디지털산업정보학회 논문지 제11권 제4호-2015년 12월
http://dx.doi.org/10.17662/ksdim.2015.11.4.047

트래픽 세션의 포트 역할을 이용한 네트워크 공격 시각화*

장 범 환**

Network Attacks Visualization using a Port Role in Network Sessions

Chang Beomhwan

〈Abstract〉

In this paper, we propose a simple and useful method using a port role to visualize the network attacks. The port role defines the behavior of the port from the source and destination port number of network session. Based on the port role, the port provides the brief security features of each node as an attacker, a victim, a server, and a normal host. We have automatically classified and identified the type of node based on the port role and security features. We detected and visualized the network attacks using these features of the node by the port role. In addition, we are intended to solve the problems with existing visualization technologies which are the reflection problem caused an undirected network session and the problem caused decreasing of distinct appearance when occurs a large amount of the sessions. The proposed method monitors anomalies occurring in an entire network and displays detailed information of the attacker, victim, server, and hosts. In addition, by providing a categorized analysis of network attacks, this method can more precisely detect and distinguish them from normal sessions.

Key Words : Security Visualization, Network Security, Network Traffic Visualization, Node Identification

Ⅰ. 서론

네트워크 트래픽 시각화 기술은 네트워크에서 발생하는 트래픽 이벤트 데이터로부터 특성 정보를 추출한 후, 정보 시각화(Information Visualization) 기법을 사용하여 2차원 또는 3차원 공간상에 이벤트의 내용과 네트워크 상황을 표현하는 기법이다[1-2]. 이는 시각화 요소를 활용하여 데이터가 정보로써 쉽게 전달되도록 시각적으로 형상화하는 것으로써 사용자가 방대한 양의 트래픽 데이터들을 직관적으로 분석하는데 매우 유용하다[1-2]. 시각화 기술은 한정된 공간 내에 많은 트래픽 데이터들을 동시에 표시할 수 있고 개별적인 트래픽 이벤트에서는 볼 수 없었던 여러 네트워크 상황들이 패턴으로 형상화되므로 사용자에게 쉽게 상황 정보를 제공한다. 네트워크 보안 관점에서도 네트워크 내의 정상 상황과 공격 상황을 패턴으로 형상화하기 때문에 주목을 받고 있다[3-6].

트래픽 세션 데이터를 이용한 공격 분석 및 시각화

* 이 논문은 2015년 호원대학교 연구비 지원을 받은 것임.
** 호원대학교 사이버수사보안학부 조교수(교신저자)

방법들은 세션 데이터 내의 출발지 및 목적지 IP 주소를 이용하여 네트워크 이상 현상들을 쉽게 표시한다[7-8]. 이는 네트워크 공격이 출발지(공격자)와 목적지(피해자) 사이의 행위 표현, 즉 두 주체간의 방향성을 갖는 연결 표현으로 공격 현상을 효과적으로 표현할 수 있기 때문이다. 따라서, 대부분의 분석 도구들은 먼저 출발지 및 목적지 IP주소의 연결 관계를 기반으로 네트워크 이상 현상을 탐지하고, 포트 번호는 이상 현상을 상세하게 분석하는 용도로 사용한다[5, 9-11]. 이런 도구들에게 있어서 출발지 IP 주소, 목적지 IP 주소, 그리고 방향성은 매우 중요한 요소이다.

그러나, 원시 데이터의 방향성이 모호하거나 방향성이 없다면, 도구들이 분석한 결과와 이런 결과를 시각화한 시각적 인터페이스는 의미를 가질 수가 없다. 이것은 방향성 상실의 문제를 고려하지 않거나 또는 방향성을 강조하여 설계한 시각화 인터페이스에는 출발지와 목적지의 표현 문제, 그리고 표현된 의미를 해석하는데 난해한 혼동 문제가 발생한다. 즉, 시각화 인터페이스는 세션의 방향성을 고려 및 강조하기 위해 출발지와 목적지 IP주소를 양쪽(좌・우, 상・하) 대칭평면으로 설계하고 점 데이터 형태(노드)로 표시한다. 하지만, 세션 내의 모든 IP 주소는 출발지와 목적지로 사용될 수 있기 때문에 동일 IP주소가 대칭축을 중심으로 양쪽에 각각 점 데이터로 표시되는 시각화 인터페이스의 설계 구조적인 문제가 발생한다. 또한, 원시 세션 데이터가 왕복하는 패킷 집합을 의미하지 않고 선행 세션과 응답 세션으로 각각 제공될 경우, 선행 세션에 대한 응답 세션인 경우에는 선행 세션 내의 출발지 IP주소는 응답 세션의 목적지 IP주소가 되므로 역시 대칭축을 중심으로 양쪽이 거울처럼 반사되어 표시되는 문제가 발생한다. 결과적으로 양쪽 두 개의 노드를 단일 노드로 해석해야 하는 문제가 발생한다. 우리는 이것을 단방향성 원시 데이터 문제 또는 원시 데이터의 방향성 상실에 의한 문제라고 규정하고 그 결과로써 나타나는 현상을 반사현상이라고 정의한다.

이와같이 반사현상은 트래픽 세션 데이터가 단방향인 특성 때문에도 발생하고, 방향성을 이용한 시각화 구조 때문에도 생겨난다. 대부분의 도구들은 원시 데이터로써 단방향 세션 데이터인 Cisco의 Netflow를 사용하고 있다. 실제 세션은 출발지와 목적지 간의 양방향으로 왕래하는 패킷들로써 연결 요청에 대한 방향성이 있지만, Netflow와 같이 실제 세션을 독립적인 2개의 단방향 세션으로 구분하여 제공할 경우, 선행 세션과 응답 세션은 출발지와 목적지가 상호 변경되므로 각 세션의 방향성은 가치가 없어진다. 따라서, 기존의 단방향 세션 데이터를 이용하는 분석 도구들과 방향성을 강조하여 설계된 분석 도구들은 반사현상이라는 심각한 문제를 내재하고 있다.

본 논문에서는 포트 역할 기반의 분석을 제안하고자 한다. 트래픽 세션 데이터에는 IP 주소 이외에도 이상현상 또는 노드 유형을 검출할 수 있는 중요한 요소로써 출발지 포트번호와 목적지 포트번호가 있다. 이 방법은 세션 데이터 내의 출발지와 목적지 IP 주소는 노드 식별자로 활용하고 출발지 및 목적지 포트번호는 역할로써 해석하여 노드의 유형을 분석하는 방법이다. 여기서, 출발지 및 목적지 포트번호는 방향성을 의미하는 것이 아니고, 노드 내의 어떤 포트번호가 서버, 클라이언트, 공격자, 피해자 역할로 활동했는지를 의미하는 것이다. 포트번호는 출발지와 목적지 용도에 따라서, 사용된 번호 범위에 따라서, 구성비와 빈도수에 따라서, 그리고 대표 포트번호의 존재 유무에 따라서 서버와 클라이언트 역할 또는 공격자와 피해자 역할을 구분하는 용도로 활용될 수 있기 때문이다. 결과적으로, 본 논문은 세션 데이터 내의 포트번호를 활용한 네트워크 노드 역할 분류 및 공격 탐지, 그리고 반사현상 문제를 근본적으로 해결하기 위한 시각화 인터페이스 구조에 대해 기술하고자 한다.

Ⅱ. 관련 연구

2.1 네트워크 공격 시각화 기술

네트워크 세션 기반 시각화 기술에는 NVisionIP[9], VisFlowConnect-IP[12], FloVis[13], SecVis[14], Visual Fingerprinting[15] VisCat[5-6] 등이 있다. NVisionIP와 VisFlowConnect는 NCSA (National Center for Supercomputing Applications) 프로젝트의 일부분으로 개발된 시각화 도구이다.

NVisionIP는 Netflow 데이터를 이용하여 B클래스 대역의 트래픽을 하나의 화면에 시각화하는 도구이다. NVisionIP는 크게 Galaxy View, Small Multiple View, Machine View로 구성된다. Galaxy View는 B클래스 서브넷 주소를 가로축으로 설정하고 호스트 주소를 세로축으로 설정하여 호스트에서 사용된 포트의 개수를 화면상에 표현한다. Small Multiple View는 네트워크 관리자에 의해 선택된 서브넷과 호스트의 특정 포트별 플로우 개수를 표현한다. Machine View는 특정 호스트에서 송·수신되는 포트별 트래픽 양을 표현한다. 또한, NVisionIP는 드릴다운(drill-down) 기능을 통하여 전체 관리 네트워크를 감시하면서 이상 현상이 발견되는 경우 상세한 화면으로 이동하는 기능을 제공한다[9]. 하지만, 관리 대상이 되는 목적지 B클래스 네트워크에 초점을 맞추고 있기 때문에 트래픽의 근원지 및 근원지 포트 정보를 표현하지 못한다. 또한, 하나의 화면에서 트래픽과 관련된 모든 상황 정보를 표현하지 못하므로 원하는 정보를 얻기 위해서는 다른 화면으로의 전환이 필요하고, 이는 관리자가 이상 현상을 파악하는데 소요되는 시간을 증가시키는 요인이 된다.

VisFlowConnect-IP는 트래픽 세션의 연결 정보에 초점을 맞추고 외부 도메인 시스템들과 내부 시스템들 간의 트래픽 흐름을 '수직의 평행 축(parallel axes)'을 사용하여 표현하는 시각화 도구이다. 입력 데이터로는 NVisionIP와 동일한 NetFlow 데이터를 사용하지만 다른 소스의 데이터도 사용 가능하다 정의된 평행 축에는 좌측부터 외부 도메인에서 데이터를 송신하는 근원지를 표현하는 축, 내부 관리 도메인의 호스트를 표현하는 축, 데이터를 수신하는 외부 목적지를 표현하는 축으로 구성된다. 그리고, 사전에 정의된 임계치를 초과하는 트래픽에 대해서 호스트들을 각각의 평행 축에 매핑하고 연결선을 이용하여 연결 관계를 표현한다. VisFlowConnect- IP는 외부 도메인과 내부 도메인 간의 연결 정보를 보여주는 External View, 선택된 외부 도메인과 내부 도메인 간의 연결 정보를 보여주는 Domain View, 내부 도메인 간의 연결 정보를 보여주는 Internal View 등으로 구성된다. 트래픽의 흐름을 나타내는 연결선은 사전에 정의된 도메인을 의미하고 트래픽의 양이 많아질수록 어두운 색으로 표현된다[12]. VisFlowConnect-IP는 종단간의 연결 정보를 제공한다는 점에서 네트워크 상황을 쉽게 이해할 수 있다는 장점이 있지만, 한정된 평행 축 상에 호스트 및 도메인을 표현하기 때문에 특정 호스트 및 도메인을 직관적으로 인지하기 어렵고, 트래픽이 다량 발생할 경우에는 너무 많은 연결선들이 화면을 가득 채워서 식별력이 떨어진다.

FloVis는 네트워크 트래픽 흐름을 여러 화면을 통해 다양한 측면으로 보여주는 시각화 도구이다. FlowBundle 화면은 전체 통신(종단간의 연결) 패턴을 내역을, NetBytes Viewer 화면은 노드들(서버, 클라이언트)의 활동 내역을, NetBytes Viewer 화면은 포트별 사용 내역을 표시하며, 각각의 화면들은 서로의 단점을 보완한다. FloVis의 기본 화면인 FlowBundle은 내부와 외부 네트워크 시스템들 간에 발생하는 NetFlow 데이터를 연결선을 이용하여 표현한다. 시스템의 위치는 IP 주소를 이용하여 방사형의

원테두리 상에 점으로 표시하고, 해당 시스템을 클릭하면 NetBytes Viewer 화면이 실행되면서 시간에 따른 포트별 변화와 데이터 량을 보여준다. FloVis는 하나의 원 위에 관찰 대상 시스템들 간의 트래픽 흐름을 표시하므로 사용자가 네트워크의 상황을 한 눈에 볼 수 있도록 해준다는 장점이 있다[13]. 하지만, VisFlowConnect-IP와 동일하게 화면에 다수의 연결선이 표시되는 문제와 너무 많은 연결선들로 식별력이 떨어지는 문제가 발생한다.

VisCat은 한국전자통신연구원에서 개발한 트래픽 세션 기반의 시각화 도구로써 네트워크 상황 정보를 제공하는 VisNet과 네트워크 공격을 상세하게 분석하는 VisMon으로 구성된다. VisNet의 IPGrid는 전체 IP주소 공간을 4개의 2차원 그리드에 근원지와 목적지를 점으로 표시 및 연결하여 트래픽의 흐름을 표시한다. 다수의 점들과 연결선들이 만드는 군집 또는 발산 모양을 활용하여 각종 네트워크 서비스들과 공격들을 시각화한다. 또한, 근원지와 목적지 IP 주소로부터 추출된 소속 국가, 소속 기관의 정보를 동일 화면상에 표시함으로써 공격자와 피해자에 대한 상세 정보를 제공한다. VisNet의 Center는 이상 현상이 발생한 IP 주소뿐만 아니라 호스트의 위치를 전자지도 상의 실제 위치로 표현함으로써 이상 현상이 발생한 호스트의 논리적 위치와 물리적 위치를 빠르게 인지하도록 돕는다[5]. VisMon은 2차원 쿼드(Quad)와 3차원 큐브(Cube)를 이용하여 네트워크 공격을 상세 분석하는 도구이다. 공격을 포함한 대부분의 네트워크 이상 현상들은 종단간의 세션 패턴들이 발산 또는 수렴 형태를 보이는데, VisMon은 5-tuple의 다양한 변화 모습을 2차원 평면과 3차원 공간상에 효과적으로 시각화하여 해당 공격을 표시하고 탐지한다. 대부분의 시각화 도구들이 단일 공격이 아닌 복수 개의 공격이 진행되거나 세션 개수가 많은 웜 공격 또는 DDoS 공격이 진행될 경우 소수의 공격들은 은닉되어 표시되지 않는 경우가 많은데 VisMon은 이와같은 경우에도 공격 표시와 탐지를 수행한다[6]. 트래픽이 다량 발생할 경우에는 VisCat 기술도 VisFlowConnect-IP나 FloVis와 같이 너무 많은 점들과 연결선들이 생겨나서 화면의 식별력이 떨어지는 단점이 있다.

이상과 같이, NVisionIP, VisFlowConnect-IP, 그리고 VisCat과 같은 Netflow를 이용하는 트래픽 세션 기반 시각화 기술들은 다량의 트래픽이 발생할 경우에 식별력이 떨어지는 공통의 문제를 가지고 있다. 하지만, 가장 큰 문제점은 세션 데이터의 방향성 상실이 가져오는 반사현상(거울효과) 문제이다.

반사현상은 두 가지 이유에서 발생한다. 제공되는 세션 데이터인 Netflow가 단방향 플로우 데이터이므로 세션 데이터를 변환없이 그대로 사용하기 때문에도 발생하고, 세션 데이터의 방향성을 고려하여 발신지와 수신지의 위치를 대칭 평면 상에 구성하여 표시하기 때문에도 발생한다. 반사현상은 방향성을 고려하여 발신지와 수신지의 위치를 대칭적으로 설계하고 표시했던 기존의 시각화 기술들에게 있었던 고질적인 문제였다. 이것은 종단간 연결[3, 16]의 고유성을 훼손할 뿐만 아니라 시각화 기술의 직관성을 떨어뜨리며, 정상 서버와 공격 현상의 구분을 모호하게 만든다.

2.2 기존 기술의 문제점 – 반사현상

통신의 종단간 연결을 의미하는 세션은 출발지와 목적지를 왕래하는 트래픽 집합을 의미한다. 대표적인 세션 데이터에는 Cisco의 netflow가 있다[3]. netflow는 단일 방향으로 전송되는 트래픽 집합을 하나의 독립된 세션으로 정의하고 있기 때문에, 실제 종단간의 연결을 방향이 다른 2개의 세션 정보로 제공한다. 따라서, netflow와 같은 단방향 플로우를 이용하여 시각화할 경우 출발지가 목적지로 목적지가

출발지로 표시되는 반사현상이 생긴다. 이는 출발지와 목적지간의 관계를 모호하게 만들며 종단간 연결의 고유성을 훼손시키는 원인이 된다.

대부분의 트래픽 세션 기반 시각화 기술들은 세션을 출발지와 목적지 간의 연결선 형태로 시각화함으로써 직관성을 높인다. 연결선은 점 사이에 짝 관계를 표현하는데 유용하고 점들 사이의 관계를 조망하는데 많은 도움을 주기 때문이다. 일반적으로 시각화 기술은 트래픽 세션 정보 내의 IP주소를 출발지와 목적지로 각각 매핑한 후, 평면 상의 점 형태로 표시하고 연결선을 이용하여 출발지와 목적지 간의 연결 관계를 표시한다. 특히 출발지와 목적지 사이의 연결 방향은 네트워크 상황을 이해하거나 네트워크 공격을 판단할 수 있기 때문에, 트래픽 세션 시각화에서 출발지와 목적지 간의 연결 방향 표시는 직관성을 높이는 중요 요소로 작용한다. 따라서, 좌. 우 또는 상. 하 공간에 출발지와 목적지의 해당 점좌표를 배치하고 연결하여 세션의 방향을 표시하는 방법을 취한다. 이와같은 방법은 동일한 IP주소가 출발지일 수도 목적지일 수 있기 때문에 양쪽 공간에 모두 존재하는 반사현상 문제가 발생한다. 이는 출발지와 목적지의 구분을 어렵고 만들어 네트워크 상황을 이해하거나 네트워크 공격을 판단하는데 있어서 오히려 직관성을 떨어뜨리고 방해하는 요소로 작용한다.

반사현상은 정상 서버와 네트워크 공격의 구분을 모호하게 만든다. 일상적으로 빈번히 발생하는 네트워크 공격에는 DoS, DDoS, 호스트 스캐닝, 포트 스캐닝 공격이 있다[17]. 트래픽 세션 기반 시각화 기술은 네트워크 공격을 탐지하기 위해 특정 IP주소(출발지 또는 목적지)에 대한 트래픽 세션의 집중 또는 발산 현상을 이용한다. 일례로 DDoS 공격은 다수의 출발지에서 목적지로 집중되는 모습을, 호스트 스캐닝 공격은 특정 출발지에서 다수의 목적지로 발산하는 모습을 보인다. 하지만, 정상 서비스일 경우에도 트래픽 세션은 클라이언트에서 서버로 연결이 집중되는 모습을 보이고, 역으로는 서버에서 클라이언트로 발산하는 모습을 보여준다. 이는 네트워크 공격과 정상 서버와의 구분을 모호하게 만드는 원인이 되며 공격과 정상 서버를 구분하는 추가적인 기술이 필요하다.

Ⅲ. 트래픽 세션 시각화 시스템

3.1 트래픽 세션 데이터 분석

통신의 종단간 연결을 의미하는 트래픽 세션 또는 플로우는 일정 시간 동안 관찰 지점을 통과하는 IP 패킷들 중에서 공통 속성 갖는 패킷들의 집합으로 정의된다. 공통 속성에는 출발지 IP주소(sip: Source IP Address), 목적지 IP주소(dip: Destination IP Address), 출발지 포트번호(spt: Source Port Number), 목적지 포트번호(dpt: Destination Port Number), 프로토콜 번호(prt: Protocol Number)가 있고, 이를 세션 식별자로 지칭하며 트래픽 5-tuple(<sip, dip, spt, dpt, prt>)이라고 정의한다[3, 6]. 전송되는 수많은 패킷들이 동일한 공통 속성들을 갖는 하나의 플로우로 정의되기 때문에, 현재와 같은 패킷 개수와 바이트량이 방대한 네트워크 환경에서는 개별 패킷 보다는 세션 데이터를 활용하는 것이 트래픽 분석에 효율적이다.

트래픽 5-tuple은 종단간 연결, 즉 출발지와 목적지 간의 연결을 의미하고 공통 정보를 포함하므로 다수의 트래픽 5-tuple들을 군집하고 분석하면 네트워크의 각종 상황 파악 또는 공격 유무를 판단하는데 유용한 정보로 활용할 수 있다. 즉, 트래픽 세션 집합은 클라이언트와 서버 간의 일반적인 연결 활동, 특정 프로토콜이나 서버 특징에 따른 특이한 연결 활동, 그리고 정상적이며 평범한 연결 활동 등 전반적인 크

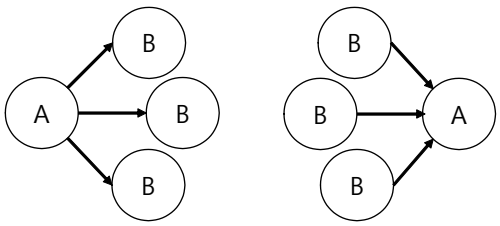
고 작은 네트워크 상황들 뿐만 아니라, 매우 작은 크기의 무의미한 연결 활동, 정상적이지 못한 이상 연결 활동, 그리고 공격자와 피해자 간의 공격 활동 등 비정상적인 네트워크 상황들을 대변한다.

대표적인 네트워크 공격에는 서비스거부(DoS) 공격, 분산서비스거부(DDoS) 공격, 인터넷 웜(Worm) 공격, 호스트 스캐닝(Scan. H) 공격, 포트 스캐닝(Scan. P) 공격이 있다[3, 6]. 다수의 사례에서 알수 있듯이, 분산서비스거부 공격과 인터넷 웜 공격을 비롯한 대부분의 네트워크 공격들은 출발지 IP주소와 목적지 IP 주소의 방향과 분포 관계, 그리고 공격에 사용되는 출발지 및 목적지 포트 번호 관계가 정상적인 호스틀과는 다른 특이한 특징을 갖는다. 예를 들면, 분산서비스거부 공격은 다수의 출발지 IP주소들이 특정 IP 주소 및 목적지 포트번호로 집중되며 며 세션의 빈도수가 크고, 호스트 스캔 공격은 특정 출발지 IP주소가 특정 IP주소 범위의 목적지들과 연결되며 목적지 포트번호는 고정되는 관계가 있다[3, 6]. 이와같이 트래픽 5-tuple은 네트워크 공격을 파악하는데 중요한 정보로 활용될 수 있다. 네트워크 공격들과 트래픽 세션 데이터와의 일반적인 관계는 <표 1>과 같다.

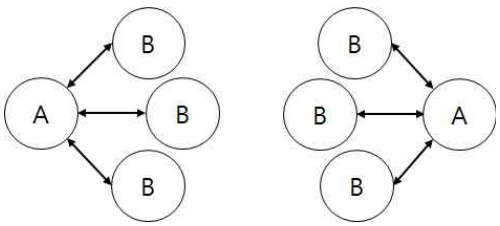
<표 1> 공격 유형과 5-tuple과의 관계
I:ICMP, T:TCP, U:UDP, ◉:특정, R:임의, W:Well-known, a-b:구간

공격 유형	sip	dip	prt	spt	dpt
DoS	1	1	T/U	R	W
DDoS	N	1	T/U	R	W
Worm	1	N	T/U	◉/R	◉
Scan. P	1	1	T/U	◉	a-b
Scan. H	1	N	I/T/U	0/◉	2048/◉
Server	1	N	T/U	W	R

<표 1>에서 보는 바와 같이, 세션 데이터 내의 출발지 IP주소와 목적지 IP주소 관계는 1차적으로 대부분의 공격들을 분류하는데 매우 유용할 수 있다. 하지만, 이는 세션 데이터의 방향성이 유효하다는 가정하에서만 성립된다. 원본 세션이 독립적인 2개의 단방향 세션 데이터, 즉 송신과 회신이 다른 두 개의 데이터로 제공될 경우에는 무의미한 결과가 된다. 예를 들면, DDoS와 Server의 출발지와 목적지 관계를 보면 N:1과 1:N을 갖게 되어 구분이 가능할 것 같지만, 원본 세션이 두 개의 단방향 세션 데이터 각각 제공된다면 DDoS의 응답 세션들은 1:N되어 Server의 모습을 보이고 Server는 N:1이 되어 DDoS의 모습을 보이는 동일한 결과를 얻게 된다.



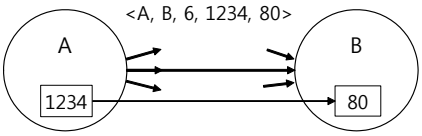
(a) DDoS (b) Server
<그림 1> 방향성이 있는 트래픽 데이터에서 노드 연결 관계



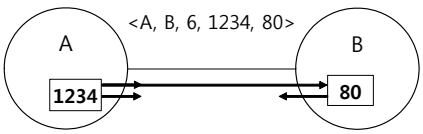
(a) DDoS (b) Server
<그림 2> 방향성이 없는 트래픽 데이터에서 노드 연결 관계

<그림 1>은 방향성이 있는 트래픽 데이터에서 노드 연결 관계를, <그림 2>는 방향성이 없는 트래픽 데이터에서 노드 연결 관계를 나타낸 것이다. 공격을 분류하는데 있어서 <표 1>를 참고하면, 상대적으로 저평가되었던 출발지 포트번호와 목적지 포트번호를 중요한 특징 요소로 사용할 수 있다. 특히, 개별 노드 내의 well-known 포트번호의 사용 빈도수와 구성비

는 노드의 유형을 규정할 수 있는 특성 인자로써 사용 가능함을 볼 수 있다. 이것은 더 이상 세션 데이터를 방향성을 중시한 이벤트 데이터로 취급하지 않고, 세션 데이터 내의 포트번호를 각 노드 내의 역할을 규정하는 요소로써 취급해야 한다는 것이다. 이것은 노드 자체의 방향성 보다는 노드를 구성하는 포트들의 역할을 중심으로 노드의 유형을 규정하기 위한 개념이다.



<그림 3> 노드의 방향성 개념



<그림 4> 노드의 포트 역할 개념

<그림 3>은 세션 데이터가 갖는 기존의 방향성 개념을, <그림 4>는 새로운 세션 데이터의 포트 역할 개념을 나타낸 것이다. 포트 역할 개념은 노드를 구성하는 포트들이 공격자, 피해자, 서버, 그리고 일반 호스트 중에서 어떤 용도와 역할로 사용되고 있는 지를 검사하는 개념이다. 따라서, 트래픽 5-tuple은 <표 2>와 같이 활용되어야 한다.

<표 3>은 일반 프로토콜인 TCP 또는 UDP 경우에 적용되는 포트와의 관계를 나타낸 것이다. 보안 관점에서 네트워크 노드들을 공격자, 피해자, 서버, 그리고 호스트로 역할을 분류한다면, 포트 역할에 따른 노드의 역할은 연결 세션의 개수와 well-known 포트의 구성 비율에 따라 구분이 가능하므로 이를 활용하면 노드 유형을 분류할 수 있다. 하지만, 포트의 역할은 모든 프로토콜에 대해 일반적으로 적용될 수 없기 때문

<표 2> 트래픽 5-tuple과 보안 활용 분야

5-tuple	특징 및 활용 분야
sip	노드 식별자, 연결관계
dip	노드 식별자, 연결관계
prt	프로토콜별 특징, 예외 사항
spt	노드 특성/역할(공격자, 서버, 피해자, 호스트)
dpt	노드 특성/역할(공격자, 서버, 피해자, 호스트)

에 프로토콜별로 예외 사항을 적용하여 노드 내의 포트 역할이 의미를 갖도록 해야한다. 예외 사항을 적용해야하는 대표적인 프로토콜에는 ICMP가 있다.

<표 4> 노드 유형과 세션개수/포트번호 관계

노드 유형	세션 개수	spt		dpt	
		<1024	≥1024	<1024	≥1024
Attacker	△	▽	△	△	▽
Victim	△	△	▽	▽	△
Server	△	△		▽	△
Host	▽		△	△	▽

ICMP 트래픽 세션은 TCP 또는 UDP 세션과 동일하게 해석해서는 문제가 발생한다. ICMP 메시지에는 출발지와 목적지 포트 번호가 없기 때문에 트래픽 세션 데이터에 포함되어 있는 출발지와 목적지 포트번호는 다른 방법으로 처리해야 한다. ICMP 메시지는 네트워크 상태나 오류 등에 관한 보고 메시지로써 type과 code를 갖는다. 이 정보는 트래픽 데이터의 목적지 포트번호인 dpt에 내재되어 <표 4>와 같이 제공된다. 그러므로 세션 데이터 내의 dpt만이 의미가 있는 값이고, 이 값을 출발지 및 목적지 노드에 모두 동일하게 적용해야 한다. 예를 들면, dpt가 2048인 트래픽 세션 데이터는 출발지 노드가 Echo Request 메시지를 보내고 목적지 노드는 Echo Request 메시지를 수신한 것으로 처리해야 한다.

<표 5> 세션 데이터 내의 목적지 포트로 표현되는 ICMP 메시지

dpt	type	code	RFC792
0	0	0	Echo Reply
768	3	0	net unreachable
769	3	1	host unreachable
770	3	2	protocol unreachable
771	3	3	port unreachable
2048	8	0	Echo Request
2816	11	0	time to live exceeded in transit

3.2 노드 유형 및 특징인자 정의

노드의 유형을 결정할 수 있는 중요한 특성 정보를 갖는 인자들을 특징인자라고 정의한다. 트래픽 세션 데이터을 이용하여 노드 유형을 식별하고자 할 때, 노드 유형과 포트 번호들은 <표 3>과 같은 관계를 가지므로 단순한 방법이지만 출발지 및 목적지 포트번호의 역할과 구성비율을 특징인자로 하는 <표 5>를 이용하면 노드 유형을 쉽게 파악할 수 있다. 다음은 본 논문에서 제안하는 특징인자들이다.

- $r_{s.w}$: 출발지로 사용된 well-known 포트 비율
- $r_{d.w}$: 목적지로 사용된 well-known 포트 비율
- r_s : 출발지로 사용된 포트 비율
- r_d : 목적지로 사용된 포트 비율
- r_x : 대표 포트의 비율

<표 6> 노드 유형과 특징인자 관계

노드 유형	N	$r_{s.w}$	$r_{d.w}$	r_s	r_d	r_z
Attacker	○		○		○	△
Victim	△	○				△
Server	○	○		△		○
Host			○		○	

특징인자들을 도출하기 위해 수집되는 세션 데이터를 식(1), (2)와 같이 (IP주소, 프로토콜번호, 포트번호, 방향, 빈도수)를 원소로 갖는 전체 세션 집합과 $(a,p,x,d)\rightarrow y$로 사상되는 함수 f로 정의한다.

$$S=\{(a_1,p_1,x_1,d_1,y_1),\ldots,(a_n,p_n,x_n,d_n,y_n)\} \quad \cdots\cdots(1)$$

$$where,\ 0\le a\le 2^{32}, 0\le p\le 255, 0\le x\le 65535, 0\le d\le 1, y\ge 1$$

$$y=f(a,p,x,d) \quad \cdots\cdots(2)$$

여기서, a는 IP주소, p는 프로토콜번호, x는 포트번호, d는 방향(출발지=0, 목적지=1), y는 빈도수를 의미한다. 그리고, 사상함수 f는 (IP주소, 프로토콜번호, 포트번호), (IP주소, 프로토콜번호), (IP주소)를 이용하여 빈도수를 구하기 위한 식(3), (4), (5)와 같이 정의한다.

$$y=f(a,p,x),\ f(A,P,X)=\sum_{d=0}^{1} f(A,P,X,d) \quad \cdots\cdots(3)$$

$$y=f(a,p),\ f(A,P)=\sum_{x=0}^{65535}\sum_{d=0}^{1} f(A,P,x,d) \quad \cdots\cdots(4)$$

$$y=f(a),\ f(A)=\sum_{p=0}^{255}\sum_{x=0}^{65535}\sum_{d=0}^{1} f(A,p,x,d) \quad \cdots\cdots(5)$$

그러면, IP주소가 A인 노드에서 최대 빈도수는 식(6), (7)과 같이 구할 수 있으며, 이때 p^*를 노드의 대표 프로토콜번호, x^*를 대표 포트번호로 정의한다. 대표 포트번호인 x^*는 프로토콜별로 각각 존재하지만, 여기서는 p^*일 조건을 충족할때로 식(8)과 같이 한정한다. A노드의 세션 개수는 식(9)과 구할 수 있다.

$$\max(y)=f(A,p^*) \quad \cdots\cdots(6)$$

$$\max(y)=f(A,P,x^*) \quad \cdots\cdots(7)$$

$$\max(y)=f(A,p^*,x^*) \quad \cdots\cdots(8)$$

$$n(S_A)=f(A) \quad \cdots\cdots(9)$$

노드 유형을 분류하기 위해 각 노드의 세션 개수를 식(10)와 같이 N으로 할당하고, 대표 프로토콜의 출발지와 목적지로 사용된 전체 포트번호들의 빈도수, 그리고 well-known 포트번호들의 빈도수를 식(11),

(12), (13), (14)과 같이 계산한다. 식(15)는 대표 포트 번호에 의한 빈도수이다.

$$N = n(S_A) = f(A) = \sum_{p=0}^{255}\sum_{x=0}^{65535}\sum_{d=-1}^{1} f(A,p,x,d) \quad \cdots(10)$$

$$N_s = \sum_{x=0}^{65535} f(A,p^*,x,0) \quad \cdots\cdots(11)$$

$$N_{s.w} = \sum_{x=0}^{1023} f(A,p^*,x,0) \quad \cdots(12)$$

$$N_d = \sum_{x=0}^{65535} f(A,p^*,x,1) \quad \cdots\cdots(13)$$

$$N_{d.w} = \sum_{x=0}^{1023} f(A,p^*,x,1) \quad \cdots\cdots(14)$$

$$N_x = f(A,p^*,x^*) \quad \cdots\cdots(15)$$

마지막으로 특징인자들($r_{s.w}$, $r_{d.w}$, r_s, r_d, r_x)은 식 (16), (17), (18), (19), (20)에 의해 계산된다.

$$r_{s.w} = \frac{N_{s.w}}{N_s} \quad \cdots\cdots(16)$$

$$r_{d.w} = \frac{N_{d.w}}{N_d} \quad \cdots\cdots(17)$$

$$r_s = \frac{N_s}{N} \quad \cdots\cdots(18)$$

$$r_d = \frac{N_d}{N} \quad \cdots\cdots(19)$$

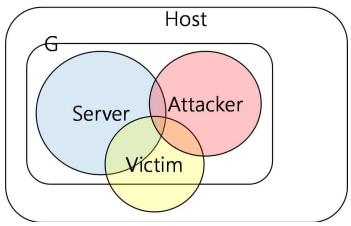
$$r_x = \frac{N_x}{N} \quad \cdots\cdots(20)$$

ICMP의 경우에는 일반적인 프로토콜과 다르게 예외 사항을 적용해야 한다. 하지만, 전술한 노드 유형 구분을 위한 특성인자들의 일관성을 유지하고, 노드 유형을 분류하는데 있어서 ICMP를 TCP나 UDP와 같은 방법으로 처리하는 것이 속도나 기능 측면에서 유리하다.

따라서, ICMP 세션의 목적지 포트번호(dpt), 즉 ICMP 메시지의 type과 code를 출발지 노드의 출발지 포트번호로 목적지 노드의 목적지 포트번호로 대입하면 일관성을 유지할 뿐만 아니라, 동일한 방법으로 노드 유형을 분류할 수 있다. 이후에 일반 프로토콜의 well-known 포트가 갖는 의미처럼 각각의 노드에서 ICMP 프로토콜에 할당된 포트번호들을 type과 code로 해석하면 된다. 예를 들면, 세션 집합 내의 어떤 노드에서 ICMP 프로토콜의 출발지 포트번호가 2048이면 Echo Request를 보낸 것이고, 출발지 포트번호가 0이면 Echo Reply를 보낸 것이다.

3.2 네트워크 공격과 정상 서비스 구분

보안 관점에서 네트워크 노드는 <그림 5>와 같이 일반 호스트, 서버, 공격자, 피해자로 분류할 수 있다. 노드 구분의 궁극적인 목표는 네트워크 상황 감시로써 정상적인 서버 활동과 네트워크 공격의 탐지이다. 사용자의 관심 사항은 공격자 및 피해자, 그리고 서버들의 정상적 또는 비정상적 활동이고 이를 통해 전반적인 네트워크 상황을 감시하는 것이다. 노드 그룹 G는 사용자의 관심 노드인 공격자, 피해자, 서버 노드들의 집합이다.



<그림 5> 네트워크 노드 유형

네트워크 공격과 정상 서비스를 구분하기 위해서 각각의 프로토콜별로 아래 분석 과정 [Step. 1]~[Step. 4]를 수행한 후, 중간 결과로써 노드의 유형을 구분한다. 이때, 각각의 프로토콜별로 다른 결과가 나타날 수 있는데, 최종 노드의 유형은 Attacker를 최우선으

로 하는 우선순위(Attaker > Victim > Server > Host)에 따라 결정한다.

[Step. 1] Host vs. G(Attacker/Server/Victim) 구분
- N : G ($\triangle$)
- x^* : G ($x^* < 1024$)
- r_x : G ($r_x \simeq 0$ or $r_x \simeq 1$)
- $x^* | r_x$: Host ($x^* \geq 1024$ and $r_x \ll 1$)

[Step. 2] G (Server/Victim) vs. Attacker 구분
- $r_s | r_d$: Attacker ($r_s \gg r_d$)
- $r_{s.w} | r_{d.w}$: Attacker ($r_{s.w} \simeq 0,\ r_{d.w} \simeq 0$)
- x^* : Attacker ($x^* \geq 1024$, ICMP:dpt=2048)
- r_x : Attacker ($r_x \simeq 0$)
- $x^* | r_x$: G ($x^* < 1024$ or $r_x \simeq 1$)
- neighbour: Server($\leftrightarrow$ Host)

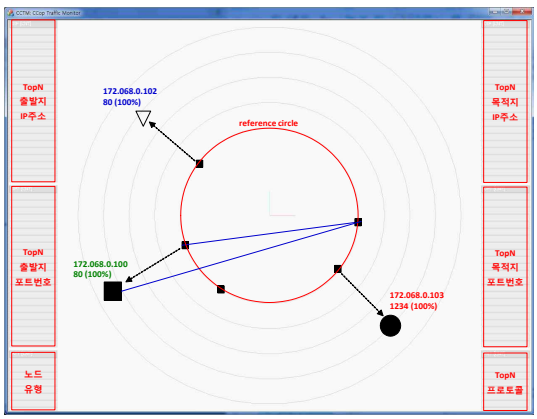
[Step. 3] Attack 종류 구분
- neighbour 관계 (1:1, N:1, 1:N)

[Step. 4] Server vs. Victim 구분
- $r_s | r_d$: Victim ($r_s \leq r_d$)
- $r_{s.w} | r_{d.w}$: Victim ($r_{s.w} \leq r_{d.w}$)
- $r_s | r_d$: Server ($r_s \simeq r_d$)
- $r_{s.w} | r_{d.w}$: Server ($r_{s.w} \simeq r_{d.w}$)
- x^* : Victim ($x^* < 1024$, ICMP:dpt=0)
- neighbour : Victim($\leftrightarrow$ Attacker)
- $x^* | r_x$: Server ($x^* < 1024$ and $r_x \simeq 1$)

3.3 시각화 인터페이스 설계

트래픽 세션의 포트 역할을 이용한 네트워크 공격 시각화 인터페이스 VisAttack의 프로토타입은 <그림 6>과 같다. VisAttack의 중앙에는 네트워크 전체의 상황을 감시하는 부분을 배치하고 좌·우측에는 세션 데이터의 통계와 주요 정보들을 필터링하는 메뉴들로 구성한다.

네트워크 상황을 감시한다는 것은 정상적인 서버 활동과 네트워크 공격을 탐지하는 행위로써, 공격자 및 피해자, 그리고 서버들의 정상적 또는 비정상적 활동을 감시하는 것이다. 모든 노드들은 기본적으로 일반 호스트 속성에서 시작하여 공격자, 서버, 그리고 피해자로 역할이 정해지는데, <그림 6>과 같이 기본 참조 원에서 시작하여 노드 유형에 따라 모양은 변화하고 세션 개수에 따라 점차 외곽으로 뻗어나가도록 강조하여 표현함으로써 직관성을 갖도록 한다. IP 주소에 따른 노드 위치는 실시간 렌더링을 제공하기 위해 고정 좌표, 즉 기본 참조 원의 원주 상에 배치하도록 한다. 여기서, 중요 노드들은 참조 원의 외곽으로 표시되기 때문에, 세션 개수가 작은 호스트들은 기본 참조 원 상에 중복되어 표시되더라도 공격을 인지하는 데에는 무리가 없도록 유지한다. 노드 유형에 따른 시각화 요소들은 <표 6>와 같으며, 노드간의 연결선은 종단간 연결 관계를 나타낸다.



<그림 6> VisAttack

세션 개수가 많은 중요 노드들은 유형에 따라 색상을 달리하여 IP 주소, 대표 포트번호, 대표 포트의 노드 내의 점유율 등을 함께 표시하여 시각화된 정보의 정확성을 갖도록 한다.

<표 6> 노드 유형에 따른 시각화 요소

노드 유형	위치	모양	크기	색(노드)	색(연결선)	색(설명)
Attacker	IP주소	●	N	x^*	x^*	Red
Victim	IP주소	▽	N	x^*	x^*	Blue
Server	IP주소	■	N	x^*	x^*	Green
Host	IP주소	·	N	x^*	x^*	Gray

주요 통계 정보 제공하고 통계 정보를 이용하여 필터링하기 위한 필터 메뉴는 트래픽 5-tuple의 속성별로 선택하거나 또는 노드 유형별(공격자, 서버, 피해자)로 선택하여 해당 정보만을 상황 감시에 활용하는 필터링 기능들로 구성된다. 특히 노드 유형별로 필터링하는 메뉴는 대량의 세션이 발생할 경우에도 관련 노드들의 연결만을 표시하기 때문에 공격 식별력의 감소 문제를 해결한다. VisAttack의 통계 정보 및 필터링 메뉴로는 다음과 같은 것이 있다.

- 출발지/목적지 IP 주소
- 출발지/목적지 포트 번호
- 프로토콜 번호
- 노드 유형 (공격자, 서버, 피해자, 호스트)

3.4 시험 데이터

시험에 사용된 데이터는 한국과학기술정보연구원의 라우터(KREONET과 미국의 StarTap 구간)를 통해 수집한 Netflow v5 데이터이다. 시험 데이터에는 UDP 1434 포트를 이용한 Slammer Worm 공격을 비롯한 각종 스캐닝 공격들이 포함되어 있으며, DDoS 공격은 자체 제작한 공격 도구를 이용하여 발생시켰다.

<표 7> 시험 데이터

파일명	이벤트 수	파일 크기	설명
N07270701	703,331	55M	일상적인 데이터 (1434 슬래머 웜 포함)
N07200703	73,572	6M	일상적인 데이터 (호스트 스캐닝 포함)

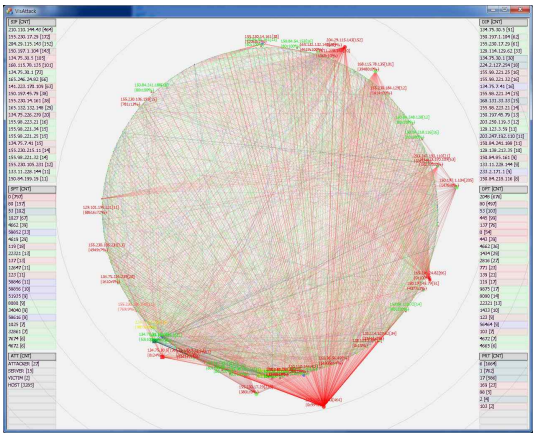
3.5 트래픽 세션 시각화

<그림 8>는 시험 데이터를 이용하여 시각화를 수행한 화면이다. (a)는 전체 노드들과 일상적인 네트워크 상황을, (b)는 호스트 스캔 및 포트 스캔 공격을, (c)는 DDoS 공격과 인터넷 웜 공격을, (d)는 정상 서버와 공격자 노드들의 세션 데이터를 각각 VisAttack로 시각화하고 감시하는 화면이다. 시험 데이터에 내에 포함되어 있었던 무분별한 노드들과 공격들이 정확하게 분류되고 시각화되었다. 이 방법은 매우 단순하기 때문에 실시간 감시에도 성능 상의 문제가 발생하지 않고, 시각적 분석을 통해 사용자의 직관성과 인지력을 높인다.

Ⅳ. 결론 및 향후 계획

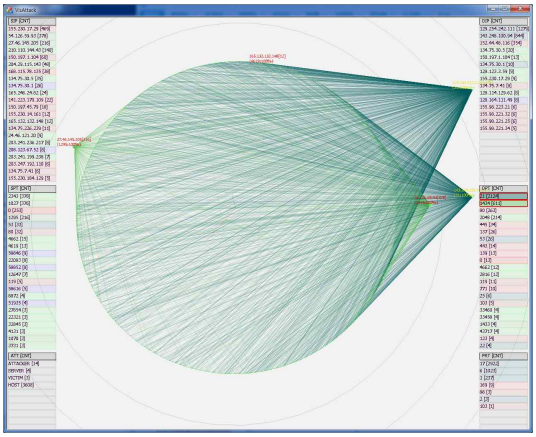
본 논문은 세션 데이터 내의 포트번호를 활용한 네트워크 공격 시각화 방법과 시각화 도구인 VisAttack에 대해 설계 및 구현하였다. 이 방법은 Netflow 데이터 기반의 시각화 기술들이 가지고 있던 반사현상 문제를 근본적으로 해결하고자 포트 역할이라는 개념을 정립하였고, 단순한 포트 역할과 구성비를 활용하여 자동적으로 노드들의 유형을 분류하고 식별하였다. 또한, 시각화 인터페이스는 대량의 세션 데이터가 발생할 경우 식별력이 떨어지던 기존 인터페이스 문제를 중요한 노드는 강조하는 방법으로 해결하였고, 종단간의 연결선은 유지하여 사용자의 직관성을 높였다.

본 논문에서 제안한 노드 분류 방법은 네트워크 관리자에게는 트래픽을 이용한 자동 자산 식별 방법으

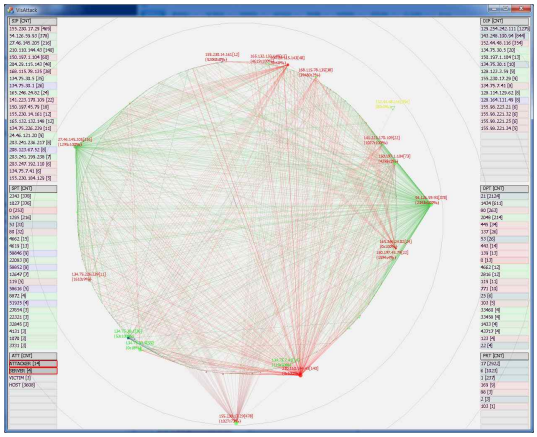


(a) 네트워크 전체 트래픽 세션 감시

(b) 호스트 스캔 및 포트 스캔 공격 세션



(c) DDoS 공격 및 인터넷 웜 공격 세션



(d) 공격자 및 정상 서버 세션

<그림 8> 트래픽 세션의 포트 역할을 이용한 네트워크 공격 시각화 (VisAttack)

로써 활용될 수 있고, 시각화 인터페이스는 다수의 보안시스템들(공격탐지시스템, 공격대응시스템)과 병행 동작하여 정확한 공격 탐지 및 대응에 활용될 수 있다. 또한, 보안관제 분야에 있어서는 네트워크의 상황을 인지하고 분석을 수행하는데 기여할 것으로 사료된다. 끝으로, 향후 VisAttack과 네트워크관리시스템, 공격탐지시스템, 공격대응시스템, 그리고 보안관제시스템과의 연동을 위해 연동 인터페이스를 정의 및 구현하고 실제 연동을 통해 효용성을 극대화할 계획이다.

참고문헌

[1] A. Oline and D. Reiners, “Exploring Three-Dimensional Visualization for Intrusion Detection,” Proceedings of the IEEE Workshop on Visualization for Computer Security, Oct. 2005, pp. 113-120.

[2] E. L. Malécot, M. Kohara, Y. Hori, and K. Sakurai, “Interactively Combining 2D and 3D Visualization for Network Traffic Monitoring,”

Proceedings of the 3rd International Workshop on Visualization for Computer Security, Nov. 2006, pp. 123-127.

[3] 장범환, "종단간의 유사 연결 패턴을 갖는 정상 서버 활동과 공격의 구분 및 탐지 방법," 정보보호학회논문지, 22(6), 2012, pp. 1315-1324.

[4] 장범환, 나중찬, 장종수, "보안 이벤트 시각화를 이용한 보안 상황 인지 기술," 정보보호학회지, 16(2), 2006, pp. 18-25.

[5] 정치윤, 손선경, 장범환, 나중찬, "시각화 기반의 효율적인 네트워크 보안 상황 분석 방법," 한국정보보호학회논문지, 19(3), 2009, pp. 107-117.

[6] Beom-Hwan Chang and Chi-Yoon Jeong, "An Efficient Network Attack Visualization using Security Quad and Cube," ETRI Journal, vol. 33 no 5, Oct. 2011, pp. 770-779.

[7] A. Giani, I. G. D. Souza, V. Berk, and G. CybenkoI, "Attribution and Aggregation of Network Flows for Security Analysis," Proceedings of the 2006 CERT FloCon Workshop, Oct. 2006, pp. 1-4.

[6] E. W. Bethel, S. Campbell, E. Dart, K. Stockinger, and K. Wu, "Accelerating Network Traffic Analytics Using Query-Driven Visualization," Proceedings of the 2006 IEEE Symposium on Visual Analytics Science and Technology, Oct. 2006, pp. 115-122.

[9] K. Lakkaraju, W. Yurcik, and A. J. Lee, "NVisionIP: Netflow Visualizations of System State for Security Situational Awareness," Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security, Oct. 2004, pp. 65-72.

[10] R. Ball, G. A. Fink, and C. North, "Home-Centric Visualization of Network Traffic for Security Administration," Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security, Oct. 2004, pp. 55-64.

[11] Y. Hu, "Adaptive Flow Aggregation-A New Solution for Robust Flow Monitoring under Security Attacks," Proceedings of the 10th IEEE/IFIP on Network Operations and Management Symposium, Apr. 2006, pp. 424-435.

[12] X. Yin, W. Yurcik, and A. Slagell, "The Design of VisFlowConnect-IP: A Link Analysis System for IP Security Situational Awareness," Proceedings of the 3rd IEEE International Workshop on Information Assurance, Mar. 2005, pp. 141-153.

[13] T. Taylor, D. Paterson, J. Glanfield and et al., "FloVis: Flow visualization system," Conference For Homeland Security, 2009. CATCH'09. Cybersecurity Applications & Technology, Mar. 2009, pp. 186-198.

[14] S. Krasser, G. Conti, J. Grizzard, J. Gribschaw, and H. Owen, "Real-Time and Forensic Network Data Analysis Using Animated and Coordinated Visualization," Proceedings of the 2005 IEEE Workshop on Information Assurance Workshop, Jun. 2005, pp. 42-49.

[15] G. Conti, and K. Abdullah, "Passive Visual Fingerprinting of Network Attack Tools," Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security, Oct. 2004, pp. 45-54.

[16] 장범환, "스마트그리드 전력망의 NSM 기반 보안

관리시스템 설계 및 구현," 디지털산업정보학회 논문지, 제9권, 제3호, 2013, pp. 107-117.
[17] 최희식, 전문석, 박재표, "SIP 플러딩 탐지 차단 실험방법에 대한 연구," 디지털산업정보학회논문지, 제7권, 제2호, 2011, pp. 39-46.

■ 저자소개 ■

장 범 환
Chang Beomhwan

2012년 3월~현재
호원대학교 사이버수사보안학부
교수
2003년 4월~2012년 2월
한국전자통신연구원
정보보호연구단 선임연구원
2003년 2월 성균관대학교 컴퓨터공학과
(공학박사)
1999년 2월 성균관대학교 컴퓨터공학과
(공학석사)
1997년 2월 성균관대학교 전자공학과(공학사)

관심분야 : 네트워크보안, 보안정보시각화,
보안상황인지, 제어시스템 보안
E-mail : bchang@howon. ac. kr

논문접수일: 2015년 11월 13일
수 정 일: 2015년 11월 22일
게재확정일: 2015년 11월 26일