Journal of the Ergonomics Society of Korea
Vol. 30, No. 2 pp.319-330, April 2011 DOI:10.5143/JESK.2011.30.2.319

An Effective Information Visualization Technique for Intrusion Detection: Hyperbolic View Intrusion Visualizer

Yunseok Jeong, Rohae Myung

Department of Industrial Management Engineering, Korea University, Seoul, 136-713

ABSTRACT

In computer forensics investigation, the investigators collect, protect, analyze and interpret massive amount of data which were used in cyber crime. However, due to its huge amount of information, it takes a great deal of time and errors often result even when they use forensics investigation tool in the process. The information visualization techniques will greatly help to improve the information processing ability of human when they deal with the overwhelming amount of data and have to find out significant information in it. The importance of Intrusion Detection System(IDS) among network forensics is being emphasized in computer forensics. In this study, we apply the information visualization techniques which are proposed to be a great help to IDS and carry out the usability test to find out the most effective information visualization techniques for IDS.

Keywords: Computer forensic, Network security, Intrusion detection, Information visualization, Focus+Context

1. 서 론

정보화 시대가 도래하면서 우리는 정보의 홍수 속에서 살고 있다. 이러한 대량의 정보 속에서 컴퓨터의 사용은 일상 생활과 사무환경에서 정보의 검색과 작업수행을 용이하게 하는 순기능을 가지고 있으나, 컴퓨터를 이용한 범죄의 증가와 같은 역기능도 나타나고 있는 실정이다(박구, 2006). 특히 최근 악의적인 해킹이나 산업 스파이에 의한 정보 유출 등의 컴퓨터 범죄가 날로 늘어나고 있는 추세이다. 이러한 컴퓨터 범죄 수사를 수사하는 데에 있어서 Encase, Helix, 또는 ILOOK과 같은 컴퓨터 포렌식 수사 지원 소프트웨어 사용의 중요성이 대두되고 있지만, 수사관이 이들 소프트웨어를 사용하더라도 여전히 수사를 하는데 많은 양의 데이터를 분석해야 해야 하며 데이터의 형식이 그림 1과 같이 텍스트 형식으로 되어 있기 때문에 수사관들의 가독성이 떨어지며, 수사에 있어서 많은 시간이 소요되며, 에러가 발생할 수 있다(조선영, 2006).

현재 주로 사용되고 있는 포렌식 수사 도구로는 국내 검찰청에서 사용하는 D.E.A.S.와 Final data사의 Final Forensics, A3Security사의 Auto Watch가 있으며 해외의 포렌식 수사 도구로는 Encase, SafeBack, ProDiscover가 있다(조선영, 2006). 그 중 제일 많이 사용되고 있는 포렌식 수사 도구인 그림 2의 Encase는 텍스트 형식으로 되어 있는 다른 수사 도구에 비해 그래픽 인터페이스를 많이 제공하는 편이기는 하지만, 표 1에서 볼 수 있는 바와 같이 그 적용된 시각적인 기법은 미비한 실정이다.

Encase 뿐만 아니라 기존에 나와 있는 수사 도구들의 문제점은 모든 증거자료가 존재할 수 있는 영역을 수사해야 하는 포렌식 수사에서 각 영역별 구분이 어렵다는 점과 Raw data(사람이 이해하기 어려운 데이터 형태)로만 표현되어 수사관의 가독성이 떨어진다는 점이다. 또한 방대한 양의 데이터를 감시하고 분석하는 것은 수사관의 인지적 부하를 증가시키며 수사를 하는데 많은 시간이 소요되게 하며, 에러 발생 또한 증가할 수 있는 어려움이 있다.

따라서 이런 어려움과 문제점을 해결하기 위해서 포렌식

Corresponding Author: Rohae Myung. Department of Industrial Management Engineering, Korea University, Seoul 136-713.
Phone: 02-3290-3392, E-mail: rmyung@korea.ac.kr
Copyright@2011 by Ergonomics Society of Korea(pISSN:1229-1684 eISSN:2093-8462). All right reserved.

15:54:29.650359 IP (tos Oxc0, ttl 2, 1d 0, offset 0, flags [none], length: 48) 128.173.55.254.1985 > 224.0.0.2.1985: [udp sum ok] HSRPv0-hello 20: state-active group-0 addr-128.173.48.1 hellotime-is holdtime-4s priority-200 auth-"clsco-0-0-0"
15:54:29.667909 arp who-has 128.173.54.230 tell 128.173.55.254
15:54:29.754435 IP (tos Ox0, ttl 58, 1d 23447, offset 0, flags [DF], length: 10 0) 128.173.37.81.42981 > 128.173.54.135.22: P 1:49(48) ack 48 win 65535 <nop, nop, timestamp 1011302959 14759405>
15:54:29.789216 IP (tos Ox10, ttl 64, 1d 15519, offset 0, flags [DF], length: 5 2) 128.173.54.135.22 > 128.173.81.42981: . [tcp sum ok] 48:48(0) ack 49 win 16704 <nop, nop, timestamp 14759489 1011302959>
15:54:29.790098 IP (tos Oxc0, ttl 2, 1d 0, offset 0, flags [none], length: 48) 128.173.55.253.1985 > 224.0.0.2.1985: [udp sum ok] HSRPv0-hello 20: state-stand by group-0 addr-128.173.48.1 hellotime-is holdtime-4s priority-100 auth-"clsco-0-0-0"
15:54:29.989909 arp who-has 128.173.55.109 tell 128.173.55.253
15:54:29.995008 arp who-has 128.173.52.245 tell 128.173.55.253

그림 1. 컴퓨터 포렌식 수사에 사용되는 데이터의 예 (Ball et al., 2004)

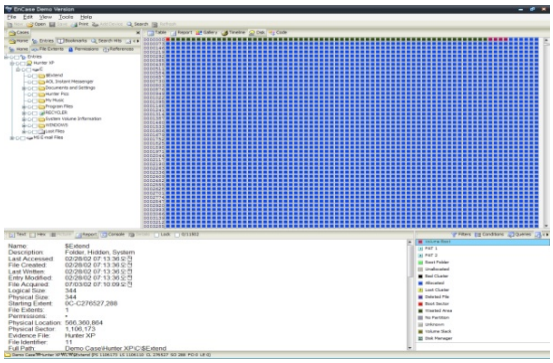

그림 2. Encase의 데이터 표현(Bunting and Wei, 2006)

표 1. 현재 Encase에서 사용된 시각적 요소

포렌식 수사에 필요한 정보	적용된 시각적인 기법
디스크 정보	텍스트
디스크에서의 데이터의 위치	블록으로 표현
데이터 계층 구조	윈도우와 같은 형태로 표현
데이터의 속성(생성 시간, 파일 형태, 사용 시간, 삭제 여부, 삭제 시간)	텍스트
삭제된 데이터의 복구가능 여부	아이콘
은닉된 데이터의 유무 검색	아이콘
타임라인	Zooming으로 표현
네트워크 접속 여부	표현되지 않음
네트워크 연결 상태	표현되지 않음

수사, 특히 현재 그 중요성이 더욱 부각되고 있는 네트워크 포렌식 수사 중 침입탐지에 있어 수사관들의 분석 시간을 단축하고, 이해하기 쉽고, 에러를 줄이고 수행도를 높일 수 있는 효율적인 포렌식 수사를 할 수 있는 방안을 연구하고, 더 나아가 이해하기 쉬운 법정자료 생산을 모색하고자 한다.

이러한 컴퓨터 포렌식 수사의 문제점 해결에 관해서는 Teerlink 등(2006)에 의해서, 많은 정보 속에서 원하는 정보를 찾아내기 위한 방법으로 가장 유용하게 쓰이는 정보 시각화가 그 해결책이 될 수 있음을 시사한 바 있다. 그러므로 본 연구에서는 실제 침입 상황에 사용할 수 있는 새로운 시각화 기법인 Hyperbolic View Intrusion Visualizer (HVIV)를 디자인하여, 이에 대한 사용성 평가를 실시하여 그 효용성을 알아보고자 한다.

2. 이론적 배경

2.1 정보 시각화

Card 등(1999)은 정보 시각화란 인간의 인지적 능력을 향상시키기 위해 컴퓨터를 이용하여 추상적인 데이터를 상호작용이 있는 시각적 표현으로 나타내는 것이라고 하였다. 이런 정보 시각화의 장점은 데이터에 대한 통찰력을 높여줌으로써 방대한 양의 데이터에 대해 빠른 이해를 가능하게 해주고, 데이터의 특정한 패턴을 인식하는데 매우 유용하다는 것이다. 그리고 예상치 못했던 데이터의 특성에 대한 사용자의 지각을 빠르게 하여 에러가 발생했을 시 즉각적인 인식을 가능하게 해줌으로써 에러를 줄여주고 작업에 있어서 수행도를 향상시켜 준다(Ware, 2004).

이런 정보 시각화의 기법에는 여러 가지가 있고, 이들 정보 시각화 기법을 구분할 수 있는 기준은 여러 가지가 있으나 널리 알려진 방법으로는 Nondistortion-oriented Technique과 소위 Focus+Context라고 불리는 Distortion-oriented Technique으로 구분할 수 있다(Leung and Apperley, 1994). Nondistortion-oriented Technique은 디스플레이에 정보의 일부분을 보여준 후 Scrolling 등을 통하여 다른 정보를 볼 수 있게 하거나, 간략한 정보를 보여준 후 계층적 구조를 만들어서 점점 자세히 보여주거나 Zooming하는 방법으로 거대한 정보 공간에서 사용자가 항행(navigating)할 때 적당한 크기의 화면을 보여줄 수 없다는 것이다. 반면에 Distortion-oriented Technique (Focus+Context)은 이런 Nondistortion-oriented Technique의 단점을 극복할 수 있는 방법으로써 사용자가 특정 부분의 정보를 자세하게 봄과 동시에 전체적인 모습을 보여줘서 쉽게 항행할 수 있으며, 그래픽적 디스플레이에서 적은 비용으로 높은 수행도를 얻을 수 있다(Leung and Apperley, 1994). 그림 3과 그림 4는 각각 Nondistortion-oriented Technique과 Distortion-oriented Technique의 하나의 예인 Scatter Plots와 Perspective Wall을 보여주고 있다.

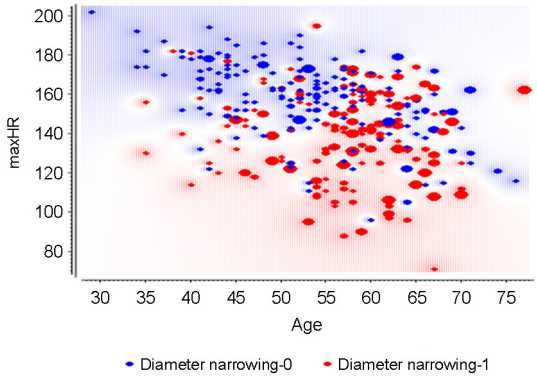

그림 3. Scatter plots

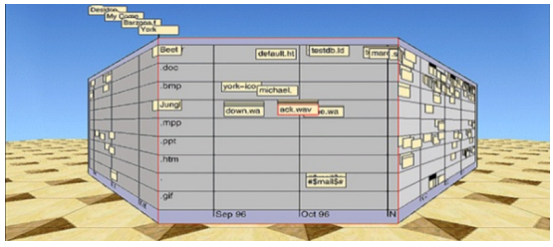

그림 4. Perspective wall

2.2 컴퓨터 포렌식

컴퓨터 포렌식이란 컴퓨터 범죄에 있어 법적으로 유효한 전자적 증거를 수집하고 분석하는 제반 행위이다(전상덕 등, 2006). 컴퓨터 포렌식은 시스템 포렌식과 네트워크 포렌식으로 분류된다(박종성, 2004). 전자는 피해 시스템의 초기 증거(Live Data)를 수집하고 분석하는 행위 또는 디스크 포렌식 수행을 통해 얻은 증거자료를 법적으로 사용하거나 사고 처리와 사건재발방지 차원의 수사에 사용하는 기술 및 절차를 말하며, 후자는 침해 사고와 관련된 네트워크 이벤트를 수집, 분석, 저장, 응용하는 일련의 기술 및 절차를 말한다.

현재까지의 포렌식은 사이버 경찰청과 같은 수사기관에서 아동 포르노나 인터넷 사기, 도박 등의 증거 확보 및 분석을 위해 사용되어 왔다. 그렇기 때문에 네트워크 포렌식보다 시스템 포렌식에 대한 연구만이 활발히 이루어졌고 많이 진척되어 있지만, 현재 컴퓨터 기술 및 네트워크 기술이 발전하면서 이제 모든 정보는 네트워크를 통해 움직인다고 해도 과언이 아니다(박종성, 2004). 또한 인터넷 사용이 활발해지면서 네트워크를 통한 불법침입이 늘어나고 있는 추세이지만 방대한 양의 네트워크 데이터를 감시하는 것은 상당히 어렵고 시간이 오래 걸리는 작업이어서 이에 대한 효과적인 방법이 필요하다. 따라서 효율적인 네트워크 포렌식을 위한 정보 시각화 기법의 중요성이 더욱 부가되고, 이러한 요구에 따라 본 연구에서는 네트워크 포렌식, 특히 그 중에서도 가장 대표적이며 위험성 및 중요성이 날로 증대되고 있는 침입탐지에 주안점을 두었다.

네트워크 데이터는 헤더파일에 중요한 내용을 표시하게 되는데 이는 Source IP address, Destination IP address, Source port, Destination port, 그리고 패킷이라는 기본 정보 단위를 갖는 정보 전송량 등이다. 이 데이터는 사용하고 있는 컴퓨터에서 어떤 프로그램과 포트를 통해 다른 컴퓨터와 연결되어 있는지를 나타내고, 불법적인 네트워크 연결이 구성되어 있는지 확인할 수 있는 기본 정보가 된다(배준우, 2008). 이는 TCPDUMP나 Ethereal과 같은 네트워크 모니터링 툴에서 그림 5와 같은 텍스트 형식의 데이터로 나타난다.

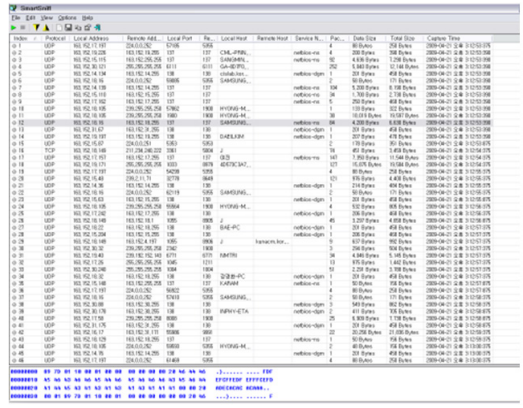

그림 5. Ethereal 등에서의 텍스트 기반의 데이터

2.3 네트워크 포렌식 수사에 적용 가능한 정보 시각화 기법: Focus+Context

현재 네트워크 포렌식 수사 도구에 그 적용이 미흡한 정보 시각화 기법을 적용하게 되면 제한된 시간 안에 수사관의 능력을 초과하는 많은 정보 처리의 요구 속에서 데이터의 분석 시간이 단축될 뿐만 아니라, 이해하기 쉬운 법정자료를 생산할 수 있으며 에러를 줄이고, 수행도를 높일 수 있는 장점이 있다(Teerlink and Erbacher, 2006).

이러한 네트워크 데이터를 분석함에 있어 고려되어야 할 중요한 속성은 상호간의 연결, 접속 빈도, 정보 전송량 등이 있는데, 이러한 네트워크 데이터 속성에 맞는 시각화 기법으로써 여러 방법이 제안되었는데, 일반적으로 많이 쓰이는 기법은 Parallel coordinates(Ball et al., 2004; Yin et al., 2004; Choi and Lee, 2005; Conti, 2007)이다. 그림 6은 실제 Conti(2007)이 제안한 Rumint에서 Source IP address,

Destination IP address, Source port, Destination port, Packet Length 등을 Parallel coordinates로 나타낸 것이다.

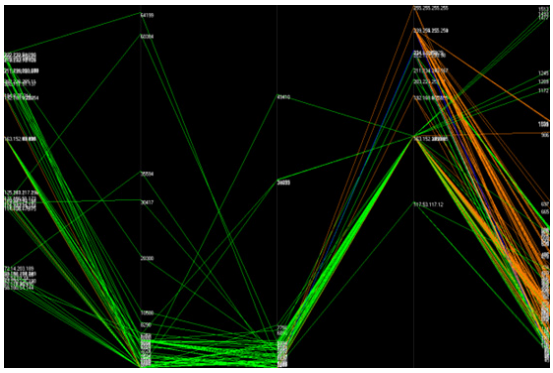

그림 6. 패킷 데이터의 Parallel coordinates(Conti, 2007)

하지만 이런 Parallel coordinates는 정보 시각화 기법 중 Nondistortion-oriented Technique으로써 데이터의 양이 많을 때 사용자는 Pan, Zoom, Scroll 등을 통해 정보를 검색해야 하고, 이에 따라 사용자는 정보 데이터 속에서 항행 시 전체 구조를 볼 수 없어서 그 방향을 잃게 되고, 현재 보여지지 않은 정보와의 연관성을 알 수 없다. 또한 이 기법의 문제점은 황보택근과 이수진(2005)의 연구에서 제안되었듯이 Parallel Coordinates의 세로축에서 데이터의 값들이 교차되면 선의 연결성과 시작점 등을 파악할 수 없게 된다.

따라서 많은 양의 정보를 화면에 전부 나타내면서도 어느 특정 부분에 더 많은 Display space를 주어서 크기를 확대하는 기술의 적용이 필요하다고 볼 수 있으며, 이러한 문제점을 해결하기 위해서는 전체적인 문맥을 살피면서 관심을 가지려는 데이터를 확대하여 볼 수 있는 Focus+Context 기법이 네트워크 포렌식 수사에 매우 적합할 것이다.

Focus+Context 기법은 Funas(1997, 2006)의 연구에서 알 수 있듯이 많은 정보 속에서 사용자가 필요로 하고, 원하는 정보가 있는 곳은 확대하여 보여주고, 나머지 부분은 전체 데이터의 관계와 흐름을 파악할 수 있도록 삭제하지 않고 특정 비율로 줄여주는 기법이므로 많은 정보 속에서 사용자가 원하는 정보를 찾아야 하는 컴퓨터 포렌식의 취지상 잘 부합한다고 할 수 있다.

인간은 local detail과 global context를 사용하여 실세계를 인지하는데 생물학적으로 우리의 눈은 focus된 부분을 통해 사물을 자세히 보며 이 focus된 부분을 인지하고 그 주변을 연결하여 전체적인 모습을 이해하게 된다(Schaffer et al., 1996). Focus+Context 기법은 이런 focus된 영역을 자연스럽게 확장해주는 기능을 함으로써 사용자의 인지 능력을 기존의 다른 시각화 기법보다 향상시켜준다(김기준 등, 1999).

Focus+Context 기법의 대표적인 예로는 Munzner와 Burchard(1995, 1997), Lamping 등(1996)의 Hyperbolic View가 있으며, 그래프(Sarker and Brown, 1992)에서 트리 구조(Lamping et al., 1995), 메뉴(Bederson, 2000; 황규민, 2004)에 이르기까지 많은 분야에 적용된 Fisheye View(Furnas, 1986)가 있다. 또한 Schaffer et al.(1996)은 다초점 줌을 Focus+Context 기법에 접목시켜 의미론적 Focus+Context 기법을 시도해 보았고, Keahey와 Robertson(1997)은 다초점 Focus+Context 기법을 적용하여 사용자가 동시에 여러 지역을 살펴볼 수 있도록 하는 시스템을 개발하였다.

이러한 Focus+Context 기법을 네트워크 포렌식 수사에 적용하면 표현성과 효율성을 높이며 수사관의 수행도 향상에 도움을 줄 수 있을 것이다. 가령 Focus+Context 기법 중 Lamping 등(1996)이 제안한 Hyperbolic View(그림 7 참조)를 이용하여 인터페이스를 디자인하면 수사관은 디스크에서 원하는 데이터의 위치와 계층 구조 등 전체적인 정보를 파악하면서 바로 원하는 데이터의 속성, 관계, 불법행위 여부 등을 알 수 있을 것이다.

그림 7. Hyperbolic View의 예시
(마우스 클릭에 의해 Focus가 달라짐)

3. HVIV (Hyperbolic View Intrusion Visualizer)

본 HVIV(Hyperbolic View Intrusion Visualizer)는 한정된 디스플레이 안에 많은 정보를 표현하여 관심 있는 영역을 자세히 볼 수 있는 Focus+Context 기법 중 네트워크 데이터에 가장 적합하다고 생각되는 Hyperbolic View를 이용하여 디자인되었다.

본 연구에서 시각화하고자 하는 것은 TCPDUMP나

Ethereal과 같은 네트워크 감시 툴의 텍스트 기반의 데이터다. 윤성종과 김정호(2006)의 연구에 의하면 네트워크에서 소통되는 packet 중에서 해킹 공격에 사용되는 유해 packet은 Source IP address, Destination IP address, Destination port의 3가지 필드를 기준으로 분석하여 탐지할 수 있고, HVIV는 이를 참고하여 이들 데이터의 시각화에 주안점을 두었다. HVIV의 주요 기능은 다음과 같다.

• Overview

HVIV의 중심(root)을 마우스로 클릭하면 전체적인 Overview을 볼 수 있다(그림 8 참조).

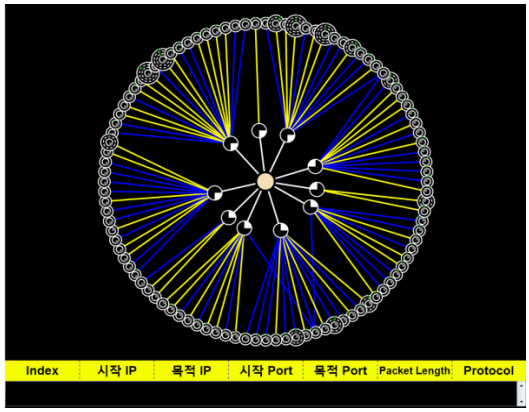

그림 8. 정상적인 상황에서의 HVIV의 Overview(root 클릭 시)

• 외부(Source) Host

HVIV에서 Overview을 나타내는 root를 지나 제1계층에 외부 Host를 나타낸다. 이는 외부 Host가 내부 Host와 어떤 정보를 얼마나 주고 받는지 알 수 있게 한다. 중요한 정보인 Source IP address는 어느 정도의 범위에 속하는지 알기 위해서 원의 4분면 안에 해당 영역을 나타내었다. 관심이 있는 Host를 클릭하면 클릭한 Host(노드)가 확대되면서 화면의 중심에 오고, 나머지 노드들은 없어지지 않고 일정 비율로 축소가 되면서 화면의 가장자리로 이동하게 된다. 그리고 이와 관련된 자세한 부가 정보가 아래의 창에 나타나게 된다(그림 9 참조).

• 내부(Destination) Host

HVIV 상에서 제2계층에 내부 Host를 나타낸다. 내부 Host와의 관계를 알 수 있게 하며 호스트 부분의 Destination IP address가 어느 정도의 범위에 속하는지 알 수 있도록 하기 위해 원의 4분면 안에 해당 영역을 나타낸다. 마찬가지로 사용자가 관심 있는 Host를 클릭하면 클릭한 Host(노드)가 확대되면서 화면의 중심에 오고, 나머지 노드들은 없어지지 않고 일정 비율로 축소가 되면서 화면의 가장자리로 이동하게 된다. 그리고 이와 관련된 자세한 부가 정보가 아래의 창에 나타나게 된다(그림 10 참조).

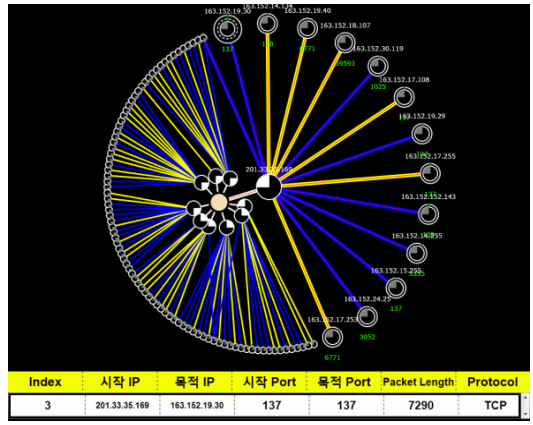

그림 9. 그림 8에서 제1계층(외부 Host)을 클릭했을 때의 HVIV

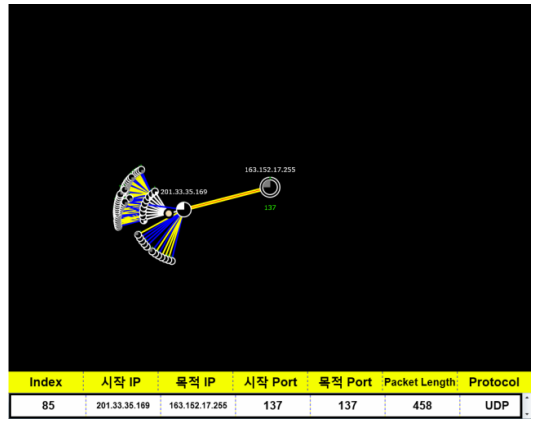

그림 10. 그림 8에서 제2계층(내부 Host)을 클릭했을 때의 HVIV

• Destination port

네트워크 상에서의 침입 행위나 공격 행위 탐지에서 중요한 정보인 Destination port는 내부 Host의 바깥에 원을 하나 더 둘러서 Port의 크기에 해당하는 영역을 선으로 표시하였다. 이는 Port의 사전적 의미인 '입구'에서 연상되는 Metaphor를 활용하였다. 일반적인 정상의 Port는 초록색으로 시각화되지만, Worm Port와 같이 비정상적인 Port를 통해 정보를 전송 받았을 때에는 빨간색으로 나타난다(그림 8, 9, 10 참조).

• Packet Length

외부 Host와 내부 Host가 주고 받은 정보의 양은 내부

Host를 두른 바깥의 원의 크기로써 나타내었다. 이는 사용자가 인지하기 쉽도록 3단계(대 · 중 · 소)로 표현하였고, 자세한 정보는 관심 있는 Host를 클릭하면 이와 관련된 자세한 정보가 아래의 창에 나타난다(그림 8, 9, 10 참조).

• Protocol

외부 Host에서 내부 Host로 정보를 전송 시, 양방향성을 갖는 TCP와 같은 Protocol은 제1계층과 제2계층을 연결하는 파란색으로 시각화를 했으며, 단방향성을 갖는 UDP, ICMP와 같은 Protocol은 노란색으로 나타내었다(그림 8, 9, 10 참조). 또한 연결선의 두께를 통해 통신의 횟수를 시각화하였다. 이에 따라 Port Scan과 같이 외부 Host에서 어떤 내부 Host로 통신 횟수가 많아질 때는 이 선이 두꺼워지게 된다(그림 8, 9, 10 참조).

4. 연구 방법

4.1 피실험자

피실험자는 컴퓨터와 무관한 전공을 갖고 있는, 네트워크에 대해 트레이닝을 받지 않은 비전문가 9명(남자 7명, 여자 2명)과 컴퓨터나 네트워크와 관련된 전공을 이수한 석 · 박사 연구원과 같은, 네트워크에 대한 트레이닝을 받은 전문가 7명(남자 5명, 여자 2명)으로 구성되었다. 비전문가의 평균 나이는 24.8세(SD=1.4), 전문가의 평균 나이는 28.3세(SD=1.1)이며 비전문가의 평균 컴퓨터 사용 경력은 10.8년(SD=1.6), 전문가의 평균 컴퓨터 사용 경력은 13.0년(SD=1.2)이다.

실제로 HVIV와 같은 네트워크 감시 툴은 비전문가가 사용하기 보다는 네트워크 관리자와 같은 전문가가 사용하여 네트워크 상황을 효율적으로 판단하는 도구로 사용될 수 있을 것이지만, 네트워크 전문가는 네트워크의 상황에 익숙하며, 네트워크의 커뮤니케이션 패턴이 '어떠해야 한다'는 고정된 Mental model을 갖고 있기 때문에 비전문가보다 쉽게 정상적인 상황과 비정상적인 상황을 구별할 수 있다. 그리고 이전까지 사용하던 텍스트 형식의 툴에 익숙하기 때문에 정보 시각화 기법이 적용된 네트워크 모니터링 툴에 대한 거부감이나, 훈련 부족 등으로 인해 공정한 평가가 될 수 없을 가능성이 존재하는데 반하여 네트워크 상황을 많이 접해보지 못한 비전문가 사용자는 그러한 편견으로 인한 결론을 쉽게 내지 않기 때문에 전문가와 비전문가 두 집단을 통해 실험을 실시하게 되었다.

4.2 실험 절차

실험은 피실험자들이 실제 네트워크 Security Operator가 되어 어떠한 침입을 탐지한다는 시나리오를 가정하고, 한 개의 정상적인(네트워크 상에서 침입 행위가 일어나지 않은) 상황과 두 개의 비정상적인 상황(Port Scan, Worm Scan)에 대해 현재 네트워크 관리자들이 제일 많이 사용하고 있는 텍스트 형식으로 된 네트워크 모니터링 툴(그림 11 참조), 그리고 Visual Fingerprinting(Conti and Abdullah, 2004), Rumint(Conti, 2007), PVF(Stolze et al., 2003), VisFlowConnect(Yin et al., 2004), PCAV(Choi and Lee, 2005) 등 정보 시각화 기법이 적용된 네트워크 모니터링 툴에서 제일 많이 쓰이고 있는 방식인 Parallel coordinates가 적용된 툴(그림 12, 13 참조), 그리고 Focus+Context 기법이 적용된 HVIV(그림 14 참조)를 통하여 각각 침입 행위를 탐지하는데 걸리는 수행 시간을 측정하고 주관적 만족도 및 선호도를 측정하였다.

Index	시작 IP	목적 IP	시작 Port	목적 Port	Packet Length	Protocol
1	34.46.63.176	163.152.17.9	5355	57185	250 Bytes	UDP
2	111.222.69.35	163.152.19.216	137	137	390 Bytes	TCP
3	201.33.35.159	163.152.19.20	137	137	7,290 Bytes	TCP
4	111.222.69.35	163.152.30.111	6111	6111	12,144 Bytes	UDP
5	201.33.35.159	163.152.14.124	138	138	458 Bytes	ICMP
6	21.205.39.189	163.152.18.3	5355	59085	171 Bytes	TCP
7	193.108.92.143	163.152.14.129	137	137	8,190 Bytes	ICMP
8	193.108.92.143	163.152.15.100	137	137	2,730 Bytes	UDP
9	21.205.39.189	163.152.17.152	137	137	468 Bytes	TCP
10	111.222.69.35	163.152.18.95	1900	57862	322 Bytes	UDP
11	111.222.69.35	163.152.18.94	1900	1900	19,597 Bytes	UDP
12	48.2.1.73	163.152.18.7	137	137	6,630 Bytes	ICMP
13	111.222.69.35	163.152.31.57	138	138	458 Bytes	TCP
14	21.205.39.189	163.152.19.187	138	138	470 Bytes	UDP
15	111.222.69.35	163.152.15.77	5353	5353	351 Bytes	UDP
16	83.43.189.162	163.152.18.139	5004	3361	3,459 Bytes	TCP
17	32.13.243.3	163.152.17.147	137	137	11,544 Bytes	ICMP
18	111.222.69.35	163.152.19.161	8979	1033	19,584 Bytes	UDP
19	83.43.189.162	163.152.17.187	5355	54299	250 Bytes	TCP
20	83.43.189.162	163.152.15.38	8649	32778	4,400 Bytes	UDP
21	21.205.39.189	163.152.14.26	138	138	484 Bytes	TCP
22	48.2.1.73	163.152.18.98	5355	62119	171 Bytes	TCP
23	83.43.189.162	163.152.15.53	138	138	458 Bytes	UDP
24	111.222.69.35	163.152.18.96	1900	55564	805 Bytes	TCP
25	21.205.39.189	163.152.17.232	138	138	468 Bytes	UDP
26	32.13.243.3	163.152.18.137	8905	1055	4,658 Bytes	TCP
27	21.205.39.189	163.152.18.12	138	138	458 Bytes	UDP
28	83.43.189.162	163.152.15.194	138	138	468 Bytes	TCP
29	111.222.69.35	163.152.18.140	8906	1055	992 Bytes	TCP

그림 11. 실제 실험에 사용된 텍스트 기법이 적용된 네트워크 감시 툴

4.2.1 수행 시간 측정

피실험자에게 신상 및 컴퓨터 사용 경력에 대한 간단한 설문조사 후 네트워크에 대한 기본적인 지식과 침입 행위의 유형들에 대한 설명을 한 후, 텍스트, Parallel coordinates 그리고 HVIV 각각의 기법을 통해 하나의 정상적인 상황의 데이터로 네트워크 상황 판단에 대해 1시간의 트레이닝을 하도록 하고(각각의 기법당 20분씩 소요), 이에 익숙해지면

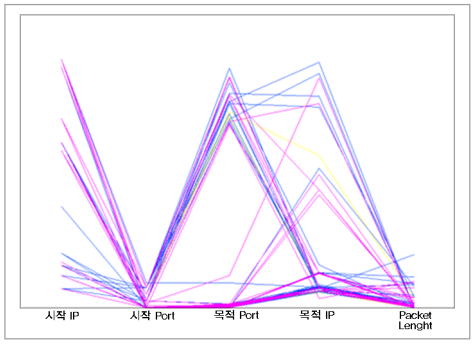

그림 12. 실제 실험에 사용된 Parallel coordinates 기법이 적용된 네트워크 감시 툴

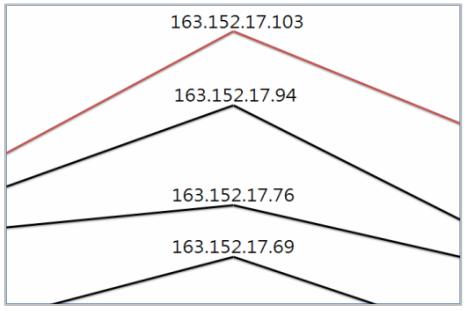

그림 13. 그림 12를 마우스 드래그를 통해 Zoom-in한 모습

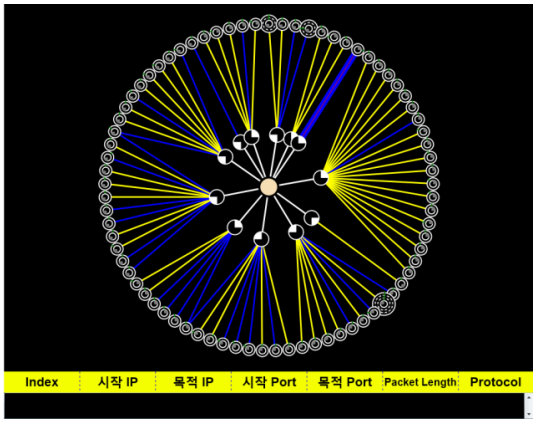

그림 14. 실제 실험에 사용된 HVIV(Port Scan일 경우)

실험용 데이터를 보여주고, 표 2와 같은 질문을 주고 답을 찾는데 걸리는 시간을 측정하였다. 실험은 within-subject design으로 진행하였으며 학습 효과를 방지하기 위해 텍스트, Parallel coordinates, HVIV가 같은 상황(정상, 비정상1(Port Scan), 비정상2(Worm Scan))에 대해서는 네트워크 데이터의 수와 형태는 같지만 Source IP address와 Destination IP address의 실제 숫자는 무작위로(random) 하였으며, 실험 순서를 counter-balancing하였다(그림 15, 16 참조).

표 2. 수행시간 측정에 사용된 질문 예시

I. 해킹 발생의 유무 판단 질문
1-1. 이 데이터에서 침임 행위가 발생하였는가.
1-2. 침입 행위가 발생했다면 어떤 경우인지 표현하시오.
II. 해킹 상황에서의 네트워크의 세부 정보 판단 질문
2-1. 이 때 공격받고 있는 Host(들)는 무엇인가.
2-2. 공격받고 있는 Host(들)는 무엇인가.
2-3. 공격받고 있는 Host는 어떤 Port(들)를 통해 공격받고 있는가.
III. 네트워크의 해킹 상황과는 별개로, 주어진 네트워크의 상황을 얼마나 잘 이해하고 있는지에 대한 부수적인 질문
3-1. 가장 큰 양의 packet length를 받고 있는 내부 host는 무엇인가.
3-2. (3-1.에 이어)그 때의 큰 양의 packet length를 내부 host에 주고 있는 외부 host는 무엇인가.
3-3. 가장 많은 수의 내부 host와 연결되어 있는 외부 host는 무엇인가.
3-4. 가장 많은 수의 외부 host와 연결되어 있는 내부 host는 무엇인가.

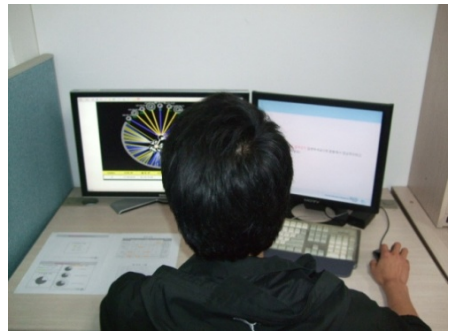

그림 15. 피실험자가 실험을 수행하고 있는 모습(1)

그림 16. 피실험자가 실험을 수행하고 있는 모습(2)

질문에 대한 내용을 읽고 이해하는 개인간의 시간차는 실험에 고려하지 않기 위해서 듀얼 모니터를 사용하여 한쪽의 모니터에 질문이 하나씩 뜨고, 피실험자가 그 질문을 다 읽고 이해하면 손을 들고 다른 쪽의 모니터에 있는 네트워크 모니터링 툴을 보고 실험을 실시하였으며, 수행 시간 분석을 위해서는 Camtasia program(v.6.02)를 통해 작업수행을 디지털로 녹화하고 이를 재생시켜 각 시간을 분석하였다. 트레이닝과 수행 시간 측정 평가 실험은 피실험자에 따라 전문가일 경우 최저 80분에서 비전문가일 경우 최고 180분이 소요되었다.

4.2.2 주관적 만족도, 주관적 선호도 측정

수행 시간 측정 실험이 끝난 직후 피실험자에게 각각의 네트워크 모니터링 툴에 대해서 네트워크 상황을 판단하는 기법에 대한 주관적 만족도, 만일 자신이 네트워크 Security Operator로서 네트워크 데이터를 모니터링하고 위험 상황을 판단해야 한다면 가장 선택하고 싶은 방식(주관적 선호도)를 물었다. 주관적 만족도는 100점 척도로 답하게 하였고, 주관적 선호도는 본 연구에서 평가한 3개의 기법 중 하나를 선택하였다.

5. 실험 결과

실험 데이터는 표 2의 각각의 문항에 대해 질문을 답하는데 걸리는 수행 시간과 같은 객관적 척도와 만족도, 선호도 등의 주관적 척도가 종속 변수로 사용되었다.

5.1 수행 시간 측정

5.1.1 총 수행 시간 비교

비전문가 집단

9명의 비전문가 집단에서 텍스트, Parallel coordinates, HVIV, 각각의 기법을 통한 총 수행 시간에 대한 ANOVA (Analysis of Variance) 결과 세 가지 기법은 총 수행 시간에 각각 유의한 차이를 보였다($F(2,213)=81.03$, $p<0.0001$). 평균 총 수행 시간은 HVIV가 4.39min(SD=0.89)로 가장 짧게 나왔고, 텍스트가 20.39min(SD=8.25), Parallel coordinates는 32.81min(SD=8.67)로 가장 길었다. SNK(Student-Newman-Keuls) 검정에서는 세 기법은 각각 다른 그룹으로 나누어졌다(그림 17 참조).

전문가 집단

7명의 전문가 집단에서 각각의 기법을 통한 평균 총 수행 시간에 대한 ANOVA 결과 세 가지 기법은 총 수행 시간에 각각 유의한 차이를 보였다($F(2,165)=26.15$, $p<0.0001$). 총 수행 시간은 HVIV는 2.47min(SD=1.68)로 가장 짧았으며, 텍스트는 10.28min(SD=5.32) 그리고 Parallel coordinates는 13.95min(SD=5.37)로 가장 길었다. SNK 검정에서 세 기법은 각각 다른 그룹으로 나누어졌다(그림 17 참조).

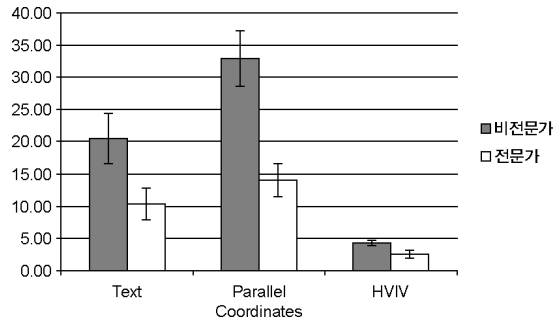

그림 17. 텍스트, Parallel coordinates, HVIV의 총 수행 시간 비교 (bar는 표준편차를 나타냄)

5.1.2 데이터의 패턴 파악에 대한 수행 시간 비교

비전문가 집단

비전문가 집단에서 Port Scan과 같이 시각화 데이터 상에서 특정한 패턴이 나타나는 경우, 이러한 침입 행위를 탐지하는 과제에 대한 평균 수행 시간은 세 가지 기법이 유의한 차이를 보였다($F(2,24)=8.35$, $p<0.0001$). 평균 수행 시간은 HVIV가 0.14min(SD=0.09)로 가장 짧았으며, Parallel coordinates는 0.91min(SD=0.45) 그리고 텍스트는 2.68min(SD=2.30)로 가장 길었다. SNK 검정 결과, HVIV와 Parallel coordinates 그룹과 텍스트 그룹의 두 그룹으로 나누어졌다(그림 18 참조).

전문가 집단

위와 마찬가지의 경우에 있어서 전문가 집단에서의 평균 수행 시간 분석 결과 세 집단은 각각 유의한 차이를 나타내었다($F(2,18)=27.19$, $p<0.0001$). HVIV의 수행 시간은 0.12min(SD=0.07)로 가장 짧았고, Parallel coordinates는 0.35min(SD=0.18), 그리고 텍스트는 0.68min(SD=0.16)로 가장 길었다. SNK 검정 결과 세 기법은 각각 다른 그룹으로 나타났다(그림 18 참조).

5.1.3 데이터의 세부 정보 파악에 대한 수행 시간 비교

비전문가 집단

비전문가 집단에서 네트워크 데이터의 세부 정보를 파악

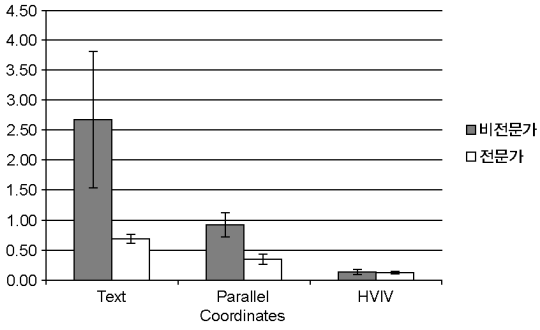

그림 18. 텍스트, Parallel coordinates, HVIV의 Port Scan 탐지 수행 시간 비교(bar는 표준편차를 나타냄)

하는 과제(표 2의 질문 II, III)를 수행하는 데 걸린 시간은 세 가지 기법이 각각 유의한 차이를 보였다($F(2,78)=62.45$, $p<0.0001$). 평균 수행 시간은 HVIV가 3.81min(SD=0.82)로 가장 짧게 나왔고, 텍스트는 13.52min(SD=5.31), 그리고 Parallel coordinates는 25.03min(SD=6.49)로 가장 길게 나왔다. SNK 검정 결과 세 기법은 각각 다른 그룹으로 나누어졌다(그림 19 참조).

전문가 집단

마찬가지로 전문가 집단에서 네트워크의 세부 정보를 파악하는 과제에 대한 평균 수행 시간은 세 가지 기법이 각각 유의한 차이를 보였다($F(2,60)=16.02$, $p<0.0001$). HVIV는 2.09min(SD=1.52)로 가장 짧게 나왔고, 텍스트는 8.10 min(SD=5.33), 그리고 Parallel coordinates는 12.08min (SD=4.73)로 가장 길게 나왔다. SNK 검정 결과 세 기법은 각각 다른 그룹으로 나타났다(그림 19 참조).

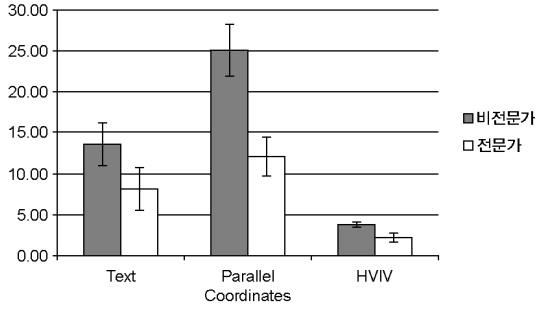

그림 19. 텍스트, Parallel coordinates, HVIV의 네트워크 세부 정보 파악 수행 시간 비교(bar는 표준편차를 나타냄)

5.2 주관적 만족도

피실험자에 네트워크 상황을 파악할 수 있는데 대한 주관적 만족도는 비전문가 집단의 경우 HVIV가 평균 87.22 (SD=7.95)로 제일 높았고, 텍스트가 62.78(SD=16.03), Parallel coordinates가 53.33(SD=18.87)으로 제일 낮았으며, 전문가 집단의 경우 HVIV는 92.44(SD=2.67), Parallel coordinates는 77.14(SD=8.59), 텍스트는 59.29 (SD=7.87)로 나타났다(그림 20 참조).

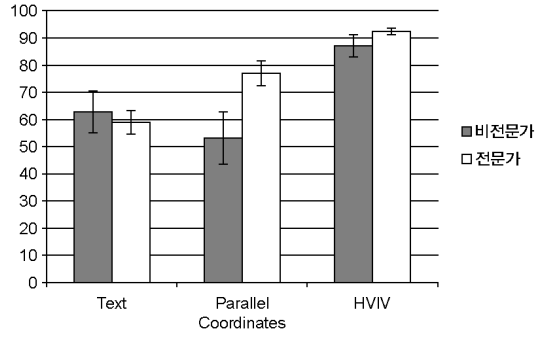

그림 20. 텍스트, Parallel coordinates, HVIV의 주관적 만족도 (bar는 표준편차를 나타냄)

5.3 주관적 선호도

각각의 기법이 적용된 네트워크 감시 툴의 주관적 선호도는 9명의 비전문가 집단에서는 8명이 HVIV를 선택했으며, 한 명이 Parallel coordinates를 선택했고, 텍스트를 선택한 피실험자는 없었다. Parallel coordinates를 선택한 피실험자에게 선택 이유를 물으니 Parallel coordinates의 경우는 만족도가 낮고 툴 사용법을 배우기가 어렵기 때문에 수행 과제를 완료했을 경우의 성취감이 크기 때문이라고 답했다 (그림 21 참조).

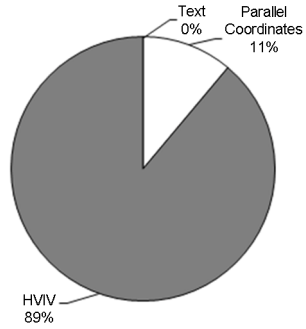

그림 21. 텍스트, Parallel coordinates, HVIV의 주관적 선호도 (비전문가 그룹)

7명의 전문가 집단에서는 6명이 HVIV를 선택했으며 1명이 텍스트를 선택했고, Parallel coordinates를 선택한 피실험자는 없었다. 텍스트를 선택한 피실험자의 선택 이유는

텍스트 사용시 HVIV보다 수행 시간은 더 많이 걸리지만, 툴 사용법이 배우기가 쉽고, 이때까지 자신이 텍스트 적용 툴에 적응되어 있기 때문이라고 답했다(그림 22 참조).

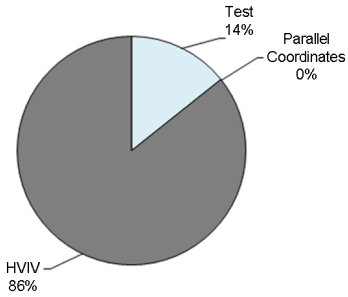

그림 22. 텍스트, Parallel coordinates, HVIV의 주관적 선호도 (전문가 그룹)

6. 토 의

총 수행 시간과 네트워크 데이터가 정상일 경우, 비정상1(Port Scan), 비정상2(Worm Scan)일 경우의 수행 시간 모두 HVIV가 제일 짧게 나왔다. 이는 HVIV가 인간의 Preattentive Processing을 통해 전체적인 데이터를 조망하면서 필요한 특정 데이터를 빨리 찾을 수 있도록 디자인되었음에 기인했기 때문으로 보인다.

총 수행 시간은 텍스트보다 Parallel coordinates가 더 길었지만 Port Scan일 경우 해킹 상황의 유무 파악에 대한 수행도는 HVIV > Parallel coordinates > 텍스트의 순으로 나타났다. 이는 Port Scan일 경우 Parallel coordinates는 시각화 데이터 상에서 특정한 패턴이 나타나기 때문에 텍스트보다 그 수행도가 좋았으나 정상일 경우와 Worm Scan일 경우는 특정한 시각화 패턴보다는 모든 데이터를 총괄하여 전체적으로 파악해야 빨리 수행을 할 수 있기 때문에 텍스트의 수행도가 Parallel coordinates보다 좋게 나온 것으로 보인다. 만일 Port Scan뿐 아니라 Host Scan이나 DOS와 같이 시각화했을 경우 특정한 패턴이 나타나는 데이터로 실험을 더 실시하였다면 Parallel coordinates의 수행도가 텍스트보다 더 좋을 것으로 예상되어진다.

그리고 전문가의 수행도는 비전문가의 수행도보다 텍스트를 사용했을 때뿐만 아니라 시각화가 적용된 HVIV나 Parallel coordinates를 사용했을 때도 좋았는데, 그 이유는 실험 전에도 예상했던 바와 같이 전문가는 비전문가보다 네트워크에 대해 더 정확하고 이해하기 쉬운 mental model을 갖고 있기 때문에 시각화 기법이 적용된 도구를 사용하더라도 비전문가보다 더 빨리 필요한 정보를 찾는 것으로 보인다.

침입 행위의 발생 유무 판단에 이어 침입 상황에서의 네트워크 세부 정보를 파악하는 수행이나, 그 외 네트워크의 부수적인 정보를 확인하는 수행에서는 HVIV > 텍스트 > Parallel coordinates 순으로 수행도가 나타났다. 이러한 과제는 패턴을 파악하는데 효과적이라는 시각화의 장점과는 별개로, 세부적인 정보를 확인해야 하는 작업이기 때문에 텍스트가 Parallel coordinates보다 더 수행도가 좋았고, Focus+Context 기법에 의해 전체적인 데이터를 파악하면서 동시에 개개의 세부적인 정보를 파악할 수 있는 장점을 가지고 있는 HVIV가 수행도가 제일 좋게 나왔다고 볼 수 있다.

또한 전문가가 현재 네트워크 모니터링 툴에서 제일 많이 사용되고 있는 텍스트를 사용했을 시의 총 수행 시간(평균: 10.28min)보다 비전문가가 HVIV를 사용했을 시의 총 수행 시간(평균: 4.39min)이 57.30%가 감소한 것으로 보았을 때, HVIV를 사용하면 쉽고 효율적으로 비전문가의 수행도를 전문가 수준의 수행도 이상으로 향상시킬 수 있을 가능성을 시사한다(그림 23 참조).

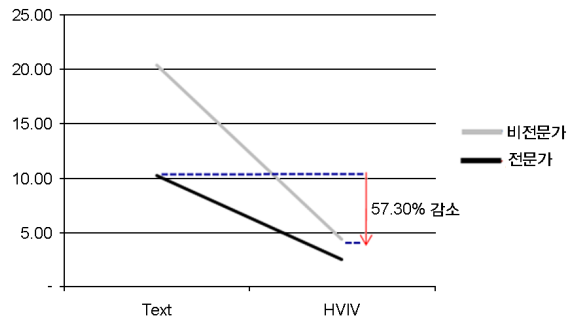

그림 23. 비전문가 집단의 텍스트, HVIV 사용 시 수행 시간 감소와 전문가 집단의 텍스트, HVIV 사용 시 수행 시간 감소의 비교

그리고 본 실험 시, HVIV는 피실험자에게 사전에 설명한 침입 행위의 유형뿐만 아니라, 일반적이지 않은 데이터 패턴도 쉽게 감지해냄을 보였다. 이를 통해서 HVIV는 알려진 침입 행위나 공격 행위뿐 아니라, 알려지지 않은 침입 행위나 공격 행위도 쉽게 발견하여, 신속하게 조치할 수 있을 것으로 보인다.

7. 결 론

본 연구를 통한 결론은 다음과 같다.

결론적으로, 효과적인 침입탐지를 위해서는 전체적인 정보를 파악하면서 사용자가 관심을 갖는 특정 정보를 파악할 수 있는 Focus+Context 기법을 사용하여 디자인한 HVIV는 Parallel coordinates보다 데이터의 전체적인 패턴을 파악하는데 효과적이고, 텍스트보다 데이터의 세부 정보를 파악하는데도 효과적임이 판명되었다.

그러므로 본 연구에서 제안한 HVIV를 사용한다면 현재 많이 발생하고 있는 해킹과 같은 사이버 범죄에 대해 수사관의 인지적 부담을 감소시키고, 수행도를 크게 향상시켜 줄 것이며, 사용자의 Preattentive Processing을 통해 데이터의 패턴을 쉽게 인지할 수 있기 때문에 현재 알려진 해킹 뿐 아니라, 알려지지 않은 침입 행위에 대해서도 이를 쉽게 발견하고 조기에 그에 대한 조치를 실시하여 문제를 해결할 수 있을 것이며, 결과적으로 네트워크 포렌식의 해킹 탐지 분야에 크게 도움을 줄 수 있을 것으로 보인다.

Acknowledgements

본 논문은 '2단계 BK21사업'과 2010년 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 기초연구사업 지원을 받아 수행된 것임(Grant No. 2010-0013188).

References

김기준, 류수연, 이장호, 김지인, Hyperbolic Tree를 이용한 Bookmark의 시각화, *한국정보과학회 봄 학술발표논문집*, 26(1), 422-424, 1999.

박구, *컴퓨터 포렌식스 관점의 해킹수사 방법에 관한 연구*, 석사학위논문, 고려대학교대학원 디지털정보공학과, 2007.

박종성, *자동화된 침해사고대응시스템에서의 네트웍 포렌식 정보에 대한 정의*, 석사학위논문, 고려대학교대학원 정보보호학과, 2004.

배준우, *정보시각화의 새로운 분류법에 관한 연구*, 석사학위논문, 고려대학교대학원 산업시스템정보공학과, 2008.

전상덕, 홍동숙, 한기준, 디지털 포렌식의 기술 동향과 전망, *정보화 정책*, 13(4), 3-19, 2006.

조선영, *포렌식 수사 관점에서의 디스크 정보의 효율적인 표현에 관한 연구*, 석사학위논문, 고려대학교대학원 정보보호학과, 2006.

윤성종, 김정호, 방화벽 로그를 이용한 침입탐지 기법 연구, *Journal of Information Technology Applications & Management*, 13(4), 141-153, 2006.

황규민, *Fisheye view를 이용한 PDA의 메뉴 디자인에 대한 정보 시각화 방안*, 석사학위논문, 고려대학교대학원 산업시스템정보공학과, 2004.

황보택근, 이수진, 인터넷 쇼핑몰에서 효율적인 상품 검색을 위한 가시화 방법, *한국 인터넷 정보 학회*, 6(5), 111-121, 2005.

Ball, R., Fink, G. A. and North, C., "Home-Centric Visualization of Network Traffic for Security Administration", *ACM Workshop Visualization and Data Mining for Computer Security(VizSEC/DMSEC)*(pp.55-64), Washington D.C., USA, 2004.

Bederson, B. B., Fisheye Menus, *ACM Symposium on User Interface Software and Technology(UIST 2000) CHI Letters*, 2(2), 217-225, 2000.

Bunting, S. and Wei, W., *The Official EnCase Examiner Study Guide*, Sybex, 2006.

Card, S., *et al*., *Readings in Information Visualization-Using Vision to Think*, Morgan Kaufmann, 1999.

Choi, H. and Lee, H., PCAV: Internet Attack Visualization on Parallel Coordinates, *Information and Communications Security*, 3783, 454-466, 2005.

Conti, G. and Abdullah, K., "Passive Visual Fingerprinting of Network Attack Tool", *VisSEC/DMSEC '04*(pp. 45-54), NY, USA, 2004.

Conti, G., *Security Data Visualization*, No Starch Press Inc., 2007.

Furnas, G. W., Generalized Fisheye Views, *Proceedings of Human Factors in Computing System(CHI 86) ACM Press*, 16-23, 1986.

Furnas, G. W., "Effective View Navigation", *SIGCHI 97 Conference Proceeding(ACM)*(pp. 367-374), Atlanta, GA, 1997.

Furnas, G. W., "A Fisheye Follow-up: Further Reflection on Focus + Context", *SIGCHI 2006 Proceedings*(pp. 999-1008), Montreal, Quebec, 2006.

Keahey, T. A. and Robertson, E. L., Techniques for Non-Linear Magnification Transformation, *Proceedings of the IEEE Symposium on InfoVis '97*, 38-45, 1997.

Lamping, J., Rao, R. and Pirolli, P., A Focus + Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies, *Proceedings of Human Factors in Computing Systems(CHI 95) ACM Press*, 401-408, 1995.

Lamping, J. and Rao, R., The Hyperbolic Browser: Focus + Context Technique for Visualizing Large Hierarchies, *Journal of Visual Languages and Computing*, 7, 33-55, 1996.

Leung, Y. K. and Apperley, M. D., A Review and Taxonomy of Distortion-Oriented Presentation Techniques, *ACM Transactions on Computer-Human Interaction*, 1(2), 126-160, 1994.

Munzner, T. and Burchard, P., "Visualizing the Structure of the World Wide Web in 3D Hyperbolic Space", *Proceedings of the VRML '95 Symposium*, San Diego. CA. 1995.

Munzner, T., H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space, *Proceeding of IEEE Symposium on InfoVis* '97, 2-10, 1997.

Sarker, M. and Brown, M. H., Graphical Fisheye Views of Graphs, *In Proceedings of Human Factors in Computing Systems(CHI 92) ACM Press*, 83-91, 1992.

Schaffer, D., *et al*., Navigating hierarchically clustered networks through fisheye and full-zoom methods, *ACM Transactions on Computer-Human Interaction*, 3(2), 162-188, 1996.

Stolze, M., Pawlitzek, R. and Wespi, A., "Visual problem-solving support

for new event triage in centralized network security monitoring: challenges, tools and benefits", *GI-SIDAR Conf. ITIncident Management & IT-Forensics(IMF)*(pp.67-76), Stuttgart, Deutschland, 2003.

Teerlink, S. and Erbacher, R. F., Improving the Computer Forensic Analysis Process Through Visualization, *Communication of the ACM*, 49(2), 2006.

Teerlink, S., *et al*., "Foundations for Visual Forensic Analysis", *Proceeding of the 7th IEEE Workshop on Information Assurance*, NY, USA, 2006.

Ware, C., *Information Visualization: Perception for Design*, 2nd ed., Morgan Kaufmann, 2004.

Yin, X., *et al*., "VisFlowConnect: netflow visualization of link relationships for security situational awareness", *Proc. of VizSEC 2004*(pp. 26-34), NY, USA, 2004.

Author listings

Yunseok Jeong: jys612@korea.ac.kr

Highest degree: Bachelor of Engineering, Department of Industrial Management Engineering, Korea University
Areas of interest: UI, HCI

Rohae Myung: rmyung@korea.ac.kr

Highest degree: Ph.D., Industrial Engineering, Texas Tech University
Position title: Professor , Dept. of Industrial Management Engineering, Korea University
Areas of interest: Cognitive Engineering

Date Received : 2010-11-05

Date Revised : 2010-12-08

Date Accepted : 2010-12-10