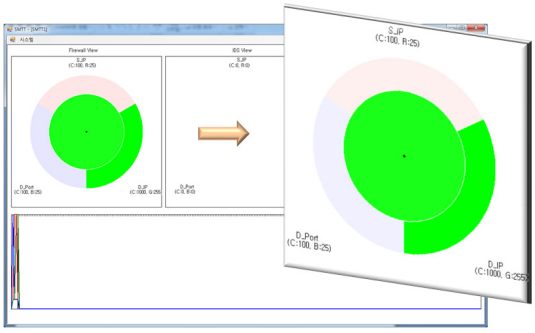


Fig. 6. Recognition of host scanning

3.4 실제 데이터 테스트

이 절에서는 'RGB Palette'에 실제 데이터를 적용하여 앞 절에서 예상한 패턴을 인식할 수 있는지를 알아보고자 한다.

3.4.1 데이터 생성

테스트 데이터는 기관에서 실제 운영중인 방화벽을 이용하여 생성되었으며, 스캐닝 도구는 nmap 6.47 버전을 이용하였다[1]. 테스트 데이터는 15시부터 21시 사이에 6시간 동안 수집하였다. 포트 스캐닝 공격은 19시 30분경에, 호스트 스캐닝 공격은 20시경에 진행하였다.

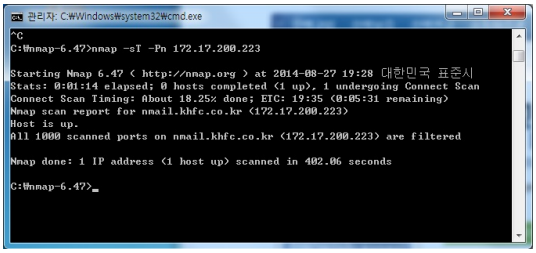


Fig. 7. Port Scanning attack(19:28~19:35)

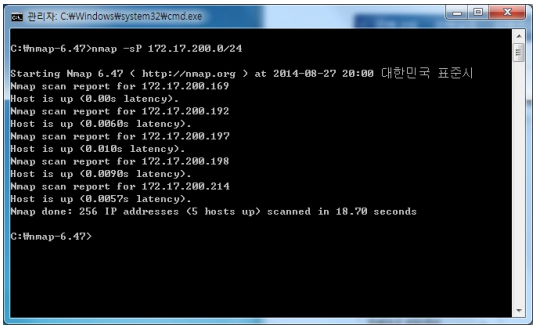


Fig. 8. Host Scanning attack(20:00)

3.4.2 Port Scanning 탐지

테스트 데이터를 RGB Palette를 이용하여 분석한 결과 포트 스캔 공격이 시작된 19시 20분대부터 Destination Port 정보를 의미하는 Blue 색상이 짙어지는 것을 확인 할 수 있었으며, 19시 30분대에 Blue 색상이 가장 진해짐을 확인 할 수 있었다. 즉, RGB Palette의 색 표현이 포트 스캐닝 공격을 효과적으로 표현하여 주고 있음을 알 수 있었다.

1) 실제 운영중인 근무 환경에서 생성된 데이터이기 때문에 비교적 유해성이 낮은 단순 스캐닝 공격만 실시함

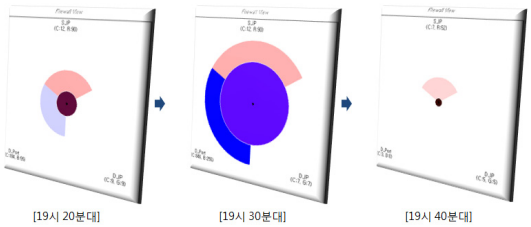


Fig. 9. Port Scanning attack recognition by RGB Palette

3.4.3 Host Scanning 탐지

테스트 데이터를 RGB Palette를 이용하여 분석한 결과 호스트 스캔 공격이 진행된 20시 00분대에 Destination IP 정보를 의미하는 Green 색상이 짙어짐을 확인 할 수 있었다. 이것은 평상시에 비하여 접근이 이루어지는 목적지가 다양해 졌음을 의미하며 RGB Palette의 색 표현이 호스트 스캐닝 공격을 효과적 보여주고 있음을 알 수 있었다.

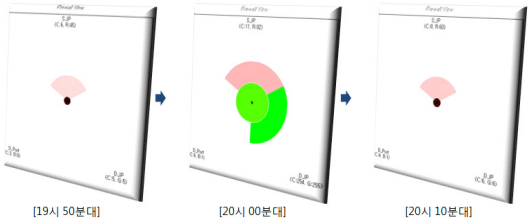


Fig. 10. Host Scanning attack recognition by RGB Palette

IV. VAST 2012 데이터 적용 및 결과 분석

지금까지 우리가 사용한 시각화 전략과 그 시각화 전략을 구현한 도구에 대해서 살펴보았다. 그리고 그 도구를 이용하여 인식할 수 있는 보안 위협 상황에 대해서도 알아보았다. 이번 장에서는 실제로 Vast 2012 데이터세트에 우리가 구현한 RGB Palette을 적용하여 그 유용성에 대해서 살펴볼 것이다.

VAST 2012 데이터세트에서 제공되는 2일간의 방화벽과 IDS 로그를 10분 단위로 데이터를 분할하였다. 그리고 각 구간에서 발생하는 Source IP, Destination IP, Destination Port 숫자를 분석에 활용하였다. 점선의 원으로 표시되는 로그 발생량의 평균치는 2일간 발생한 데이터의 평균값을 활용하였다.

4.1 Port Scanning 탐지

RGB Palette를 이용하여 VAST Challenge 2012 데이터세트를 분석해보면 우선 6일 오후 2시 30분경부터 방화벽 그래프에 Blue(Destination Port) 값이 상대적으로 높으나, Destination 및 Source IP의 수는 상대적으로 낮음을 발견할 수 있다. 그리고 중심 원이 점선의 원보다 작은 것으로 보아 전체 로그 발생량도 적다는 것을 알 수 있다. 이것은 3.4.2 절에서 살펴보았던 포트 스캐닝 공격 유형과 일치한다.

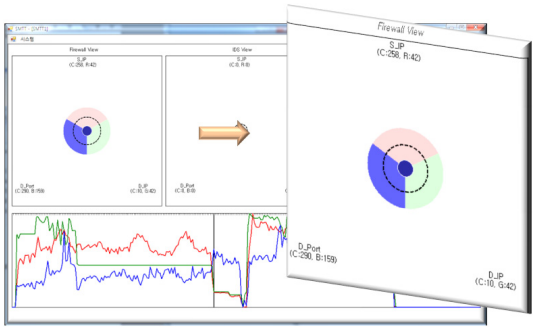


Fig. 11. Port Scanning detection

4.2 웜 활동 탐지

Fig. 12.를 보면 모니터링 전 구간에 걸쳐서 일정하게 30분 주기로 IDS의 Red(Source IP) 색상 농도가 짙어짐을 발견할 수 있다. 일반 사용자들의 정상적인 사용이라고 판단하기에는 너무나 규칙적인 활동 양상을 보여준다. 이것은 웜이 일정한 주기로 활동을 시도하는 것이라고 의심해볼 수 있을 것이다.

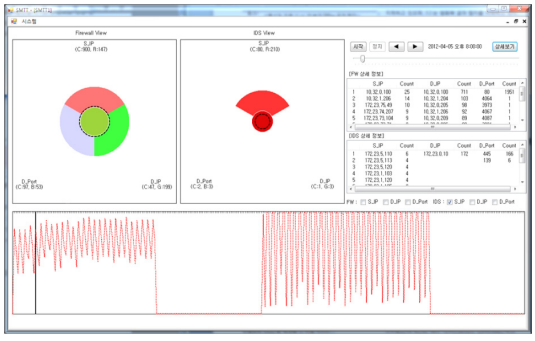


Fig. 12. Worm activity detection

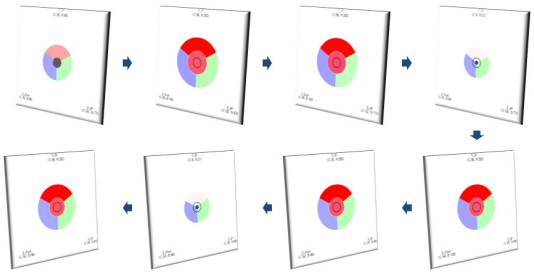


Fig. 13.. Periodic concentration change of Red color(IDS logs)

4.3 방화벽 또는 네트워크 스위치 장애 탐지

Fig. 14.를 보면 방화벽 수치가 급격하게 감소하고, 방화벽 로그가 감소하기 직전 IDS로그 발생량이 과다하게 많은 것을 발견할 수 있다. 또, Fig. 15.를 보면 10분 후에는 IDS 로그도 급격하게 감소한다. 이것은 특정 공격에 의한 방화벽 또는 네트워크 스위치의 장애 발생을 의심하게 한다.

Fig. 14. Firewall or network switch failure 1

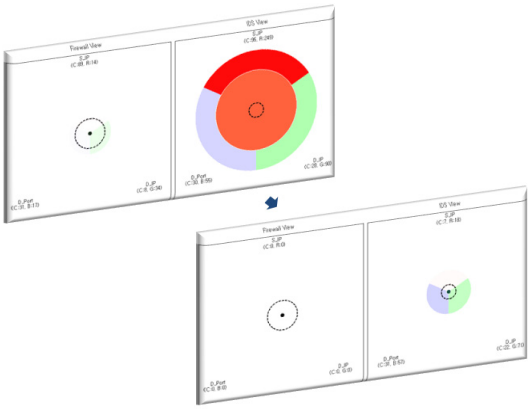


Fig. 15. Firewall or network switch failure 2

4.4 웜 또는 바이러스 확산 시도 탐지

Fig. 16.을 보면 IDS로그에서 Destination IP의 수와 Destination Port의 수가 급격하게 증가하기 시작하는 것을 발견할 수 있다. 이것은 웜 또는 바이러스의 확산시도 또는 활동량의 증가를 의심해 볼 수 있다. 추가적인 조사를 통해서 확산의 근원지가 되는 곳을 찾아야 할 것이다.

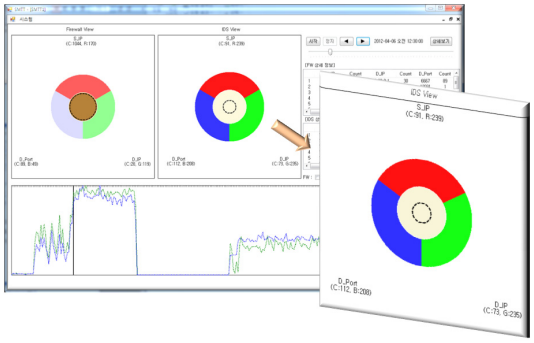


Fig. 16. Worm or virus outbreak attempts detection

4.5 DDos 공격 의심

Fig. 17.을 보면 방화벽 로그에서 Source IP, Destination IP, Destination Port가 모두 증가하여 중앙원이 흰색(255, 255, 255)에 가까워지는 것을 볼 수 있다. 이러한 Source IP, Destination IP, Destination Port의 무차별적인 증가는 DDos 공격 시도를 의심할 수 있게 한다.

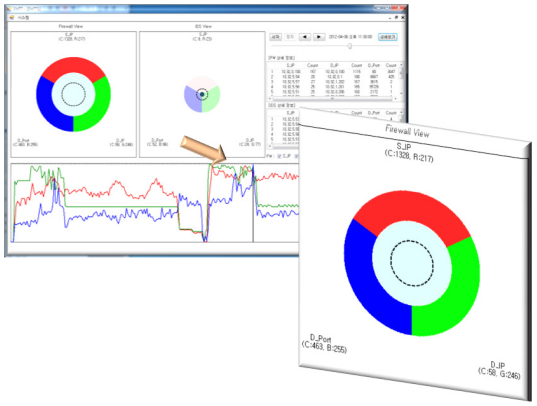


Fig. 17. DDos attack suspect

4.6 VAST 2012 우수 방법론 비교

앞 절에서 우리는 RGB Palette를 이용하여 VAST 2012 Mini-Challenge2 데이터세트를 RGB Palette를 이용하여 분석한 결과 Table. 3.과 같은 보안 위협을 발견을 할 수 있었다. 이번 절에서는 이 결과를 VAST 2012 우수 방법론을 통해 발견된 공격 내용과 비교해 봄으로써 RGB Palette의 효과성을 확인해보고자 한다.

Hildenbrand[10]는 'Chart- and Matrix-based Approach to Network Operations Forensics'을 통해서 VAST 2012 데이터세트의 공격 분석에 다음과 같은 결과를 도출하였다.

1) 분석 시점 이전부터 malware의 존재를 의심하고 있으며, 분석 구간 기간 중 6667포트를 사용하는 웜의 활동 가능성을 제기
2) 4월 6일 오전 0시 4분경 방화벽의 공격을 의심
3) 4월 6일 오후 6시에 DDos 공격을 의심
4) 4월 6일 오후 2시 24분 로그의 급 감소를 언급하며, 이상 행위로 지적

Table. 3. Detection of attack patterns to RGB Palette

Separator	Detected attack
R-1	4/6 pm 2:30 Port Scanning detection
R-2	worm activity detection
R-3	4/6 pm 5:20 network failure detection
R-4	4/6 am 00:30 Worm outbreak attempts detection
R-5	4/6 pm 11:30 DDos attack detection

Table. 4. Presented attack patterns in Hildenbrand's paper

Separator	Detected attack
J-1	worm(using the 6667 port) activity detected
J-2	4/6 am 00:04 firewall attack detection
J-3	4/6 pm 06:00 DDos attack detection
J-4	4/6 pm 02:24 anomaly detection

이 내용을 표로 정리해보면 다음과 같다.

본 논문에서 제시한 공격패턴과 VAST 2012 우수 방법론을 통해 발견된 공격패턴을 비교해보면 다음과 같다.

1) R-1은 포트 스캐닝 공격을 지적하고 있으며, J-4는 이상 행위의 발생 가능성을 지적하고 있다. 그리고 Shurkhovetskyy[3]도 이 시점에서의 포트스캐닝 또는 DDos 공격의 발생을 공통적으로 지적하고 있다.
2) R-2와 J-1은 이 분석 데이터의 전 구간에 걸쳐서 웜(6667포트 사용)의 활동을 공통적으로 지적하고 있다. 6667포트를 이용한 웜의 활동은 이 외에도 shi[4], zhao[6] 등의 논문에서도 공통적으로 언급되고 있다. 특히 zhao가 제안한 시각화 도구에서도 RGB Palette와 같이 주기적인 웜 활동을 시각적으로 보여주고 있다.
3) R-3는 네트워크 장비의 장애를, J-3는 DDos 공격의 발생을 지적하고 있다. 이는 DDos 발생으로 인한 네트워크 장비의 장애 현상 발생을 의심할 수 있을 것이다.
4) R-4는 웜바이러스의 확산 시도를 지적하고 있으며, J-2는 방화벽 공격 탐지를 지적하고 있다. 이는 공통적으로 유해 활동의 증가를 지적하고 있는 것이다.

위 비교 내용을 검토해 볼 때 상당 부분에서 탐지 결과의 유사성을 확인할 수 있었다. 이로써 보안에 대한 전문 지식이 없는 일반 관제 요원들이 1차적인 이상 현상 발생을 인지할 수 있도록 하겠다는 RGB Palette의 목적은 달성되었다고 볼 수 있을 것이다.

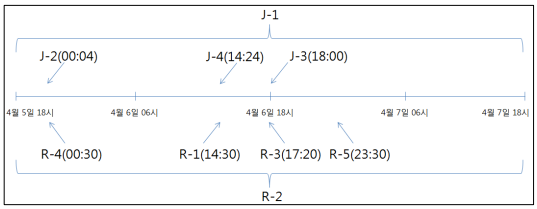


Fig. 18. Comparison of analysis results

V. 결론 및 향후 계획

앞 장에서 우리는 VAST Challenge 2012 데이터세트에 실제로 RGB Palette를 적용하여 RGB Palette의 유효성 검증을 시도하였다. 그 결과 Port Scanning, 웜의 활동, DDos 공격 등 보안 위협 요소들을 발견할 수 있었다. RGB Palette의 급격한 색상 변화를 인지함으로써 네트워크상에서 무엇인가 유해한 상황이 발생하고 있다는 것을 쉽게 인지할 수 있다는 것을 알 수 있었다.

하지만 이 도구는 관제 인원에게 1차적인 보안 위협상황 발생 경고를 제공하는 것을 목적으로 하였기 때문에 상세한 로그 분석을 위해서는 추가적인 분석 장비를 필요로 한다는 한계가 존재한다.

그리고 VAST Challenge 2012 데이터세트에서는 이틀간의 로그 데이터만 제공되어 평소 정상 상황의 데이터를 활용한 학습값의 생성에 한계가 있었다. 향후에는 실제 네트워크 시스템 운영 환경에서 평상시의 각 시간대별 수치를 학습하여 그 차이 부분을 RGB Palette를 이용하여 표현한다면 더욱 차별화된 경고(alarm)를 제공할 수 있을 것이다.

References

[1] Gwang-sun Choi, "Bigdata Visuali- zation", Korea Society of Computer Information, 2013

[2] Say Min, "What is data visualization," TED, 2012

[3] G. Shurkhovetskyy, "Visual Analytics for Network Security," VAST challenge 2012, 2012

[4] L. Shi, "Investigating Network Traffic Through Compressed Graph Visuali- zation", VAST challenge 2012, 2012

[5] H. Gibson, "Network Infrastructure Visualization Using High-Dimensional Node-Attribute Data," VAST challenge 2012, 2012

[6] Y. Zhao, "A Real-Time Visualization System for Network Security", VAST challenge 2012, 2012

[7] Y. Cao, "Dynamic Analysis of Large Datasets with Animated and Correlated Views", VAST challenge 2012, 2012

[8] Y. Barcelos, "Combining traditional and high-density visualizations in a dash-board to network health monitoring", VAST challenge 2012, 2012

[9] L. Harrison, "situ: Situational Understanding and Discovery for Cyber Attacks", VAST challenge 2012, 2012

[10] J. Hildenbrand, "Chart- and Matrix-based Approach to Network Operations Forensics," VAST challenge 2012, 2012

[11] M. Zhao, "Federating Geovisual Analytic Tools for Cyber Security Analysis", VAST challenge 2012, 2012

[12] Hyun-sang Choi, "Fast detection and visualization of network attacks on parallel coordinates", Computer & Security 28, pp.276-288, 2009

[13] Chi-yoon Jeong, "An Efficient Method for Analyzing Network Security Situation Using Visualization," Journal of The Korea Institute of information Security &Cryptology, 19(3), pp. 107-117, June 2009

[14] W. Urbanski, "Cover-VT: Converged Security Visualization Tool," IEEE, pp.714-717, May 2011

[15] T.J. Jankun-Kelly, "A Visual Analytic Framework for Exploring Relationships in Textual Contents of Digital Forensics Evidence," IEEE, pp.39-44, Oct. 2009

[16] K. Lakkaraju, "Closing-the-Loop: Discovery and Search in Security Visualizations," IEEE, pp.58-63, June 2005

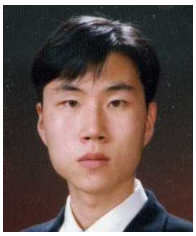
[17] R.M. Savola, "A Visualization and Modeling Tool for Security Metrics and Measurements Management," IEEE, pp.1-8, Aug. 2011

[18] C. Muelder, "A Visualization Methodology for Characterization of Network Scans," IEEE, pp.29-38, Oct. 2005

[19] D. Barrera, "Accommodating IPv6 ad-

dresses in security visualization tools," Information Visualization (2011) 10, pp.107-116. Nov. 2010
〔20〕 G. Conti, "Countering Security Information Overload through Alert and Packet Visualization," IEEE, pp.60-70, Mar. 2006
〔21〕 R.F. Erbacher, "Designing Visualization Capabilities for IDS Challenges," IEEE, pp.121-127, Oct. 2005
〔22〕 P. Ren, "IDGraphs: Intrusion Detection and Analysis Using Histographs," IEEE, pp.39-46, Oct. 2005
〔23〕 M. Alsaleh, "Improving Security Visualization with Exposure Map Filtering," IEEE, pp.205-214, Dec. 2008
〔24〕 J. Steele and N. Illinsky, Jin-hong Kim, "Beautiful Visualization," pp.2-4, pp.24-25, 2012

〈저자소개〉



이 동 건 (Dong-gun Lee) 정회원
2003년 3월: 고려대학교 컴퓨터교육과 졸업
2013년 9월~현재: 고려대학교 정보보호대학원 석사과정
2007년 4월~현재: 한국주택금융공사 근무
〈관심분야〉 정보보호, 금융보안, 데이터 시각화

김 휘 강 (Huy Kang Kim) 종신회원
1998년 2월: KAIST 산업경영학과 학사
2000년 2월: KAIST 산업공학과 석사
2009년 2월: KAIST 산업및시스템공학과 박사
2004년 5월~2010년 2월: 엔씨소프트 정보보안실장, Technical Director
2010년 3월~현재: 고려대학교 정보보대학원 조교수
〈관심분야〉 온라인게임 보안, 네트워크 보안, 네트워크 포렌식

김 은 진 (Eunjin Kim) 정회원
1999년 2월: KAIST 산업경영학과 졸업
2001년 2월: KAIST 경영공학과 석사 졸업
2007년 8월: KAIST 경영공학과 박사 졸업
2008년 9월~현재: 경기대학교 국제산업정보학과 부교수
〈관심분야〉 경영정보시스템, 보안경제학