5

개인정보 저장 형태에 따른 유출 탐지 방안

고려대학교 ▌한승원 · 이상진
한국인터넷진흥원 ▌이강신 · 차윤호

1. 서 론

개인정보는 국내법에서 "2008년 12월 14일부터 시행한 대한민국 정보통신망 이용촉진 및 정보보호 등에 관한 법률 제2조 6항"에 따르면 다음과 같이 정의하고 있다[1]. 개인 정보란 생존하는 개인에 관한 정보로서 당해 정보에 포함되어 있는 성명 · 주민등록번호 등의 사항에 대하여 당해 개인을 식별할 수 있는 부호 · 문자 · 음성 · 음향 및 영상 등의 정보(당해 정보만으로도 특정 개인을 식별할 수 없더라도 다른 정보와 용이하게 결합되어 식별할 수 있는 것을 포함한다)이다.

최근 GS 칼텍스, 옥션, LG 텔레콤, 하나로 텔레콤 등에서 대량의 개인정보 유출 사건이 발생함에 따라 개인정보 유출에 대한 우려의 목소리가 높아지고 있다. 개인정보를 다루는 단체나 기업에서는 개인정보 유출 탐지 및 차단 시스템을 도입하거나 대응 방안을 마련하고 있다. 그러나 많은 영세한 업체의 경우 개인정보 유출에 대해 적절한 대응을 하지 못하고 있다. 이와 같은 현실을 감안하면 사회적으로 문제가 되는 개인정보의 유출은 국가적인 차원의 대응이 필요하다.

개인정보 유출의 경우 내부자에 의한 개인정보 유출과 외부의 공격에 의한 유출로 크게 나눌 수 있는데 우리는 외부에서 웹 공격을 통해 개인정보가 유출되는 경우에 초점을 맞추어 개인정보 유출 탐지 방안을 모색하고자 한다. 그리고 기존에는 웹 서버 공격이나 웹 어플리케이션 공격을 탐지하는 것과 다르게 개인정보가 유출되는 경우에 신속히 탐지하여 대응하는 아웃-바운드 탐지의 접근 방법을 사용한다.

본 논문은 국가차원의 개인정보 유출을 효과적으로 대응하기 위한 기반 연구로써 개인정보의 저장 형태를 조사하고 유출되는 파일 포맷을 분석하고 개인정보 유출 탐지 시스템의 네트워크 구성에 대해 조사하였다.

본 논문의 구성은 다음과 같다. 2장에서는 시중에 유통되고 있는 개인정보 유출 탐지 시스템의 탐지 방식과 구조를 조사하고 3장에서는 웹 공격을 통해 유출되는 개인정보의 형태를 조사하고 분석하였다. 4장에서는 국가차원의 효율적인 탐지를 위한 탐지 구성에 대해 제안한다. 마지막 5장에서는 결론 및 향후연구에 대해서 논한다.

2. 관련연구

개인정보 유출 탐지 시스템은 크게 웹 페이지의 게시물이나 문서 파일 형태의 개인정보 유출 탐지 시스템과 데이터베이스 유출 탐지 시스템으로 나눌 수 있다. 두 가지 형태의 개인정보 유출 탐지 시스템을 조사하고 기능을 비교 분석하였다.

2.1 문서 파일 및 텍스트 형태의 유출 방지 시스템

현재 시중에 나와 있는 문서 파일 및 텍스트 형태 개인정보 유출 방지시스템은 표 1과 같이 크게 개인정보 유출 차단과 탐지 기능을 제공하고 있다. 차단

표 1 파일 및 텍스트 형태의 개인정보 유출 방지 시스템 비교

기능	펜타 시큐리티	위즈 디엔에스	컴트루 테크놀로지	엑스큐어넷
차단 기능				
게시물 차단	×	○	○	×
키워드 차단	-	○	○	○
첨부파일 차단	-	×	○	○
탐지패턴 설정	-	○	×	×
탐지 기능				
게시물 스캔기능	×	○	○	×
웹페이지 대한 검색 기능	×	○	○	×
기타				
웹 공격 탐지	○	×	×	×
비정상적인 트래픽 탐지	○	×	○	×

기능은 첨부파일을 포함한 웹 페이지에 대해 특정 키워드 검색을 이용하여 개인정보의 업로드를 차단하는 것을 말한다. 그리고 탐지 기능은 현재의 웹 페이지에 대해서 개인정보 패턴을 스캔하여 개인정보의 노출 가능성을 탐지하고 제거하기 위해 사용된다. 그 외에 웹 공격 패턴 기반의 탐지와 비정상 트래픽 기반의 탐지 기능을 제공한다[2-5].

기존의 개인정보 유출 탐지 시스템의 구성은 단일 조직의 망에서 개인정보의 유출을 탐지하는 구성을 가지고 개인정보가 포함된 파일의 업로드를 금지하거나 개인정보 노출을 탐지하는 방식이다.

웹을 통한 개인정보 유출에 대한 국가 차원의 대응에서는 기존의 웹 게시물 업로드 차단과 웹 공격의 탐지 방법은 많은 부하와 오탐이 발생하므로 적합하지 않다.

본 논문에서는 개인정보의 저장 형태를 분석하여 네트워크 병목 구간에서 개인정보의 유출을 탐지한다. 그렇기 때문에 수많은 웹서버를 대상으로 개인정보 유출을 탐지할 수 있다. 그리고 실제 개인정보 유출에 초점을 맞추어 탐지하기 때문에 부하가 적고 웹 공격을 탐지하는 경우보다 오탐의 발생률이 현저히 낮아지는 이점이 있다.

2.2 데이터베이스 정보 유출 방지 시스템

데이터베이스 정보 유출 방지시스템은 데이터베이스의 접근을 제어하는 방식과 데이터베이스 자체를 암호화하는 방식으로 나눌 수 있다. 또한 최근에는 사용자의 데이터베이스 접근과 작업에 대한 로그 정보를 저장하고, 저장된 로그를 분석 및 감사하는 기능도 포함하고 있다[2,6-8]. 데이터베이스 정보 유출 방지 솔루션은 표 2와 같다.

데이터베이스 정보 유출 방지 시스템은 데이터베이스 서버 앞단에 독립적으로 설치되어 클라이언트와 서버의 환경을 변경하지 않으면서 데이터베이스의 유출을 모니터링 한다. 그 외에도 세션 모니터링과 같은 다양한 기능을 제공하고 있다.

국가 차원의 개인정보 데이터베이스 유출을 탐지하기 위한 방법으로 기존의 접근 제어와 방어적인 방식은 적합하지 않다. 본 논문에서는 실제 데이터베이스의 유출 형태를 분석하여 데이터베이스가 유출되는 것을 네트워크 단에서 탐지하는 방안을 제시한다.

3. 개인정보 저장 형태 분석

개인정보는 평문, 데이터베이스, 복합 문서, OOXML 형태로 저장된다. 네트워크를 통한 개인정보 유출을 탐지하기 위해서는 개인정보가 저장되는 형태를 분석할 필요가 있다. 본장에서는 개인정보 유출을 효율적으로 탐지하기 위해서 개인정보가 저장되고 유출되는 데이터 형태를 분석한다.

3.1 평문 형태

평문 형태로 개인정보가 유출되는 경우는 개인정보가 기록된 텍스트 형식의 파일이 유출되거나 SQL Injection과 같은 공격을 통해 네트워크상에 평문 데이터가 전송되는 경우이다.

평문 형태의 개인정보는 텍스트 인코딩 방식에 따라 ASCII나 Unicode로 표현되는데 전송 계층에서도 같은 방식으로 전송된다. 따라서 전송 계층의 세그먼트를 확인하여 개인정보를 탐지할 수 있다. 하지만 응용 계층의 데이터가 세그먼트 단위로 분할되면서 개인정보 데이터가 분할되는 경우가 발생할 수 있다. 이러한 경우에는 그림 1과 같이 두 개의 세그먼트를 연속적으로 탐지하는 2-gram 분석 방법을 사용하여 탐지가 가능하다.

표 2 데이터베이스 정보 유출 방지 솔루션 비교

기능	웨어밸리	피앤피시큐어	신시웨이	펜타시큐리티
접근제어				
DB 접근 감시	○	○	○	○
작업 정보 감시	○	○	○	○
보안정책 접근 통제	○	○	○	○
컬럼 단위 접근 제어	×	×	×	○
암호화				
중요 데이터 암호화	×	×	×	○
컬럼 단위 암호화	×	×	×	○
감사				
데이터베이스(SQL) 로깅 및 감시 정책 관리	○	○	○	○
서버 로깅 및 감시정책 관리	○	○	○	×

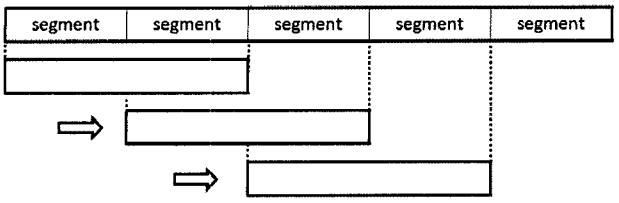

그림 1 세그먼트의 2-gram 분석 방법

3.2 데이터베이스 덤프 파일

다량의 개인정보를 다루는 기업, 기관 및 단체에서는 데이터 관리를 효과적으로 하기 위해 데이터베이스를 사용한다. 이러한 데이터베이스는 저장된 데이터의 백업 목적으로 덤프 기능을 지원한다. 덤프 기능을 사용하여 공격자가 의도적으로 덤프 파일을 생성한 경우나 백업 목적으로 생성한 덤프 파일이 적절히 관리되지 않을 경우 외부로 유출될 가능성이 있다.

데이터베이스 덤프 파일에서 개인정보를 탐지하기 위해서는 덤프 파일에서 개인정보가 저장되는 형식을 알아야 한다. 현업에서 많이 사용하는 6가지 데이터베이스를 선정하였고, 개인정보의 저장용도로 많이 활용되는 문자열 데이터타입[9-14]을 추출하였다. 그리고 데이터베이스에 개인정보를 입력한 후, 덤프 파일을 생성하여 저장형식을 살펴보았다. 표 3은 데이터베이스 덤프 파일별로 문자열 데이터타입의 저장형식을 보여준다.

표 3 덤프 파일 내부에 문자열 데이터타입 저장형식

데이터베이스	데이터타입	저장형식
Oracle 11g	char	ASCII
	varchar	
	nchar	
	nvarchar	
MS-SQL Server 20505	char	ASCII
	varchar	
	text	
	nchar	Unicode
	nvarchar	
	ntext	
MySQL 5.1	char	ASCII
	varchar	
	tinytext	
	text	
	mediumtext	
	longtext	
DB2 9.5	char	ASCII
	varchar	
Sybase ASE 15.0	char	ASCII
	varchar	
	text	
	nchar	
	nvarchar	
PostgreSQL 8.3	char	ASCII
	varchar	
	text	

결과에서 확인할 수 있듯이 별도의 암호화나 인코딩 방식을 사용하지 않고 ASCII나 Unicode 형식으로 저장되는 것을 확인할 수 있다. 이러한 경우에는 앞서 언급한 2-gram 분석 방법을 사용하여 개인정보를 탐지할 수 있다.

데이터베이스 덤프 파일에 대해 좀 더 효과적으로 탐지하기 위해서는 데이터베이스 덤프 파일을 특정할 수 있어야 한다. 덤프 파일은 각 데이터베이스 별로 고유한 시그니처(Signature)를 가지고 있다. 표 4는 각 데이터베이스 덤프 파일의 시그니처를 보여준다.

전송 계층의 세그먼트에서 알려진 데이터베이스 시그니처가 탐지될 경우 해당 세션을 통해 전달되는 세그먼트들은 덤프 파일의 데이터일 가능성이 높다. 따라서 전송 계층으로 전달되는 많은 세그먼트들에서 데이터베이스 덤프 파일의 세그먼트들만 선택적으로 탐지할 수 있다. 이러한 선택적인 탐지 방안은 탐지 시스템의 부하를 줄일 수 있다.

3.3 복합 문서 파일

복합 문서(Compound Document)는 하나의 통합된 지각 환경을 이루는 사용자 인터페이스의 조직화된 모음으로 텍스트, 오디오, 동영상 등 서로 다른 데이터 형식을 포함할 수 있는 구조의 문서를 의미한다[15]. 복합 문서 형식을 사용하는 대표적인 응용프로그램은 Microsoft Office 97-2003, Haansoft 흔글 2000 이상 등으로 데이터를 저장하는 문서 파일의 대부분이 복합 문서 형식을 사용한다. 따라서 데이터베이스가 아닌 파일로 개인정보를 기록하여 활용할 경우 복합 문서에 개인정보가 기록될 가능성이 매우 높다.

복합 문서 파일의 구조는 운영체제에서 사용하는 파일 시스템(예: FAT 등)과 유사하다[16-8]. 복합 문

표 4 데이터베이스 덤프 파일의 시그니처

데이터베이스	시그니처
Oracle 11g	"..NEXPORT:"
MS-SQL Server 2005	"TAPE"
MySQL 5.1	"-- MySQL dump"
DB2 9.5	"SQLUBRMEDHEAD"
Sybase ASE 15.0	"PGDMP" or "toc.dat"
PostgreSQL 8.3	"VOL1"

표 5 복합 문서와 파일 시스템 비교

구분	파일 시스템	복합 문서
메타데이터	부트 섹터	헤더 블록
저장 방식	클러스터 할당으로 파일저장	블록 할당으로 스트림 저장
데이터 단위	클러스터	블록, 소형 블록
데이터 형식	폴더 및 파일	스토리지, 스트림
데이터 형식 정보	MFT	디렉터리 개체
할당 정보 관리	파일 할당 표	블록 할당 표
포함관계	트리 구조	트리 구조
최상위 저장소	루트 디렉터리	루트 디렉터리
하위 저장소	디렉터리	스토리지

서 파일은 스토리지와 스트림의 계층 구조로 구성되며, 이들을 관리하기 위한 메타데이터가 존재한다[17,18]. 메타데이터란 사용자가 입력한 데이터를 설명하거나 관리하기 위한 데이터로 응용 프로그램에서 자동적으로 생성된다. 표 5는 복합 문서와 파일 시스템을 비교한 것이다.

복합 문서에서 개인정보를 탐지하기 위해서는 복합 문서의 구조를 알아야 한다. 그림 2는 복합 문서 파일의 추상화된 구조를 나타내는데, 루트(Root), 스토리지(Storage)와 스트림(Stream)의 트리(tree) 형태로 데이터를 저장한다. 최상위 저장소는 루트이며, 루트는 스트림 또는 스토리지를 가질 수 있다. 스토리지는 파일시스템의 디렉터리와 같이 스트림을 저장하는 역할을 수행하고, 스트림은 복합 문서에서 사용하는 문서구조, 글자 및 문단 형식, 속성, 본문 텍스트 등의 실제 데이터를 저장한다. 따라서 문서 종류별로 본문 텍스트를 저장하는 특정한 스트림이 존재하는데 해당 스트림을 확인해서 본문에 저장된 개인정보를 탐지할 수 있다.

표 6은 복합 문서 종류에 따라 본문 텍스트가 저장되는 스트림과 저장형식을 확인한 것이다. 결과에서 확인할 수 있듯이 복합 문서의 본문 텍스트는 ASCII나 Unicode로 저장되는 것을 확인할 수 있다. 따라서 복합 문서가 외부로 유출될 경우 본문 텍스트가 저

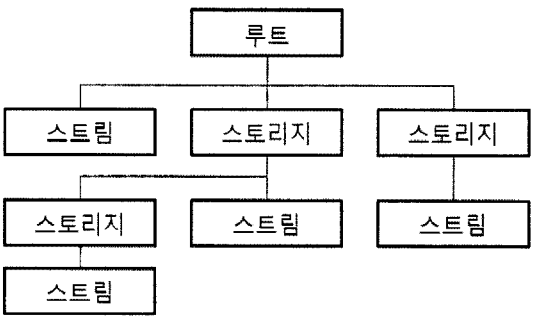

그림 2 복합 문서 파일 구조

표 6 복합 문서에서 본문 텍스트 저장형식

복합 문서 종류	스트림	저장형식
Microsoft Word 97-2003	WordDocument	ASCII 또는 Unicode
Microsoft Excel 97-2003	Workbook	ASCII 또는 Unicode
한글 2002-2007	SectionN(N=0 .. N)	Unicode

장되는 스트림 영역에 대해 2-gram 분석 방법을 사용하여 개인정보를 탐지할 수 있다. 단, 한글 파일의 경우에는 기본적으로 본문 텍스트 영역이 DEFLATE 알고리즘으로 압축되어 저장된다. DEFLATE 압축 방식은 압축된 전체 데이터가 없이도 순차적으로 압축 데이터를 해제할 수 있다[19]. 따라서 압축이 해제될 때마다 본문 텍스트 영역을 확인하여 개인정보의 포함 여부를 탐지할 수 있다. 이는 응용 계층에서 사용하는 전체 데이터가 조합되지 않은 상태에서 빠르게 탐지가 가능하다.

복합 문서 파일로 유출되는 개인정보를 좀 더 효과적으로 탐지하기 위해서는 복합 문서를 특정할 필요가 있다. 복합 문서에도 복합 문서를 판별할 수 있는 고유한 시그니처가 존재한다. 복합 문서의 시그니처는 "0xD0CF11E0A1B11AE1" 값을 가진다[15]. 따라서 복합 문서의 시그니처가 탐지된 세션의 세그먼트들은 복합 문서일 가능성이 높기 때문에 해당 세션 세그먼트를 대상으로 탐지할 수 있다. 이는 복합 문서에 해당하는 세그먼트만 선택적으로 탐지할 수 있기 때문에 전체 세그먼트를 대상으로 하는 것보다 효과적이다.

3.4 OOXML 파일

OOXML(Office Open XML)은 복합 문서 파일 형식을 대체하는 새로운 표준 문서 파일 형식이다. OOXML은 문서의 내용을 XML 형식으로 표현하여 구조적으로 기록한다. 기록한 이후 파일의 크기를 줄이기 위

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	0123456789ABCDEF
50	4B	03	04	14	00	06	00	08	00	00	00	21	00	2B	75	PK.....!.+u
AD	8C	94	01	00	00	AC	06	00	00	13	00	08	02	5B	43	.....[C
6F	6E	74	65	6E	74	5F	54	79	70	65	73	5D	2E	78	6D	ontent_Types].xm
6C	20	A2	04	02	28	A0	00	02	00	00	00	00	00	00	00	l ...(.....
00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....

그림 3 OOXML의 헤더 형식

해 DEFLATE 압축 알고리즘을 사용한 ZIP 아카이브(Archive)형태로 압축하여 저장한다[19]. OOXML 형식은 현재 Microsoft Office 2007에서 사용되고 있기 때문에 해당 응용프로그램을 사용하는 경우 OOXML 파일에 개인정보가 기록될 가능성이 매우 높다.

OOXML 파일에서 개인정보를 탐지하기 위해서는 먼저 압축된 데이터를 압축 해제해야 한다. 압축을 해제하면 구조화된 XML 형식의 데이터가 나타나는데 본문 텍스트를 확인하면 ASCII나 Unicode를 사용해 표현한 것을 확인할 수 있다. 이 경우 2-gram 분석 방법을 사용하여 개인정보 유무를 판별할 수 있다. 그리고 전체 데이터 없이 순차적으로 압축 해제할 수 있는 DEFLATE 방식이므로 전송 계층의 세그먼트를 대상으로 실시간으로 압축을 해제하면서 본문 텍스트를 확인하면 효과적으로 탐지가 가능하다[20].

OOXML 파일에서 개인정보를 좀 더 효과적으로 탐지하기 위해서는 OOXML 파일을 특정할 필요가 있다. OOXML 파일은 ZIP 파일의 시그니처인 "0x04034B50"를 가진다[4]. 하지만 해당 시그니처는 실제 ZIP 압축 파일의 시그니처에도 사용되므로 좀 더 정확한 탐지를 위해서는 파일 헤더의 파일 이름을 확인해야 한다. OOXML 파일의 경우에는 ZIP 파일 헤더의 파일 이름으로 그림 3과 같이 "[Content_Types].xml"을 갖는다. 이처럼 OOXML 파일을 특정하게 되면 해당 파일이 전송되는 세션 세그먼트만을 대상으로 탐지를 수행할 수 있다는 장점이 있다.

4. 국가 차원의 개인정보 유출 탐지 구성도

지금까지 개인정보가 저장되는 데이터 형태에 대해서 알아보았다. 개인정보가 저장되는 데이터의 분석 정보를 바탕으로 네트워크 레벨에서 개인정보 유출을 탐지할 수 있다. 본 장에서는 다수의 웹 서버를 대상으로 개인정보 유출을 탐지하기 위한 네트워크 구성에 대해서 알아본다. 탐지 시스템의 구성은 크게 시스템 기반의 탐지와 네트워크 기반의 탐지로 나눌 수 있는데 본 논문에서는 네트워크 기반의 탐지를 중심으로 한다. 국가 차원의 개인정보 유출 탐지 구성을 게이트웨이 방식의 탐지와 역 프록시 방식 탐지로 나누어 알아본다.

4.1 게이트웨이 방식의 탐지 구성

게이트웨이 방식의 탐지 시스템 구성은 다수의 웹 서버의 병목구간인 ISP 단에 탐지 시스템을 설치하는 방법을 말한다. 이와 같은 구성을 통해서 다수의 웹 서버에 대한 네트워크 패킷을 모니터링 할 수 있고 개인정보 유출을 탐지할 수 있다. 게이트웨이 방식의 탐지 구성은 그림 4와 같다.

그림 4와 같이 다수의 웹 서버가 모이는 ISP의 병목구간에 개인정보 유출 탐지 시스템을 설치하여 아웃-바운드 패킷을 모니터링하고 개인정보의 패턴을 신속히 탐지하고 대응할 수 있다.

게이트웨이 방식의 탐지시스템을 구성할 경우 몇 가지 고려해야 할 사항이 있다. 우선 ISP의 네트워크 장비에 탐지시스템을 구성해야 하기 때문에 ISP의 협조가 전재되어야 한다. 그리고 탐지 대상이 되는 웹서버의 IP를 확인하여 해당 아이피에 대해서만 탐지를 수행해야 한다. ISP에서 네트워크 트래픽을 모니터링하는 것은 시스템의 과부하가 발생할 수 있는데 해당 도메인의 트래픽 량을 추정하고 적절한 탐지 시스템을 구성해야 한다. ISP에서 한 대의 시스템으로 대량

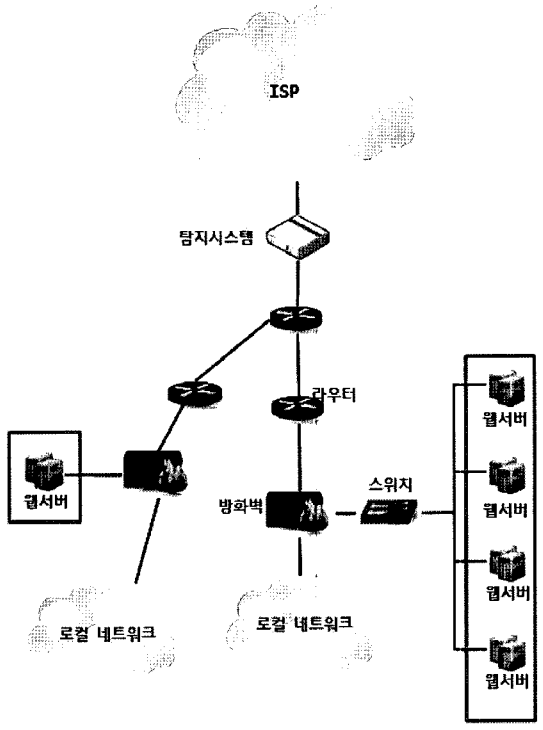

그림 4 ISP 게이트웨이에 위치한 탐지시스템

의 패킷을 분석해야 하기 때문에 네트워크 프로세서(Network Processor 또는 Network Processor Unit)와 같은 고성능 네트워크 카드 사용을 고려해야 한다. 네트워크 프로세서는 병렬처리와 멀티쓰레딩이 가능하기 때문에 높은 대역폭을 통해 이동하는 다량의 패킷 등을 정체 없이 처리할 수 있다.

4.2 역 프록시(Reverse Proxy) 방식의 탐지 구성

프록시 방식의 구성은 일반 프록시와 역 프록시로 나눌 수 있다. 일반 프록시 방식은 웹 브라우저에서 프록시 서버의 IP를 설정한 클라이언트가 불특정 다수의 웹 서버를 접근할 경우 프록시 서비스를 제공받을 수 있다. 반면 역 프록시 방식의 경우는 불특정 다수의 클라이언트에서 한정된 웹 서버로의 프록시 서비스를 제공할 수 있다. 역 프록시 방식의 탐지 구성을 이용함으로써 모든 인터넷 사용자에 대해서 지정한 웹 서버의 개인정보 유출을 탐지할 수 있다.

역 프록시를 이용한 국가차원의 개인정보 유출 탐지 구성은 그림 5와 같다. 그림과 같이 클라이언트는 웹 서버에 접속을 요청하면 맨 처음 클라이언트는 웹 서버를 찾기 위해 DNS에 DNS Query를 보낸다. 사전에 DNS에는 관리 대상의 웹 서버의 도메인에 대한 아이피가 탐지 시스템을 경유하는 프록시 서버의 아이피로 설정을 한다. 프록시 요청을 받은 DNS는 실제 웹 서버의 주소가 아닌 역 프록시 서버의 아이피 주소를 응답하게 된다. DNS로부터 응답을 받은 클라이언트는 DNS에 등록된 프록시 서버로 접근하게 되고 역 프록시 서버는 요청된 컨텐츠가 캐쉬에 있는지 검사하고 만약 그렇지 않으면 실제 웹 서버로 연결하고 디스크 캐쉬에 요청된 컨텐츠를 다운로드 한다. 프록시 서버의 캐쉬 용량을 적게 함으로써 캐쉬의 기능보다 트래픽을 우회하는 기능으로 사용할 수 있다.

결과적으로 클라이언트가 요청한 패킷은 프록시 서버 앞단에 위치한 탐지 서버를 통과하게 된다. 따라서 프록시 서버와 DNS의 설정 변경으로 여러 대의 웹 서버에서 개인정보의 유출을 효율적으로 탐지할 수 있다.

게이트웨이 방식의 개인정보 유출 탐지 구성은 ISP의 장비에 탐지 시스템을 구성해야 하기 때문에 ISP의 협조가 필요하다. 그러나 역 프록시 방식의 개인정보 유출 탐지 구성은 탐지해야 하는 웹서버가 모이는 네트워크 도메인의 특정 위치에 시스템을 설치하여 트래픽을 우회시키는 방법으로 ISP 네트워크 구성에 영향을 받지 않는 이점이 있다.

네트워크 기반의 장점은 특정 호스트에 설치되어 해당 시스템만 보호하는 호스트 기반 탐지 시스템과

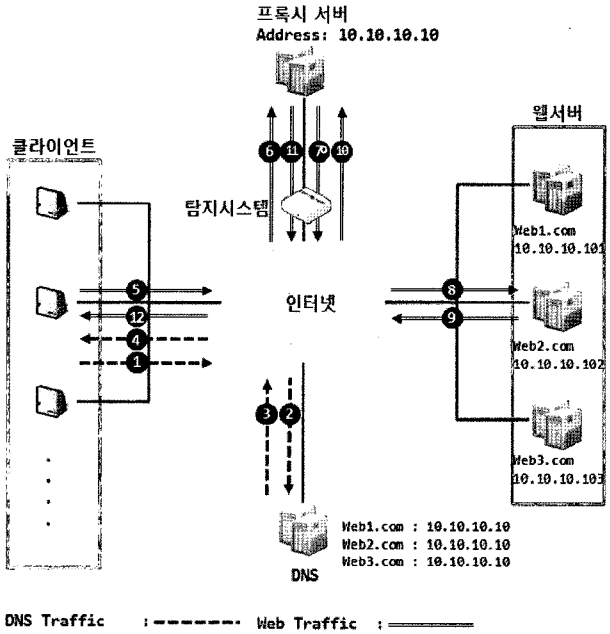

그림 5 역 프록시 기반의 개인정보 유출 탐지 구성도

달리 하나의 시스템으로 여러 대의 웹 서버를 보호할 수 있다. 호스트 기반의 경우 운영체제에 맞춰 설치해야 하는 것에 비해 특정 네트워크의 입구가 되는 게이트웨이에 설치되어 그 네트워크에 해당하는 웹 서버를 보호할 수 있다.

네트워크 기반 탐지 시스템은 보호하기 위한 웹 서버들과는 별도로 설치되기 때문에 웹 서버 자체의 성능에 영향을 미치지 않는다. 하지만 네트워크 구성이 추가되는 경우 탐지 시스템을 구입해야 하기 때문에 추가비용이 발생한다. 또한 탐지시스템에 오류가 발생 했을 때 그 이하에 있는 모든 웹 서버에 대한 장애 우려가 있다. 그렇기 때문에 네트워크 기반의 탐지 시스템을 구성할 때는 네트워크 환경과 탐지 대수, 네트워크 대역폭, 시스템의 성능 등을 고려해야 한다.

5. 결 론

최근 웹 서버 해킹을 통한 개인 정보의 유출에 대한 우려의 목소리가 높아지고 있다. 규모가 큰 조직의 경우에는 웹 서버의 보안이 강하게 적용되어 있어 개인정보 유출 사고에 대비할 있지만, 작은 규모의 영세한 중소기업에서 운영하는 웹 서버는 형편이 다르다. 대량의 개인정보가 저장된 웹 서버의 보안 관리가 미흡하여 개인정보가 유출되는 사고가 종종 발생한다. 이런 영세한 기업을 대상으로 국가 차원의 개인정보 유출 대응이 필요하다.

개인정보 유출 탐지에 대한 연구에 앞서 개인정보 유출 방지 시스템의 기술 동향을 조사하였다. 이를 통

해서 개인정보의 저장과 유출되는 형태를 두 가지로 요약할 수 있는데 하나는 데이터베이스내의 개인정보이고 다른 하나는 문서파일내의 개인정보이다.

네트워크 레벨에서 개인정보 유출의 탐지 가능성에 대한 확인을 위해 데이터베이스내의 개인정보와 문서파일내의 개인정보가 어떤 형태로 저장되는지 분석하였다. 대중적으로 많이 사용되는 6개의 데이터베이스를 선정하여 개인정보가 포함된 데이터베이스를 생성하였다. 데이터베이스가 유출되는 경우 백업파일 형태로 유출되기 때문에 데이터를 백업하여 개인정보의 데이터 저장 형태를 분석하였다. 대부분의 개인정보는 아스키나 유니코드로 저장되기 때문에 네트워크 패킷에서 개인정보를 탐지할 수 있다. 그리고 문서파일의 경우 문서파일 내의 개인정보 저장형태를 분석하기 위해 한글과 MS Office 엑셀, 워드(97-2003, 2007)를 대상으로 하였다. MS Office 97-2003의 엑셀과 워드 문서는 개인정보가 아스키와 유니코드로 저장되기 때문에 네트워크 패킷에서 탐지가 가능한 반면 한글 파일과 MS Office 2007 엑셀, 워드파일의 경우에는 기본으로 압축되어 저장되기 때문에 네트워크 패킷에서 탐지할 수 없다. 그러므로 압축된 문서 파일은 별도로 패킷을 조합하여 텍스트를 추출하고 개인정보를 탐지하는 과정을 거쳐서 개인정보 유출 여부를 탐지해야 한다.

다수의 분산된 웹 서버의 개인정보 유출을 효율적으로 탐지하기 위해 네트워크 구성을 조사하였다. 다수의 웹 서버가 모이는 ISP의 병목 구간에 탐지 시스템을 구성하는 방법과 역 프록시를 기반으로 네트워크 구성의 변경 없이 탐지 시스템을 구성하는 방법에 대해서 알아보았다. 탐지 대상이 되는 웹 서버의 위치와 분포를 고려해서 구성 방법을 선택할 수 있고 이를 통해서 효율적으로 개인정보 유출을 탐지할 수 있다.

참고문헌

[1] 방송통신위원회, "대한민국 정보통신망 이용 촉진 및 정보보호 등에 관한 법률", 제2조 6항, 2008.

[2] 펜타시큐리티, URL : http://www.pentasecurity.com/

[3] 위즈디엔에스, URL : http://www.weeds.co.kr/

[4] 컴투루테크놀로지, URL : http://www.comtrue.com/

[5] 엑스큐어넷, URL : http://www.xcurenet.com/

[6] 웨어벨리, URL : http://www.warevalley.com/

[7] 피엔피시큐어, URL : http://www.pnpsecure.com/

[8] 신시웨이, URL : http://www.sinsiway.com/

[9] Oracle Datatype, URL : http://www.ss64.com/ora-syntax/datatypes.html, SS64

[10] MS-SQL Datatype, URL : http://webcoder.info/reference/MSSQLDataTypes.html, Webcoder

[11] MySQL Datatype, URL : http://dev.mysql.com/tech-resources/articles/visual-basic-datatypes.html, MySQL

[12] DB2 Datatype, URL : http://www.michael-thomas.com/tech/db2/db2_survival_guide.htm, Michael-thomas

[13] SYBASE Datatype, URL : http://www.sfu.ca/sas-doc/sashtml/accdb/z0439559.htm, SFU

[14] PostgreSQL Datatype, URL : http://www.postgresql.org/docs/7.4/interactive/datatype.html, PostgreSQL

[15] Daniel Rentz, "Microsoft Compound Document File Format", OpenOffice.org's Documentation, Aug 2007.

[16] Brian Carrier, File System Forensic Analysis, Addison Wesley, 2005.

[17] Daniel Rentz, OpenOffice.org's Documentation of the Microsoft Compound Document, The Spreadsheet Project, OpenOffice.org, 2007.

[18] Microsoft Corporation, Windows Compound Binary File Format Specification, Microsoft Corporation, 2008.

[19] Microsoft Corporation, "Standard ECMA-376 Office Open XML File Formats", 2nd Edition, Dec 2006.

[20] RFC 1951, "DEFLATE Compressed Data Format Specification version 1.3", May 1996.

한 승 원

2003 건양대학교 정보전산학과 학사
2008 고려대학교 정보경영공학전문대학원 정보경영공학과 석사수료
관심분야 : 디지털 포렌식, 악성코드 사고 조사, 봇넷 분석 등
E-mail : normalhan@korea.ac.kr

이 상 진

1987 고려대학교 수학과 이학사
1989 고려대학교 수학과 이학석사
1994 고려대학교 수학과 이학박사
1989~1999 ETRI 연구원 역임
1999~현재 고려대학교 정교수
관심분야 : 디지털 포렌식, 모바일 포렌식, 심층 암호, 해쉬 함수 등
E-mail : sangjin@korea.ac.kr

이 강 신

1989 한양대학교 수학과 이학석사
2005 고려대학교 정보보호대학원 공학박사
1990~1992 데이콤 종합연구소 연구원
1992~2000 한국전산원 정보화표준부장
2000~현재 한국인터넷진흥원 개인정보보호기획팀장
관심분야 : 개인정보보호, 네트워크보안, 정보보호아키텍처
E-mail : kslee@kisa.or.kr

차 윤 호

2002 전북대학교 컴퓨터공학과 석사
2002~현재 한국인터넷진흥원 선임연구원
관심분야 : 개인정보보호, 네트워크보안
E-mail : yhcha@kisa.or.kr