情報保護學會論文誌
第19卷 第6號, 2009. 12

웹 로그 데이터에 대한 개인정보 위협분석 및 보안 가이드

여 성 구,† 심 미 나, 이 상 진‡
고려대학교 정보경영공학전문대학원

The Threat Analysis and Security Guide for Private Information in Web Log

Sung-Koo Ryeo,† Mi-Na Shim, Sang-Jin Lee‡
Graduate School of Information Management and Security CIST, Korea University

요 약

본 논문은 오늘날 정보사회의 핵심 정보 자원인 개인정보가 웹 로그를 통해 누출될 수 있는 보안 위협의 심각성을 재인식시키고, 이를 근원적으로 예방하기 위한 대응방안을 제시한다. 최근 개인정보는 정보사회의 발전과 함께 범위 및 종류가 확대되고 그 중요성이 매우 커지게 되었다. 웹 로그는 법·제도적으로 규정된 개인정보가 저장되는 개인정보 파일임에도 불구하고, 웹 서비스의 부산물 정도로만 인식되어 충분한 보호조치가 이루어지지 못하고 있다. 웹 로그를 통해 노출될 수 있는 개인정보를 개발 단계에서 통제하여 웹 로그에 개인정보가 저장되는 것을 최소화하고, 운영 단계에서 적용되어야 하는 기술적 대안을 제시한다. 근본적 보호체계를 통해 법·제도적 규제를 준수하고 개인정보를 효과적으로 보호할 수 있다.

ABSTRACT

This paper discusses an issue of serious security risks at web log which contains private information, and suggests solutions to protect them. These days privacy is core information to produce value-added in information society. Its scope and type is expanded and is more important along with the growth of information society. Web log is a privacy information file enacted as law in South Korea. Web log is not protected properly in spite of that has private information It just is treated as residual product of web services. Many malicious people could gain private information in web log. This problem is occurred by no classified data and improper development of web application. This paper suggests the technical solutions which control data in development phase and minimizes that the private information stored in web log, and applies in operation environment. It is very efficient method to protect private information and to observe the law.

Keywords: Private Information, Web Log, Threat Analysis, Security Guide

I. 서 론

오늘날 개인정보는 과거와 달리 부가가치를 창출하는 핵심 정보 자원으로서 매우 중요한 가치를 가지고 있다. 정보사회의 발전으로 인해 GPS에 의한 개인 위치 정보, CCTV에 의해 수집되는 화상정보 등 과거에는 존재하지 않았던 새로운 유형의 개인정보가 등장하여 개인정보의 범위도 확대되었다.

개인정보가 웹 서비스에서 다루어짐으로 인해 과거 오프라인에서와는 다른 보안 위협들이 발생하고 있다. 이로 인해 발생하는 개인정보의 침해는 '수집 목적 이외의 사용', '개인정보의 비밀수집', '개인정보의 남용', '기술적·관리적 조치 미흡으로 인한 개인정보 누출' 등이며, 한국인터넷진흥원(KISA)의 조사에 따르면 2005년 18,206건, 2006년 23,333건, 2007년 25,965건, 2008년 39,811건으로 지속적으로 증가하고 있는 추세이다[1,2]. 이러한 보안 사고의 증가로 인해 정부 및 기업, 개인은 개인정보보호의 심각성을

접수일(2009년 9월 3일), 게재확정일(2009년 10월 5일)
† 주저자, bar4mi@gmail.com
‡ 교신저자, sangjin@korea.ac.kr

인지하고, 정보제공과 이용에 기술적, 관리적, 법·제도적 차원의 보호 노력을 지속적으로 하고 있다. 그러나 다각적인 보안 조치에도 불구하고 그 중요성이 간과되어 보안 조치가 충분하지 못한 영역이 존재한다. 웹 로그는 웹 서비스를 제공할 때 자동적으로 생산되고 저장되는 데이터로, 개인정보가 저장될 수 있다. 그러나 이러한 사실에 대한 간과와 중요성에 대한 인식 부족으로 법·제도적 규제, 그리고 기술적 보안 조치가 보안적 관점에서 미흡하다.

웹 서비스 기술과 개인정보 관련 법·제도적 규제를 이해하고, 웹 로그로 인해 발생할 수 있는 웹 서비스의 보안 위협과 법·제도적 규제의 준수사항을 검토함으로써, 웹 로그의 중요성과 보안 위협을 재인식하고 웹 로그 속에 존재하는 개인정보를 안전하게 보호하는 방안을 제안하고자 한다.

II. 웹 서비스와 개인정보 법·제도적 규제의 이해

웹 로그로 인해 발생하는 개인정보의 보안위협을 이해하기에 앞서, 웹 애플리케이션의 전송방식과 이로 인해 생산되는 부가 데이터에 대해 살펴보고, 법·제도적으로 규정된 개인정보의 정의와 개인정보에 대한 보호대책 요건에 대해 기술한다.

2.1 웹 서비스 기술의 이해

웹 애플리케이션의 요청은 일반적으로 GET과 POST 메소드(Method)를 통해 전송되며, 사용자 측 요청과 서버 측 응답 과정에서 웹 로그를 비롯한 다양한 부가 데이터가 사용자 측과 서버 측에 생산 및 저장된다.

2.1.1 HTTP 요청 전송 방식

HTTP 프로토콜에서 지원하는 요청 메소드는 GET, POST, HEAD, TRACE 등 다양하다[3]. 이 중에서 GET과 POST 메소드가 사용자 측 요청을 처리하기 위해서 주로 사용된다. GET 메소드와 POST 메소드의 가장 큰 차이점은 로그 파일에 입력값(Parameter) 기록 여부이다. GET 메소드는 자원의 검색을 지원하기 위해 디자인된 것이다. 사용자는 URI-요청(Request)을 통해 원하는 웹 서비스 자원(/dir/board.jsp)에 입력값(id=notice&no=100)을 전송할 수 있다. 이러한 요청을 저장할 수 있기 때문에 사용자들은 이후 접속을 위해 동적 자원을 북마크 할 수 있다. GET 메소드로 전송된 URI-요청은 HTTP 헤더의 'Referer'를 통해 이후 접속하는 웹 애플리케이션(사이트)에 전달된다. POST 메소드는 폼의 Action을 수행하기 위해 디자인된 것으로, URI-요청과 메시지의 본문(Body)을 함께 이용하여 사용자 요청을 서버 측에 전송할 수 있다. URI-요청을 북마크할 수 있으나 본문에 포함된 입력값은 제외된다[4].

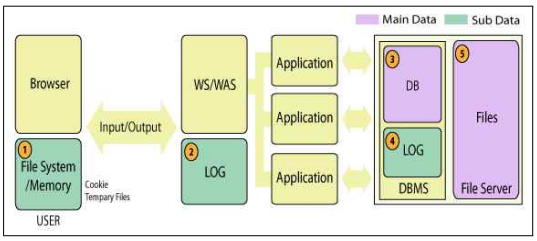
[표 1] 메소드별 웹 로그 데이터

구분	내용
GET	172.30.20.139 - - [07/Jun/2009:09:11:58 +0900] "GET /trac/wiki/ThesisPrivary?action=edit HTTP/1.1" 200 7447
POST	172.30.20.139 - - [07/Jun/2009:09:28:30 +0900] "POST /trac/wiki/ThesisPrivary HTTP/1.1" 303 14

웹 로그에 저장되는 데이터는 전송시간, 요청자의 IP, 메소드, 상태코드, 응답의 크기 등이다[5]. 웹 서버·웹 애플리케이션 서버에서 지원하는 옵션을 이용하여 '쿠키(Cookie)'를 로그 파일에 기록할 수도 있다[6]. [표 1]에서 볼 수 있듯이 POST 메소드 요청은 웹 애플리케이션까지만 저장되며, GET 메소드는 입력값까지 포함되어 저장된다.

2.1.2 생산되는 데이터

웹 서비스는 회원 가입 처리, 게시판, 구매 처리 등의 웹 애플리케이션들을 포함하고 있다. 개별 웹 애플리케이션은 파일·데이터베이스·데이터 처리라는 특정 목적을 수행하기 위해 제작된 것으로, 사용자 측에서 전송하는 데이터를 입력값으로 받아들여 [그림 1]과 같이 각 영역에서 부가적인 데이터들을 생성한다. '③ DB'는 웹 서비스의 회원 정보, 회계 정보 등의 웹 서비스 핵심 정보가 저장되며, '⑤ Files'에는 첨부 문서 및 프로그램 등과 같은 공유 목적으로 생성되는 파일



[그림 1] 웹 서비스로 인해 생성되는 데이터

들이 저장된다. '④ LOG'는 DBMS에서 SQL 질의와 사용자의 접속 이력을 감사할 목적으로 생성된다. '② LOG'는 웹 서버 및 웹 애플리케이션 서버에서 디버깅과 감사 등의 목적으로 사용자의 URI-요청, Referer 등과 같은 요청과 정보를 저장하는 것으로 웹 로그라고 명칭된다. '① File System/ Memory'는 사용자의 인증 및 편의 목적으로 생성되는 쿠키(Cookie)와 임시 파일, 히스토리 등의 파일과 웹 서비스 이용 중 메모리에 존재하는 데이터들이다. 보안적 측면에서 아이디, 권한, 성명 등과 같은 정보는 쿠키 값 대신 세션 ID 이용을 권장하고 있지만, 아직도 많은 개발자가 그대로 쿠키를 사용하고 있다. 만약 쿠키 값을 저장하도록 웹 로그가 설정되어 있다면, 사용자 측의 쿠키 값이 웹 로그에 저장될 수 있다.

2.2 법·제도적 이해

개인정보 보호는 법률에 명시되어 있는 개인정보의 정의로부터 시작된다. 개인정보를 안전하게 보호하기 위해 다양한 법률에서 개인정보의 보호대책을 규정하고 있으며, 안전한 정보통신망 운영과 사고 대응을 위해 접속기록에 대한 종류와 기한 등을 법률로 정하고 있다.

2.2.1 개인정보의 정의

개인정보는 '생존하는 개인에 관한 정보로서 당해 정보에 포함되어 있는 성명, 주민등록번호 등의 사항에 의하여 당해 개인을 식별할 수 없더라도 다른 정보와 용이하게 결합하여 식별할 수 있는 것'을 포함한다[7,8]. 이를 풀어서 보면, 개인 식별이 가능한 정보는 성명 및 주민등록번호와 같이 해당 정보에 포함된 사항에 의하여 개인을 식별할 수 있는 정보를 말하며, 주소와 같이 개인을 식별할 수는 없으나 개인 식별 정보 등과 조합할 경우 특정인을 식별할 수 있는 정보 등도 개인정보에 해당된다는 것이다[9].

[표 2] 개인정보의 구분과 예

구분		개인정보의 예
1차 개인정보	오프라인	성명,주민등록번호,핸드폰번호, 주소, 운전자등록번호, 신용카드 등
2차 개인정보	식별자	아이디·비밀번호, 보험증권번호, 사원번호 등
	성향, 상태등	IP, 웹 애플리케이션명, 입력값 등

본 논문에서는 앞서 설명한 개인정보의 정의를 [표 2]와 같이 웹 서비스 차원에서 재정의하여 오프라인에서 개인정보로 취급되는 개인정보를 1차 개인정보로 정의하고, 애플리케이션과 연계되어 사용자의 행위 패턴, 특징 등을 도출할 수 있는 정보를 2차 개인정보로 정의한다. 예를 들어 오프라인에서와 동일하게 온라인상에서도 개인을 대표하는 1차 개인정보는 성명, 주민등록번호, 핸드폰번호, 주소, 운전자등록번호, 신용카드 등이다[10]. 이는 오프라인의 개인정보가 온라인으로 그대로 전이된 것으로 온·오프라인 양측에서 개인을 식별할 수 있으며, 많은 사람들이 해당 정보가 개인정보라고 인식하고 있어 그 중요도가 매우 높다. 2차 개인정보는 온라인상에서 웹 애플리케이션과의 조합을 통해 개인을 식별할 수 있게 한다. 2차 개인정보는 오프라인 상의 개인 정보 또는 온라인상에서의 개인 식별자(아이디)가 웹 애플리케이션과 조합되어 개인의 식별, 성향, 상태 등을 나타낼 수 있다. 해당 웹 서비스에 대한 이해도가 낮은 사람은 그 중요성과 심각성을 인지하기 어려운 측면이 있으나, 해당 웹 애플리케이션에 대한 이해도가 높은 운영자·개발자 등의 이해 관계자는 해당 정보를 이용하여 개인을 식별할 수 있고 다량의 데이터를 수집할 경우 개인의 프라이버시를 침해할 수 있다. 2차 개인정보는 온라인을 이용한 기술 및 서비스가 하루가 다르게 발전하고 있기 때문에 그 종류 및 범위를 한정 짓기 어려운 특성이 있다.

2.2.2 법·제도적 규제 현황

'정보통신망 이용촉진 및 정보보호 등에 관한 법률(법률 제9637호)'과 '공공기관의 개인정보에 관한 법률(제8871호)'을 기초로 하여 동 법률의 시행령, 시행세칙, 고시 등을 통해서 개인정보에 대한 보호대책을 명시하고 있다. 이로 인해 개인정보책임관 지정과 개인정보취급자에 대한 운영방안을 수립해야하며, 개인정보에 대해 '개인정보 암호화', '송수신 정보보안', '불법접근 통제', '접속기록의 위조·변조 방지'와 같은 보호 조치를 실시해야 한다[11].

웹 로그(접속기록)에 대한 법률적 근거는 '정보통신이용촉진 및 정보보호 등에 관한 법률'과 '공공기관의 개인정보보호에 관한 법률', 그리고 '정보시스템의 효율적 도입 및 운영 등에 관한 법률' 등을 기초로 하고 있다. 법·제도적 규제로 인해 웹 로그와 같은 접속기

록은 서비스 제공자의 소속에 따라서 보관 기간이 달라질 수 있으나, 최소 6개월 이상 유지해야 한다[12].

III. 웹 로그로 인한 보안 위협

웹 로그는 웹 서비스의 부가 데이터로서 부적절한 웹 애플리케이션의 구현으로 인해 개인정보가 포함될 수 있다. 이에 반해 법·제도적 보호조치는 데이터베이스시스템에 집중되어 있으며, 웹 로그로 인한 개인정보 노출이 간과되고 있다.

3.1 웹 서비스의 보안 위협

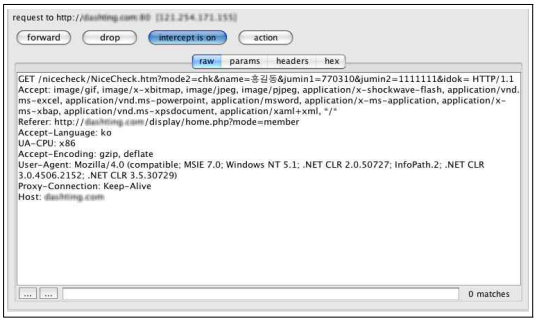
웹 서비스에서 다루어지는 개인정보는 앞서 살펴본 것과 같이 생산되는 데이터와 잘못된 HTTP 요청 방식에 의해서 웹 로그에 저장될 수 있다. GET 메소드의 URI-요청, 잘못 설정된 쿠키 값, Referer 헤더값 등 다양한 원인에 의해 개인정보는 인지하지 못하는 사이에 웹 로그에 저장된다. 이렇게 저장된 개인정보는 웹 애플리케이션의 구현에 따라서 1차 개인정보와 2차 개인정보를 포함할 수 있다. 웹 로그는 외부 공격자뿐만 아니라 내부자에 의해 노출·유출될 수 있는 보안위협이 존재한다.

3.1.1 HTTP 요청 방식에 의한 보안위협

부적절한 HTTP 요청 방식으로 인해 발생할 수 있는 보안 위협은 서비스 제공자가 과다수집 또는 불충분한 보호조치를 함으로써 발생한다. 악의적인 서비스 제공자가 남겨서는 안 되는 개인정보를 의도적으로 수집하고자 하는 경우, 웹 로그를 이용하여 수집할 수 있다. 서비스 제공자는 웹 애플리케이션에서 전송 메소드를 GET 메소드로 변경하는 간단한 조작을 하는 것만으로, 웹 로그를 통해 개인정보를 수집할 수 있다. 악의적 조작 외에도 개발자의 이해 부족이나 디버깅 목적으로 인해 GET 메소드를 사용할 경우 개인정보가 수집이 된다. 개인정보 수집을 의도하지 않은 경우에도 타 서비스의 잘못된 HTTP 요청 방식과 미흡한 개인정보 관리로 인해 개인정보가 노출될 수 있다.

한 예로 국내에서는 성인 콘텐츠에 대해 미성년자의 접근을 통제하기 위한 성인인증과 가입자의 신원을 확인하기 위한 실명인증을 실시하고 있다. 이 중 일부 사이트를 자세히 살펴보면, 인증 정보가 POST 메소드가 아닌 GET 메소드를 통해 전송되고 있는 것을

[그림 2] 회원 가입을 위한 성인인증



[그림 3] GET 요청으로 전송되는 성인인증정보

알 수 있다. 이 경우 서비스 제공자의 웹 로그에 사용자의 이름과 주민등록번호가 고스란히 저장되게 된다. [그림 2]에서 사용자의 성명과 주민등록번호를 입력한 후 '무료회원가입'을 요청하게 되면 GET 메소드를 통해서 전송되는 것을 [그림 3] 웹 프록시를 통해 알 수 있다. 해당 사이트의 웹 로그에는 URI-요청인'/nicecheck/NiceCheck.html?mode2=chk&name=홍길동&jumin1=770310&jumin2=1111111&idok='이 저장되게 된다. 이로 인해 개인정보 서비스 제공자의 웹 로그에 저장된다. 일부 벤더의 웹 서버 로그 설정은 'Referer'를 기본 저장되도록 설정되어 있다. 만약 악의적인 서비스 제공자의 사이트를 방문하여 인증을 실시한 후 생각이 바뀌어 해당 웹 브라우저에서 새로운 주소를 입력하고 타 사이트로 접속하는 경우, 타 사이트의 웹 로그 'Referer' 영역에 위와 같은 개인정보가 저장되게 된다.

3.1.2 웹 애플리케이션에 의한 보안위협

웹 로그의 노출로 인해 발생할 수 있는 또 다른 위협은 웹 애플리케이션 성격과 입력값으로 제공되는 아이디와 같은 개인 식별자가 조합되어 개인의 성향 및 상태가 노출되는 것이다. 웹 로그는 일반적으로 시스템 관리자만 읽을 수 있도록 설정되어 있으며, [표 1]

과 같이 일반 TEXT로 정보를 저장하고 있다. 따라서 해당 권한을 가지고 있다면, 임의의 텍스트 편집기를 이용하여 손쉽게 정보를 취득할 수 있다. 예를 들면, 로그인을 처리하는 웹 애플리케이션의 웹 로그에 아이디와 비밀번호가 저장되어 있다면, 공격자는 이를 획득하여 타인으로 로그인하여 해당 사용자의 개인정보를 절취할 수 있다. 웹 로그로 인한 다른 위협들을 살펴보자. 문자 메시지를 전송하는 웹 애플리케이션의 웹 로그를 통해 문자 메시지 전송 애플리케이션 이름과 수신자·송신자의 핸드폰 번호, 그리고 요청 시간 등이 유출될 수 있으며, 이를 통해서 어떤 핸드폰 번호를 가진 사용자가 어떤 사람과 몇 시에 문자 메시지를 전송하였는지를 파악할 수 있다. 포탈 서비스와 같이 여러 콘텐츠를 제공하는 웹 서비스의 웹 로그를 통해 유출되는 번호 애플리케이션 이름과 특정인의 아이디를 조합하여 특정인의 취향을 파악할 수 있다. 금융 관련 상품 가입·해지·대출금 신청 등의 웹 애플리케이션의 웹 로그를 통해 특정 업무를 수행하는 웹 애플리케이션 이름, 보험·증권번호, 또는 연락처 등이 유출될 수 있으며, 이로 인해 개인정보뿐만 아니라 서비스 제공자의 기밀정보가 유출될 수 있다.

3.2 법 · 제도적 보안 위협

개인정보파일로서 웹 로그 보호는 법·제도적 요건을 충족시키지 못하고 있으며, 법·제도적으로 충분히 반영되어 있지 못하다. 개인정보를 포함한 웹 로그는 현재 개인정보에 관한 법·제도적 규제를 준수하고 있지 못하며, 개인정보 보호 규제 자체도 데이터베이스 시스템에 한정되어 있어 법·제도적 보호조치 자체가 웹 서비스의 개인정보를 보호하기에는 미흡하다.

3.2.1 법 · 제도적 규제 미충족

개인정보를 다루고 있는 웹 서비스 제공자는 앞서 살펴본 다양한 법·제도적 규제에 의해서 최소 네 가지의 개인정보에 대한 보호장치를 충족시켜야 하지만 이를 충족시키고 있지 못하다. 첫째, 개인정보 암호화 측면에서 웹 로그에 저장된 개인정보는 암호화 되어 있지 않다. 부적절한 HTTP 전송 요청과 부적절한 웹 애플리케이션 구현으로 인해 웹 로그에는 데이터베이스시스템과 동일한 수준의 개인정보가 저장되지만, 웹 로그에 대한 암호화 방안은 존재하지 않는다. 둘째, 송수신 정보보안 측면에서는 웹 로그는 채널통신 암호화 이후에 기록이 되는 것으로 송수신 암호화 대책에는 해당사항이 없다. 셋째, 불법접근 통제 측면에서 웹 로그는 파일시스템에 대한 접근통제를 실시하고 있으나, 이는 악의적인 내부자에 대한 충분한 통제가 되질 못한다. 파일시스템 측면에서 웹 로그에 접근을 할 수 있는 내부 권한자는 시스템관리자와 적법한 권한을 부여 받은 시스템 계정 사용자이다. 악의적인 내부 권한자는 웹 로그에 대한 열람 권한을 가지고 이용하여, 웹 로그에 포함된 1차 개인정보 및 2차 개인정보를 획득하여, 개인적 또는 경제적 이익을 위해 사용할 수 있다. 이러한 악의적인 행위에 대해 파일시스템 차원의 접근제한은 한계점을 가지고 있기 때문에 이를 사전에 차단하거나 사후 감사하기 위한 조치가 이루어지지 못한다. 넷째, 접근기록 위조·변조 방지 측면에서 웹 로그는 권한이 있는 내부자에 의해 위조·변조 될 수 있다. 일부 시스템의 경우 로그 파일을 원격의 자기매체에 보관하고 있기는 하나, 일반적으로 동일 시스템의 파일시스템에 저장하고 있다. 권한이 있는 내부자는 웹 로그의 열람, 수정이 자유로우며, 해당 시스템 수행한 명령어 이력(.bash_history)을 수정함으로써 열람 흔적을 조작할 수 있다.

3.2.2 법 · 제도적 규제의 미흡

기술적 차원에서 웹 로그와 데이터베이스시스템은 동일한 수준의 개인정보가 저장됨에도 불구하고 개인정보에 대한 법적 조치를 데이터베이스시스템에 한정하고 있으며, 권한이 있는 내부자에 의해 웹 로그에 발생할 수 있는 보안 위협을 간과하고 있다.

‘정보통신망 이용촉진 및 정보보호 등에 관한 법률 시행령 제15조 2항 1호’에서는 개인정보처리시스템을 데이터베이스시스템으로 한정하고 있다. 해당 시행령의 모태가 되는 ‘정보통신망 이용촉진 및 정보보호 등에 관한 법률 제4장 2절 28조 1항’과 ‘공공기관의 개인정보보호에 관한 법률 제2장 제7조의2 1항 3호’에 의해 웹 로그는 개인정보파일에 속한다. 해당 시행령은 개인정보파일을 데이터베이스시스템에 저장이 되는 데이터로만 한정 짓고 있어, 그 이하 시행 규칙, 고시 등에서 웹 로그에 대한 보호조치가 미흡한 것이 현실이다.

개인정보 보호에 대한 목적과 개인정보처리시스템으로 규정된 데이터베이스시스템에 대한 보호조치를 고려한다면, 웹 로그에 대한 권한이 있는 내부자에 의해 발생할 수 있는 정보유출 등의 보안위협에 대한 보

호조치가 부족하다. 권한 있는 내부자는 개인정보에 대해 임의적인 열람은 물론 열람기록 등을 수정할 수 있는 권한을 가지고 있으나, 이를 견제할 수 있는 보호조치는 고려되어 있지 않다.

IV. 웹 로그의 개인정보 보호 방안

웹 로그로 인해 발생하는 개인정보의 보안위협은 수집, 운영 단계에서 발생하는 것으로, 개발 단계에서 개인정보보호 가이드를 준수하여 웹 로그에 개인정보가 저장되는 것을 사전에 예방할 수 있다. 또한 의도하지 않은 개인정보 수집으로 인해 발생하는 보안위협을 예방하기 위해 웹 로그에 대한 정기적인 감사를 실시해야 한다.

4.1 보안 위협의 원인 분석

웹 로그의 개인정보를 안전하게 보호하기 위해서는 모든 이해관계자를 위협원으로 정의하고, 각 보안 위협에 대해 핵심 개인정보파일인 데이터베이스와 동일한 수준의 보호조치가 이루어져야 한다. 하지만 웹 서비스에서 웹 로그는 감사, 사용자 행위 분석 등을 위한 것으로, 비즈니스 측면에서 부가적인 것이다. 부가적인 가치에 대해 비즈니스의 핵심인 데이터베이스와 동일한 수준의 비용을 지불한다는 것은 현실적이지 못하다. 또한 각종 보안 솔루션과 보안 절차를 도입한다고 하더라도 이는 근본적인 해결책이 되지 못한다.

웹 로그에서 발생할 수 있는 보안 위협을 서비스제공자, 개발자, 권한 있는 내부자, 공격자의 입장에서 종합해보면 [표 3]과 같이 정리된다. 현재의 법·제도적 규제로 통제가 가능한 위협은 서비스 제공자에 의한 악의적 개인정보 수집, 개발자의 개인정보 처리 방침에 대한 이해 부족, 공격자에 의한 악의적 개인정보 노출이다. 앞서 살펴본 보안 위협은 웹 서비스 기술에 대한 이해 부족과 법·제도적 규제가 충분하지 못해 발생하는 보안위험으로, 의도하지 않은 개인정보 수집, 개인정보의 웹 로그 축적, 악의적인 권한자에 의한 개인정보 유출이 존재한다. 이러한 보안위협은 수집 및 운영 단계에서 발생하는 것이다. 수집 단계에서 서비스 제공자에 의해 악의적·비의도적인 개인정보 수집이 이루지고, 운영 단계에서는 내부자에 의한 개인정보 유출, 불법적 공격자에 의한 개인정보 노출 등이 존재할 수 있다. 특히 로그 열람에 대한 감사 기능 및 절차가 존재하지 않기 때문에 개인정보에 대한 침해가 얼마만큼 발생했는지조차 알기 어렵다.

[표 3] 개인정보 이해관계자별 발생 가능한 보안 위협

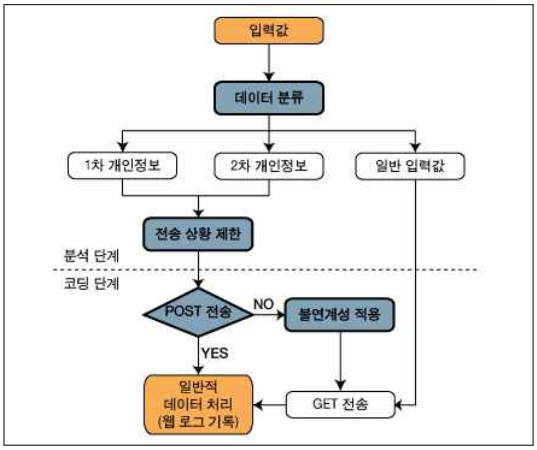
구분	발생가능한 위협
서비스 제공자	• 악의적인 개인정보 수집 • 의도하지 않은 개인정보 수집
개발자	• 개인정보 처리 방침에 대한 이해 부족 • 웹 서비스 기술에 대한 이해 부족(부가 데이터)
권한있는 내부자	• 악의적인 권한자에 의한 개인정보 유출
공격자	• 악의적인 내·외부자에 의한 공격으로 인한 개인정보 노출

현재까지 웹 애플리케이션 보안은 잘 알려진 웹 애플리케이션 공격기법에 대응하는 보안 코딩 기법이나 웹 애플리케이션의 프로그래밍 구현 측면에서의 안전한 웹 애플리케이션을 개발하는 것에 초점이 맞추어져 있다. 웹 애플리케이션의 보안을 강화함으로써 침해의 위협을 제거하면 개인정보가 잘 보호될 것이라는 가정을 하고 있는 것이다. 실제 웹 애플리케이션에서 다루게 되는 데이터들은 개인정보 측면에서 어떻게 다루어져야 하는 지는 개발자 개인의 몫으로 남겨 놓고 있다. 웹 로그의 보안을 강화하기 위해 서버 측에 접근제어 솔루션을 도입하여 시스템 계정의 웹 로그 접근과 명령어 실행을 통제하거나 원격 로그 백업 시스템을 도입할 수 있다. 하지만 이는 비용이 많이 들면서 근원적인 해결방안은 되지 못한다. 근본적인 해결책을 위해 웹 애플리케이션 개발 단계에서 데이터를 통제하고 가이드 함으로써, 의도하지 않은 개인정보 수집을 최소화하고, 웹 로그를 통해 수집되는 개인정보를 통제할 수 있기 때문에 예상치 못한 보안 사고를 예방할 수 있음은 물론 이에 대한 적극적인 대응도 할 수 있다. 이로써 비용 효과적으로 웹 로그를 통해 발생하는 개인정보 유출·노출 위협을 제거할 수 있다. 또한 의도하지 않은 개인정보 수집 및 부적절한 개인정보 수집 여부를 확인하기 위해, 웹 로그에 대한 정기적인 감사를 실시함으로써 개인정보 유출 위협을 사전에 통제할 수 있다.

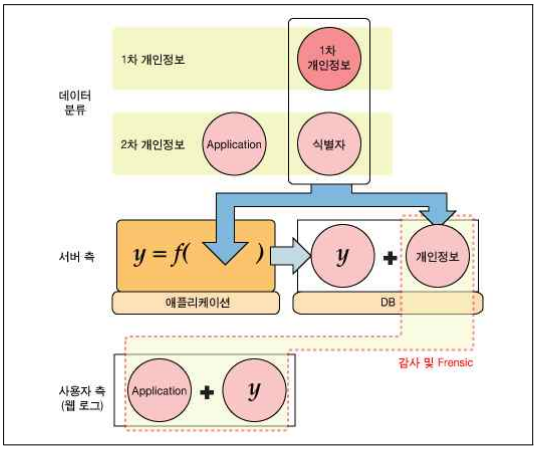
4.2 웹 로그의 개인정보 보호 가이드

웹 애플리케이션를 통해 생산되는 웹 로그의 개인정보를 안전하게 보호하기 위해서는 [그림 4]와 같은 절차에 따라서 데이터가 분류되고 구현되어야 한다.

첫째, 데이터 분류를 실시한다. 소프트웨어 개발 수명 주기(SDLC)의 '분석(Analysis) 단계'에서 데이



[그림 4] 웹 로그 데이터 처리 절차



[그림 5] 웹 로그의 불연계성 구현

터 분류를 수행할 때, 웹 애플리케이션에서 처리하는 입력값 중에서 1차 개인정보와 2차 개인정보, 그리고 일반 입력값으로 데이터를 분류한다. 데이터 분류 시에는 웹 애플리케이션과 1·2차 개인정보를 함께 목록화하여 어떤 애플리케이션에서 해당 입력값을 사용하는지 기록한다.

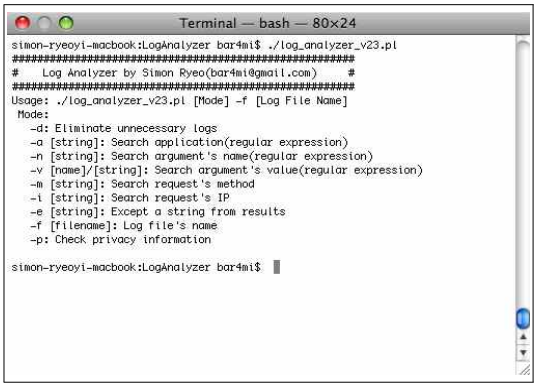
둘째, 입력값 위치(상황)별 처리 방법을 규정한다. 개인정보는 크게 3가지 상황으로 구분될 수 있다. 첫 번째는 사용자가 최초 입력하는 상황이며, 두 번째는 개발자가 사전에 정의한 입력값을 처리하는 상황이다. 세 번째는 저장된 개인정보가 사용자 측에 전송되었다가 재전송 받아 처리하는 상황이다. 두 번째의 경우 내부 연산으로서 웹 로그에 저장되지 않기 때문에 배제할 수 있다. 문제가 되는 것은 첫 번째와 세 번째 상황으로 실제적인 예는 회원가입 폼에서 사용자 정보를 입력받는 경우와 회원 정보 수정, 사용자 인증 토큰이 쿠키로 저장되어 있는 경우 등이다. 웹 애플리케이션 또는 웹 서비스 연계 등의 특별한 상황을 제외하고 1차 개인정보가 URI-요청을 통해 전송되어서는 안 된다. 쿠키에 개인정보가 포함되는 것을 제한해야 하며, 다른 보안 위협을 제거하기 위해서라도 쿠키 대신 세션 ID를 이용하여 인증 토큰을 처리해야 한다.

셋째, 전송 방식을 준수해야 한다. 1차 개인정보는 웹 로그의 특수한 목적이나 상황을 이유로 의도적으로 웹 로그에 개인정보를 남기는 것을 제외하고 , POST 메소드를 이용하여 입력값을 전송해야 한다[13]. 의도적으로 GET 메소드를 이용해 개인정보 또는 개인 식별자를 전송하는 경우에도 개인정보와 웹 애플리케이션 간의 불연계성(Unlinkability)를 보장하여 개인정보가 노출되는 위험을 제거해야 한다.

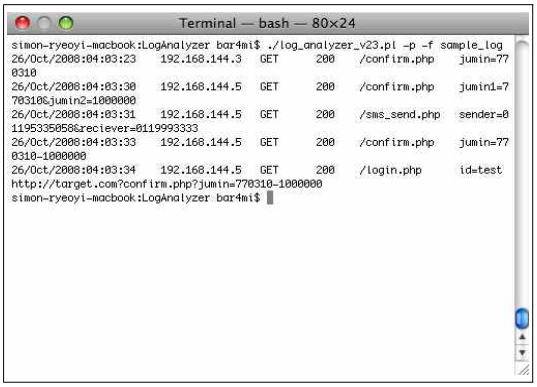
넷째, 불연계성을 보장해야 한다. 웹 애플리케이션의 특성 및 감사, 사용자 성향 분석, 포렌식이라는 목적을 위해 의도적으로 개인정보를 남겨야 하는 경우가 있다. 이때는 웹 로그만으로 개인정보가 노출되지 않도록 정보와 권한을 분리시켜 권한이 있는 내부자에 의한 정보 누출을 예방해야 한다. [그림 5]에서와 같이 개인정보는 서버 측에서 단방향 함수 $f()$를 이용하여 암호 식별자 y로 변경, 데이터베이스에 대한 개인정보 보안 조치가 적절히 이루어지고 있다는 가정 하에 기존의 개인정보와 함께 암호 식별자 y를 연계하여 저장한다. 웹 애플리케이션에서는 기존의 1차 개인정보나 아이디와 같은 개인 식별자 대신 암호 식별자 y를 이용하여 GET 메소드 등과 같이 웹 로그에 저장되는 요청들을 처리하도록 한다. 이로 인해 서버 관리자는 데이터베이스에 존재하는 개인정보와 암호 식별자 y의 연관 관계를 파악하지 못하는 한 특정 개인을 식별할 수 없게 된다.

사용자 행위 분석을 실시해야 할 경우는 데이터베이스에서 사용자의 개인 식별자와 암호 식별자 y의 정보를 넘겨받아, 기존과 동일한 사용자 행위 분석을 수행할 수 있다. 보안 감사와 포렌식을 수행할 경우에도 규정에 따라 데이터베이스 관리자에게 해당 데이터들을 넘겨받아 공격자 및 규정 위반자를 추적할 수 있다.

다섯째, 정기적인 감사 실시와 관리적 보안 규제를 강화한다. 권한이 있는 내부자에 의한 보안 위협을 예방하기 위해서는 정기적으로 웹 로그를 분석하여 개인정보가 존재하는 지를 점검해야 한다. 이를 통해 불필요한 정보의 저장을 최소화하고 인지하고 있지 못한



[그림 6] 웹 로그 분석기



[그림 7] 로그 분석기를 통한 개인정보 존재 확인

개인정보가 저장되고 있는지를 반드시 파악해야 한다. 이러한 활동을 통해 비의도적으로 개인정보가 수집되는 것을 탐지할 수 있으며, 개인정보 보호를 위해 보다 강화된 접근제어 정책과 애플리케이션의 수정을 통해 개인정보가 의도치 않게 노출·누출되는 것을 예방할 수 있다. [그림 6]의 웹 로그 분석기(Log Analyzer)는 웹 로그에 대한 감사를 실시하기 위해 perl 언어를 이용하여 제작한 것이다.

해당 웹 로그 분석기는 개인정보 탐색 옵션(-p)를 이용하여 웹 로그에 포함되어 있는 주민등록번호, 핸드폰번호, 신용카드번호와 같은 1차 개인정보를 정규표현식으로 검출할 수 있으며, 증권번호 등과 같이 특정 조직에서 사용되는 개인정보는 [표 4]와 같이 설정파일(privacy.txt)의 정규표현식을 이용하여 정의 가능하다.

또한 IP를 기반(-i)으로 한 사용자 행위 분석과 특정 웹 애플리케이션(-a)에 대한 분석, 특정 변수(-n) 또는 입력값(-v)을 가지고 있는 사용자 요청을 분석할 수 있다. [그림 7]과 같이 웹 서버 아파치의 로그파일에 존재하는 개인정보를 탐색한 결과, 패턴으로 설정된 개인정보가 검출되었다. 분석도구를 이용하여 정규화할 수 있는 개인정보 뿐만 아니라 사용자의 요청에 따른 사용자 행위에 포함되는 개인정보를 분석할 수도 있다.

[표 4] 개인정보 정규표현식의 예(privacy.txt)

```
# For Korean Social Security Number
JUMIN1=[0-9]{2}[01][0-9][0123][0-9]-[1234][0-9]{6}
JUMIN2=[0-9]{2}[01][0-9][0123][0-9]
JUMIN3=[1234][0-9]{6}
# For Cellphone Number
PHONE1=01[016789]-(\d{3}|\d{4})-\d{4}
PHONE2=01[016789][0-9]{3,4}[0-9]{4}
# For Credit Card
# 3: American Express, JCB
# 4: Visa
# 5: Master
# 6: China UnionPay
# 9: Etc
CREDIT1=(3|4|5|6|9)[\d]{15}
```

4.3 개선결과

웹 애플리케이션의 개발 단계에서 개인정보를 분류하고 통제함으로써 웹 로그에 1차 개인정보와 2차 개인정보 식별자가 저장되는 것을 예방하고 [표 5]와 같이 법·제도적 규제를 충족시킬 수 있다. 이로 인해 개인정보 유출·노출로 인해 발생할 수 있는 개인의 피해는 최소화 할 수 있다. 그러나 '불법접근 통제'와 '접근기록 위·변조 방지'의 요건은 웹 로그로 인해서 노출될 수 있는 성향과 특징 등의 보안 위협이 그대로 잔존한

[표 5] 개선 된 웹 로그 개인정보 보호 현황(TO-BE)

구분	충족여부	적용 가이드
개인정보 암호화	충족 (1차,2차 식별자)	데이터 분류, 전송 상황 제한, 불연계성
송수신 정보보안	해당사항 없음	
불법접근 통제	일부 충족 (개인식별 불가)	데이터 분류, 전송 상황 제한, 불연계성
접근기록 위·변조 방지	일부 충족 (개인식별 불가)	데이터 분류, 전송 상황 제한, 불연계성

다. 하지만 데이터 통제와 암호화 조치로 인해 특정 개인을 지정하여 한 개인의 성향과 상태를 파악할 수 없다.

법·제도적으로 미흡한 부분은 상위 법률에서 개인정보보호 규제를 명시하고 있는 만큼 시행령과 시행규칙 등에서 개인정보처리시스템의 범위를 처리·저장시스템으로 확대해야 할 것이다. 본 논문에서 구체적인 법·제도적 대안을 제시하지 않겠지만, 개인정보 처리 보안가이드를 준수하게 되면, 법·제도적 규제 속에서 [표 5]와 같은 요건들을 충족시킬 수 있다.

V. 결론 및 향후 방향

웹 로그는 보안 사고의 감사 및 사용자 행위 분석 등에 사용되는 중요한 데이터이다. 웹 애플리케이션의 잘못된 구현 및 악의적인 구현으로 인해 개인정보가 웹 로그를 통해 노출될 수 있다. 개인정보 보호를 강화하기 위해서는 개발 단계에서 개인정보의 불필요한 저장을 사전에 차단함으로써 수집과 운영 단계에서 발생하는 부담을 줄이고 보안 사고를 예방할 수 있다. 이러한 관점에서 웹 로그의 개인정보 보호 가이드는 설계 단계에서 개인정보를 분류하고, 구현 단계에서 정보를 통제함으로써 운영 단계에서 개인정보 보호에 드는 비용을 최소화하고 관리 및 통제의 부담을 최소화할 수 있다.

본 논문의 대응방안의 불연계성 가이드는 데이터베이스시스템에 대한 개인정보 보호조치가 충분히 이루어지고 있음을 가정한다. 개인 식별자의 노출로 인한 특정 개인의 개인정보가 노출되는 것을 예방할 수 있으나, 전체 사용자의 행위나 서비스 상태를 분석하는 행위 자체를 차단하는 것에는 한계가 존재한다. 또한 악의적인 서비스 제공자에 의한 개인정보 수집은 법·제도적 차원에서 제재가 이루어져야 한다. 이러한 제약사항에도 불구하고 개인정보 처리 보안가이드는 기술적인 차원에서 전반적인 개인정보 수집·저장 등에 대한 보안 수준을 향상시킬 수 있는 이점이 있다. 보다 안전한 개인정보 보호를 위해서는 법·제도적인 노력도 함께 이루어져야 한다.

웹 서비스는 다양한 분야로 확장되고 있으며, 이에 따라서 개인정보의 범위 및 종류도 계속 확장되고 있다. 웹 서비스를 통해 노출될 수 있는 개인정보의 종류와 개인정보 보호차원에서 중요한 웹 서비스·애플리케이션 성격을 분류하고, 사용자 개인정보를 전송 상황별로 제안하는 상세 가이드와 웹 애플리케이션에서의 불연계성 모델 등을 정립하는 추가적인 연구가 필요하다.

참 고 문 헌

[1] 국가정보원, “2008 국가정보보호백서, 제2편 제6장 개인정보보호 활동,” 2008년 4월.

[2] 국가정보원, “2009 국가정보보호백서, 제2편 제6장 개인정보보호 활동,” 2009년 4월.

[3] W3C, “Hypertext Transfer Protocol – HTTP/1.1,” RFC 2616, June 1999. http://www.w3.org/Protocols/rfc2616/rfc2616-sec9.html

[4] D. Stuttard and M. Pinto, The Web Application Hacker′s Handbook, Wiley Publishing, Inc., pp. 38-39, Oct. 2007.

[5] “Apache HTTP Server Log Files,” http://httpd.apache.org/docs/2.2/logs.html

[6] “W3C Extended Log File Format,” http://technet.microsoft.com/en-us/library/cc786596(WS.10).aspx

[7] 정보통신망 이용촉진 및 정보보호 등에 관한 법률(법률 제9637호) 제1장 제2조(정의) 제1항 제6호.

[8] 공공기관의 개인정보보호법에 관한 법률(법률 제8871호) 제1장 제2조(정의) 제2호.

[9] 김현수, “정보화와 개인정보보호의 현황 및 과제,” 국민윤리연구, 제63호, p. 177, 2006년 5월.

[10] “Personal Identifiable Information,” http://en.wikipedia.org/wiki/Personally_identifiable_information

[11] 정보통신부, “정보시스템 운영 보안-로그-개발가이드,” 정보통신부고시 제2006-37호, 2006년 9월.

[12] 정보통신망 이용촉진 및 정보보호 등에 관한 법률 제4장 제2절 28조(개인정보의 보호조치).

[13] W3C, “RFC 2612 – 15 Security Considerations,” http://www.w3.org/Protocols/rfc2616/rfc2616-sec15.html

<著者紹介>

여 성 구 (SungKoo Ryeo) 학생회원
2003년 2월: 울산대학교 경영학과·전자계산학과 (경영학사·공학사)
2003년 4월 ~ 2004년 7월: STG시큐리티 보안컨설턴트
2004년 7월 ~ 2008년 11월: 안철수연구소 보안컨설턴트
2008년 3월~현재: 고려대학교 경영정보공학전문대학원 석사과정
<관심분야> 정보보호, 해킹, 디지털 포렌식

심 미 나 (MiNa Shim) 정회원
1996년 2월: 성신여자대학교 전산학과 (이학사)
2006년 2월: 고려대학교 정보보호대학원 (공학석사)
2008년 2월: 고려대학교 정보경영공학전문대학원 (박사수료)
2008년 3월 ~ 현재: 고려대학교 정보보호기술연구센터 연구원
<관심분야> 정보보호정책, 프라이버시, 개인정보보호, 위험관리, 개인정보위험평가

이 상 진 (SangJin Lee) 종신회원
1987년 2월: 고려대학교 학사 졸업
1989년 2월: 고려대학교 석사 졸업
1994년 8월: 고려대학교 박사 졸업
1989년 10월 ~ 1999년 2월: ETRI 연구원 역임
1999년 3월 ~ 2001년 8월: 고려대학교 자연과학대학 조교수
2001년 9월 ~ 현재: 고려대학교 정보경영공학전문대학원 교수
<관심분야> 디지털 포렌식, 모바일 포렌식, 심층 암호, 해쉬 함수