지능형 SQL Query 분석을 통한 Application Layer 역추적 연구

백종일* · 박대우*

*호서대학교 벤처전문대학원 IT응용기술학과

A Study of Application Layer Traceback Through Intelligent SQL Query Analysis

Jong-Il Baek* · Dea-Woo Park*

*Dept. of IT Application Technology, Hoseo Graduate School of Venture

E-mail : *jibaig101@empal.com · *prof1@paran.com

요 약

현재의 IP 위주의 역추적은 Proxy와 우회기법의 발달로 인하여 Real IP 역추적에 어려움이 있다. 또한 IP역추적 후에도 실제 Source IP인지 확인이 어렵다. 따라서 본 논문에서는 지능형 SQL Query에 대한 field, column, table 등의 요소값과 매칭되는 key값을 분석하고 여기에서 사용되는 Data값의 hit point를 분석하여 최초 사용자에 대한 Application Layer를 분석함으로써 IP 역추적에 대한 포렌식 증거로 삼는다. 본 연구는 포렌식과 DB보안 등 전자거래 발전에 기여할 것이다.

ABSTRACT

Current Traceback is difficult due to the development of bypass technique Proxy and IP-driven to trace the real IP Source IP is the IP traceback after the actual verification is difficult. In this paper, an intelligent about SQL Query field, column, table elements such as analysis of the value and the matching key values and Data used here to analyze source user hit point values for the user to trace the Application Layer IP for the analysis of forensic evidence guided by In this study, including forensic DB security will contribute to the development of electronic trading.

키워드

IP Traceback, DB Security, SQL(Structured Query Language) Query, Application Layer, Forensic

Ⅰ. 서 론

그림 1은 지난 2006년부터 2010년까지 웜 바이러스, 해킹신고처리, 악성 봇의 관련한 침해사고 통계 자료이다. 인터넷 침해사고는 2010년에도 월 1,000건 이상씩 발생하여 신고 되고 있다[1].

구분	2006년 총계	2007년 총계	2008년 총계	2009년 총계	2010년			2010년 총계
					1월	2월	3월	
웜/바이러스	7,789	5,996	8,469	10,395	932	1,302	1,085	3,319
해킹신고처리	26,808	21,732	15,940	21,230	898	1,076	1,053	3,027
- 스캔클레이	14,055	11,668	6,490	10,148	154	317	222	693
- 피싱경유지	1,266	1,095	1,163	988	78	106	116	300
- 단순침입시도	3,711	4,316	3,175	2,743	232	230	345	807
- 기타해킹	4,570	2,360	2,908	3,031	223	233	267	723
- 홈페이지 변조	3,206	2,293	2,204	4,320	211	190	103	504
악성 봇(Bot)	12.50%	11.30%	8.10%	1.00%	0.60%	0.60%	0.70%	0.60%

그림 1. 인터넷 침해사고 통계

각종 해킹 및 침해사고의 위협에 있어 근본적인 문제점은 바로 사고 근원지에 대한 정확하고 빠른 색출과 법적인 증거자료의 분석에 있다.

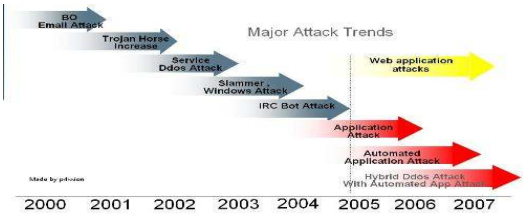

그림 2. 침해사고 공격의 추세 분석

그림 2는 침해사고를 발생시키는 몇 년간의 공격의 추세를 보여주고 있다. 특히 최근 몇 년 간

은 Application Layer를 겨냥한 공격이 추세적으로 보여주고 있다.

이 결과는 인터넷 사용이 높아지면서 트래픽의 절대 다수를 차지하는 HTTP기반 프로토콜이 증가와 비례해서, 사회 및 기업의 업무 시스템 전반들이 대부분 HTTP 프로토콜 기반의 "웹 어플리케이션(Web Application)"의 형태로 구성되어 운용되고 있다.

따라서 본 논문에서는 Application Layer의 공격을 분석하고, Application Layer에서 암호화 되지 않고 전송되는 SQL Query에 대한 field, column, table 등의 요소 값과 매칭 되는 key값을 지능적으로 분석한다.

SQL Query 결과의 분석에 사용되는 Data값의 hit point를 분석하여 최초 공격자나 사용자에 대한 확증적인 증거자료를 획득하여 Application Layer를 분석 대상으로 하고, Real IP 역추적을 실시하여 법적인 책임 소재의 증거가 되는 포렌식 증거로 보고서로 작성하게 한다.

본 연구는 기존 TCP/IP의 한계성을 초월하는 IP역추적을 기반으로 한 "사용자 식별" 기법 중에 하나이며, 연구 결과는 Real IP역추적을 실시하는 수사기관과 법정에서 증거자료의 책임을 판단하는 사법기관기관의 포렌식 기술발전에 기여할 것이다[2].

본 논문은 Ⅰ장, 서론에서 침해사례 통계 및 본 논문의 필요성을 설명하고, Ⅱ장, 관련연구에서는 역추적 기법을 소개한다. Ⅲ장에서는 지능형 SQL Query 분석을 실시하고, Ⅳ장에서는 지능형 SQL Query 분석을 통한 Application Layer의 역추적 방안을 연구하고, 기존 방식과 비교한다. 마지막으로 Ⅴ장에서 결론과 향후 연구를 설명한다.

Ⅱ. 관련연구

기존의 역추적 방식들의 문제점은 TCP/IP의 구조적 한계로 인해, 실제 사용자의 추적이 쉽지 않다는 기술적인 한계가 있다.

2.1. SQL Query

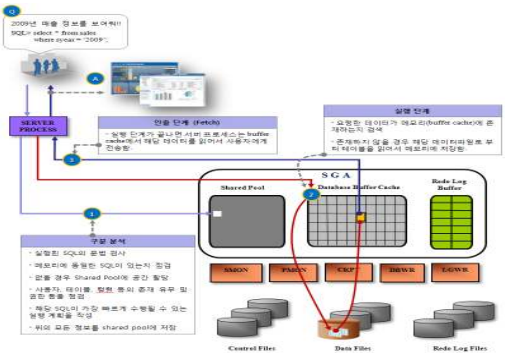

그림 3. SQL SELECT 구문 처리 단계

SQL은 구조화 질의 언어로써 관계형 데이터베이스에 접근하여 데이터를 조작할 수 있는 언어이다. RDBMS(Relational database management system) 표준 언어이기 때문에 조금의 차이는 있어도 DBMS 종류에 상관없이 RDBMS인 경우에는 대부분 공통적으로 사용가능한 언어이다. SQL은 데이터 정의어(Data Definition Language), 데이터 조작어(Data Manipulation Language), 데이터 제어어(Data Control Language) 및 데이터 질의어(Data Query Language) 등으로 구분된다.

2.2. Host 기반의 사용자 역추적

■ CIS

CIS(Caller Identification System)는 사용자가 특정 시스템에 접속하고자 할 때, 해당 시스템은 한 시스템에 접속하기 위해서 자신이 경유한 모든 시스템에 대한 목록을 제공해야 하는 것이다.

■ AIAA[4]

AIAA(Autonomous Intrusion Analysis Agent) 시스템은 침해를 당한 서버의 해킹 피해 분석과 추적을 위한 로그 분석을 에이전트를 이용해 자동화한 역추적 시스템이다.

■ 패킷손실기반의 논리적 전송경로 추정

패킷손실기반의 논리적 전송경로 추정 방법은 다수의 호스트가 임의의 IP주소로 목적지 호스트를 공격하는 분산서비스공격의 경우에 패킷의 전송경로를 역추적하기 위한 방법이다.

■ Log

로그 분석은 TCP/IP 구조상 직전 단계 IP밖에 수집할 수밖에 없어, 위장 및 우회접속 시 효과를 상실한다. 변경 및 조작이 용이하여 신뢰성이 점차 낮아져 사고발생시 입증 책임에 어려우며 위협 및 패턴분석이 별도로 필요하다. 또한 실시간 분석 및 대응이 어렵다는 단점을 가지고 있다. 대표적으로 유닉스 계열 시스템의 시스로그와 IIS나 아파치의 웹 로그가 대표적이라 할 것이다.

2.3. Network 기반의 사용자 역추적

■ TCP 시퀀스 넘버의 증가정도를 통한 알고리즘[5]

TCP 시퀀스 넘버를 이용하는 알고리즘은 비록 송수신되는 데이터가 암호화 되더라도 데이터의 양은 크게 변하지 않는다는 점에 착안하여 Sequence Number의 증가 정도를 변동 폭의 조정을 통해 비교하고 연결 체인을 구성하는 알고리즘이다. 즉, TCP Sequence Number 를 이용하는 시스템은 비록 송수신되는 데이터가 서로 상이하더라도 데이터의 양은 크게 변동되지 않는 다는 점을 착안하여 Sequence Number 의 증가 정도를 변동 폭의 조정을 통해 비교하고 Connection Chain 을 구성하는 시스템이라 할 것이다.

■ TCATS

TCATS(TCP Connection based Attack Trace-Back System)는 일종의 SWT(Sleepy Watermark

Tracing) 방식인 사용자에 대한 응답 패킷에 워터마크를 삽입하여 역추적을 수행하는 방식의 연장선상에서 발전시킨 개념이라 할 것이다.

방식은 크게 3가지로 구분되며, 첫째 Agent에 대한 분석, 둘째, Request Manager에 대한 분석 , 셋째, Reply Manager에 대한 분석 이다.

Ⅲ. 지능형 SQL Query 분석

3.1. Application Layer의 서명과 비서명 인증방식

Application Layer기반 사용자 추적 기법은 "브라우저(Browser)의 플러그인(Plug-in)을 이용하는 방법으로, 브라우저가 지원하고 있는 Active-X, 자바 등을 이용하여 추적을 하는 것이다. Application Layer 방식의 사용자 추적 기법은 서명기반과 비서명 기반으로 나누어 질 수 있다.

서명 기반은 사용자의 동의하에 인증 창을 통해 수행되며 대표적으로 Active-X가 있으며, 비서명 기반은 사용자의 동의 없이 이뤄지며 대표적으로 JVM 등이 존재한다. 서명 기반의 방식은 인증을 요구하기 때문에 사용자가 쉽게 밝혀지므로 공격자 입장에서는 은익성이 보장되지 않는 관계로 사용이 적게 된다.

비서명 기반은 은익성이 향상되나 공격자 입장에서는 은익성은 높아지나 사용제한으로 인하여 로컬권한이 존재하지 않기 때문에, 허용된 사이트만 사용하게 된다. 이 결과 하드웨어 정보와 같은 증거자료들을 수집 또는 저장할 수 없다.

HTTP 프로토콜 기반 사용자 식별 및 추적 기법은 네트워크상의 웹서비스 응용 단계에서 고유의 기법인 Browser Plug-in 방식으로 이를 식별하고 추적할 수 있다. 실제업무에서 HTTP 프로토콜을 기반으로 하는 WWW(World Wide Web)이 주요 통신수단으로 활용되어 지고 있으며, 웹 어플리케이션 기반을 구성하는 HTTP 프로토콜을 중심으로 어플리케이션 레이어의 역추적 연구가 실효성을 거둘 수 있다[6].

3.2. 비서명기반의 경우 Real IP 역추적에 대한 문제점

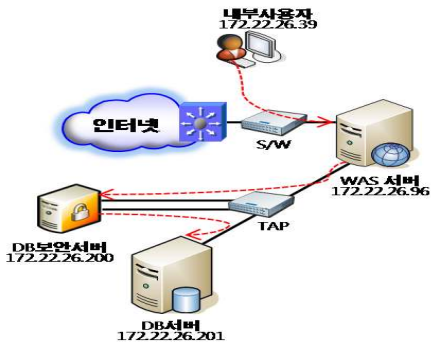

그림 4. 웹 어플리케이션을 통한 접근 경로

그림 4처럼 SQL Query가 DB 보안시스템을 통해 접근 및 권한제어가 구현되어있는 웹 어플리케이션을 통해 DBMS 서버로 접근할 때 172.22.26.39의 사용자가 네트워크를 통해 172.22.26.96의 WAS서버를 경유해 172.22.26.200의 DB보안 Gateway 서버의 정책에 의해 172.22.26.201의 DB서버로 login하게 된다. 이때 DB보안서버는 자체 DB를 통해 로그정보를 저장하게 된다[7].

그림 4에서 공격자가 웹 어플리케이션을 경유해서 DB로 접근을 시도했을 때 DBMS 보안서버에 저장되는 IP, 시간, 포트 등의 패킷 정보는 출발지가 공격자가 아닌 웹 어플리케이션 서버의 정보가 남게 된다.[3] 따라서 이러한 문제점을 해결하기 위해서는 지능형 SQL Query를 통한 분석이 Application Layer에서 이루어져야 한다.

3.3. Traceback 시스템의 Key값 발행과 요건

사용자 식별 키 생성 방안으로서 유일한 키 정책을 자동 배분하고 이를 해당 사용자가 읽고/쓰기가 가능해야 한다. 이를 위해 키 값은 확률적 통계상 중복성이 극히 낮아야 하므로 최소 8자리 이상의 난수발생 알고리즘을 적용한다. 더불어 해당 키 값은 절대적인 기준 값으로 ID와 동일한 개념으로 간주하여 기록/분류한다.

Ⅳ. 지능형 SQL Query 분석을 통한 Application Layer 역추적

4.1. Query와 Key값의 일치도(hit point) 분석

추적 정밀성과 성공률을 향상시키기 위한 방법의 핵심은 키 값에 대한 관리와 통제이다. 이를 위해 키 값을 적용하기 위한 알고리즘으로 사용자가 접속 시, 키 값을 읽고 키 값이 존재하면 이를 타임스탬프와 함께 DB 기록만 하며, 만약 키 값이 미 존재 시 상기 키 생성 정책에 입각하여 새로운 키 값을 사용자에게 배부하고 이를 DB에 기록한다. 아울러 추적된 Real IP 정보와 함께 기록하여 키 값을 기준으로 IP변동을 추적 관리할 수 있게 된다. 또한 상기 서술한 비서명 기반 브라우저 플러그인을 통한 다중 플러그인을 동시에 통제하여 성공률을 높인다.

공격자인지 판단되면 Application Layer에서 공격자를 판단하고 Query를 날린다. 이때 요청사항은 Key값 등이다. 이리하면 Real IP와 인증 값 등으로 공격자의 원래 정보를 파악할 수 있다. 로그 기록과 부인방지를 위해 타임스탬프나 해시값을 사용한다.

4.2. Real IP 역추적

Real IP를 역추적하기 위해서는 위와 같이 생성된 키 값을 통해 변동된 IP 정보를 기록하고, 세션의 Packet Head Key값 분석하여 패킷의 내용을 파악한다. 이때 반드시 Real IP를 가지고 와

야만 키 값을 발행하여 준다. 패킷의 내용까지 파악이 되면 수집된 Real IP 정보가 Source IP 정보가 확실한지 판단하기위해 Three hand shake 기법을 이용하여 인증값과 Real IP Matching 작업을 실시한다. 이러한 인증과 더불어 그림 5와 같이 L7스위치를 통한 Service영역인 Application Layer의 패킷분석으로 Real IP 역추적 시 침입자의 정확한 근원지를 파악할 수 있다.

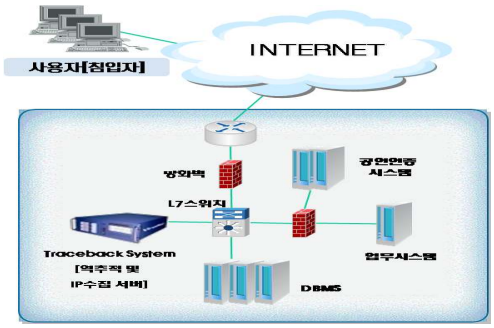

그림 5. 지능형 역추적 시스템 구성도

표 1은 Traceback 시스템의 기존 제품의 단점을 기준으로 개선시스템의 향상된 기능 구현 결과를 요약하였다.

표 1. Traceback 시스템의 개선체계 결과 요약

구분		기존 시스템	개선 시스템
모니터링 정보(프락시 IP 등 접속 위장 후 접속 시)	직전 IP(프락시) 식별	O	O
	공인 IP 식별	X	O
	사설 IP 식별	X	O
	브라우저 및 호스트 정보 식별	X	O
	키 값 생성 및 히스토리 관리	X	O
은폐 및 보안성 확보		X	O
위협분석 및 패턴분석		X	O
실시간 대응 (경고 및 차단 여부)		X	O

4.3. 포렌식자료 생성

지능형 SQL Query 분석을 통한 Application Layer 역추적을 통해 공격자로 판단되면 Query를 날려 Key값 등을 요청한 후, Real IP와 인증값 등으로 공격자의 원래 정보를 파악할 수 있게 되고, 이렇게 수집된 로그 기록을 통해 감사 자료인 포렌식 보고서를 생성한다.

위의 보고서 자료를 통해 사고 분석 및 증거자료로 활용하고, ESM 및 TSM 등 통합보안관제 시스템과 연동을 통해 조기경보, 위험관리, 위협관리를 실시한다. 또한 위치 기반 지리정보와 연동하여 위치정보까지 정확히 수집하게 된다. 연동 방안으로는 “WHOIS” 및 “DHCP 기반 유동 IP에 대한 지역 정보 시스템” 서비스를 기반으로 인터넷상의 지도정보와 API 등을 통해 연계하여, 고정IP 뿐만 아니라 유동IP의 경우에도 위치추적이 가능하게 된다.

Ⅴ. 결 론

본 논문에서 연구의 핵심은 불특정 다수를 대상으로 하는 인터넷 환경에서 사용자를 추적하여 식별하는 진보된 방법을 제시함으로써, 지능적으로 통제하고 추적하기 위함이다. 즉, 추적(프락시 서버 경유) 및 식별(쿠키 차단)이 제한된 상태라 하더라도 도출하고자 하는 사용자 식별 키 값을 포함한 정확한 사용자 측 각종 IP 정보(사설 IP, 공인 IP, 프락시 IP 등) 일체를 획득할 수 있었다.

향후 연구 방향은 다양화되는 각종 플러그인들에 대한 추가적인 적용, 변화되는 네트워크 및 시스템 환경에 따른 적절한 변형작업이 요구될 것이다. 아울러 실 IP를 찾아내고 지능형으로 해커 시스템과 사용자를 실시간으로 검색 추적하는 3D 차원 연구가 필요하다.

참고문헌

[1] 한국정보보호진흥원, “인터넷침해사고 동향 및 분석 월보”, http://www.krcert.or.kr, 2010. 3.

[2] 이인희, “IP역추적 설계 및 보안감사 자료생성에 관한 연구”, 한국컴퓨터정보학회논문지, 제 15권 제 1호, pp.53-64, 2007. 6.

[3] 윤병선, “우회적인 공격에 대한 실제 IP 역추적 실시와 포렌식 자료 생성”, 한국컴퓨터정보학회논문지, 제 13권 제 1호, pp.143-151, 2008. 1.

[4] Chaeho Lim, "Semi-Auto Intruder Retracing Using Autonomous Intrusion Analysis Agent", FIRST Conference on Computer Security Incident Handling & Response 1999, 1999.

[5] K. Yoda and H. Etoh, "Finding a Connection Chain for Tracing Intruders", In F. Guppens, Y. Deswarte, D. Gollamann, and M. Waidner, editors, 6th European Symposium on Research in Computer Security - ESORICS 2000 LNCS -1985, Toulouse, France, Oct 2000.

[6] 김태봉, “역추적 기술의 동향 및 적용 사례 분석”, 정보보호학회지, 제 15권, 제 1호, pp.98-103, 2005. 2.

[7] 백종일, “DBMS WAS 우회접속의 쿼리정보 역추적 연구”, 한국컴퓨터정보학회논문지, 제 14권 제 3호, pp.173-181, 2009. 12.