비정상 웹 세션 탐지 및 지역 기반 시각화

김상록[O] 이준섭 서정석 차성덕
한국과학기술원
{srkim[O], jslee, jsseo, cha}@dependable.kaist.ac.kr

Detection and Location-based Visualization of Anomalous Web Sessions

Sangrok Kim[O] , Junsup Lee, Jeongseok Seo, Sungdeok Cha
Div. of Computer Science Dept. of EECS, KAIST and AITrc/DSRC/IIRTRC/SPIC

요 약

한 해에도 수많은 해킹 사고가 발생하고 있고, 이 중에서 웹 해킹이 차지하는 비율은 급격하게 증가하고 있다. 또한 최근의 해킹 동향을 분석해 보았을 때 웹 해킹의 비율은 더욱 증가할 것이라고 예상된다. HTTP 프로토콜을 이용한 공격의 특성 상 정상행위와 비정상 행위의 구분이 어렵다. 따라서 웹 서비스에 특화된 침입탐지 시스템이 요구된다. 또한 웹 사이트 관리자는 빠른 탐지와 대응을 위해 이상 행위에 대한 신속하고 정확한 인식을 필요로 한다. 본 논문에서는 이러한 필요성을 기반으로 Location-based Visualization Tool을 제안한다. 웹 사용 현황 및 이상행위에 대해 시각적인 정보를 제공하기 위해 웹 서버의 access log를 분석하여 이상 행위를 탐지하였고, IP정보를 기반으로 지역 정보의 시각화를 구현하였다.

1. 서론

미국의 Gartner Group에 의하면 전체 해킹 공격의 75% 이상이 웹 어플리케이션이라는 결과가 발표되었다. 이 수치는 해킹 공격의 지난 통계에 비춰보았을 때 앞으로의 동향을 예측할 수 있는 실마리를 제공한다. 그동안 보안에 대한 이슈가 부각되면서 대부분 회사의 시스템은 방화벽으로 외부와 격리시키고 네트워크는 침입탐지 시스템(IDS)을 통해 감시하는 등 많은 노력과 투자를 아끼지 않고 있다. 그러나 앞의 통계에서 알 수 있듯이 원격 접속 서비스들(Telnet, FTP 등)은 외부와 격리되거나 감시를 하고 있지만 웹 서비스(port 80)는 외부에 개방되어 있다. 공격자들은 보안이 고려되지 않은 HTTP 프로토콜을 통해서 기존의 침입 탐지 시스템을 우회하여 공격하고 있다.

이에 따라 웹 서비스를 좀 더 안전하게 제공하고 웹 서비스를 이용한 공격을 방어하는 연구의 필요성이 커지고 있다. 더욱이 최근에는 웹 공격에 대해 보다 빠르게 반응할 수 있고, false alarm을 줄일 수 있도록 웹 서비스에 대해 특화된 침입탐지 시스템의 필요성이 급증하고 있다.

웹에 대한 위협 및 공격행위를 대처하기 위해 해결해야 될 중요한 이슈는 어떻게 이상행위를 탐지를 것인가와 탐지된 정보를 어떻게 보고할 것인가의 2가지 큰 이슈가 있다. 이상행위에 대한 탐지는 이미 많은 연구가 이루어지고 있으나 웹의 특성 상 수많은 정보들 속에서 관리자가 원하는 정보를 추출하는 것은 쉽지 않다. 이것을 어떻게 효과적으로 추출하고 관리자에게 보고할 수 있는가는 중요한 문제이다.

기존의 연구들은 통계적 자료나 패턴을 기반으로

하여 시각화 문제에 접근하였으나, 대부분 텍스트 기반으로 제공되었기 때문에 직관적인 이해가 쉽지 않았고, 신속한 대응을 어렵게 하는 원인이었다. 또한 대부분 시그너처 기반이기 때문에 방대하고 복잡하며 급변하는 현재 웹 환경에서는 적합하지 않다[1]. 본 연구에서는 세션을 이용한 베이지언 추정을 기반으로 하여 이상행위를 탐지하고 관리자에게 시각화된 정보를 제공하는 것을 목표로 한다.

본 논문의 구성은 다음과 같다. 본론의 1장 관련연구에서는 시각화를 위한 기존의 여러 기법 등에 대해 소개한다. 2장은 SAD server및 viewer에 대해 설명하고 location을 기반으로 시각정보를 제공하기 위한 시각화 도구에 대한 설명은 3장에서 할 것이다. 그리고 마지막으로 결론 및 향후 연구방향에 대해 소개한다.

2. 본론

2.1 관련연구

이 장에서는 웹 사용 현황의 시각화를 위한 기존의 연구와 웹 공격의 보고기법 연구들에 대해 알아본다.

2.1.1 Tivoli 웹 침입탐지 시스템 - IBM[2]

Tivoli 웹 침입탐지시스템은 웹 로그(access log) 파일을 기반으로 웹 공격을 탐지하는 침입탐지시스템이다. 침입탐지는 지식 기반(knowledge-based)의 오용탐지 기법을 사용한다. 실시간 모드와 배치 모드에서 운영이 가능하다. 실시간 모드에서는 웹 로그 엔트리가 새로 생성될 때마다 IDS는 새로운 이벤트를 침입탐지의 데이터 소스로 사용하는 방식이다. 반면에 배치 모드는 웹 IDS가 웹 서버가 설치된 서버에 같이 설치되지 못하였거나 실시간으로 웹 공격을 탐지할 필요가 없을 때 사용하는 방식으로 웹 로그 파일을 웹 IDS의 입력으로 주면 해당 로그 파일에서 웹 공격을 탐지하게 된다.

2.1.2 웹시큐어 - 아이자이어 로보텍스[3]

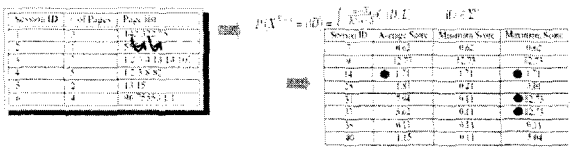
웹시큐어 웹 침입탐지시스템은 네트워크 기반의 웹 패킷을 침입탐지의 소스 데이터로 사용하여 웹 공격을 탐지한다. 또한 웹 공격에 대한 정형 패턴을 가지고 오용탐지 기능을 수행하며 일부는 웹 응답을 기반으로 이상탐지를 수행하고 있다. 기존의 방화벽이나 DS, IPS 등에 적용되었던 Stateful inspection기법이나 Deep Packet Inspection기법이 아닌 Streaming Inspection 기법을 이용하여 웹 어플리케이션을 지원하고 있다.

그러나 앞서 설명한 웹 어플리케이션 침입탐지시스템은 국내외적으로 이제 연구가 시작되고 있는 초기단계이다. 그렇기 때문에 아직 오탐율이 높고 모든 웹 공격을 탐지하지 못하는 단점을 가지고 있다. 또한 웹 공격들은 컨텐츠에 따라 다양한 변화를 보이고 있는 반면 초기의 침입탐지 시스템은 대부분 시그니처 기반의 패턴 매칭 기법을 이용하고 있기 때문에 공격의 다형성을 따라가지 못하고 있다

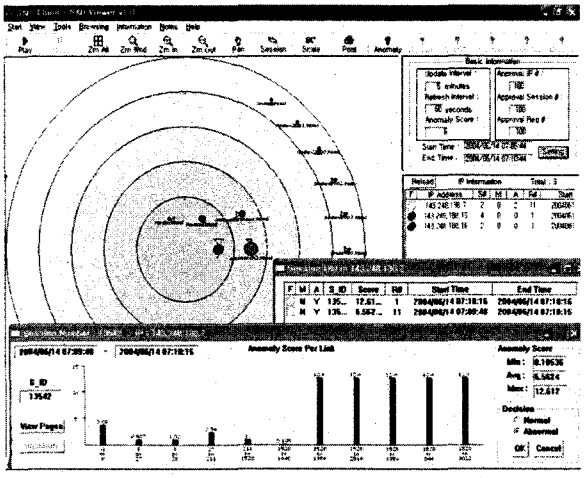
2.2 SAD(Session Anomaly Detection)[4]

본 논문의 목적은 사용자의 웹 사용 현황을 효과적으로 시각화하고 웹 공격과 같은 이상 행위를 실시간으로 탐지하여 관리자에게 빠르게 보고할 수 있는 tool의 제공에 있다. 이 장에서는 KAIST에서 제안한 SAD server및 view에 대해 설명하고 다음 장에서는 탐지된 결과를 효율적으로 시각화하기 위해 본 논문에서 제안한 Location Based Tool에 대해 설명하도록 하겠다.

SAD 시스템은 웹 로그 기반의 침입탐지시스템이다. Snort 룰 형식의 패턴을 사용하여 오용탐지를 수행하며, 웹 세션을 추출하고 이를 통해 사용자의 웹 행위를 모델링하고 비정상 행위를 탐지하는 이상탐지를 수행한다. 이상탐지를 위해 베이지언 추정 기법[그림 1][5]을 이용하였고, 추정을 통해 anomaly score를 계산하여 비정상 여부를 판단한다. 그리고 효율적인 시각화를 위해 계층적 동심원 구조를 이용하여 웹 페이지를 표현하였다. 또한 바인딩 기법을 적용하여 표현되는 웹 페이지 및 링크의 크기, 색깔 등의 변화를 통해 관리자에게 시각화된 정보를 제공한다.



[그림1] 베이지언 추정 기법



[그림 2] SAD system

제안된 SAD 시스템은 기본적으로 웹 서버의 로그에 기반한다. 일반적인 웹 서버의 경우 하루에 생성되는 로그가 Giga Byte이상이 되며 최악의 경우 수백 Giga byte의 로그가 생성된다. 따라서 이러한 대용량의 로그를 신속히 처리하여 관리자에게 실시간으로 제공할 수 있는 성능이 요구되었다. Access log를 SAD에 맞는 소스로 가공하기 위해 SAD server라는 preprocessor를 구현하였고 파일시스템 기반에서 DB 및 stored procedure 기반으로의 수정을 통해 성능 개선을 이루어 내었다.

2.3 Location Based Visualization Tool

Location based Visualization Tool은 현재 웹 서버에 접근하고 있는 모든 웹 사용자의 상황을 사용자의 IP에 근거하여 위치를 기반으로 시각화하는 도구이다. 표, 그래프, 차트를 이용하여 시각화하려는 다른 도구들 보다 사용자의 전체 현황을 효율적으로 파악할 수 있다는 장점이 있다. 그 이유는 사람이 가장 많은 정보를 단 시간 내에 얻을 수 있는 감각인 시각에 초점을 두고, 그 효율성을 극대화하고자 그래프와 표가 가지는 추상성을 배제 하고 직관적인 시각적 해석이 가능한 인터페이스와 정보 표현을 제공했기 때문이다.

웹의 특성상 웹 서비스에 접근하는 사용자는 전세계적 혹은 지역적으로 폭넓게 분포되기 마련이다. Microsoft에서 제공하는 Map point의 월드뷰를 통해 웹 사용 현황에 대한 정보를 표현함으로써 현재 서버로 요청되는 페이지들이 어느 지역으로부터의 요청인지 직관적으로 알 수 있다. 또한 웹 서버에서 제공하는 서비스에 따라 정도의 차이가 있겠지만, 높은 빈도를 가진 주 사용자는 주로 특정 지역에서 지속적인 접근을 하게 된다는 점에 착안하여, 지역별 웹 사용현황을 확인할 수 있도록 하였다. 즉 Location based Visualization Tool은 웹 공격자의 위치와 공격 경로를 지도를 통해 즉시 파악하고 효과적인 대책을 세울 수 있도록 관리자에게 시각화 정보를 제공한다. 또한 공격자의 공격 패턴에 따라 그 심각한 정도와 공격에 사용된 방법에 근거하여 색채 정보를 다르게 함으로써 관리자에게 직관적으로 이해할 수 있는 시각정보를 제공한다.

Location based Visualization Tool을 구현하면서 다음과 같은 요구사항이 도출되었다.

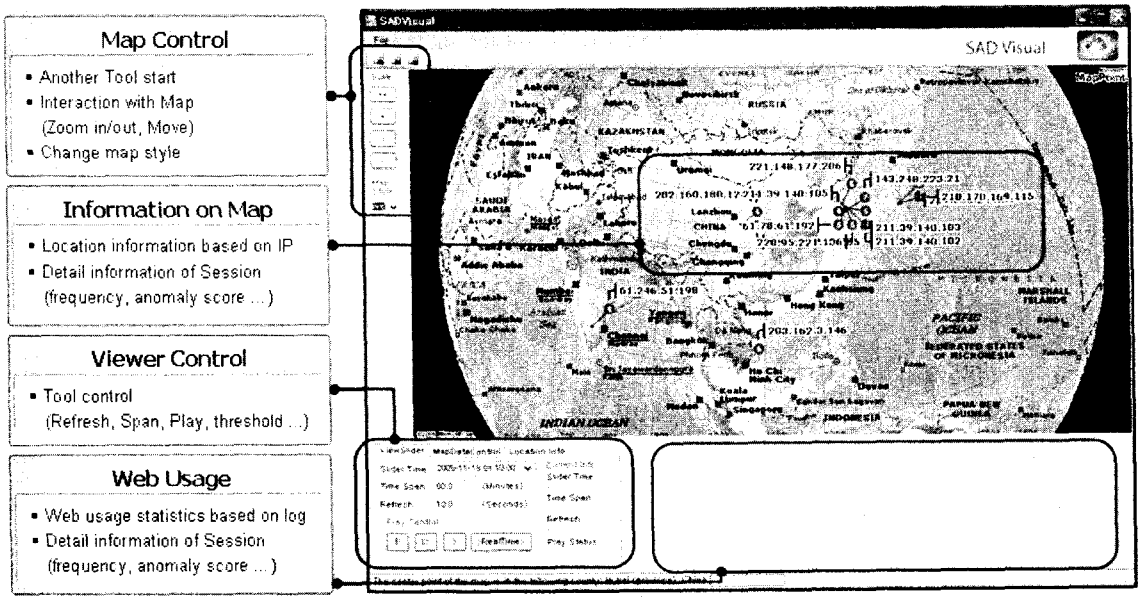
Location 정보의 신뢰성

세션 정보를 Location을 기반으로 표현하기 위해서는 세션의 IP정보를 바탕으로 정확한 지도 정보를 얻을 수 있어야 한다. 이를 위해서 Microsoft MapPoint Web Services[6]를 선택하였는데, 이유는 MapPoint가 전세계에 걸쳐 신뢰할 만한 GIS를 구축해 놓았기 때문이고, 또한 단순한 지리적 정보뿐 만이 아니라 MapPoint service내 POI(Point of Interest) 서비스를 폭넓게 지원하여 방대한 데이터를 가지고 있기 때문이다. 그리고 이러한 Map정보를 Web Services를 통해 실시간으로 이용할 수 있고 Location 정보를 위해 Map Database를 생성하고 유지하는 등의 부담을 덜어주는 등의 장점이 있기 때문에 MapPoint 서비스를 이용했다.

지리 데이터 위에 표시해줄 IP 위치 정보는 MaxMind GeoIP City[7]라는 데이터베이스를 이용하였다. 이 서비스는 주어진 IP 정보에 대하여 그 위치 정보를 위도, 경도 값으로 반환 해주는 서비스로 ISP정보 등 기타 IP와 관련된 정보를 제공하고 있기 때문에 다양한 방법으로 Web Security에 적용이 가능하다.

세션에 Location 개념 적용

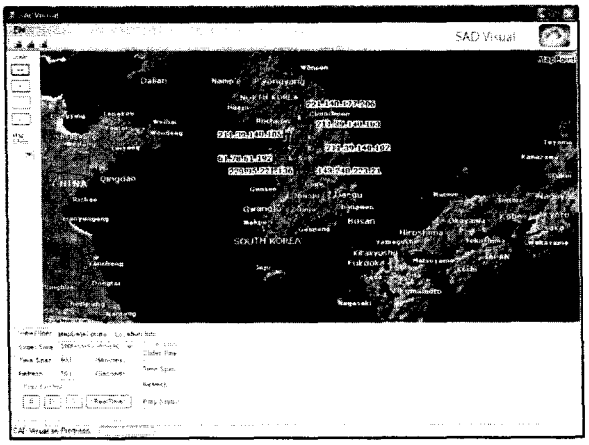
세션에 Location 개념을 적용하는 요구사항은 세션의 IP정보를 통해 구현하였다. SAD server에서 분석된 세션 정보는 IP정보를 포함하고 있다. 위의 요구사항에



[그림 3] Location Based Visualization Tool

서 구현된 바와 같이 MaxMind의 GeoIP 서비스를 통해 해당 세션의 IP의 위도, 경도 정보를 얻어오고, 이 정보를 바탕으로 지도 위에 세션을 표현하게 된다.[그림 3,4]

지도 위 원하는 위치에 세션을 표현하는 방법은 Microsoft MapPoint에서 제공하는 Pushpin 서비스를 이용한다. 위도, 경도 정보를 가지고 Pushpin 서비스에 요청하게 되면 Map상에 원하는 정보를 원하는 위치에 배치할 수 있다.



[그림 4] Location Based Visualization Tool

3. 결론

정보의 시각화는 이용자가 원하는 정보에 쉽고 적은 노력으로 접근할 수 있도록 데이터를 재정리하고 요약하여 보여주는 기법이다[8]. 지금처럼 방대한 웹 환경에서는 관리자가 관리해야 될 정보도 많고 복잡하다. 뿐만 아니라 인터넷이 발전할수록 관리해야 될 정보 또한 더욱 많아지고 복잡해질 것이다. 따라서 관리자의 부담을 줄이고 효율적은 정보의 인지와 사용을 위해 텍스트 형태의 정보제공 보다는 시각적인 형태로 정보를 제공해야 된다.

본 연구가 제공한 성과는 다음 2가지이다. 첫째, 웹 서버의 로그를 바탕으로 웹 사용현황 및 이상행위에 대한 효과적인 시각화 문제를 개선하였다. 기존에 많이 사용되었던 그래프나 트리 형태가 아닌 SAD viewer를 이용하여 다른 형태의 시각화 정보를 제공하였다. 또한 세션의 시각화 개념에 지역정보라는 개념을 도입하여 Location based Visualization Tool을 구현하였다. 두 번째는 실시간 모니터링 기능이다. 웹 서버에서 생성된 로그정보를 실시간으로 처리하여 시각 정보로 표현함으로써 관리자로 하여금 빠른 인식과 탐지, 대응이 가능하게 하도록 한다. 이를 위해 SAD server라는 pre processor를 구현하였다.

향후 연구 계획은 다음과 같다. 먼저 SAD viewer에서 발견된 새로운 요구사항을 적용하는 것이다. 웹 서버에 존재하지 않는 페이지의 경우 동심원 밖에 복잡한 형태로 표현되는 문제와 같이 추가적으로 발견되는 요구사항을 해결하기 위한 연구가 필요하다. 두 번째로 새로운 웹 공격을 탐지하기 위한 알고리즘의 개발이다. 현재의 시그너처 방식이 갖고 있는 문제점을 개선하고, 다른 웹 공격을 탐지하기 위한 연구가 우선적으로 이루어져야 한다. 마지막으로 새로운 시각화의 방법에 대한 연구가 필요하다. 마지막으로 본 논문에서 제안한 Tool에 웹 사용현황을 시각화하여 보여주는 기능을 추가한 뒤 SAD View와의 통합이 요구된다. 그리고 다른 시각화 방법에 대한 새로운 연구가 필요하다.

Microsoft에서 실제 발생하는 Log를 바탕으로 소규모의 웹 서버가 아닌 실제 많이 사용되고 있는 웹 서버에서 우리가 제안한 방법과 도구들이 얼마나 효율적인지를 사례연구를 통해 알아볼 예정이다.

4. 참고 문헌

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

[2] http://www-306.ibm.com/software/tivoli/

[3] ttp://www.websec.co.kr/product/pro_pg01.htm

[4] S.H.Cho, S.D.Cha, "SAD:Web Session Anomaly detection based on parameter estimation", Computer & Security, Vol. 36, no.3, pp89-100, 2004.8

[5] Friedman N, Singer Y. Efficient Bayesian parameter estimation in large discrete domains. Advances in neural information processing systems 11. Cambridge, Mass: MIT Press; 1999.

[6] http://www.msdn.com Microsoft MapPoint Web Services 2006.

[7] MaxMind GeoIP® City Database http://www.maxmind.com/app/city.

[8] Jee Yeon Lee, "An analysis of Information Visualization Problems using User Interface Design Principles", 정보관리 연구, Vol. 34, no. 2, 2003.