情報保護學會論文誌
第14券 第4號, 2004. 8

실시간 웹 사용 현황과 이상 행위에 대한 시각화*

이 병 희†, 조 상 현†, 차 성 덕†
한국과학기술원 전자전산학과

Real-Time Visualization of Web Usage Patterns and Anomalous Sessions

Byung-Hee Lee†, Sang-Hyun Cho†, Sung-Deok Cha†
KAIST

요 약

현재의 웹 규모는 과거와 비교할 수 없을 만큼 복잡해지고 사용자의 패턴 또한 다양해지고 있다. 웹에 대한 공격은 점차 증가하고 있으며 이에 대한 탐지는 점점 어려워지고 있다. 이러한 웹 사이트의 효과적인 관리를 위하여 시각화를 통한 사용자들의 사용패턴과 보안 측면에서 이상행위 발생에 대한 신속하고 적절한 정보전달이 필요하다. 본 연구에서는 이러한 필요성에 기반을 두어 웹 서버의 access log를 분석하여 웹 사용 현황과 이상행위에 대한 효율적인 실시간 시각화를 위한 요구사항을 제안하고 이를 만족시키기 위해 SAD Viewer라는 툴을 개발하였다. 그리고 실제 시그네춰 위반 공격, DoS 공격, 코드레드 공격, Whisker 공격에 대한 실험을 통하여 구현된 Viewer가 효율적으로 사용자들의 사용패턴과 이상행위를 시각화함을 보여주었다.

ABSTRACT

As modern web services become enormously complex, web attacks has become frequent and serious. Existing security solutions such as firewalls or signature-based intrusion detection systems are generally inadequate in securing web services, and analysis of raw web log data is simply impractical for most organizations. Visual display of "interpreted" web logs, with emphasis on anomalous web requests, is essential for an organization to efficiently track web usage patterns and detect possible web attacks. In this paper, we discuss various issues related to effective real-time visualization of web usage patterns and anomalies. We implemented a software tool named SAD (session anomaly detection) Viewer to satisfy such need and conducted an empirical study in which anomalous web traffics such as Misuse attacks, DoS attacks, Code-Red worms and Whisker scans were injected. Our study confirms that SAD Viewer is useful in assisting web security engineers to monitor web usage patterns in general and anomalous web sessions in particular.

Keywords : *web visualization, anomaly detection, web usage patterns, anomalous sessions, visualization principles, SAD viewer, DoS attack, Code-Red worm, Whisker scans*

1. 서 론

인터넷 사용의 급증으로 인해, 웹을 이용한 정보의 전달 능력은 과거와 비교할 수 없을 정도로 발전했다. 특히 멀티미디어 콘텐츠의 등장으로 텍스트 기

접수일 : 2004년 4월 12일 ; 채택일 : 2004년 8월 5일
* 본 연구는 첨단정보기술연구센터(AITRC), 소프트웨어 프로세스 개선센터(SPIC), 인터넷 침해대응기술 연구센터(IIRTRC)의 지원을 받아 수행하였습니다.
† 주저자 : {bhlee, shcho, cha}@salmosa.kaist.ac.kr

반의 정보전달 능력을 극복함으로써 정보의 양이 많아지고 그로 인해 웹의 구성은 더욱 복잡해졌다. 웹 사이트는 이제 단순히 페이지와 링크의 구성이 아닌 사용자에게 인지적 가상공간으로 변화하고 있다. 웹 사이트 관리자는 다양한 웹 마이닝 기법[1]을 이용하여 웹 페이지들의 구조, 사용자들의 경향을 분석하고 이를 기반으로 효율적인 웹 페이지의 구성을 원한다. 또한 지난 2000년에 발생한 세계 유명 인터넷사이트 연쇄 해킹(yahoo, 아마존, e베이, 바이닷컴, CNN)의 예를 볼 때, 웹 서버의 보안은 앞으로 더욱 중요한 요소가 될 것이다. 따라서 웹 서버 관리를 위해서는 사용자들의 방문 경향뿐만 아니라 이상 행위에 대한 정보도 함께 제공해야 한다.

여기서 중요한 문제는 습득된 정보와 탐지된 이상 행위를 어떻게 보고하느냐 하는 것이다. 웹이라는 특성상 방대한 양의 데이터로부터 정보를 추출해 내기 때문에 정보의 양도 많고 복잡하다. 따라서 이를 어떻게 효과적으로 관리자에게 제공하는가 하는 것은 중요한 문제이다. 기존의 대부분의 연구에서는 이 두 가지(웹 사용 현황의 시각화, 이상 행위의 보고)를 별개의 주제로 연구하였다. 먼저 웹 사용 현황의 시각화에 관한 연구는 사용자들의 통계적인 경향이나 이동 패턴을 시각적으로 보여준다. 하지만 이상 행위에 대한 시각화나 시간의 흐름에 따른 사용자들의 이동 경향은 보여주지 못하고 있다. 반면에 이상 행위의 보고에 관한 연구는 주로 텍스트 형태로 이상 행위가 보고되기 때문에 관리자가 직관적으로 이해하기 어렵고 전문적인 지식을 요구하며, 이는 신속한 대응 방법을 찾는데 어려움을 주게 된다.

이러한 문제점들을 해결하기 위해서 본 연구는 다음의 두 가지 목적으로 진행되었다. 첫째는 웹 사용 현황의 실시간 시각화이다. 웹 서버의 access log로부터 마이닝 기법을 이용하여 사용자들이 주로 사용하는 경로나 자주 방문하는 페이지들에 대한 정보들을 얻을 수 있다. 이런 정보들을 실시간으로 웹 서버로부터 가져와서 그래프 레이아웃을 이용하여 변화하는 과정을 시각적으로 표현하는 것이 웹 사용 현황의 시각화이다. 두 번째는 이상 행위의 시각화이다. 현재까지의 IDS(Intrusion Detection System : 침입탐지시스템)는 시그네춰에 기반을 둔 오용 탐지 시스템이다. 하지만 웹 서비스는 다른 인터넷 서비스와는 다르게 개방적 서비스이고 복잡하기 때문에 특정 시그네춰를 만들기 어려운 문제가 있다.[2] 따라서 웹 환경에서는 이상 탐지 기법의 사용이 불가피하다. 본 연구에서는 사용자들의 웹 페이지 방문 순서를 프로파일링 하고 이와 다른 방문형태를 찾아 시각화함으로써 관리자가 직관적으로 웹 사이트의 이상 행위를 탐지할 수 있도록 하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 관련 연구로써 현재까지 연구된 웹 사용 현황 시각화와 웹 공격 보고에 대해 알아보고 선행 연구로써 SAD(Session Anomaly Detection) System에 대해 설명하도록 하겠다. 3장은 효율적인 시각화를 위한 요구사항들에 대해서 설명하고, 4장에서 본 연구의 결과물인 SAD Viewer가 제안된 시각화 요구사항들을 어떻게 구현하였는지 보여주겠다. 5장에서는 SAD Viewer의 전체적인 구성도와 실제 실험을 통해 SAD Viewer의 성능에 대해서 평가한다. 그리고 마지막으로 6장에서 결론을 맺도록 하겠다.

II. 관련 연구

이 장에서는 웹 사용 현황의 시각화와 웹 사이트의 공격에 대한 보고로 구분하여 관련 연구들에 대해서 알아본다. 또한 SAD Viewer가 효율적인 시각화를 이룰 수 있도록 필요한 정보를 제공해 주는 SAD System에 대해서 설명하도록 하겠다.

2.1 웹 사용 현황 시각화

표 1. 웹 사용 현황의 시각화

이 름	특 징	단 점
WebViz[3]	웹 페이지들과 사용자들의 이동 경로를 시각화	데이터 마이닝 기법 적용의 부재
WUM[4]	MINT라는 마이닝 툴을 이용하여 웹 페이지 방문 경로를 표시	시각적 표현의 부족
VISVIP[5]	3D 기법을 이용하여 웹 페이지 방문 경로 시각화	사용자 로그인이 필요 (client customization)
WebQuilt[6]	중간에 프록시(proxy) 서버를 두어 사용자 로그인 과정을 제거	이동 특성(traversal property)나 동적 페이지(dynamic page)들에 대한 고찰 부재
Naviz[7]	이동 특성(traversal property)의 적용과 동적 페이지(dynamic page)들에 대한 시각화	실시간 변화과정에 대한 시각화 부재 이상 행위에 대한 시각화 부재

〔표 1〕은 웹 사용 현황의 시각화에 대한 연구의 변화를 보여주고 있다. 변화의 핵심은 〔그림 1〕에서처럼 랜덤하게 페이지들의 좌표를 설정하여 화면에 보여주는 형태에서 〔그림 2〕에서와 같이 계층적으로 페이지들을 배치시키는 형태로 발전하였다는 것이다. 또한 〔그림 1〕의 경우에는 로그에 있는 정보를 단순히 시각화 한 것인데 반해, 〔그림 2〕의 경우에는 데이터 마이닝 기법을 적용하여 관련 있는 페이지들은 근접하여 위치시키고 색깔에 따라 방문 빈도를 차별화하여 구성하였다. 하지만 이러한 시각화의 문제점은 사용자의 일반적인 행위에 대해서는 시각화하여 보여주지만 이상행위에 대해서는 시각화하지 못한다는 점이다. 또한 실시간으로 사용자의 웹 사용 현황을 보여주는 것이 아니라, 어느 정도 시간이 경과한 다음에 웹 로그 분석을 통해 정적인 사용자 경향만을 보여주고 있다.

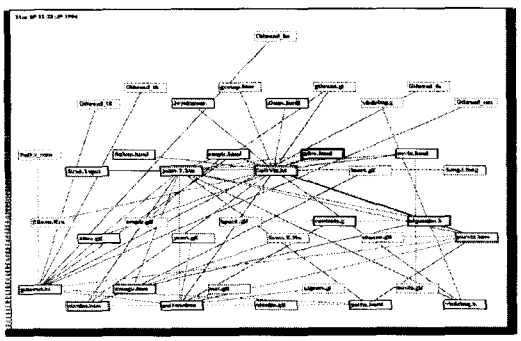


그림 1. WebViz

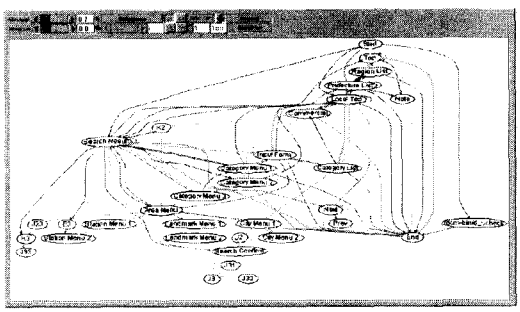


그림 2. Naviz

2.2 웹 공격의 보고

〔그림 3〕은 대표적인 네트워크 IDS인 Snort의 탐지 결과를 웹 페이지로 보여주는 SnortSnarf[8]

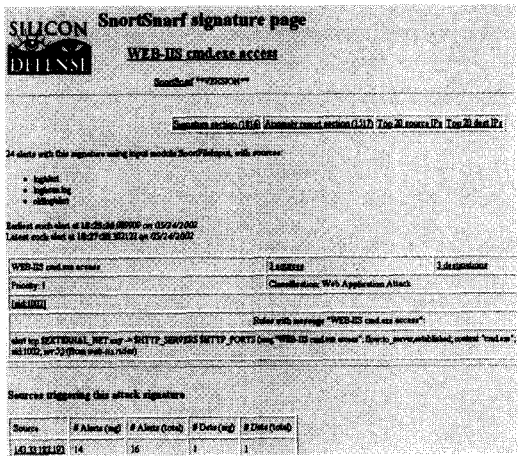


그림 3. SnortSnarf

의 예이다. 〔그림 3〕에서 보여주는 것과 같이 대부분의 공격에 대한 경보는 누가, 어떤 페이지에, 어떤 이상 행위를 했는지를 텍스트 형태로 보여주고 있다. Snort의 탐지 결과에 대해서 웹을 통해 정보를 보여주기는 하나 텍스트 형태로 보여주기 때문에 직관적으로 이해하기는 어려운 단점이 있다. 또한 탐지 기법 역시 과거 문제가 되었던 CGI에 대한 요청이나 특정 웜의 패턴의 탐지에만 국한되고 있기 때문에 새로운 형태의 웜 공격이나 다양한 형태의 웹 기반 공격에는 대응하기 어렵다는 문제가 있다. 따라서 이상 탐지 기법의 적용과 시각화를 위한 웹 공격의 보고에 대한 연구가 필요하다.

2.3 SAD(Session Anomaly Detection) System

SAD System은 웹 서버 로그를 기반으로 일반적인 웹 페이지 방문 시퀀스의 프로파일을 만들고, 현재의 행위가 만들어진 프로파일과 비교하여 정상사용인지 이상사용인지를 구분하는 이상탐지시스템이다.[9] 〔그림 4〕를 보면 SAD System은 웹 로그로부터 사용자들의 세션[1]을 찾아내고 이 세션의 발생 가능성을 기존의 프로파일과 비교/분석하여 이상인지 아닌지를 판단하는 시스템이다. 판단의 결과는 이상점수로 나타나게 되는데 이상점수가 높을수록 기존 패턴에 없는 형태이기 때문에 관리자의 분석이 필요할 것이다. 현재까지 SAD System의 보고 방식은

1) 사용자가 웹 사이트에 들어와서 나가기까지의 일련의 방문 순서 .

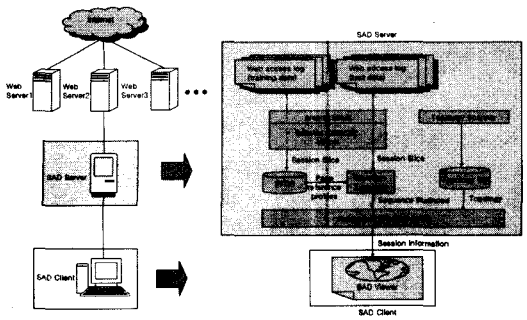


그림 4. SAD System

텍스트 형태로 구현되어 있으며 세션 내 페이지들의 이상 경로 및 페이지와 링크에 대한 정보를 제공한다. 시스템의 효율적인 이상탐지 성과에도 불구하고 현재의 텍스트기반의 이상 경보체계는 보안 관리자로 하여금 시기적절한 대처를 할 수 없도록 만들며 또한 다량의 이상 경보는 정확한 판단을 저해하게 된다. SAD System의 적절한 운영을 위해서도 좀 더 유용한 경보시스템이 필요하다.

III. 효율적인 시각화 요구사항

본 연구의 목적은 웹 사용자의 웹 사이트 접속 유형을 세션별로 시각화하여 웹 사이트의 관리를 효율적으로 할 수 있는 정보를 제공하고 침입탐지 관점에서 이상세션이 발생할 경우 이를 효과적으로 경보 할 수 있는 SAD Viewer의 완성에 있다. 이 장에서는 이를 구현하기 위해 필요한 시각화 요구사항을 일반적인 웹 사용 현황의 시각화와 이상 행위의 시각화로 구분하여 알아보도록 하겠다.

3.1 일반적인 웹 사용 현황의 시각화를 위한 요구사항

[그림 5]는 약 40명의 사용자를 갖고 있는 대학의 한 연구실의 웹 서버에서 10분가량의 로그를 분석하여 그래픽 파일들(예: *.jpg, *.gif)에 대한 요청 부분을 제거하고 나머지 페이지들에 대한 접근 경로들을 그려준 것이다. [그림 5]를 보게 되면 너무 복잡해서 관리자는 어떤 정보도 얻을 수 없을 것이다. 이것은 기존의 텍스트형태로 표현하는 것보다 오히려 효율성이 떨어짐을 알 수 있다. 이처럼 단순히 사용자들의 경로를 화면상에 펼쳐 놓는 것은 무의미하다. 따라서 효율적인 웹 사용 현황의 시각화를 위해서는 다음 사항들을 고려해야한다.

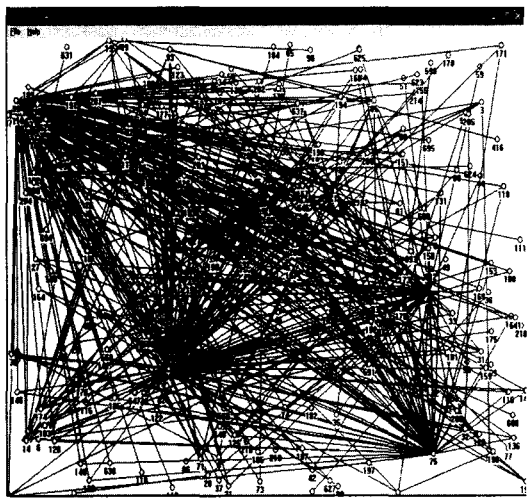


그림 5. 효율적이지 못한 웹 시각화

요구사항 1. 제한된 영역(화면)내에 웹 페이지들의 체계적인 배치

[그림 5]에서처럼 단순히 페이지들을 랜덤하게 위치시키는 것은 복잡성을 증가시킨다. 또한 웹 서버내의 모든 페이지를 한 화면에 나타내는 것은 불가능한 일이다. 따라서 효율적인 시각화를 위해서는 어떤 페이지들을 어떻게 위치시킬지 결정하는 작업이 필요하다.

요구사항 2. 방문 빈도수에 의한 페이지와 링크의 차별 시각화 정책

단순히 페이지와 그들 간의 링크를 화면에 표현해 준다고 하는 것은 관리자에게 많은 양의 정보를 제공하지 않는다. 관리자가 관심을 갖고 있는 것은 페이지와 링크뿐만 아니라, 사용자들이 어떤 페이지를 많이 방문하고 적게 방문하는지에 대한 정보와 어떤 경로를 주로 이용하는지에 대한 정보도 포함될 것이다. 따라서 이러한 정보들도 효율적인 시각화를 위해 필요한 요소들이다.

요구사항 3. 사용자 세션의 실시간 업데이트

관리자는 사용자들의 페이지 방문 경로를 실시간으로 확인하고 싶을 것이다. 특히 이것은 실시간 침입 탐지의 관점에서 볼 때, 반드시 필요한 요소라고 할 수 있다. 만약 이 요구사항이 만족된다면 관리자는 사용자들의 페이지 방문 경향이 시간에 따라 어떻게 변화하는지 확인할 수 있게 된다. 이는 효율적인 페이지 관리에 도움을 줄 것이다.

요구사항 4. 원하는 정보를 제공하기 위한 관리자와의 효율적인 상호작용(interaction)

시각화 툴이 관리자가 원하는 정보를 완벽하게 찾아서 보여줄 수는 없을 것이다. 따라서 관리자와의 계속적인 상호작용을 통하여 원하는 정보에 가까이 접근하는 과정이 필요하다. 관리자와의 효율적인 상호작용의 구현은 효율적인 시각화의 시작이라고 할 수 있다.

요구사항 5. 전체적인 웹 사용 현황을 보여주는 통계적인 정보

관리자는 특정 사용자의 웹 사용 현황뿐만 아니라 전체 사용자들의 웹 사용 현황을 확인하고 싶을 것이다. 이를 위해서 가장 좋은 방법이 통계적 기법을 이용하는 것이다. 이를 통해 전체 사용자의 패턴을 확인할 수 있고 웹사이트의 효과적인 운영에 도움을 준다. 효율적인 통계 기법의 사용과 보고의 결정이 효율적인 웹 사용 현황의 시각화에 큰 영향을 줄 것이다.

3.2 이상 행위의 시각화

효율적인 이상 행위 시각화를 위해서는 위에 제시된 요구사항들 외에도 다음의 요구사항들이 추가로 필요하다. 즉 아래에 제시되는 요구사항들은 이상 행위의 시각화와 밀접하게 관련된 부분이다.

요구사항 6. 정상적인 사용자 세션과 비정상적인 사용자 세션의 구분

데이터 마이닝 기법을 이용하여 정상 세션과 비정상 세션을 구분하였다면 이를 시각적으로 어떻게 보여줄 지에 대한 고찰이 필요하다. 이상 행위 시각화의 목적이 이상 행위에 대한 빠른 탐지와 신속한 대응에 도움을 주는 것이므로 좀 더 쉽고 명확하게 이상 세션을 구분해 주는 과정이 필요하다.

요구사항 7. 이상세션의 경우, 세션을 만든 사용자에 관한 정보 및 세션을 구성하는 페이지에 관한 상세 정보

이상 세션의 경우에는 공격으로 분류될 확률이 다른 세션에 비해 높기 때문에 이 세션을 누가 생성했고 어떤 경로로 들어왔는지 신속히 확인해야 할 필요가 있다. 이러한 정보들을 기존의 연구에서처럼 단순히 텍스트 형태로 보여주는 것보다 시각적으로 보여주는 것이 효율적일 것이다. 이런 정보의 제공을 통해서 관리자는 이상 행위에 대해 빠르게 대응 방안을 모색할 수 있다.

요구사항 8. 이상 세션으로 판단한 근거를 보여줄 수 있는 정보

관리자는 이상 세션이 발견 되었을 때 왜 이상 세션이 되었는지 원인을 알고 싶을 것이다. 따라서 해당 침입 탐지 메커니즘에서 어떤 사항을 위반하여 이상 세션으로 간주되었는지에 대한 정보를 제공할 필요가 있다. 만약 이러한 정보가 제공되지 않는다면 관리자는 직접 원인을 밝혀야 하고 이에 따른 시간의 경과로 많은 피해를 받을 수 있게 된다.

위의 **요구사항 1~8**이 만족된다면 관리자는 어떤 사용자가 어떤 페이지들에 대하여 어떤 의도로 접근 하였는지를 확인할 수 있을 것이다. 세부적으로는 정보 확인을 통해 침입인지 아니면 평소에 사용되지 않던 페이지에 대한 단순 접근인지 또는 바이러스나 웜에 의한 이상 발생인지를 알 수 있다.

〔표 2〕는 기존 연구들의 요구사항 만족도를 나타낸 것이다. **요구사항 6~8**은 이상 행위의 시각화와 관련되어 있기 때문에 기존의 시각화 툴들이 전혀 만족하지 않는 것을 확인할 수 있다.

표 2. 기존 연구의 요구사항 만족도 (O : 만족, △ : 부분 만족, X : 비만족)

이 름	요구사항1	요구사항2	요구사항3	요구사항4	요구사항5	요구사항6	요구사항7	요구사항8
WebViz[3]	X	X	X	△	X	X	X	X
WUM[4]	X	X	X	O	X	X	X	X
VISVIP[5]	O	X	X	O	X	X	X	X
WebQuilt[6]	O	O	X	O	△	X	X	X
Naviz[7]	O	O	X	O	O	X	X	X
SAD Viewer	O	O	O	O	O	O	O	O

표 3. Alberto가 규정한 시각화 원칙(Principles)

이 름	정 의	종 류
레이아웃(layout)	페이지와 링크를 화면에 어떻게 나타낼 지를 결정하는 것	spring-based[12], hierarchical
추상화(abstraction)	필요한 정보만 추출해서 화면에 보여주는 것	structural, content-related
포커스(focus)	원하는 정보만을 강조해서 화면에 보여주는 것	zooming, 특정부분 강조
상호작용(interaction)	사용자와 필요한 정보를 주고받는 것	various

Ⅳ. SAD Viewer

기존의 연구에서는 효율적인 웹 행위 시각화를 만족시키기 위해 [표 3]에서와 같은 네 가지 원칙을 제시하였다.[10] 본 연구에서는 위의 4가지 원칙 외에 바인딩(binding) 개념을 추가하였다. 그 이유는 시각화의 목적이 단순히 페이지와 링크의 관계만을 표현하는 것이 아니라 그 이상의 정보 (예: 사용자들의 방문 빈도수에 따라 다른 색깔로 페이지들을 표시함으로써 페이지 관리에 도움을 주는 것)도 시각적으로 보여주는 것이기 때문에 어떻게 바인딩을 구현하였는지도 시각화를 위한 중요한 원칙이 되기 때문이다.

요구사항 1은 효율적인 레이아웃의 구성을 통해 달성될 수 있다. 레이아웃을 결정하기 위해서는 다음의 두 가지에 대해 생각해 봐야 한다. 첫째는 스케일(scale) 문제이다. 화면의 제약 때문에 모든 페이지들을 표현할 수는 없다. 따라서 최대한 많은 페이지를 표현할 수 있도록 효율적인 레이아웃을 정의해야 한다. SAD Viewer에서는 페이지 레이아웃[2](page layout)과 깊이 레이아웃(depth layout)을 이용하여 페이지들이 중첩되지 않도록 구성하였다. 두 번째는 작업의 특성이다. 이것은 관리자가 의도에 따라서 그래프의 구조를 바꿔야 함을 의미한다. 예를 들어 관리자가 특정 사용자의 이동 경로를 보고 싶다면 계층적 레이아웃(hierarchical layout)이 좋을 것이다. 계층적 레이아웃은 트리 형태처럼 페이지들을 순서화하여 배치시키는 방법이다. 하지만 페이지들 간의 연관성이 얼마나 있는지를 보고 싶을 때에는 스프링 기반 레이아웃(spring-based layout)이 효율적일 것이다. 스프링 기반 레이아웃은 연관관계에 따라 관련 있는 페이지들을 근접해서 위치시키는 방법이다. 예를 들면 A페이지를 방문한 대부분의 사용자가 B페이지를 방문 하였다면 A와 B를 물리적인 구조에 상관없이 인접해서 위치시키는 경우이다. SAD Viewer에서는 계층적 레이아웃을 사용하였다. 왜냐하면 SAD Viewer의 목적이 웹 사용 현황과 이상 세션을 보여주는 것이므로 일련의 방문 순서가 중요하기 때문이다. 또한 페이지 간의 연관 관계는 해당 페이지간의 링크 두께로도 충분히 예측 가능하다.

계층적 레이아웃을 어떻게 구성할 지에 대해서는 수평 분할, 수직 분할, 동심원 분할 등이 있을 수 있다. SAD Viewer의 초기 버전에서는 [그림 6-(가)]처럼 수평 분할을 이용하였는데 다음과 같은 문제가 나타났다. 실험 환경의 깊이(depth)에 따른 페이지 개수와 방문 빈도를 분석한 결과, 방문 페이지의 개수가 점점 증가하다가 감소하는 것을 알 수 있었다. 이것은 웹 사이트의 특성에 따라 다소 차이는 있겠지만 대체적으로 이러한 분포를 따를 것으로 예상된다. 왜냐하면 웹 사이트의 중요 페이지들은 대부분 중간 깊이에 위치하기 때문이다. 예를 들어 어떤 쇼핑몰 사이트가 있을 때, 전체 상품에 대한 설명은 상위 깊이에 위치하겠지만 자세한 설명은 해당 상품을 따라 내려간 하위 깊이에 위치하게 된다. 하지만 그보다 더 하위 깊이의 경우에는 실제 상품을 구입한다거나 더 자세한 정보를 얻고 싶을 때 방문하게 되므로 상대적으로 방문 빈도수는 적게 된다. 이를 통해 중간 depth의 페이지들을 사용자들은 가장 많

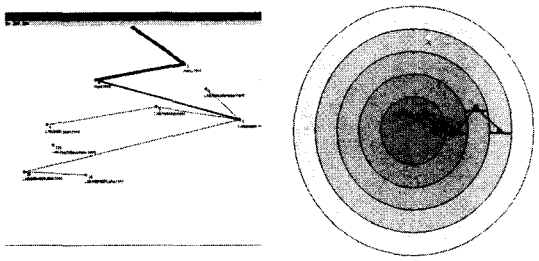


그림 6 (가) 수평 분할의 예 (나) 동심원 분할의 예

2) 각 페이지마다 URI 이름내의 "/" 개수에 따라 페이지들을 계층화한 것

이 방문할 것이다. 따라서 SAD Viewer에서는 화면의 중앙을 중심으로 깊이에 따라 동심원을 그리며 깊이가 증가 되도록 구성하였다.[그림 6-(나)] 깊이가 증가함에 따라 동심원에 들어갈 수 있는 페이지도 증가하게 되어 많은 양의 페이지를 효율적으로 보여줄 수 있게 된다. 또한 웹 페이지에 없는 페이지에 대한 과도한 요청으로 웹 서비스를 마비시키려는 공격의 탐지를 위해 존재하지 않는 페이지에 대한 요청은 동심원 외부에 랜덤하게 노드를 위치시키도록 구성하였다. 이를 통해 존재하지 않는 페이지에 대한 과도한 요청을 시각적으로 확인할 수 있다.

요구사항 2는 바인딩(binding)을 통해 달성될 수 있다. 바인딩은 웹 페이지의 구성뿐만 아니라 관리자가 좀 더 알고 싶은 정보(예: 가장 많이 가는 페이지, 주로 이용하는 경로 등)들을 화면에 어떻게 보여줄지 결정하는 것이다.[11] SAD Viewer서는 페이지와 링크에 대해 색깔과 두께를 이용하여 바인딩을 구현하였다. 페이지와 링크의 두께는 1~10단계로 구분하였는데, 이유는 페이지 방문 빈도수가 드문(sparse) 경우, 빈도수가 적으면 상대적으로 너무 작게 그려지기 때문이다. 이것은 효율적인 정보 전달이 아니다. 따라서 비율에 따라 두께를 구분하여 표현하는 것이 필요하다. 색깔에 관해서도 10단계로 구분하였고 각 단계 당 구분이 쉽도록 보색 관계를 이용하였다.[그림 7]

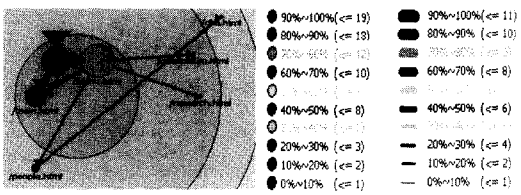


그림 7. 바인딩 적용의 예

요구사항 3은 시간의 흐름을 고려한 레이아웃이 되어야 한다는 것이다. 웹 사용 분석에서는 시간의 흐름이 중요하기 때문에 시간 요소를 레이아웃에 반영해야 한다. SAD Viewer에서는 [그림 8]과 같이 시간적 요소를 추가함으로써 관리자는 시간의 흐름에 따른 사용자의 경향을 확인할 수 있고, 실시간으로 이상 행위에 대한 탐지를 할 수 있다. [그림 8]을 보면 1분 사이에 새로운 세션이 추가된 것을 확인할 수 있다.

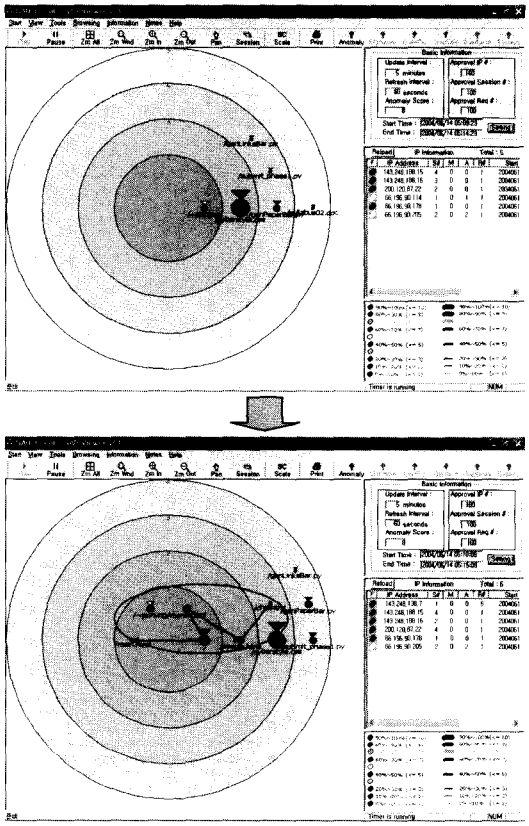


그림 8. 시간의 흐름에 따른 세션의 변화

요구사항 4는 효율적인 추상화(abstraction)와 포커스(focus)의 구현으로 달성된다. 추상화는 필요한 정보만 추출해서 보여주고 불필요한 부분은 화면에서 제거함으로써 복잡성을 감소시키고 가독성을 증가시키는 작업이다. SAD Viewer에서는 세션의 생성 시간, 세션의 상태(정상, 이상), 사용자, 페이지 이름, 페이지와 링크의 방문 빈도수 등 총 5가지 카테고리로 추상화를 구현하였다. 예를 들면, [그림 9]는 전체 세션, [그림 10]은 전체 세션 중 이상 세션만을 시각화한 것이다. 그림에서 알 수 있듯이 추상화에 의해 복잡성이 줄어드는 것을 확인할 수 있다.

관리자는 특정 사용자의 페이지 방문 순서, 즉 세션을 보고 싶어 할 것이다. 하지만 세션의 개수가 많아서 잘 구분이 안 될 때는 포커스 기법을 이용할 수 있다. [그림 11]을 보면 관리자가 특정 세션을 선택했을 때 해당 세션에 포함되는 페이지와 링크만 하이라이트(highlight) 시켜줌으로써 포커스 기법을 구현하였다. 이를 통해 관리자는 선택한 세션에 대한 구별을 쉽게 할 수 있게 된다.

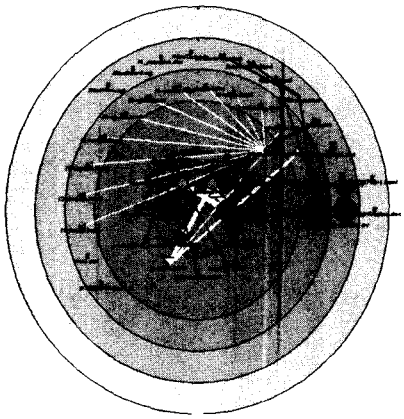


그림 9. 전체 세션

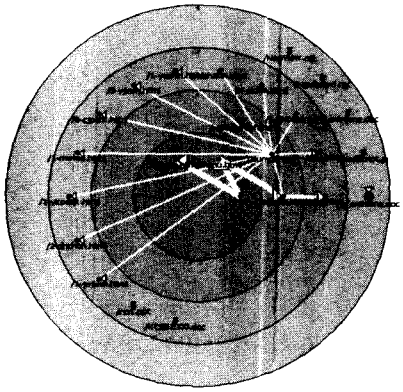


그림 10. 이상 세션

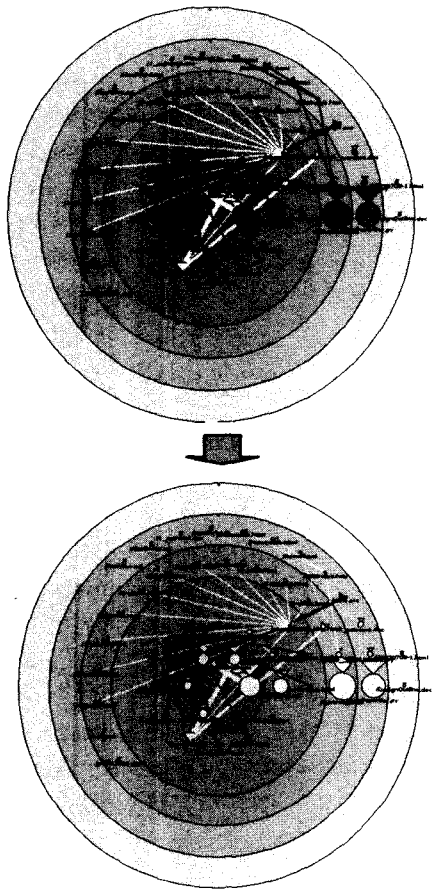


그림 11. 특정 세션에 대한 focus 적용

요구사항 5와 관련된 문제에서 SAD Viewer는 Session Statistics View를 통해 웹 행위에 대한 통계적인 정보를 제공한다. [그림 12]에서 보는 바와 같이 기준 시간 내의 전체 세션, 정상 세션, 이상 세션에 대한 통계적인 분포를 그래프 형태로 확인할 수 있다. 이 통계적인 그래프는 DoS(Denial of Service) 공격과 같이 과도한 요청이 많은 공격의 탐지에 효율적일 것이다.

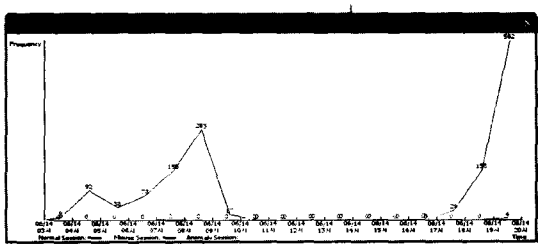


그림 12. 세션 통계

요구사항 6은 시각화 원칙 중 바인딩을 통해 구현 되며 SAD Viewer에서는 이상 세션인 경우에는 세션에 포함된 링크가 노란색과 붉은색으로 표현되게 함으로써 정상 세션과 이상 세션을 구분하였다.

요구사항 7과 **요구사항 8**은 Session Information View를 통해 구현되었다. [그림 13]은 Session Information View의 예를 나타낸다.

먼저 **요구사항 7**과 관련해서는 해당 세션의 방문 순서가 [그림 13]의 우측 상단에 표시됨으로써 관리자는 공격자의 공격 경로를 쉽게 파악할 수 있다. 또한 우측 하단에 각각의 링크에 대해 이상 점수를 알려 줌으로써 왜 해당 세션이 이상 세션으로 간주 되었는지에 대한 정보를 보여준다. [그림 13]을 보면 우측 중앙에 143.248.138.7의 IP가 일련의 페이지를 방문 하였는데 그 중 10번의 페이지 요청이 이상 점수가 높게 나온 것을 확인할 수 있다. 이런 경향으

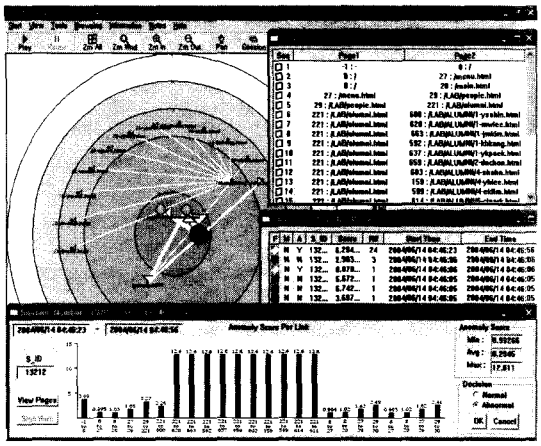


그림 13. Session Information Viewer

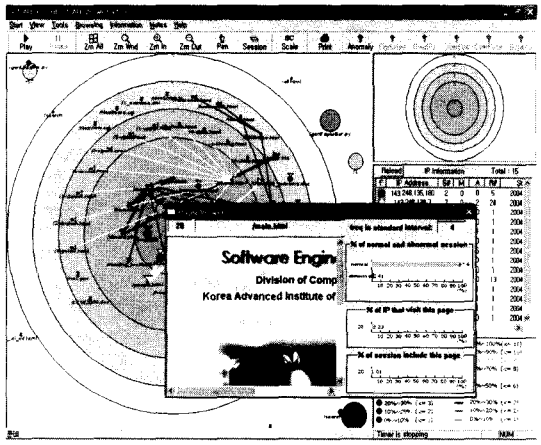


그림 14. Page Information Viewer

로 전체 세션의 이상 점수가 기준 점수인 6보다 높은 6.29가 되었기 때문에 이상 세션으로 간주하여 노란색 링크로 화면에 그려진 것이다. 또한 [그림 14]를 보면 특정 페이지에 대한 세부 정보(예: 페이지 내용, 해당 페이지가 이상 세션에 포함될 확률 등)를 보여주고 있다.

요구사항 8과 관련해서는 [그림 13]의 중앙 하단에 있는 이상 점수 그래프를 통해 확인할 수 있다. 이 부분은 세션내의 각 링크의 이상 점수를 나타낸다. 베이지언 추정은 조건부 확률을 이용하기 때문에 이상 점수가 높다는 것은 해당 링크를 방문할 확률이 적다는 의미이다. [그림 13]의 경우에는 이 세션이 기준 이상 점수인 6보다 높은 6.29의 이상 점수를 갖고 있기 때문에 이상으로 간주되어 노란색으로 화면에 나타난 것이다. 만약 로그인이 필요한 페이지를 우회하려는 시도가 있었다면 해당 링크에 대한 부분에 막대그래프가 최대가 될 것이다. 왜냐하면 프로파일링 기간 동안 이러한 시도가 없었기 때문이다.

V. 실험 및 성능 평가

SAD Viewer의 전체적인 구성과 웹 사용 현황과 이상 행위의 탐지 결과를 어떻게 효율적으로 보여주는 지에 대해 알아보겠다. 실험은 약 40명의 사용자가 있는 연구실의 웹서버에서 생성된 로그를 대상으로 수행되었다. 한 대의 아파치 웹 서버를 사용하였으며, 1개월의 로그를 바탕으로 사용자들의 이용패턴을 구축하고, 하루 동안 수집된 로그를 바탕으로 - 모의 공격을 포함하여 - 이상탐지 여부를 실험하였다.

실험 당일에는 웹 공격에 대한 시각화를 실험하기 whisker라는 툴을 이용하여 웹 서버에 대한 공격을 시도하였고 코드레드 공격도 시도하였다. 이 실험은 베이지언 추정 기법을 이용한 이상 탐지가 얼마나 정확한가 하는 것을 보는 것이 아니라, SAD Viewer가 얼마나 웹 사용 현황과 이상 세션을 효율적으로 표현하는가에 초점을 맞추고 있다.

5.1 SAD Viewer 구성도

[그림 15]는 SAD Viewer의 전체적인 구성을 나타낸다. SAD Viewer Preprocessor는 클라이언트 데이터베이스와 Misuse Detection Module로부터 각 viewer에게 필요한 정보를 제공해 주는 모듈이다. Misuse Detection Module은 SAD Viewer가 제공하는 오용 탐지 알람들을 탐지하는 모듈이 된다. Main Viewer는 일정 기간 내의 페이지와 링크의 이용 결과를 시각적으로 보여주는 부분

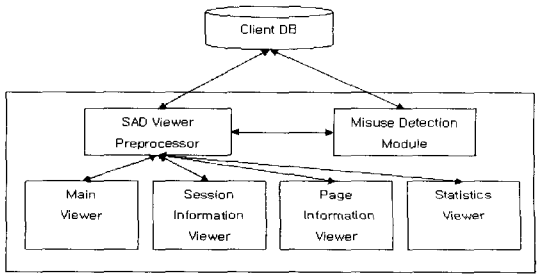


그림 15. SAD Viewer 구성도

이다. 앞에서 설명한 페이지와 링크의 바인딩 메커니즘에 따라 페이지와 링크의 크기 및 색깔이 다르게 표현되고 이상 점수에 따라 이상 세션은 붉은색으로 표시 된다. 이 부분을 통해서 관리자는 사용자들의 웹 사용 경향이나 이상 행위를 확인할 수 있다. 또한 현재 그려진 세션에 대한 시작 시간과 종료 시간, update interval, refresh interval, 이상 점수 등의 정보들도 포함한다. 관리자는 이상 세션에 대해서 왜 이상이 있는지에 대해서 분석을 할 것이다. 이런 분석에 도움을 주기 위한 부분이 Session Information Viewer이다. 또한 관리자가 페이지를 관리하는데 도움을 주기 위한 부분이 Page Information Viewer이다. 이 부분을 통해 관리자는 특정 페이지에 대한 사용자들의 경향을 예측할 수 있다. 마지막 Statistics Viewer는 통계적인 정보를 제공함으로써 전체적인 웹 사이트 관리에 도움을 줄 수 있다. SAD Viewer는 윈도우 기반에서 동작하며 Visual C++를 이용하여 구현되었다.

5.2 알람 모드

SAD Viewer에서는 두 가지 알람 모드가 있다. 하나는 SAD의 결과로 나온 이상 탐지에 대한 알람 모드이다. 이 모드에 대해서 SAD Viewer는 노란색을 이용하여 경고 메시지와 링크를 표현하였다.[그림 16] 다른 하나는 오용 탐지에 대한 알람 모드이다. 오용 탐지 알람은 시그네춰 위반, 우회 공격, 최대 사용자 초과, 최대 세션 초과, 최대 페이지 요청 초과의 5가지 알람으로 구분된다. 먼저 시그네춰 위반은 URL 부분에 특정 시그네춰를 삽입하여 공격하는 경우를 탐지하기 위해서 snort에서 제공하는 웹과 관련된 시그네춰를 기반으로 알람이 발생된다. 다음 우회 공격은 관리자가 중요 페이지에 대한 링크 규칙을 설정함으로써 만약 공격자가 해당 경로를 경유하지 않을 경우 알람이 발생된다. 최대 사용자 초과, 최대 세션 초과, 최대 페이지 요청 초과 부분은 DoS 공격과 관련된 부분으로 관리자가 설정한 기준 이상의 요구가 발생했을 경우에 알람이 울리게 된다. 오용 탐지에 대한 알람 모드는 이상 탐지의 경우 보다 위험도가 높기 때문에 구별을 위해서 붉은색으로 경고 메시지와 링크를 표현하였다.[그림 15] 이렇게 색깔을 구별한 이유는 이상 탐지 기법이 상대적으로 거짓 경보(false alarm) 비율이 높기 때문에 둘 사이의 구별이 필요하기 때문이다.

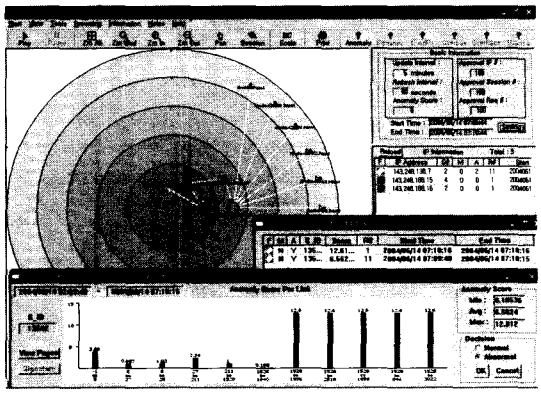


그림 16. 이상 탐지 시각화

5.3 이상 탐지 시각화

사용자들이 주로 방문하는 경로가 아닌, 다른 경로를 이용하는 세션에 대해서 알람을 울리게 된다. 만약 어떤 사용자가 이런 경로를 이용한다는 것은 웹 사이트를 공격할 가능성이 높기 때문에 관리자는 주의를 해야 할 것이다. [그림 16]을 보면 임의의 사용자가 보통의 경로로 들어와서 자주 방문하지 않는 경로로 이동한 것을 알 수 있다. 해당 경로로의 이동 자체가 공격은 아니지만 공격이 될 수 있는 이상 행위기 때문에 우측 상단의 이상 탐지 알람이 울리게 된 것이다.

5.4 오용 탐지 시각화

5.4.1 특정 시그네춰 위반

오용 탐지 기법의 장점은 특정 시그네춰가 있기 때문에 거짓 경보(false alarm)의 비율이 낮다는

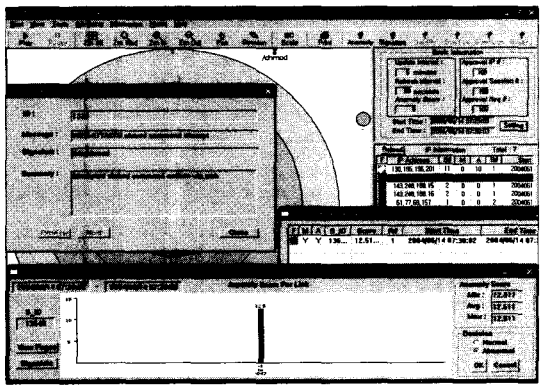


그림 17. 시그네춰 위반

것이다. 상대적으로 이상 탐지 기법은 상대적으로 거짓 경보 비율이 높다. 따라서 SAD Viewer에서는 SAD의 이상 탐지 결과뿐만 아니라 자체 시그네춰 테이블을 통해 오용 탐지 기법도 동시에 적용할 수 있도록 하였다. 해당 시그네춰는 Snort에 있는 시그네춰들 중에서도 웹과 관련된 부분만 가져온 것이다. [그림 17]을 보면 임의의 사용자가 웹을 통해서 /bin/chmod 명령을 실행하려 했다는 것을 알 수 있다. SAD Viewer를 통해 실시간으로 이러한 공격을 탐지함으로써 신속한 대응을 할 수 있다.

5.4.1 DoS (Denial of Service) 공격

DoS 공격이란 공격자가 시스템의 하드웨어나 소프트웨어 등을 무력하게 만들어, 시스템이 정상적인 수행을 하는데 문제를 일으키는 모든 행위들을 일컫는다. 얼마 전 뉴스에도 크게 보도 되었던 Yahoo를 비롯해서 CNN, NBC 등의 거대 인터넷 홈페이지의 공격도 바로 이 DoS 공격이었다. DoS 공격은 특별한 해킹 지식이 없이 수행할 수 있기 때문에 많이 사용되고 있는 공격 중 하나이다. 본 연구에서는 실제 특정 웹 페이지만을 과도하게 요청하는 DoS 공격을 수행하여 SAD Viewer가 어떻게 시각적으로 보여주는지 실험을 하였다. DoS 공격에 대한 세 가지 알람 모드가 있다. 먼저 처음은 기준 시간 내에 사용자의 수가 기준 사용자 수를 초과하는 경우이다. IP를 변조하면서 DoS 공격을 하는 경우가 많기 때문에 기준 시간 내의 사용자 수를 통해서 DoS 공격을 탐지할 수 있다. 두 번째 모드는 최대 세션의 개수를 기준으로 하는 알람 모드이다. 대부분의DoS 공격이 이에 해당되는데 공격자가 변조(spoofing)한 IP를 통해 임의의 페이지들을 과도하게 요청하는 경우 세션이 과도하게 증가하여 웹 서버를 마비시킬 수가 있게 된다. 따라서 최대 허용 세션을 설정해 놓으면 기준을 넘는 사용자에 대해서는 알람을 발생하여 관리자가 신속히 대응할 수 있게 된다. 마지막 세 번째는 한 세션 내의 최대 요청 개수를 설정하는 것이다. 프로파일을 통해서 사용자들의 평균 세션 길이에 대한 정보를 얻을 수 있을 것이다. 이 정보를 통해서 평균을 따르지 않고 과도하게 많은 경로를 이동하는 경우에 대해서 관리자는 관심을 가질 필요가 있다. [그림 18]은 두 번째 모드인 과도하게 많은 세션을 요구한 경우이다. 기준 시간 내의 최대 허용 세션을 30으로 설정하였는데 특정 사용자가 기준 세션을 초과하여 세션을 만든 경우이다.

5.2 코드 레드 공격 시각화

코드 레드 공격[13]을 잘 탐지해 내고 효율적으로 보여주는지 확인하기 위해서 코드 레드 데이터를 실험 당일의 로그에 삽입하여 실험을 하였다. [그림 19]를 보게 되면 동심원의 바깥 부분에 하나의 노드가 상대적으로 나머지 노드보다 큰 것을 확인할 수 있다. 보통 잘못된 페이지에 대한 요구는 빈도수(frequency)가 한번인데 반해, 이 노드는 크기로 봐서 여러 IP가 웹 서버에 존재하지 않는 특정 페이지를 과도하게 요청한 것임을 확인할 수 있다. 또한 해당 페이지의 URL 값에 많은 특정 문자(N)가 나타남으로써 코드레드 공격을 시도했음을 나타낸다. 이상 행위 관점에서 보면 과거 훈련 데이터 내에 코드 레드 공격이 존재하지 않았기 때문에 이상 점수가 최대가 된 것을 확인 할 수 있다.

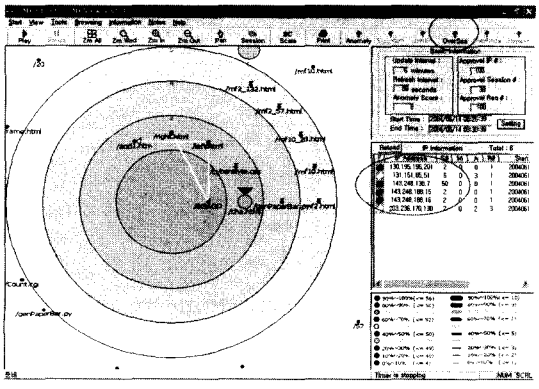


그림 18. DoS 공격 시각화

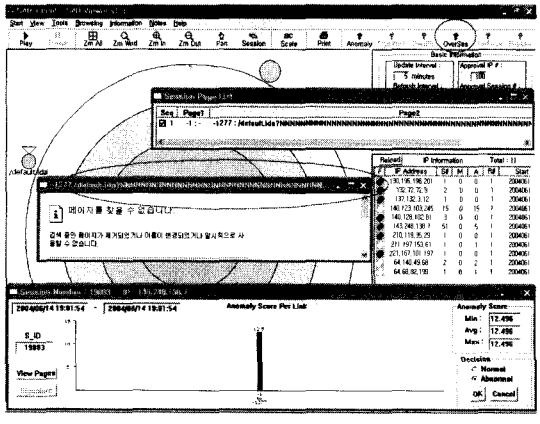


그림 19. 코드레드 공격

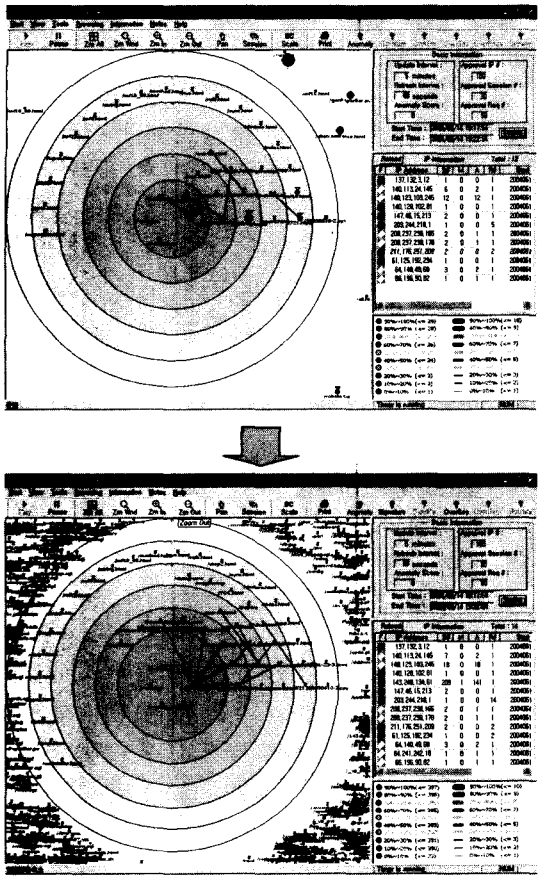


그림 20. Whisker Scans 시각화

5.3 Whisker Scans 시각화

Whisker는 다양한 IDS 우회 기법을 활용하여 취약 CGI 스캐닝을 하는 네트워크 도구이다. [그림 20]은 정상적인 경우와 whisker 공격이 있는 시간대의 세션들을 나타낸 것이다. [그림 20]을 보면 whisker 공격이 있었을 때 동심원 외부에 노드들이 많이 증가되고 커진 것을 확인할 수 있다. 이것은 정상 사용과 비교하여 짧은 시간 동안에 존재하지 않는 페이지 요청이 많았다는 것을 의미한다. 또한 [그림 12]를 보면 그래프가 갑자기 상승한 것으로 보아 특정 시간대에 과도하게 많은 페이지 요청이 있었음을 확인할 수 있다.

Ⅳ. 결 론

정보시각화는 이용자가 원하는 정보에 쉽고 적은 노력으로 접근할 수 있도록 데이터를 재정리하고 요약하여 보여주는 기법이다.[14] 현재 인터넷의 발전과 더불어 웹상에서 제공해야 하는 정보의 양이 증가하고 복잡해지고 있는 게 사실이다. 따라서 웹 환경에서 정보의 제공 또한 시각화를 통해서 보여줘야 할 것이다. 기존의 텍스트 형태로의 정보 제공은 적은 양의 정보를 제공하는 데에는 적합하지만 정보의 양이 많아지면 복잡성으로 인해 효율성이 떨어지게 된다. 본 연구에서 보여주고자 하는 것은 다음의 세 가지 이다. 첫째는 웹 서버의 로그를 바탕으로 웹 사용 현황을 시각화 하는 것이다. 기존의 연구와 차별화된 그래프 레이아웃을 사용함으로써 좀 더 효율적으로 사용자 행위를 시각화 하였다. 두 번째는 이상 행위에 대한 시각화 이다. 본 연구는 전체 SAD System 개발의 마지막 단계로써 SAD System의 탐지 모듈에서 탐지해낸 이상 세션에 대해서 시각적으로 관리자에게 보여줌으로써 빠르게 대처하는데 도움을 주고자 하였다. 세 번째는 실시간 모니터링 기능이다. 이 기능은 이상 행위를 탐지하기 위해서는 반드시 필요한 기능이다. 비록 SAD System이 웹 서버 로그를 이용하므로 어느 정도의 시간차는 있겠지만 생성되는 로그를 빠르게 분석하여 화면에 보여준다면 사람이 하는 것보다 신속히 이상 행위를 탐지할 수 있을 것이다. 본 연구를 통해 시그네춰 위반 공격, DoS, 코드 레드 공격, Whisker 공격에 대해 시각화할 수 있었으며 텍스트 형태의 보고 보다 효율적임을 확인하였다.

향후 연구 계획은 다음과 같다. 첫 번째는 웹 서버에 존재하지 않는 페이지의 layout 문제이다. 현재 SAD Viewer에서는 웹 서버에 존재하지 않는 페이지 요청의 경우에는 동심원 외부에 랜덤 위치로 그려지도록 하였는데 이는 효율적이지 않다. 따라서 존재하지 않는 페이지를 어떻게 시각화할지에 대한 연구가 필요하다. 두 번째는 동적 페이지(dynamic page)들에 대한 처리이다. 현재의 SAD Viewer에서는 동적 페이지들을 하나의 노드로 통합하였는데 각 페이지들이 각각의 특성이 있기 때문에 Bowo가 제시한 Naviz 툴[7]처럼 동적 페이지들에 대한 시각화 문제에 대한 처리가 필요하다. 마지막으로 실제 환경에의 적용이다. 현재는 대학교의 한 실험실 내의 웹 서버만을 대상으로 하였는데 실제 많은 웹 페이지들을 갖고 있고 많은 사람들이 사용하는 웹 서버에 대해서 SAD Viewer가 얼마나 효율적인지 실험할 필요성이 있다.

참 고 문 헌

[1] Srivastava J, Cooley R, Deshpande M, Tan P N. "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data." *ACM SIGKDD*, Jan 2000.

[2] J.S. Seo, H.S. Kim, S.H. Cho and S.D. Cha. "Web Server Attack Categorization based on Root Causes and Their Locations." *Internation Conference on Information Technology*, April, 2004.

[3] James E. Pitkow & Krishna A.Bharat. "WEBVIZ : A Tool For World-Wide Web Access Log Analysis." *First International World-wide Web Conference*, 1994.

[4] Myra Spiliopoulou and Lukas C. Faulstich. "WUM : A Tool for Web Utilization Analysis." *EDBT Workshop WebDB'98*, Valencia, Spain, Mar. 1998.

[5] http://www.itl.nist.gov/iaui/vvrg/cugini/webmet/visvip/webvis-paper.html.

[6] Jason I. Hong, James A. Landay. "WebQuilt: a framework for capturing and visualizing the web experience." *International World Wide Web Conferences*, pp 717-724, 2001.

[7] Bowo Prasetyo, Iko Pramudiono, Katsumi Takahashi, Masaru Kitsuregawa. "Naviz : Website Navigational Behavior Visualizer." *Pacific-Asia Conference on Knowledge Discovery and Data Mining* pp 276-289, 2002.

[8] http://www.silicondefense.com/software/snortsnarf/

[9] Sanghyun Cho and SungDeok Cha. "SAD : Web Session Anomaly detection based on parameter estimation." *Computers & Security Journal*, Elsevier, 2004.

[10] Alberto O. Mendelzon. "Visualizing the World Wide Web." *Proc. AVI'96*, May 1996.

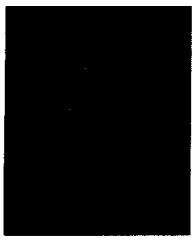
[11] Sougata Mukherjea, James D. Foley. "Visualizing the World-Wide Web with the Navigational View Builder." *Computer Networks and ISDN Systems 27(6)*, pp 1075-1087, 1995.

[12] http://www.research.att.com/sw/tools/graphviz/

[13] David Moore, Colleen Shannon, Jeffery Brown. "Code-Red: a case study on the spread and victims of an Internet worm." *Internet Measurement Workshop*, 2002.

[14] Jee Yeon Lee. An Analysis of Information Visualization Problems using User Interface Design Principles. *정보관리연구*, vol. 34, no. 2. 2003.

〈著者紹介〉



이 병 희 (Byung-Hee Lee) 정회원
2002년 2월 : 동국대학교 컴퓨터공학과 학사 졸업
2002년 8월 : 한국과학기술원 전산학과 석사 졸업
2002년 8월~현재 : 삼성전자 정보통신 총괄 근무
〈관심분야〉 정보보호, 네트워크 보안, 침입 탐지

조 상 현 (Sang-Hyun Cho) 정회원
1997년 2월 : 고려대학교 이과대학 컴퓨터학과 학사 졸업
1999년 2월 : 한국과학기술원 전산학과 석사 졸업
1999년 3월~현재 : 한국과학기술원 전산학과 박사 과정
〈관심분야〉 정보보호, 네트워크 보안, 침입 탐지

차 성 덕 (Sung-Deok Cha) 정회원
1983년 : UC Irvine 전산학과 학사 졸업
1986년 : UC Irvine 전산학과 석사 졸업
1991년 : UC Irvine 전산학과 박사 졸업
1994년~2001년 : 한국과학기술원 조교수
2001년~현재 : 한국과학기술원 부교수
〈관심분야〉 정형기법 및 명세, 정보보호, 네트워크 보안, 침입 탐지