ISSN 1598-0170 (Print)
ISSN 2287-1136 (Online)
http://www.jksii.or.kr

개선된 ATMSim을 이용한 DDoS 공격 분석☆

DDoS Attack Analysis Using the Improved ATMSim

정 해 덕[1] 류 명 운[1] 지 민 준[1] 조 유 빈[1] 예 상 국[2] 이 종 숙[3*]
Hae-Duck J. Jeong Myeong-Un Ryu Min-Jun Ji You-Been Cho Sang-Kug Ye Jong-Suk R. Lee

요 약

최근 정보통신망의 발전과 스마트 폰의 대량 보급으로 인하여 인터넷 트래픽이 기하급수적으로 증가하고 있다. 이와 관련하여, 본 논문은 증가하고 있는 인터넷 침해사고와 네트워크 공격 중 대표적인 DDoS 공격에 대해서 탐지 및 분석한다. 이를 위해 네트워크 플로우 정보를 바탕으로 동작할 수 있도록 기존의 ATMSim 분석 패키지의 기능과 GUI를 개선하고, 이를 이용하여 캠퍼스 내부 LAN을 통해 대량으로 유입되는 정상적인 트래픽과 DDoS 공격이 포함된 비정상 트래픽을 생성한다. 수집 • 생성된 정상 • 비정상 트래픽의 특성을 분석하기 위해서 자기유사성 추정 기법을 이용하여, 그래픽 분석 및 Hurst 파라메터 (자기유사성 파라메터) 추정량 분석결과 정상 트래픽과 비정상 트래픽이 자기유사성 관점에서 추정치 Hurst 값이 높음을 보여 주고 있다.

☞ 주제어 : 비정상 트래픽, 자기유사성, Hurst 파라메터, ATMSim, DDoS 공격

ABSTRACT

Internet traffic has been significantly increasing due to the development of information and communication networks and the growing numbers of cell phone users that access networks. This paper connects to this issue by presenting a way to detect and analyze a typical DDoS attack that results in Internet breaches and network attacks, which are on the increase. To achieve this goal, we improve features and GUI of the existing ATMSim analysis package and use it. This package operates on a network flow-based analysis method, which means that normal traffic collected through an internal LAN at the Korean Bible University campus as well as anomaly traffic with DDoS attacks are generated. Self-similarity processes are used to analyze normal and anomaly traffic that are collected and generated from the improved ATMSim. Our numerical results obtained from three Hurst parameter estimate techniques show that there is quantitatively a significant difference between normal traffic and anomaly traffic from a self-similarity perspective.

☞ keyword : Anomaly traffic, self-similarity, Hurst parameter, ATMSim, DDoS attack

1. 서 론

최근 정보통신망의 발전과 스마트 폰의 대량 보급으로 인하여 인터넷 트래픽이 기하급수적으로 늘어나고 있다. 이와 더불어 각종 인터넷 침해사고와 네트워크 공격에 대한 정보보안 문제가 심각히 대두되고 있는 상태이며, 최근 들어, 이와 관련된 연구가 지속적으로 제안되고 있다[1, 2, 3]. 하지만 이러한 연구가 지속적으로 제안되고 있음에도 불구하고, 실시간으로 비정상 트래픽을 정확하게 탐지해 내는 기술은 아직까지 부족한 실정이다. 이에 따라 피해사례는 늘어나고 있으며[15], 네트워크 공격의 종류도 다양해지고, 그 대상 또한 기존 정부기관 언론사, 금융권, 기업 등에서, 최근에는 한 유명 e-Sports 대회 결승전 도중 공격을 받아 중단되는 등[13, 16] 수법도 다양해지고 있는 실정이다. 여기에 사용된 대표적인 네트워크 공격은 DDoS(Distributed Denial of Service)로 알려져 있다. DDoS 공격은 P2P 응용서비스, e-mail 등을 통해 웜 바이러스, 백도어 등을 인터넷에 다량으로 유포하여 해당 PC를 좀비 PC로 감염시켜 피해 대상 시스템에 대하여 다양한 공격 기법을 활용하여 서비스 거부 상태로 만드는 공격이다.

본 연구에서는 기존 연구의 한계 점들을 해결하기 위해 본 연구그룹에서 지속적으로 기능을 추가 개발하고 있는 ATMSim(an Anomaly Teletraffic detection Measurement analysis Simulator)이라는 네트워크 트래픽 분석 패키지의

[1] Department of Computer Software, Korean Bible University, Seoul, 01757, Korea.
[2] Division of LBS Solution, SK MNS, Seoul, 03188, Korea.
3 Department of Computational Science & Engineering, KISTI, Daejeon, 34141, Korea.
* Corresponding author (jsruthlee@kisti.re.kr)
[Received 21 December 2015, Reviewed 5 January 2016, Accepted 27 January 2016]
☆ 이 논문은 2015년도 한국성서대학교 학술연구조성비와 한국과학기술정보연구원에 의하여 연구되었음.

기능과 GUI를 개선하고, 이를 활용하고자 한다. 개선된 ATMSim은 하둡(Hadoop)을 이용한 데이터 처리와 자기유사성 기반의 실시간 네트워크 트래픽의 모니터링을 통해 비정상적인 네트워크 트래픽 상황을 조기에 탐지할 수 있는 시스템이다.

본 논문의 구성은 다음과 같다. 이어지는 2장에서는 비정상 트래픽, 자기유사성에 대한 개념과 기존 비정상 트래픽 탐지 방법 관련 연구들을 소개하고, 3장에서는 제안하는 시스템의 설계 및 구현에 대하여 자세한 방법을 서술한다. 4장에서는 본 연구에 대한 실험 과정과 결과를 보이고, 마지막 5장은 결론 및 후속 연구과제로 결언을 맺는다.

2. 관련연구

2.1 비정상트래픽

비정상 트래픽이란 일련의 연산과정을 통해 정해진 임계치를 기준으로하여 발생한 트래픽이 초과해 비정상적 작동을 했을 때, 이 트래픽을 비정상트래픽이라 한다. 이러한 비정상트래픽은 흔히 네트워크 공격을 통해 발생할 수 있으며, 서버 장애나 접속자수가 급격하게 증가하는 경우에도 비정상 트래픽으로 분류한다.

2.2 자기유사성

이 절에서는 자기유사성(self-similarity)의 수학적인 정의에 관하여 간단히 살펴본다. 자기유사성은 크게 결정적 자기유사성 (deterministic self-similarity)과 통계적 자기유사성(stochastic self-similarity)으로 나눌 수 있다. 결정적 자기유사성은 프랙탈(fractal)이라고도 하며, 자기유사성과 순환성을 포함한다. 본 연구와 관련된 통계적 자기유사성이란 초, 분, 시간, 일, 월 등의 시간 단위 (time unit)의 크기(scale)를 변화시켜 생성한 새로운 프로세스는 통계적인 측면 (또는 분포면)에서 볼 때, 원래의 프로세스와 동일한 자기상관함수 (ACF, auto correlation function)를 갖는 특성을 말한다. 자기유사성의 특성을 분석하기 위해서는 자기유사성 모델의 파라메터인 Hurst 파라메터 (또는 자기유사성 파라메터) 값을 구하며, Hurst 파라메터 값이 1에 가까워질수록 자기유사성 정도가 높은 것을 의미한다.

그래픽 분석 및 자기유사성 파라메터 값을 구하는 방법으로 periodogram plot, R/S-statistic plot, wavelet -based H 추정법 등이 있으며, 이들의 추정방법을 이용하여 정상트래픽과 비정상 트래픽 여부를 탐지한다. 자기유사성과 추정기법들에 관한 보다 자세한 내용은 [4, 5, 6]를 참조하기 바란다.

2.3 기존 비정상트래픽 탐지 관련 연구

기존의 트래픽을 분석하는 방법은 크게 두 가지로 나눌 수 있다. 링크를 통해 흐르는 모든 패킷을 분석하는 패킷 기반의 트래픽 분석 방법과 동일한 5-tuple 을 가지는 다수의 패킷으로 이루어진 플로우를 이용하는 플로우 기반의 트래픽 분석 방법이다[1].

2.3.1 패킷 기반 트래픽 분석 방법

패킷 기반 트래픽 분석 방법의 경우 패킷의 모든 계층에 대한 분석이 가능하기 때문에 다양한 정보를 수집할 수 있으며, 플로우 기반의 트래픽 분석보다 비교적 정확한 분석이 가능하다. 하지만 패킷에 포함된 데이터 분석은 인터넷 사용자의 사생활 침해 문제가 될 수 있다. 또한 패킷 데이터를 분석하는 시간의 증가와 트래픽 데이터를 저장하기 위한 방법이 문제가 될 수 있다[1]. 이러한 문제를 해결하기 위한 최근의 연구를 살펴보면 많은 양의 데이터를 처리하는데 드는 비용과 제한성을 줄이기 위해 기존 RDBMS(Relational Database Management System)에서 NoSQL을 이용하는 방식에 대한 연구[2], 패킷 내 개인정보 유출 방지를 위한 연구[3] 등이 소개되고 있다.

2.3.2 플로우 기반 트래픽 분석 방법

플로우 기반 트래픽 분석 방법에서는 다수의 패킷으로 이루어진 플로우를 중심으로 분석이 이루어진다. 하나의 플로우는 일정 시간 이내에 지나가는 5-tuple 헤더로 이루어진 패킷들의 집합으로 정의한다[1]. 5-tuple 은 source IP address, destination IP address, protocol number, source port, destination port[7]로 구성된다. 패킷 레벨에서는 불가능한 세부적인 트래픽 분석이 가능할 뿐 아니라, 동일한 패턴을 단일 플로우로 저장하는 압축 효과가 있기 때문에 대용량 트래픽을 운용하는 네트워크 환경에서도 매우 효율적이다. 네트워크상에서 이러한 플로우 정보를 얻기 위한 방법은 크게 두 가지가 존재한다. 첫 째로 라우터나 스위치에서 생성된 플로우 정보를 받아 오는 방법이며, 대표적으로 CISCO NetFlow[17]가 이에 속

한다. 두 번째 방법으로는 모니터링 시스템 자체적으로 Probe와 같은 플로우 데이터 수집기를 이용하여 트래픽 정보를 수집해 플로우를 생성하는 방법이 있다.

본 논문에서는 이러한 플로우 기반 트래픽 분석 기법 중 ATMSim 시뮬레이터 자체적으로 플로우 데이터를 수집하여 네트워크상에서 발생하는 트래픽을 실시간으로 분석한다. 이에 관련된 보다 자세한 사항은 3장에서 다루도록 하겠다.

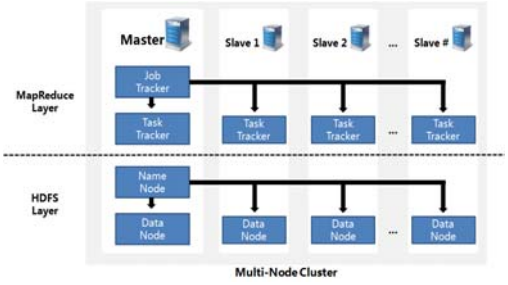
3. 실시간 비정상 트래픽 탐지 시스템 설계 및 구현

3.1 ATMSim

ATMSim은 하둡과 자기유사성 기반의 비정상트래픽을 수집, 탐지, 측정 및 분석(AT-IDS, Anomaly Teletraffic-Intrusion Detection System) 기능을 갖고 있으며, 다양한 공격 tool을 탑재하여 실제 상황을 모의 실험할 수 있는 시뮬레이터이다.

3.1.1 하둡

하둡은 대용량의 데이터를 처리하기 위한 분산 응용 프로그램을 지원하는 오픈소스 프레임워크이다. 하둡은 웹 검색 엔진 너치(Nutch)의 분산처리를 위해 개발된 것으로 핵심요소로 GFS(Google File System)을 벤치마킹하여 만든 HDFS(Hadoop Distributed File System)와 MapReduce로 구성된다. 또한 구조는 그림 1에서 보는 바와 같이 마스터-슬레이브(master-slave) 구조로써 마스터 노드(master node)는 네임노드, 2차 네임노드(secondary name node), 잡트래커(job tracker)로 구성되고, 슬레이브 노드는 데이터 노드(data node), 태스크트래커(task tracker) 등의 구조를 가지고 있다.



(그림 1) 하둡의 구조
(Figure 1) Hadoop architecture.

HDFS는 파일을 기본 64Mbyte 단위로 나누어 분산 저장하여 병렬처리를 하는데 이는 안정적이고 빠른 저장소의 역할을 한다. 이에 앞서 설명한대로 네임노드와 데이터노드가 구분되어있다[9, 10].

MapReduce는 병렬처리하기 위한 분산 프로그래밍 모델로써 대용량의 데이터를 빠르고 안전하게 처리할 수 있으며, 키 값을 통해 필터링 혹은 변환 작업을 하는 Map과 Map 함수를 통해 출력된 결과 값을 새롭게 연산과정을 거쳐 목록을 생성하는 Reduce 함수로 구성된다. 전체적인 흐름은 빅데이터 블록을 분할하고, MapReduce에 데이터를 입력한 후 출력 결과를 정렬, 분할, 병합, 셔플링 등의 단계를 거쳐 Reducer에 전송한다. 이후 최종 출력 데이터를 HDFS에 저장하는 처리과정을 거친다[11].

하둡의 성능은 2008년 4월에 있었던 테스트에서는 테라바이트 이상의 데이터 정렬을 위해 가장 빠른 시스템으로서 세계 기록을 갱신하였고, 2009년 5월에 야후는 하둡을 이용하여 62초 만에 1테라바이트 데이터를 정렬하는 쾌거를 올렸다. 하둡의 이점 중 하나인 확장성의 부분을 하둡을 사용하는 Yahoo나 Facebook이 잘 보여주고 있다[9, 10].

ATMSim은 이 하둡을 적용시켜 대용량 저장의 어려움과 빅데이터의 처리과정, 비용적인 측면을 해결하였다.

3.1.2 AT-IDS

AT-IDS는 ATMSim 상에서 개발된 자체 침입탐지시스템이다. 일반적인 IDS의 목적이 각종 해킹수법을 자체 내장하여 네트워크를 모니터링하고 시스템의 활동을 실시간으로 분석하여 침입을 탐지하며 이를 추적, 제어하는 등의 대응을 통해 보안을 강화하는 데에 있다.

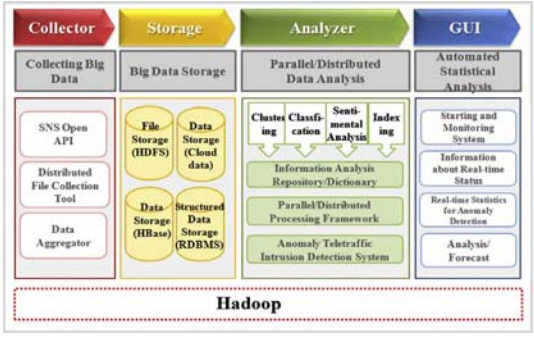
AT-IDS의 탐지 기술은 접근 방식에 따라 3가지 유형으로 볼 수 있다. 통계적 방식, 지식기반 방식, 기계학습 방식이다. 표 1은 AT-IDS의 접근방법 및 장, 단점을 보여주며, 먼저 통계적 기반 방식은 시스템이나 네트워크 트래픽의 움직임을 통해서 데이터를 수집한다. 이 움직임들을 통해 정상적인 트래픽과 비정상적인 트래픽의 편차를 확인한다. 두 번째로 지식기반 방식은 검사된 데이터들을 미리정의 된 룰을 통해 사전 데이터로 저장한다. 마지막으로 기계학습 기반 방식은 Markov models, genetic algorithms, neural networks, Bayesian networks, fuzzy logic 등을 사용하여 실제 공격을 탐지하는 방법이다.

(표 1) AT-IDS의 접근방법 및 장·단점.
(Table 1) Advantages and disadvantages of the AT-IDS techniques[8].

접근 방법	관련 기법	Advantages and disadvantages
통계적 기반 (statistical-based)	- Univariate models - Multivariate models - Time series models - Self-similar models	- 정상적인 활동에 대해 사전지식이 필요하지 않음 - 악의적 활동의 정확한 알림 - 공격자에 의해 훈련되어지는 취약점 - Parameter와 metrics의 어려운 설정 - 현실적이지 않은 quasi-stationary 과정 가정
지식기반 (knowledge-based)	- Finite state machines - Description languages - Expert systems	- 견고성 - 유연성과 확장성 - 질적으로 높은 데이터를 얻기가 어렵고, 시간 소요에 대한 효율이 떨어짐
기계학습 기반 (machine learning-based)	- Markov models - Genetic algorithms - Neural networks - Bayesian networks - Fuzzy logic - Clustering and outlier detection	- 유동성과 적응성 - 상호의존적 수집 - 높은 자원의 소모 - 시스템으로부터 획득한 동작에 추정치로부터의 높은 의존성

3.2 ATMSim의 일반적인 시스템 아키텍쳐

그림 2는 ATMSim의 일반적인 시스템 아키텍쳐를 보여주고 있으며, 수집 모듈은 웹 페이지, SNS, system log 데이터를 통해 분산 파일을 수집하며, 저장 모듈은 파일과 데이터 구조적 데이터를 저장하는 영역으로써 이를 실시간으로 분석해준다. 분석 모듈은 실제 사용되는 데이터 분석뿐만 아니라 예측 분석과 자연어 처리과정 및 텍스트마이닝 과정 등을 거치며 특히 MapReduce 프레임워크를 사용하여 병렬 및 분산된 데이터를 분석하고 모으며 분류한다[12]. 마지막으로 GUI 모듈은 시스템을 모니터링하고, 시스템의 상태나 AT-IDS를 통해서 실시간 통계 정보 등을 자동으로 통계 분석을 해준다.



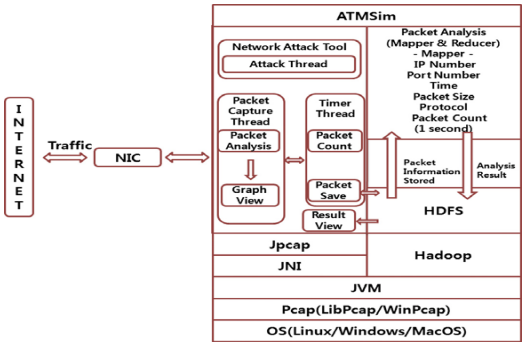
(그림 2) AT-IDS의 일반적인 아키텍쳐
(Figure 2) A proposed general architecture of AT-IDS, called ATMSim[8].

3.3 ATMSim의 상세 시스템 작동방식

ATMSim은 실시간으로 네트워크 텔레트래픽을 탐지 및 측정, 분석하는 ATMSim의 시스템구조를 보여준다. 또한, ATMSim은 자바로 구현된 패키지이고, Jpcap을 사용하여 패킷을 분석하며, Oracle에서 제공하는 JavaFx Library를 이용하여 UserInterface와 트래픽에 대한 real-time graph 및 total-time graph를 제공한다. JVM이 설치된 리눅스, 윈도우, MacOS 어디서나 사용 가능하다. 그림 3은 이런 ATMSim의 상세 시스템 구조를 보여준다.

ATMSim의 real-time graph는 매초마다 들어오는 다른 소스 주소와 목적지 주소를 가진 패킷들의 정보를 그래프로 나타낸다. 특정 프로토콜의 패킷 사이즈의 정보를 특정 기간의 그래프를 보여주고, total-time graph에서는 분석한 전체 시간동안의 그래프를 보여준다. 그리고 이와 별도로 매초마다 들어오는 모든 트래픽은 매초마다 3가지 형태로 저장되는데 정상 트래픽, 비정상 트래픽, 혼합된 트래픽의 형태로 저장 분석된다. 저장되는 데이터는 발생시간과 프로토콜의 종류, 패킷의 사이즈 및 카운트 정보를 담는다. 이는 HDFS를 기본으로 하여 이 후

HDFS에 저장된 데이터를 MapReduce 과정을 통해 비정상 트래픽을 감지해낸다. 비정상 트래픽의 프로토콜 종류에 상관없이 패킷의 사이즈를 캡처해낼 수 있다[8, 13].



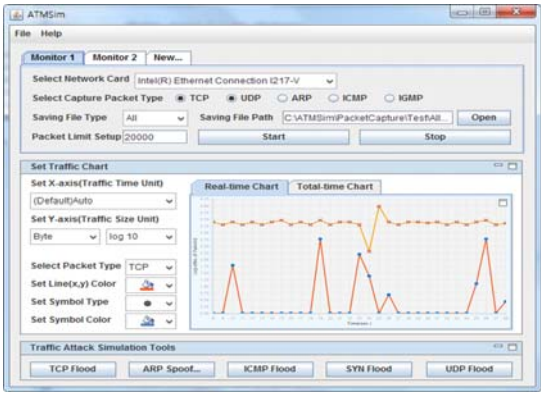
(그림 3) ATMSim의 상세 시스템 구조
(Figure 3) A detailed system architecture of ATMSim[8].

ATMSim은 비정상트래픽을 발생시킬 수 있는 다양한 공격 tool을 내장하고 있어, SYN flood, ICMP flood, ARP spoofing, UDP flood를 실제상황에 맞도록 시뮬레이션을 하기 위해서 설계되었다[8]. 이러한 공격 tool들은 IP를 스스로 변조할 뿐만 아니라 패킷을 대체하는 기능을 하는 해킹공격들이다[14].

3.4 개선된 ATMSim의 구현결과

그림 4는 앞서 시스템 설계 부분에서 설명한 부분을 구현한 ATMSim의 메인화면이다. 기존 ATMSim 화면 구성의 경우 패킷 타입과 공격 시뮬레이션을 트래픽 분석으로써 단순한 그래프만을 보여주고 있었으나 이를 GUI로 개선하였으며, 사용자가 쉽게 읽을 수 있도록 y축을 단위(byte, mega byte, giga byte / *log10*(y), *loge*(y)) 로 제공한다. 화면의 구성은 크게 3가지 나뉘게 되며, 첫째, 각 네트워크 카드에 대한 트래픽 캡처 모니터링 설정을 지원하고 있으며, 이에 관련된 세부적인 모듈들은 네트워크 카드 선택 (Intel Ethernet I217, I218 series), 수집할 패킷의 종류 (UDP packet 및 TCP packet), 초당 수집 패킷 제한 (1,500,000 limit), 저장 패킷 종류(실험의 경우 TCP, UDP 패킷 모두의 저장을 위한 ALL Setting) 및 경로에 대한 설정들을 지원하고 있다.

둘째, 캡처한 패킷에 대해 차트 설정을 지원한다. 또한 차트의 우측 상단에 새창으로 열기 버튼을 지원하여 보다 정확한 모니터링이 가능하도록 구현하였다. 마지막으로 다양한 트래픽 공격 시뮬레이션 툴을 지원하고 있다. 공격 시뮬레이션 툴들은 앞서 3.3절에서 언급한 바와같이 TCP flood, SYN flood, ICMP flood, ARP spoofing, UDP flood들을 지원한다. 각 버튼을 클릭하면 해당하는 공격 툴에 대한 시뮬레이션이 가능하다.



(그림 4) 개선된 ATMSim의 메인 메뉴
(Figure 4) Main menu of the improved ATMSim

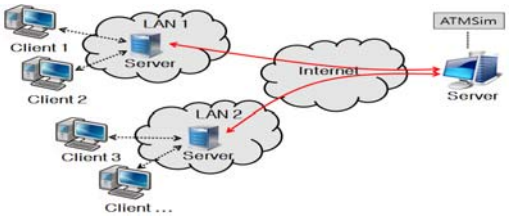
4. 실험 및 결과

4.1 실험환경

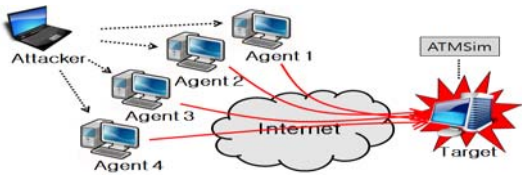
본 논문에서 제안한 ATMSim을 이용한 비정상 트래픽 탐지 시스템에 대한 검증을 하기 위해 비정상 트래픽 발생 시나리오를 작성하여 실험을 수행하였다.

4.1.1 실험방법

실제 상황과 유사한 환경을 만들기 위해 본 논문에서는 KBU(Korean Bible University) 캠퍼스 내의 내부 LAN을 통한 트래픽 수집을 진행하였다. ATMSim 내에 장착된 자기유사성을 이용한 비정상 트래픽을 탐지하기 위해서는 정상 트래픽과 공격시의 비정상 트래픽이 필요하다. 따라서 정상트래픽 수집은 2015년 7월 28일 00:00:00부터 2015년 7월 29일 23:59:59까지 진행하였고, 총 8회에 걸쳐 Non-Spoofed UDP Flood 공격기법으로 네트워크 공격이 포함된 비정상 트래픽 수집은 2015년 7월 25일 16:09:44부터 2015년 7월 27일 16:09:43까지 각각 2일 동안의 데이터를 확보하였다. 이에 대한 수집 시나리오는 다음 그림 5, 6와 같다.



(그림 5) 정상 트래픽 수집 시나리오
(Figure 5) Scenario for collecting LAN traffic.



(그림 6) 비정상 트래픽 수집 시나리오
(Figure 6) Scenario for collecting anomaly traffic.

(표 2) 실험 컴퓨터 시스템 사양
(Table 2) Specifications of experimental computer systems.

실험 시스템	구분	사양
Attacker	OS	Windows 8
	CPU	Intel Core i7-4702MQ CPU
	Memory	8GB
Agent1	OS	redhat 7.0
	CPU	Intel Core2 Duo CPU
	Memory	512MB
Agent2	OS	Windows2003 Server
	CPU	Intel Core2 Duo CPU
	Memory	1GB
Agent3	OS	Windows 7
	CPU	Intel Core i5-2410M CPU
	Memory	4GB
Agent4	OS	Windows XP
	CPU	Pentium Dual-Core CPU
	Memory	2GB
탐지시스템	OS	Windows 7
	CPU	Intel Core i5-4570 CPU
	Memory	4GB

그림 5, 6와 같이 본 논문에서 제안한 기법의 실험을 위해서 정상 트래픽과 네트워크 공격이 포함된 비정상 트래픽을 각각 수집하였다. 먼저, 정상 트래픽 수집은 외부 LAN에서 KBU 캠퍼스 내의 내부 LAN으로 들어오는 트래픽에 대하여 정상 트래픽으로 간주하고, ATMSim이 설치된 Server를 통하여 수집을 진행하였다. 다음으로 비정상 트래픽 수집의 경우에는 1대의 공격시스템과 4대의 공격 Agent(좀비PC), 그리고 네트워크 트래픽을 탐지할 수 있는 ATMSim이 설치된 피해시스템 1대로 구성된다. 실험에서 사용된 공격 Agent는 Netbot6.0[18]에 의해 검색된 Agent로 좀비 시스템이다. 공격 시스템에는 프로토콜별 공격이 가능한 Netbot6.0이 설치되어 있어 피해시스템을 대상으로 DDoS공격을 하게 된다. 이에 사용된 각 PC 시스템의 사양은 위의 표 2와 같다.

표 2와 같이 공격자의 OS는 Windows로 명령을 내릴 수 있는 tool의 화면을 볼 수 있다. 공격을 하는 Agent는 Linux, Windows 등 다양한 OS가 설치되어 있으며, 각각의 하드웨어 사양도 다양하다. 마지막으로 피해시스템은 앞서 언급했듯이 외부 LAN과 연결된 KBU 내부 호스트 중 하나를 선택하여 ATMSim을 설치하여 사용하였다.

4.1.2 UDP Flooding Attack을 이용한 실험

제안한 시스템이 실시간으로 발생하는 비정상 트래픽을 탐지할 수 있는지 검증하기 위해 공격자 PC에서 Netbot6.0에 의해 검색된 4대의 Agent을 이용해 외부 네트워크에서 KBU 캠퍼스 내부 ATMSim이 설치 된 피해자 호스트로 UDP flooding attack을 실행하였다. UDP flooding attack은 하루 중 임의의 시각에 실행되도록 미리 설정 되었고, 테스트에서는 현실적으로 심각한 피해를 초래하는 공격은 직접 실험해 볼 수 없기 때문에 단시간 동안 각 Agent가 보낼 수 있는 최대 rate으로 공격을 수행하였다. 이에 대한 자세한 사항은 아래의 표 3과 같다.

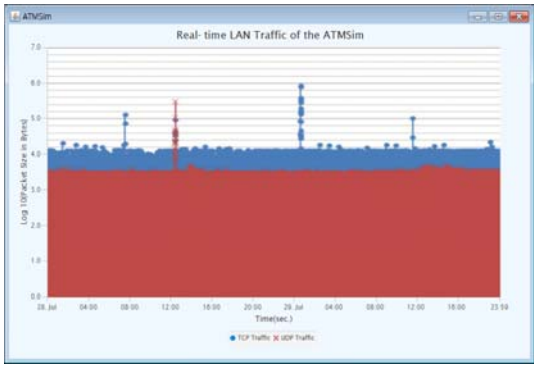
(표 3) UDP Flooding Attack 시나리오
(Table 3) Scenario for UDP flooding attack.

수집	수집 기간	2015.7.25. 16:09:44 ~ 2015.7.27. 16:09:43 (48시간)
	수집 패킷 종류	UDP Packet 및 TCP Packet
네트워크 공격	좀비 PC 수	4대
	공격 기법	Non-Spoofed UDP Flood
	공격 특징	Source IP를 변조하지 않음 고정 Packet Size(UDP 10 Bytes)
	총 공격 횟수	8 회
	각 공격 시간	[1차] 2015.7.25. 18:33:10 [2차] 2015.7.25. 23:00:20 [3차] 2015.7.26. 01:01:00 [4차] 2015.7.26. 11:40:40 [5차] 2015.7.26. 18:33:10 [6차] 2015.7.26. 23:00:20 [7차] 2015.7.27. 01:01:00 [8차] 2015.7.27. 11:40:40
	좀비 PC 당 공격 트래픽 양	전송 가능한 최대 rate

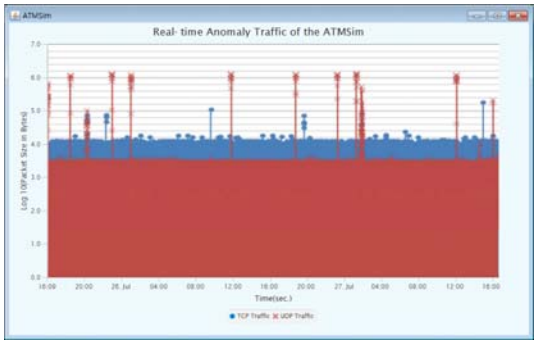
4.2 트래픽 분석

그림 7와 그림 8에서 보는 바와 같이 정상적인 LAN 트래픽과 공격이 포함된 비정상 트래픽 간에는 시각적으로 큰 차이를 보여 주고 있다. 또한 비정상 트래픽을 추정하기 위해서 periodogram plot, R/S-statistic plot과 wavelet-based H 추정기법을 사용하여 분석하였다.

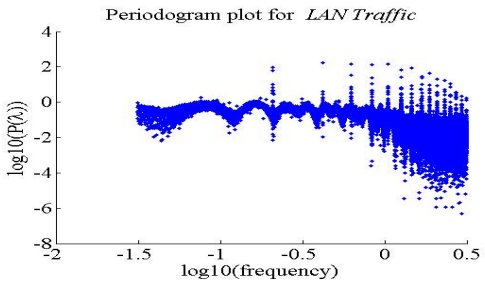
자기유사성의 정도가 높을수록 Hurst 파라메터 값이 점점 커지는 특성이 있으며, 주파수가 0으로 가까워 질 때, power spectrum의 기울기를 사용하는 periodogram plot으로부터 얻은 LAN 트래픽과 비정상 트래픽의 추정된 Hurst 파라메터 값은 각각 0.616과 1.051로 그림 9, 10에서 보는 바와 같이 나타났다. Power law를 따르는 R/S-statistic plot 방법은 y축의 log10(R/S) 값들은 x축의 log10(lags m)이 커짐에 따라 기울기가 1/2과 1사이에 나타나며, 추정치 H값은 각각 0.597과 0.631로 그림 11, 12에서 보는 바와 같이 나타났다. 그리고 점추정법의 하나인 wavelet-based H estimator 방법은 Gaussian이라는 가정 없이 일반적인 환경에서도 쓸 수 있는 추정방법으로 신뢰구간을 가지며, LAN 트래픽과 비정상 트래픽으로부터 추정된 Hurst 파라메터 값은 각각 0.734[0.728, 0.740]과 1.261[1.256, 1.266]으로 그림 13, 14에서 보는 바와 같이 나타났다. 이는 UDP flooding attack 시나리오에 의해서 생성된 비정상 트래픽의 추정된 Hurst 파라메터 값이 LAN 트래픽보다 많은 차이가 있음을 정량적으로 보여주고 있다.



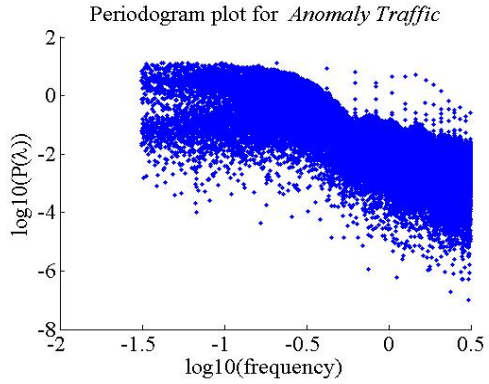
(그림 7) ATMSim을 이용한 정상 트래픽 그래프
(Figure 7) Real-time LAN Traffic of ATMSim.



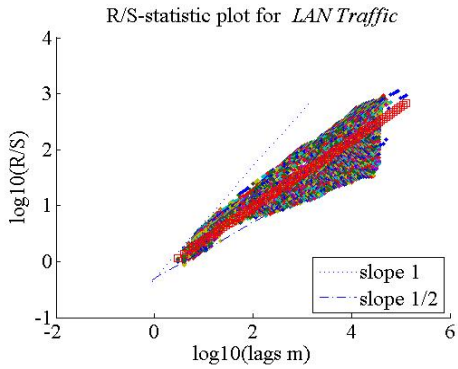
(그림 8) ATMSim을 이용한 비정상 트래픽 그래프
(Figure 8) Real-time Anomaly Traffic of ATMSim.



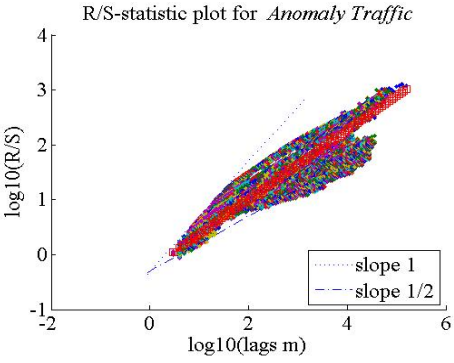
(그림 9) 정상트래픽의 periodogram plot 결과
(Figure 9) Periodogram plot for real-time LAN traffic.



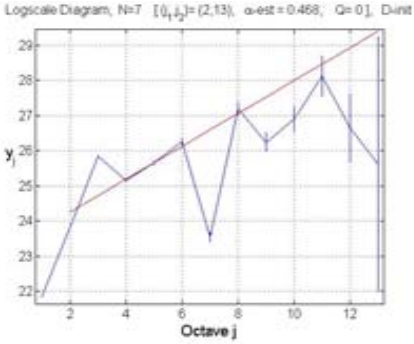
(그림 10) 비정상트래픽의 periodogram plot 결과
(Figure 10) Periodogram plot for real-time Anomaly traffic.



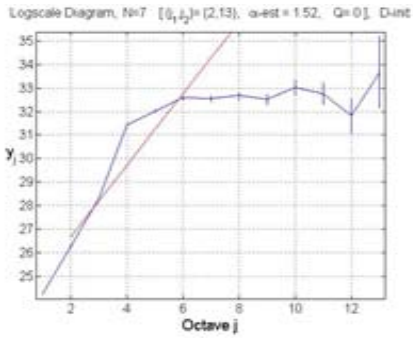
(그림 11) 정상트래픽의 R/S-statistic plot 결과
(Figure 11) R/S-statistic plot for real-time LAN traffic.



(그림 12) 비정상트래픽의 R/S-statistic plot 결과
(Figure 12) R/S-statistic plot for real-time anomaly traffic.



(그림 13) 정상트래픽의 wavelet-based H estimator 결과
(Figure 13) Wavelet-based H estimator for real-time LAN traffic.



(그림 14) 비정상트래픽의 wavelet-based H estimator 결과
(Figure 14) Wavelet-based H estimator for real-time anomaly traffic.

5. 결 론

본 논문은 최근 기하급수적으로 늘어나고 있는 인터넷 침해사고와 네트워크 공격 중 대표적인 DDoS 공격에 대해서 탐지 및 분석했다. 이를 위해 네트워크 플로우 정보를 바탕으로 동작하는 기존의 ATMSim 분석 패키지의 기능들과 GUI를 추가로 개선하였고, 이를 이용하여 캠퍼스 내부 LAN을 통해 대량으로 유입되는 정상적인 트래픽과 DDoS공격이 포함된 비정상 트래픽을 생성하였다.

수집 • 생성된 정상 • 비정상 트래픽의 특성을 분석하기 위해서 자기유사성 추정 기법을 이용하여, 그리팩 분석 및 Hurst 파라메터 추정량 분석결과 정상 트래픽과 비정상 트래픽이 자기유사성 관점에서 정량적으로 많은 차이가 있음을 보여 주었다.

참 고 문 헌 (References)

[1] W.-C. Kang, Y.-H. Lee, Y.-S. Lee, "A Hadoop-based Traffic Analysis System Architecture for Multiple Users," Proceedings of KIISE, vol. 38, no. 1D, pp.252-255, 2011. http://www.dbpia.co.kr/Journal/ArticleDetail/NODE01680166

[2] B.-M. Choi, J.-H. Kong, M.-M. Han, "The Model of Network Packet Analysis based on Big Data," Journal of Korean Institute of Intelligent Systems, vol. 23, no. 5, pp.392-39, Oct. 2013. http://www.riss.kr/link?id=A99799696

[3] T.-K. Ju, C.-M. Hong, W. Shin, "A Monitoring Tool for Personal Information Leakage Prevention in Network Packets," Journal of Information Processing Systems, vol. 2, no. 11, pp.489-494, 2013.
http://www.riss.kr/link?id=A99920290

[4] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," IEEE ACM Transactions on Networking, vol. 2, no. 1, pp. 1 – 15, 1994.
http://ecee.colorado.edu/~ecen5032/handouts/94LelandSelfSim.pdf

[5] H.-D. Jeong, J.-S. Lee, Pawlikowski, K. and McNickle, D. "Comparison of Various Estimators in Simulated FGN," Simulation Modelling Practice and Theory. vol.15, pp. 1173-1191, Oct. 2007.
http://www.sciencedirect.com/science/article/pii/S1569190X07001013

[6] H.-D. Jeong, J.-S. Lee, D. McNickle, K. Pawlikowski, Self-Similar Properties of Malicious Teletraffic, International Journal of Computer Systems Science and Engineering 28(1) (2012) 1-7.
http://dblp.uni-trier.de/db/journals/csse/csse27.html#LeeMPJ12

[7] M.-S. Kim, "Internet application traffic monitoring and analysis," PhD Thesis, Dept. of Computer Science and Engineering, Pohang University of Science and Technology (POSTECH), 2004.
http://www.riss.kr/link?id=T13645544

[8] J.-S. Lee and S.-K. Ye, H.-D. Jeong, "ATMSim: an Anomaly Teletraffic Detection Measurement Analysis Simulator," Simulation Modelling Practice and Theory, vol. 49, pp.98-109, 2014.
http://www.riss.kr/link?id=O64187481

[9] H.-J. Lee, "Uitilization of Big Data Hadoop Platform," Journal of KICS, vol. 29, no. 11, pp.43-47, 2012.
http://www.riss.kr/link?id=A100392834

[10] J.-P. Lee, "Security framework of big data distributed processing environment using Hadoop," Hannam University, 2014.
http://www.riss.kr/link?id= T13378318

[11] C.-B. Kim, J.-P. Chung, "Processing Method of Mass Small File Using Hadoop Platform," Journal of KONI, vol. 18, no. 4, pp.401-408, 2014.
http://www.riss.kr/link?id=A100111693

[12] X. Su, G. Swart, "Oracle in-database Hadoop: When MapReduce Meets RDBMS," in: SIGMOD '12: Proceedings of the 2012 International Conference on Management of Data, pp. 779-790, 2012.
http://www.cs.yale.edu/homes/xs45/pdf/ss-sigmod2012.pdf

[13] M.-J. Ji, E.-K. Cho, S.-R. Kim, I.-S. You, H.-D. Jeong, "Setting Rules for a Fraud Detection System by Applying ATMSim in Mobile Internet Environment," Proceedings of KSII, vol. 16 no. 1, 2015.
http://www.riss.kr/link?id=A100503751

[14] Financial Security Agency, "Response Manual for the Different Types of DDoS Attacks," 2008.

[15] Kaspersky. Lab, "Kaspersky DDoS Intelligence Report Q2 2015" Kaspersky, Aug. 2015.
https://securelist.com/analysis/quarterly-malware-reports/71663/kaspersky-ddos-intelligence-report-q2-2015/

[16] Chris. Plante, "Valve's $18 million Dota 2 tournament delayed by DDoS attack," THEVERGE, Aug. 2015.

[17] Cisco Systems, "NetFlow Services and Applications," White Papers.
http://www.cisco.com/warp/public/cc/pd/iosw/ioft/neflct/tech/napps_wp.htm

[18] Netbot, http://www.hackeroo.com.

◐ 저 자 소 개 ◑

정 해 덕 (Hae-Duck J. Jeong)
2003년 University of Canterbury, Department of Computer Science and Software Engineering, New Zealand(공학박사)
2008년 - 2010년 Life University, College of Mathematics and Science, Cambodia, 겸임학장
2006년, 2008년, 2010년 University of Canterbury, Department of Computer Science and Software Engineering, New Zealand, 방문교수
2004년~현재 한국성서대학교 컴퓨터소프트웨어학과 교수
관심분야 : 컴퓨터시뮬레이션, 트래픽 모델링
E-mail : joshua@bible.ac.kr

류 명 운 (Myeong-Un Ryu)
2010년~현재 한국성서대학교 컴퓨터소프트웨어학과 재학
관심분야 : 웹 프로그래밍, 모바일 플랫폼
E-mail : aparecium117@gmail.com

지 민 준 (Min-Jun Ji)
2011년~현재 한국성서대학교 컴퓨터소프트웨어학과 재학
관심분야 : 웹 프로그래밍, 모바일 플랫폼
E-mail : inciojs@gmail.com

조 유 빈 (You-Been Cho)
2013년~현재 한국성서대학교 컴퓨터소프트웨어학과 재학
관심분야 : 데이터베이스, 빅 데이터
E-mail : yubin1020@gmail.com

예 상 국 (Sang-Kug Ye)
2015~현재 SK엠앤서비스 재직
관심분야 : 딥 러닝, 인공지능
E-mail : ysk@sk.com

이 종 숙 (Jong-Suk R. Lee)
2001년 University of Canterbury, Department of Computer Science and Software Engineering, New Zealand (공학박사)
2002년~현재 한국과학기술정보연구원(KISTI) 국가슈퍼컴퓨팅연구소 계산과학공학연구실(실장)
2005년~현재 한국과학기술연합대학원대학교(UST) 슈퍼컴퓨팅 전공 (겸임 정교수)
관심분야 : 분산컴퓨팅, SW, 미들웨어, 스마트 러닝
E-mail : jsruthlee@kisti.re.kr