공격 트래픽의 효과적인 저장 및 분석 방안 연구

이기성*, 이동영**, 서희석***

요약

웹사이트 대상 공격이 증가함에 따라 웹사이트 보안에 대한 중요성이 부각되고 있으며, 이러한 공격정보를 수집 및 분석이 필요하다. 따라서 본 논문에서는 방대한 공격 트래픽을 효과적으로 저장 및 분석 방안을 연구하기 위해 Raw Traffic의 공격 유형별 저장방법, 공격 트래픽의 저장유형에 따른 분류방법으로 나뉜다. 차후 웹사이트 공격정보 수집 및 분석 모델 연구를 위해 활용하고자 할 경우 선행 연구 자료 및 향후 웹사이트 공격정보 수집 및 분석에 대한 지침서로 활용될 수 있을 것으로 기대된다.

A Study on The Methodology of Effective Storage and Analysis for Attack Traffic

Seol-Hee Lee*, Dong-Yung Lee**, Hee-Suk Seo***

ABSTRACT

As increasing the attacks for web site, the importance of security is highlighted. Also the need for collecting attack information and analysis method is increased. Therefore this paper that will be studied to storage and analysis methods of effective of enormous attack traffic. The kind of research is to methods of stored of attacks type of Raw Traffic and the classification according to the type to traffic of attacks. If you take advantage of this paper can be used as guidelines for collecting website attack information analysis in the future.

Keywords : Traffic, Attack, Analysis, Network, Storage, Sampling

* 한국기술교육대학교 인터넷미디어공학부(✉ lgomdori@kut.ac.kr)
** 명지전문대학 정보통신과
*** 한국기술교육대학교 컴퓨터공학부
· 제1저자(First Author) : 이기성 · 교신저자(Correspondent Author) : 서희석
· 접수일(2011년 2월 15일), 수정일(1차 : 2011년 3월 16일), 게재확정일(2011년 3월 21일)

Ⅰ. 서 론

최근 네트워크의 발전으로 인하여 사용자들은 인터넷이라는 기술을 이용하여 보다 다양하고 빠른 서비스를 이용하여 많은 정보를 손쉽게 접하고 활용할 수 있다. 그러나 이로 인해 사회 기반이 첨단화될수록 사이버 범죄 또한 빈번하게 발생하고 있다. 과거 공격 대상을 개별 시스템에 주안점을 뒀던 공격자들이 DDoS, 웹 공격, SQL Injection 등과 같은 웹사이트를 대상으로 공격 방법을 달리함에 따라 웹사이트의 보안이 중요한 이슈가 되고 있다.

이러한 다양한 웹사이트에 대한 공격들은 사회적으로 많은 혼란을 야기 시키고 있는데 대표적인 예로 2003년 1.25 인터넷 대란, 2009년 7.7 DDoS 공격, 최근에 발생한 DDoS 공격 등이 있다. 이러한 사이버 침해사고의 피해에 따른 사회적 파장이 커지면서 웹사이트 보안의 중요성 및 사이버 테러 대응방안의 필요성이 요구되고 있으며, 사이버 위협은 21세기 경제와 국가안보에 대한 가장 심각한 위협요인으로 등장하고 있다.

따라서 본 논문에서는 웹사이트에서 발생할 수 있는 공격 트래픽을 효과적으로 저장하고 분석할 수 있는 방안에 대하여 연구를 수행하였다.

웹 기반 공격은 웹을 기반으로 하는 악성 공격을 통칭하며, 네트워크 기반 공격은 네트워크를 기반으로 하여 네트워크 자원 고갈이나 스캐닝 등의 공격을 말한다. 이러한 웹 기반 공격과 네트워크 기반 공격에 대한 침입 탐지모델에 따라서 오용 침입 탐지와 이상 침입 탐지로 구분할 수 있다.

이러한 공격으로 인해 발생할 수 있는 공격 트래픽의 효과적인 저장 및 분석 방안을 위해 전수조사 방법, 트래픽 샘플링, 트래픽 요약정보(Flow)에 대하여 각각 분석하였다.

Ⅱ. 웹사이트 보안 동향

2009년 11월 한국인터넷진흥원의 인터넷 침해사고 동향 및 분석 월보에 따르면 국내 ISP의 일부구간(국내 인터넷망)에서 수집된 트래픽의 Top10포트의 추이를 파악한 결과, 이미 잘 알려진 TCP/80(HTTP), TCP/25(SMTP), TCP/443(HTTPS)등 외에

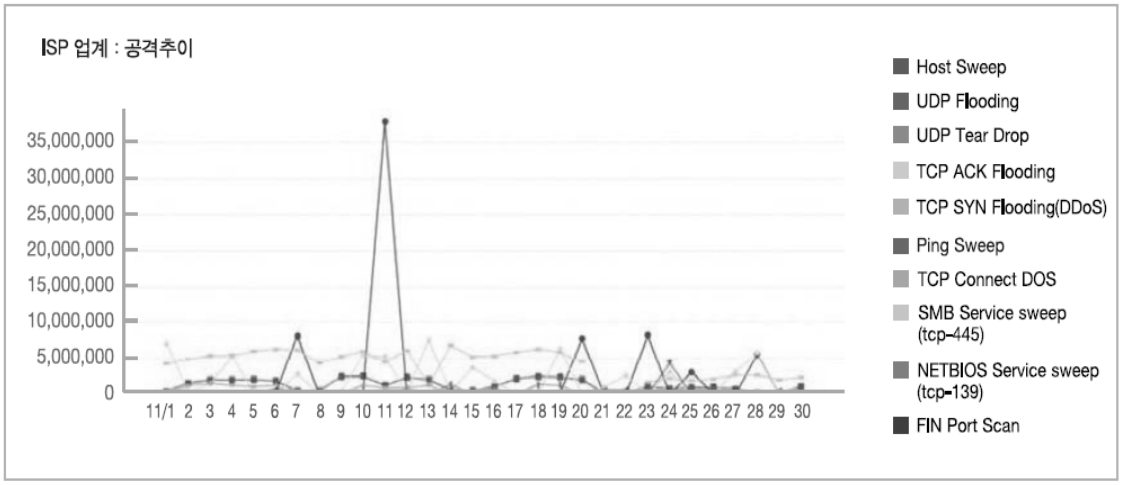

그림 1. 국내 인터넷망에 유입된 공격 유형
Fig. 1. The type of attacks flow in internet of South Korea

TCP/50001(commplex-link), TCP/2004(mailbox), TCP/5004(NateOn)포트 등에 대한 트래픽이 많이 관찰 되어 HTTP 서비스 외의 다른 서비스들도 공격에 대비한 분석과 방어 방법에 대한 분석이 필요하다.

위의 그림 1은 2009년 11월 한국인터넷진흥원의 인터넷 침해사고 동향 및 분석 월보결과를 참조한 자료이며 11월달 국내 ISP 일부구간에서 수집된 공격 유형을 분석한 결과, 실질적인 공격을 수행하기 전에 각종 정보를 수집하기 위한 Host Sweep 스캔 및 UDP Flooding, TCP ACK Flooding 등 DDoS 공격 트래픽 등이 많이 탐지 되었으며 중국으로 유입된 트래픽이 63.4%, 미국(12.0%), 대만(4.6%) 순으로 공격 트래픽이 많았다[1]. 또한 중국으로부터의 트래픽은 TCP/1433 및 TCP/2967 포트에 대한 서비스 스캔이 가장 많은 비중을 차지하였다.

아래의 표는 현재 가장 많이 사용되고 있는 웹사이트 공격 기법 중 일부분이다.

표 1. 웹사이트 공격 기법
Table 1. The technique of attack of website

공격기법	설명
Cross-site Scripting	공격자로 하여금 이용자가 방문하는 페이지에 스크립트를 심어서 페이지의 작동이나 내용을 조작
SQL Injection	웹사이트 입력란에 데이터베이스를 실행하는 SQL 문장을 입력하여 임의의 코드를 실행
File Include	PHP 애플리케이션에서 발생되는 취약점으로 애플리케이션이 코드를 검색한 후 이를 로컬 애플리케이션에서 실행할 때 발생
기타	서비스 불능화 공격과 공격자가 파일이나 디렉토리 또는 비인가된 이용자의 정보나 웹 애플리케이션의 요소들을 보거나 변경하는 기법

III. Raw Traffic의 공격 유형별 저장방법

Raw Traffic이란 네트워크를 통해 가공되지 않은 데이터의 량이나 어떤 종류의 트랜잭션 및 메시지를 말하며 본 논문에서는 웹 기반 공격과 네트워크 기반 공격으로 나뉘어, 각각의 공격 분석과 탐지 및 저장 방법에 대하여 연구하였다.

3.1 웹 기반 공격

웹 기반 공격 유형은 침입 탐지 모델에 의해서 오용과 이상 침입 탐지 모델로 분류된다[2]. 웹 기반의 오용 침입 탐지는 웹 사이트에 설치된 프로그램 내부의 디자인 결함이나 버그 또는 소프트웨어 시스템 내의 다른 구성요소들 사이의 예기치 못한 상호작용으로 인한 취약점들을 악용하는 공격을 탐지한다. 이러한 공격유형의 예로는 프로그램 내에서 함수의 리턴주소의 위치를 임의의 위치로 변경함으로써 권한을 얻게 되는 버퍼 오버플로우(Buffer Overflow)공격이나 포맷(format)문 인자의 값을 메모리에서 가져오는 취약점을 이용해 메모리 변조를 통한 불법적인 권한 획득 방법인 포맷 스트링(Format String), 버그를 갖고 있는 시스템 프로그램과 크래커(Cracker)의 취약 프로그램(Exploit Program)이 경쟁상태(Race Condition)에 이르게 하여 시스템 프로그램이 갖는 권한으로(Set-User ID가 붙은 경우 Root, Bin등) 파일에 대한 접근을 가능하게 하는 방법인 경쟁상태, 시스템의 루트권한을 획득한 뒤 차후의 침입을 위한 비밀 출입구인 백도어(Backdoor), 그리고 자체적인 복제능력은 없지만 시스템에서 몰래 정보를 빼돌리는 트로이목마 등의 공격을 들 수 있다.

웹 기반의 이상 침입 탐지는 웹 기반의 오용 침입 탐지와는 다르게 웹 사이트 내의 사용자 ID를 도용하거나 아직까지 알려지지 않았거나 알려진 공격에 대한 변형된 공격을 탐지한다. 이러한 공격으로는 취약점 분석이나 해커의 침입유형에 대해서 많은 지식을 가

진 관리자가 시스템에 남겨진 최소한의 정보만으로 침입확인을 할 수 있는 관리자 직관으로 판단이 가능한 공격, 다른 사용자의 ID나 패스워드를 도용해 마친 허가된 사용자인 것처럼 행동하는 위장(Impersonation), 그리고 지금까지는 알려지지 않았지만 새로운 소프트웨어를 사용한 공격이라든지 잘못된 명령어의 사용, 사용자의 실수에 의한 비의도적인 행위 등이 웹 기반의 이상 침입에 속한다.

- 웹기반 공격 분석

웹 기반 공격을 분석하기 위해서는 패킷의 페이로드까지의 분석을 통하여 공격을 분석할 수 있다. 공격 데이터의 페이로드를 분석하는 것은 해당 공격의 성격을 알아내는데 중요한 역할을 한다. 이렇게 분석한 정보를 이용하여서 정확한 방어 및 공격자의 정보를 얻어 낼 수 있다[3].

MS08-067 익스플로잇에 대한 페이로드를 살펴보면 아래 그림1 과 같다. 그림에서 보는 것과 같이 MS08-067 익스플로잇에 대한 페이로드를 살펴보면 \..\.. 패턴이 보이고 그 뒤에 쉘코드처럼 보이는 것이 있다. 이처럼 웹 기반 공격을 탐지하기 위해서는 패킷의 페이로드까지 분석을 해야 하며, 이러한 페이로드는 특정 공격 패턴을 가지고 공격을 시도한다[4].

- 탐지 시스템 및 저장 방법

해당 시스템은 크게 snort를 이용한 오용탐지 부분과 메시지 검증을 이용한 이상탐지 부분으로 오용탐지와 이상탐지 기법의 장점을 혼합하여 구성하였다.

```
00000000  00 00 00 00 cf 00 00 00  00 00 00 00 cf 00 00 00  |................|
00000010  5c 00 41 00 5c 00 2e 00  2e 00 5c 00 2e 00 2e 00  |\.A.\.....\.....|
00000020  5c 00 90 90 90 90 90 90  90 90 90 90 90 90 90 90  |\...............|
00000030  90 90 2b c9 83 e9 b8 d9  ee d9 74 24 f4 5b 81 73  |..+.......t$.[.s|
00000040  13 ba 1a cd 77 83 eb fc  e2 f4 46 70 26 3a 52 e3  |....w.....Fp&:R.|
00000050  32 88 45 7a 46 1b 9e 3e  46 32 86 91 b1 72 c2 1b  |2.EzF..>F2...r..|
00000060  22 fc f5 02 46 28 9a 1b  26 3e 31 2e 46 76 54 2b  |"...F(..&>1.FvT+|
00000070  0d ee 16 9e 0d 03 bd db  07 7a bb d8 26 83 81 4e  |.........z..&..N|
00000080  e9 5f cf ff 46 28 9e 1b  26 11 31 16 86 fc e5 06  |._..F(..&.1.....|
00000090  cc 9c b9 36 46 fe d6 3e  d1 16 79 2b 16 13 31 59  |...6F..>..y+..1Y|
000000a0  fd fc fa 16 46 07 a6 b7  46 37 b2 44 a5 f9 f4 14  |....F...F7.D....|
000000b0  21 27 45 cc ab 24 dc 72  fe 45 d2 6d be 45 e5 4e  |!'E..$.r.E.m.E.N|
000000c0  32 a7 d2 d1 20 8b 81 4a  32 a1 e5 93 28 11 3b f7  |2... ..J2...(.;.|
000000d0  c5 75 ef 70 cf 88 6a 72  14 7e 4f b7 9a 88 6c 49  |.u.p..jr.~O...lI|
000000e0  9e 24 e9 59 9e 34 e9 e5  1d 1f 87 c0 5e 35 dc 72  |.$.Y.4......^5.r|
000000f0  de e4 dc 49 44 96 2f 72  21 8e 10 7a 9a 88 6c 70  |...ID./r!..z..lp|
00000100  dd 26 ef e5 1d 11 d0 7e  ab 1f d9 77 a7 27 e3 33  |.&.....~...w.'.3|
00000110  01 fe 5d 70 89 fe 58 2b  0d 84 10 8f 44 8a 44 58  |..]p..X+....D.DX|
00000120  e0 89 f8 36 40 0d 82 b1  66 dc d2 68 33 c4 ac e5  |...6@...f..h3...|
00000130  b8 5f 45 cc 96 20 e8 4b  9c 26 d0 1b 9c 26 ef 4b  |._E.. .K.&...&.K|
00000140  32 a7 d2 b7 14 72 74 49  32 a1 d0 e5 32 40 45 ca  |2....rtI2...2@E.|
00000150  a5 90 c3 dc b4 88 cf 1e  32 a1 45 6d 31 88 6a 72  |........2.Em1.jr|
00000160  3d fd be 45 9e 88 6c e5  1d 77 41 41 41 41 41 41  |=..E..l..wAAAAAA|
00000170  41 41 41 41 41 41 41 41  41 41 41 41 41 41 41 41  |AAAAAAAAAAAAAAAA|
00000180  41 41 41 41 41 41 41 41  41 41 41 41 41 41 41 41  |AAAAAAAAAAAAAAAA|
00000190  41 41 41 41 41 41 41 41  41 41 41 41 41 41 41 41  |AAAAAAAAAAAAAAAA|
000001a0  41 41 41 41 41 41 41 41  41 41 cc 41 00 00 00 00  |AAAAAAAAAA.A....|
000001b0  01 00 00 00 02 00 00 00  00 00 00 00 02 00 00 00  |................|
000001c0  5c 00 00 00 01 00 00 00  01 00 00 00              |\...........|
```

그림 2. MS08-067 익스플로잇에 대한 페이로드 정보
Fig. 2. Payload information of MS08-067 exploit

외부 망으로부터 들어오는 모든 트래픽은 방화벽룰 정책을 적용하여 웹 어플리케이션 서비스 포트인 TCP 80/433port를 제외한 모든 트래픽을 Drop하고, 나머지 트래픽에 대하여 네트워크 인터페이스에서 Ipworks라는 패킷 캡처 라이브러리를 사용하여 수집한다. 패킷 캡처된 트래픽을 snort filter를 이용하여 패턴매칭을 실시한다.

아래 그림2는 웹 기반 침입탐지 시스템의 구조이다.

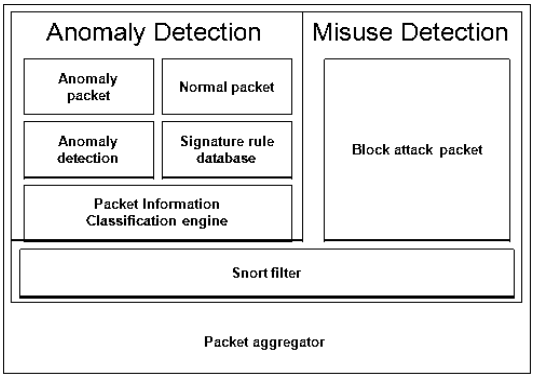

그림 3. 웹 기반 탐지 시스템 구조도
Fig. 3. The structure of Web-based detection system

3.2 네트워크 기반 공격

네트워크 기반 공격 유형 또한 침입 탐지 모델에 의해서 오용과 이상 침입 탐지 모델로 분류된다[5]. 네트워크 기반의 오용 침입 탐지는 네트워크를 지원하는 호스트의 서버 프로그램들의 설계 결함, 환경설정 결함 등의 취약점을 악용하거나 TCP/IP 프로토콜의 취약점을 이용하는 공격을 탐지할 수 있다. 이러한 공격의 형태로는 서비스를 제공하는 시스템에 대량의 서비스 요청을 보냄으로써 서비스 처리를 위해 할당된 메모리를 모두 낭비하게 해 더 이상의 서비스 제공을 불가능하게 하는 서비스 거부 공격(Denial of Service), 특정 사용자를 대상으로 대량의 메일을 보내는 메일 폭탄(Mail bomb), 자신의 IP를 속이기 위해 패킷의 근원지 주소를 타인의 IP로 패킷을 조작해서 상대방을 속이는 스푸핑(Spoofing), 공격을 위한 사전 작업으로 목적 시스템의 취약점을 찾으려는 시도로 목적 시스템의 포트를 차례대로 확인하는 방법인 탐침(Probing), 네트워크 카드를 무차별 모드(Promiscuous)로 변경해 네트워크에 흘러 다니는 모든 패킷을 분석해 정보를 얻어내려는 스니핑(Sniffing) 공격 등을 들 수 있다.

- 네트워크 기반 공격 분석

네트워크 기반 공격을 분석하기 위해서는 패킷을 분석함으로써 공격을 분석할 수 있다. 먼저 IP Fragmentation은 이기종 네트워크 환경에서 IP 패킷의 효율적인 전송을 보장해주고 있지만 몇 가지 보안 문제점을 가지고 있다.

Teardrop 공격은 Fragment의 재조합 과정 취약점을 이용한 서비스거부공격으로 두 번째 Fragment의 offset을 조작하여 Fragment들을 재조합하는 과정에서 버퍼를 넘쳐 겹쳐 쓰게 한다[6].

위와 같이 첫 번째 Fragment 크기는 36바이트인데 두 번째 Fragment의 offset이 24byte로 재조합된다. 그림 4는 아래의 패킷 정보를 통하여 Fragment 형식을 나타낸 그림이다. 따라서 아래의 패킷 정보를 분석한 결과 Fragment의 취약점을 이용한 Teardrop 공격임을 알 수 있다. 이처럼 네트워크 기반 공격을 탐지하기 위해서는 패킷 정보를 분석해야 하며, 이러한 패킷 정보를 통해 특정 공격을 탐지할 수 있다.

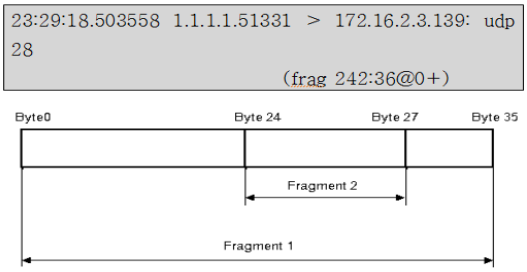

그림 4. Teardrop 공격의 Fragment 형식
Fig. 4. Fragment Format of Teardrop Attacks

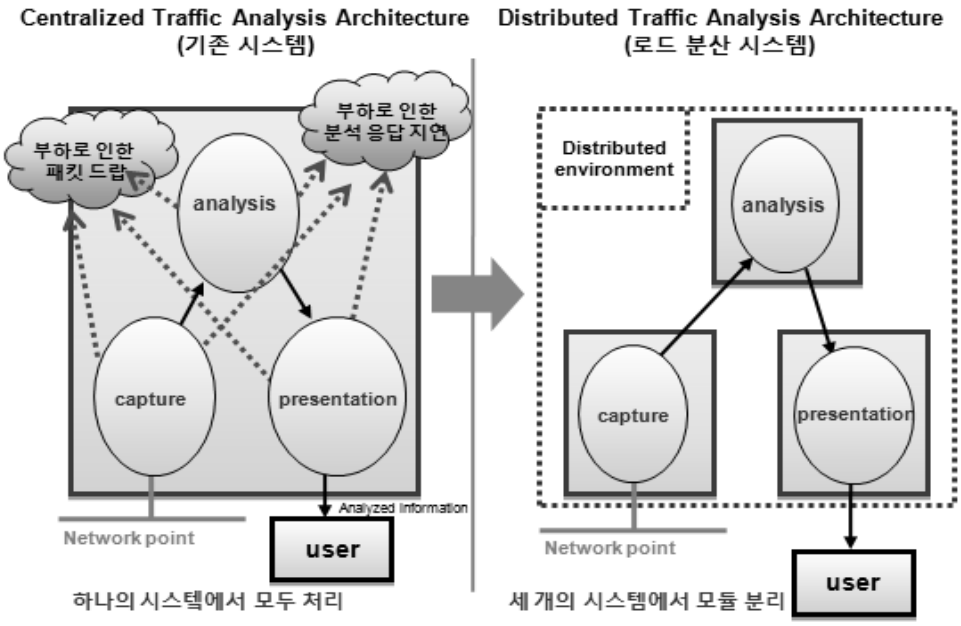

그림 5. WebTrafMon II에 적용한 로드 분산 방법
Fig. 5. Load balancing method to applied WebTrafMon II

- 탐지 시스템 및 저장 장법

기본적으로 WebTrafMon II는 아래 그림 4와 같이 로드 분산 기법을 적용하며 기존 시스템은 한 시스템에서 패킷 캡처 및 분석을 하므로 패킷 손실이 발생하고 패킷을 빠른 시간에 분석하기가 어렵다. 따라서 위의 로드 분산 시스템과 같이 각 모듈을 독립시키면 시스템 과부하로 인한 패킷 손실을 막고 빠른 분석을 할 수 있다. 또한 멀티미디어 포트 번호를 분석 시에 반영하면 멀티미디어 트래픽 분석이 가능하다[7].

Ⅳ. Raw Traffic의 공격 유형별 저장방법

4.1 전수저장

전수저장은 해당 웹사이트로 들어오는 모든 트래픽을 수집하여 분석 및 저장하는 방법이다.

전수저장은 트래픽 분석과 관리, 유해 트래픽과 공격성 트래픽의 탐지/방어를 위해 미러링이나 탭을 통해 트래픽을 저장한다. 하지만 전수저장은 ISP 전체 연동 네트워크를 모니터링 하고자 할 경우 트래픽 부하 및 하드웨어 요구가 많아 기하급수적인 비용이 발생하게 된다. 이러한 전수저장 방식은 패킷 기반 전수저장 방식과 플로우 기반 전수저장 방식으로 나눌 수 있다[8].

- 패킷 기반 전수 저장

패킷 기반 공격 탐지 방식을 이용하여 측정지점에서 흐르는 모든 트래픽을 수집하여 분석하고, 수집 및 분석에 소요되는 시간 및 저장장치의 부하가 가장 크다. 플로우는 일련의 동일한 특성을 가진 패킷의 집합으로 정의할 수 있다. 따라서 실시간성을 요하는 분석에는 활용이 힘들다. 하지만 모든 패킷을 수집함으로 가장 자세한 분석이 가능한 장점을 가진다.

표 2. 전수저장 방식의 비교·분석
Table 2. Comparison and analysis of a formula to store for all

구분	Web Log	Packet Sniffing	Script&Tagging
특징	· 웹 서버에 쌓인 로그기록을 분석프로그램으로 사후 분석	· 네트워크에서 주고받는 패킷에 담긴 정보를 Sniffing, 방문자의 트랜잭션 수집	· 웹 페이지에 사용자 정보를 인식하는 Tag Code를 삽입하여 정보수집
장점	· 네트워크 트래픽이 없음 · 보안/SSL 수집가능, 외부 DB 및 데이터와 함께 분석 가능	· 실시간 분석, 분산된 웹서버 로그 자동 수집용이	· 실시간 분석, 브라우저 캐싱 처리, 필요한 데이터만 수집, SSL 수집가능 Flash 분석 가능
단점	· 실시간 분석 불가, 분석 장비로 로그 전송 문제 · 브라우저 캐싱 해결 불가	· 보안 문제, 장애대처 어려움, 네트워크 트래픽 증가(버퍼링), SSL 측정불가	· 분석대상 모든 페이지에 Tag 삽입
제공업체	WebTrends, 넷스루(Wiselog)	소만사	로거, 에이스카운터 등
주요 이용업체	대기업 및 쇼핑몰		중·소규모 업체
비용	솔루션 판매, 초기투자비용 높음		월별 과금

또한 이러한 기능을 라우터나 스위치 등에 내장시키기에는 성능 및 경제적인 부담이 되기 때문에 일반적으로 별도의 측정 및 분석 시스템 형태로 제공된다.

- 플로우 기반 전수 저장

패킷 기반 방식의 단점을 보완하기 위해 개선된 방식으로 패킷 단위로 측정하는 대신 공통된 특성(Source IP, Destination IP, Source Port, Destination Port, Protocol)을 가지는 패킷들을 플로우라는 단위로 묶어서 처리하는 방식으로 처리시간 및 저장 공간의 현저한 감소를 가능하게 한다. 플로우 기반 전수저장 방식은 망 장비에 내장될 수 있으며, 1Gbps 이상 고속 인터페이스의 경우에는 여전히 성능의 부하가 너무 커서 샘플링 방식을 사용하기도 한다.

- 웹 분석에 따른 전수저장 방법

웹 분석에 따른 전수저장은 Web Log, Packet Sniffing, Script& Tagging로 구분된다. Web Log는 웹 서버에 쌓인 로그기록을 분석프로그램으로 사후에 분석한다. Packet Sniffing은 네트워크에서 주고받는 패킷에 담긴 정보를 Sniffing하고 방문자의 트랜잭션을 수집한다. Script&Tagging은 웹 페이지에 사용자 정보를 인식하는 Tag Code를 삽입하여 정보를 수집한다. 표 2는 전수저장에 따른 특징과 장·단점 등을 비교·분석한 내용이다[9].

일반적으로 Web Log, Packet Sniffing 방식은 초기 설치비용이 많이 들기 때문에 중·소형 업체에서는 Script&Tagging 방식의 서비스를 제공하는 방식을 주로 이용한다. 하지만 Script&Tagging 방식은 분석하고자하는 모든 웹페이지에 Tag Code를 삽입해야 하는 특징이 있기 때문에 이 부분을 주의하면서 웹 로그 분석을 이용해야 한다.

4.2 트래픽 샘플링

샘플링 기법은 시간, 비용을 고려하여 미리 정한 모집단에 있는 모든 거래들이나 사건들에 대해 100% 검증을 원치 않을 때 모집단으로부터 특정한 표본을 추출해 내는 과정을 의미한다. 패킷 샘플링 기법과 관련된 연구는 네트워크 트래픽 분석을 목적으로 1993년도부터 시작되었다. 주로 네트워크 관리를 위해 필요한 트래픽 통계를 작성하는 작업에 활용되었으며, 방대한 트래픽 양으로 인한 계산 오버헤드를 줄이려는 노력의 일환으로 샘플링 기법을 사용하였다. 최근에는 네트워크 관리나 보안을 위한 실시간 네트워크 트래픽 모니터링이나 계정관리/청구서작성(accounting/billing) 서비스를 위한 트래픽 플로우별 트래픽 양 측정을 위해 고속의 네트워크에서도 샘플링 기법을 사용하고 있다[9]. 이러한 샘플링 기법은 전체 네트워크 트래픽 양의 1/N을 확률적으로 샘플링 하는 것을 기본으로 한다. 여기서 N은 표본 집단의 개수를 모집단의 개수로 나눈 것을 말한다. 즉 평균적으로 매 N개의 패킷마다 하나씩 샘플링을 수행한다. 샘플링 방법은 패킷을 샘플링 하는 방법에 따라 구간을 나눠 패킷을 단순한 확률에 따라 무작위로 선별하는 단순 랜덤 샘플링(Simple Random Sampling), 각각의 구간 안에서 무작위로 패킷을 샘플링 하는 층화 샘플링(Stratified Sampling), 각 구간의 첫 번째 패킷을 샘플링 하는 규칙적 샘플링(Systematic Sampling), 모집단을 몇 개의 군집으로 나누고 샘플링 할 군집을 샘플링 하는 군집 샘플링(Cluster Sampling), 그리고 샘플링을 여러 단계로 나누어서 샘플링 하는 다단계 샘플링(Multi-Stage Sampling) 등이 있다.

- 단순 랜덤 샘플링

모집단(Population)을 구성하는 개체에 번호를 부여한 후, 추출할 개체의 양만큼 랜덤 번호를 생성한다. 생성된 랜덤 번호에 해당하는 개체를 모집단에서 추출한다. 매우 간단하고 적용이 쉬운 샘플링 기법이다. 대체로 모집단이 작은 경우 유용하다. 일반적인 트래픽 샘플링에서 표준적인 기법으로 많이 사용된다.

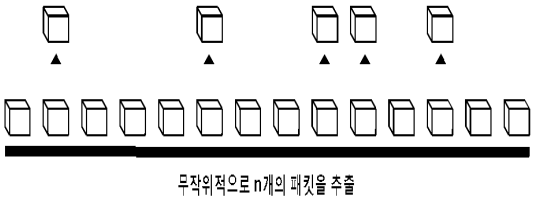

그림 6. 단순 랜덤 샘플링
Fig. 6. Simple random sampling

- 층화 샘플링

모집단에서 동일한 속성을 가지는 개체들을 묶어 여러 층(Strata)을 생성하고, 각 층으로부터 일정한 개체를 추출하는 방법이다. 트래픽을 한 덩어리로 보고 단순한 샘플링 기법을 적용하는 것보다는 효과적일 것으로 예상된다. 트래픽은 다양한 프로토콜(TCP, UDP, ICMP 등)의 특성에 따라 그 양과 형태가 다르기 때문이다. 층화 샘플링은 비례 층화 샘플링과 비비례 층화 샘플링으로 나눌 수 있다.

비례 층화 샘플링은 모집단에서 각 층이 정하는 비례에 따라 각 층의 크기를 할당하여 표본을 추출하는 방법이고, 비비례 층화 샘플링은 각 층에서 각층의 크기와는 상관없이 같은 수의 표본을 추출하는 방법이다.

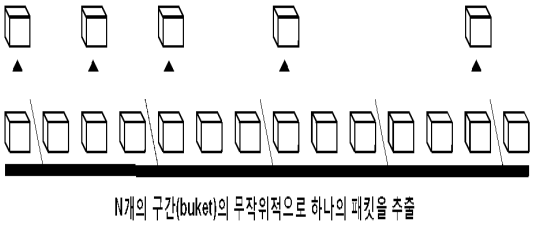

그림 7. 층화 샘플링
Fig. 7. Stratification sampling

표 3. 샘플링 방법별 장·단점
Table 3. The strengths and weaknesses of the ways sampling

구분	장점	단점
단순 랜덤 샘플링	· 심플하고 표준오차를 쉽게 구할 수 있음	· 모집단의 전체 리스트가 필요함 · 모집단 대표성이 떨어질 수 있음
규칙적 샘플링	· 전체 리스트에 대해서 대표성을 확보할 수 있음 · 실행하기 쉬움	· 일련의 순서를 매긴 리스트가 특정 사이클을 가지고 있는 경우, 표본의 정확성 떨어짐
층화 샘플링	· 모집단의 데이터 특성을 살려 비교적 모집단 대표성을 확보할 수 있음	· 표준오차를 정확히 측정하기 어려움 · 각 하위 그룹별로 추출된 표본수가 너무 적은 경우 정확성이 떨어질 수 있음
군집 샘플링	· 샘플링 리소스가 적게 들며, 전체 완벽한 모집단 리스트가 없어도 됨	· 군집 샘플 결과 각 군집들이 상이한 특징을 보이면, 표본의 대표성 떨어짐 표본 오차를 정확히 측정하기 어려움
다단계 샘플링	· 전체 모집단이 큰 경우에 가장 실행 가능한 접근 방식	· 여러 개의 샘플 리스트 필요 · 표본 오차

- 규칙적 샘플링

모집단에서 일정한 간격에 한 번씩 개체를 추출한다. 이러한 특성 때문에 규칙적 샘플링을 간격 샘플링(Interval Sampling)이라고도 한다. 간격의 설정 방식에 따라서 계수 기반 규칙적 샘플링(Count-based Systematic Sampling)과 시간 기반 규칙적 샘플링(Time-based Systematic Sampling)으로 나눌 수 있다. 규칙적 샘플링은 트리거가 동장할 때마다 해당 순서의 개체를 추출한다.

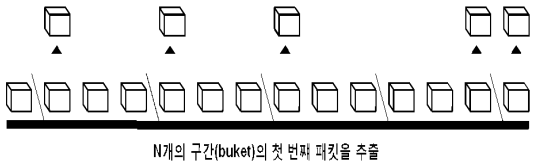

그림 8. 규칙적 샘플링
Fig. 8. Rule sampling

- 군집 샘플링

광범위한 지역 전체에서 표본을 추출하는 것이 아니라 몇 개의 지역을 추출하고, 추출한 지역 내에서만 표본을 추출하는 방법이다. 모집단을 몇 개의 군집으로 나누고 샘플링 할 군집을 단순 무작위 방식으로 선발하여 선발된 군집 안에서 무작위로 자료를 취득하는 것을 말한다.

- 다단계 샘플링

실세계의 복잡성으로 인해 위의 네 가지 방법만으로는 정확한 샘플링을 하는데 한계가 있으므로 이들 중 두 개 이상의 방법들을 적절히 조합하여 효율적이고 효과적으로 샘플링 하는 방법이다. 보통 이단계(two-stage) 또는 삼단계(three-stage) 샘플링이 많이 사용된다.

4.3 트래픽 요약 정보(Flow)

플로우는 일련의 동일한 특성을 가진 패킷의 집합으로 정의할 수 있다[10]. 또한 플로우는 네트워크 홉을 경유하는 응용서비스별 트래픽의 흐름을 모니터링하고 특정 트래픽의 급격한 증가를 파악하기 용이한 장점을 지니고 있다. 응용서비스별 플로우 정보와 버스트(burst) 정도에 따라 임계치(threshold)를 적용하

여 비정상 트래픽에 대한 관리가 가능하다. 대표적으로 시스코의 넷플로우(NetFlow)를 보편적으로 사용하며 통신사업자의 백본라우터에서 대부분 채택하고 있다. 그밖에 sFlow규격도 있지만 지원하는 장비가 극소수이기 때문에 통합시스템의 요구기능으로 검토하기 어렵다. 플로우의 주요정보는 5가지로 출발지/도착지 주소, 출발지/도착지 포트, 프로토콜이다. 플로우와 SNMP, RMON을 기능적 측면에서 비교한다면 플로우가 SNMP보다 상세한 정보를, RMON 보다 단순한 정보를 제공한다[11].

Ⅴ. 결론

본 논문에서는 공격 트래픽의 효과적인 저장 및 분석 방안에 대한 연구를 수행하기 위해 Raw Traffic의 공격 유형별 저장방법, 공격 트래픽의 저장유형에 따른 분류방법으로 분류하였다.

Raw Traffic의 공격 유형별 저장방법에서는 웹 기반 공격과 네트워크 기반 공격으로 분류하고, 각각에 대한 공격 형태에 대한 분석 방법과 공격을 탐지 및 저장할 수 있는 방법에 대해 연구를 수행하였다.

공격 트래픽의 저장유형에 따른 분류방법에서는 각각의 공격 형태에 따라 저장하는 방식과 방법 등이 달라짐에 따라 모든 트래픽에 대한 정보를 저장하는 형태인 전수저장 방식에 대한 연구를 수행하였다. 하지만 모든 트래픽에 대한 정보를 저장하는 형태인 전수저장 방식은 매우 큰 용량의 파일 형태와 분석하기 위해 정보가 많아지기 때문에 이를 해결할 수 있는 트래픽 샘플링과 Flow에 대한 연구를 수행하였다.

추후 연구방향으로는 본 논문을 바탕으로 실제 웹사이트 상에서 공격정보에 대한 수집 및 연구가 이루어져야 할 것이다. 실제 웹사이트 상에서 수집된 공격정보를 바탕으로 본 논문에서 연구한 연구 결과에 따라 공격정보를 저장하고, 저장된 공격정보를 통해 실제 웹사이트 상에서 공격 탐지 및 차단이 가능한지에 대한 연구가 필요할 것이다. 또한 제시된 공격 트래픽의 저장 및 분석 방안을 통해 실제 시스템에 적용하여 효과적인 트래픽 샘플링, Flow 등의 저장 및 분석 방법에 대한 연구가 수행되어야 한다. 마지막으로 웹사이트 상에서 발생하는 공격정보를 수집하여 저장하는 형태는 각각의 보안서비스마다 다르기 때문에 이를 해결할 수 있는 표준적인 저장방법에 대한 연구가 수행되어야 할 것이다.

참고문헌

[1] 한국인터넷진흥원, *인터넷 침해사고 동향 및 분석 월보*, 2009. 11.

[2] 한국정보보호진흥원, "*인터넷기반구조에서의 비정상트래픽 발생 징후 탐지를 위한 기법 연구*", 2004.

[3] 정보통신부, *네트워크 기반 침입패턴 자동생성 알고리즘*, 2004. 6.

[4] 정현철, "*IP Fragment를 이용한 공격기술*", 2001.

[5] 성종규, 이민형, 황찬규, 유재형, "네트워크 자원 정보를 이용한 Flow 트래픽 분석시스템 개발", *KNOM review*, Vol. 9, No. 1, pp.17-24, June 2006.

[6] 김정현, 원유집, 안수한, "수동적 인터넷 측정을 위한 샘플링 기법 비교: 사례 연구를 통한 검증", *전자공학회 논문지*, 제45권 TC편 제6호, pp.34-50, 2008. 6.

[7] 강신헌, 김재현, "네트워크 트래픽 특성 분석을 통한 스캐닝 웜 탐지 기법", *정보과학회논문지*, 제35권 제6호, pp.465-560, 2008. 12.

[8] 강길수, 이준희, 최경희, 정기현, 심재홍, "DDoS 공격 탐지를 위한 패킷 샘플링 기법들의 성능 분석", *정보처리학회논문지C*, 제11권-C권 제6호, pp.711-718, 2004. 12.

[9] http://adioshun.springnote.com/pages/413943?print=1

[10] 금융보안연구원, DDoS 공격 유형별 대응방안 설명, 2008. 4.

[11] 천인혁,엄태랑,권란,이경근,송정희,"사이버 교육 컨텐츠에 따른 트래픽 측정", *한국정보과학회 봄 학술발표논문집*, 2003

저자소개

이기성(Ki-Sung Lee)

2010년 한국기술교육대학교
인터넷미디어공학부 재학

※ 관심분야 : 암호 알고리즘, 무선 인터넷 보안

이동영(Dong-Yung Lee)

1993년 동아대학교 전자공학과(학사)
1998년 성균관대학교 정보공학(석사)
1993년-1997년 기아자동차
중앙기술연구소 연구원
2002년 성균관대학교 컴퓨터공학(박사)
2003년 2월~현재 명지전문대학 정보통신과 교수

※ 관심분야 : 네트워크보안, 홈 네트워크, USN

서희석(Hee-Suk Seo)

2000년 성균관대학교 공학사
2002년 성균관대학교 공학석사
2005년 성균관대학교 공학박사
2005년 한국시뮬레이션학회 이사
2005년~현재 한국기술교육대학교 교수

※ 관심분야 : 네트워크보안, 모델링 방법론