트래픽 폭주 공격 탐지 시스템의 의미론적 해석

유재학, 오승근, 이한성, 박준상, 김명섭, 박대희
고려대학교 컴퓨터정보학과
e-mail:{dbzzang, gmo85, mohan, runtoyou, tmskim, dhpark}@korea.ac.kr

Semantic Analysis on Traffic Flooding Attacks Detection System

Jaehak Yu, Seunggeun Oh, Hansung Lee, Jun-Sang Park, Myung-Sup Kim, Daihee Park
Dept. of Computer & Information Science, Korea University

요 약

DoS/DDoS로 대표되는 트래픽 폭주 공격은 대상 시스템뿐만 아니라 네트워크 대역폭 및 시스템 자원 등을 고갈시킴으로써 네트워크에 심각한 장애를 유발하기 때문에, 신속한 공격 탐지와 공격유형별 분류는 안정적인 서비스 제공 및 시스템 운영에 필수요건이다. 본 논문에서는 1) 데이터마이닝의 대표적인 분류 모델인 C4.5 알고리즘을 기반으로 SNMP MIB 정보를 사용하여 트래픽 폭주공격을 탐지하고 각 공격유형별 분류를 수행하는 시스템을 설계 및 구현하였다; 2) C4.5에서 추가적으로 제공하는 동작원리에 관한 규칙들을 상세히 분석함으로써 공격탐지 및 공격유형별 분류에 관한 시스템의 의미론적 해석을 시도하였다; 3) C4.5는 주어진 SNMP MIB의 속성들의 정보이익 값을 이용하여 예측모형을 구축하는 알고리즘으로, 특징선택 및 축소의 효과를 추가적으로 얻었다. 따라서 시스템의 운용 시, 제안된 모델은 전체 13개의 MIB 정보 중 5개의 MIB 정보만을 사용하여 보다 신속하고, 정확하며, 또한 가벼운 공격탐지 및 공격유형별 분류를 수행함으로써 네트워크 시스템의 자원관리와 효율적인 시스템 운영에 기여하였다.

1. 서론

DoS/DDoS로 대표되는 트래픽 폭주 공격은 대상이 되는 컴퓨터 시스템은 물론 네트워크의 자원을 고갈시킴으로써 정상적인 서비스를 수행하지 못하게 하는 공격으로 업무에 막대한 피해를 준다. 이러한 악의적인 접근이나 침입 등을 신속하게 탐지하고 대처할 수 있는 보안 기술이 학계의 최근 중요한 이슈 중 하나이다[1-2].

트래픽 폭주 공격 탐지에서의 전통적인 패킷 수집 방법[1-2]은 공격에 대한 상세한 분석은 가능하나 고가의 고성능 분석시스템이 요구될 뿐만 아니라 설치 및 운영상의 확장성이 부족하다는 단점을 가지고 있다. 따라서 이를 보완하기 위한 방법으로 최근 SNMP에서의 MIB 정보를 이용한 침입탐지 방법론[1-2]이 주목을 받고 있다. SNMP MIB 정보를 이용한 트래픽 폭주 공격 탐지는 MIB 데이터 수집을 위한 시스템 및 네트워크 리소스의 사용이 적고, 표준화된 네트워크 성능 데이터를 제공 받을 수 있기 때문에 패킷 기반 탐지 방법에 비해 보다 빠르고 효과적인 탐지가 가능하다[1-2].

SNMP MIB 정보를 이용하는 DDoS 탐지 방법은 프로토콜별 추이분석, 일주 트래픽 추이분석, 그리고 MIB에서의 특정 객체와 객체 정보간의 상관관계를 이용하는 방법 등으로 분류된다[1-2]. 그러나 이러한 방법론들은 대부분 테스트에 사용된 공격툴의 기능과 특성에 의존적으로 개발된 시스템으로, 새로운 공격유형이나 끊임없이 발전하는 공격툴에 유연하게 대처하기 어렵다. 즉 새로운 공격 형태나 툴이 발견되면 그때마다 새롭게 알고리즘 전체를 수정해야하는 단점을 가지고 있다. 따라서 최근 학계에서는 전술한 문제점의 해결 방안을 데이터마이닝 및 기계학습 기법에서 찾고자 하는 시도가 활발히 진행 중이다.

최근의 연구문헌 조사에 의하면, 기계학습 기법과 SNMP MIB 정보를 이용한 매우 흥미로운 몇 개의 침입탐지 시스템이 발표되었다: Li 등[2]은 SNMP MIB-II 데이터를 probability density function으로 변환한 후, backpropagation 기반의 인공신경망을 이용하여 침입 여부를 결정하는 시스템을 제안하였다. Puttini 등[3]은 SNMP MIB 데이터를 Bayesian 분류기에 적용하여 Mobile Adhoc NETworks(MANET)에서의 비정상 트래픽을 탐지하였다. Ramah 등[4]은 Shyu 등[5]이 제안한 Principle Component Analysis(PCA) 기반의 비정상 탐지 알고리즘을 이용한 침입탐지시스템을 제안하였다. 또한 Yu 등[1]은 Support Vector Machine(SVM)을 이용하여 트래픽 폭주공격을 탐지하고 공격유형별 분류를 수행하는 시스템을 제안하였다. 그러나 이러한 연구들은 모두 전통적인 DDoS 탐지 방법론의 단점을 해결한다는 기치아래 자칫 전통적 방법론의 장점을 간과할 수도 있다. 즉, 효율적인 시스템의 구축이라는 입장만을 견지하는 위의 기계학습론적 방법론은 시스템 작동 원리의 역학적 해석을 간과하여, 핵심 동작원리를 black-box화 하였다. 따라서 휴리스틱한 방법론이긴 하나 전통적인 DDoS 탐지 방법론의 해석학적 장점도 고려할 수 있는 보다 포괄적인 시스템이 바람직해 보인다.

본 논문에서는 데이터마이닝의 대표적인 분류 모델인 의사결정나무(decision tree) 중, C4.5 알고리즘을 기반으로 SNMP MIB 정보를 사용하여 트래픽 폭주공격을 탐지하고 각 공격유형별 분류를 수행하는 시스템을 설계 및 구현한다. 본 시스템은 C4.5를 기반으로 정상트래픽과 공격트래픽을 빠르게 탐지하는 계층과 탐지된 공격트래픽을 DDoS의 대표적 공격유형인 TCP-SYN flooding, UDP

flooding, ICMP flooding으로 분류하는 계층으로 구성된다. 공격유형에 대한 세분화된 분류는 공격이 발생한 프로토콜에 대해서만 서비스를 제한하고 관리함으로써 보다 안정적인 네트워크 환경과 원활한 자원관리를 지원할 수 있다. 또한 C4.5에서 추가적으로 제공하는 동작원리에 관한 규칙들을 상세히 분석함으로써 공격탐지 및 공격유형별 분류에 관한 시스템의 동역학의 의미론적 해석을 시도한다. 특히 C4.5는 주어진 SNMP MIB의 속성들의 정보이익(information gain) 값을 이용하여 예측모형을 구축하는 알고리즘으로, 특징선택 및 축소(feature selection & reduction)의 효과를 추가적으로 얻을 수 있다. 따라서 시스템의 운용 시, 제안된 모델은 전체 13개의 MIB 정보 중 5개의 MIB 정보만을 사용하여 보다 신속하고, 정확하며, 또한 가벼운 공격탐지 및 공격유형별 분류를 수행함으로써 네트워크 시스템의 자원관리와 효율적인 시스템 운영에 기여할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 C4.5 알고리즘과 본 논문에서 제안하는 계층적 트래픽 폭주 공격 탐지 모델을 소개한다. 3장에서는 실험결과 및 발견된 규칙에 대한 의미론적 해석을 기술하며, 마지막으로 4장에서는 결론 및 향후 연구과제에 대해 논한다.

2. 트래픽 폭주 공격 탐지 및 분류 모델

2.1 C4.5 의사결정나무

의사결정나무는 데이터마이닝 분야에서 분류 및 예측문제에 자주 사용되는 기법으로 변수들 간에 미치는 영향이나 상호작용을 누구나 쉽게 이해할 수 있는 방법론이다. 이는 신경망구조 분석과는 달리 얻어진 지식을 표현하는 것이 직관적이며 손쉽게 규칙을 생성하기 때문에 분류 및 예측 모형을 제시하는데 주로 사용되어 왔다[6-8]. 대부분의 의사결정 알고리즘은 하향식의 분할-정복(divide and conquer) 방법을 따르고 있으며, 훈련 데이터와 연관된 클래스 레이블로부터 모형을 구축한다. 의사결정나무의 대표적 알고리즘인 ID3는 많은 범위의 값을 갖는 속성이 상위 노드로 선택되는 단점을 가지고 있기 때문에 본 논문에서는 가장 진보되고 분류 및 예측 성능이 이미 검증된 C4.5 의사결정나무 알고리즘[6-7]을 사용하였다.

2.2 계층적 트래픽 폭주 공격 탐지 및 분류 시스템

본 논문에서 제안하는 계층적인 트래픽 폭주 탐지 시스템은 총 4개의 모듈로 구성된다. 즉, 1개의 오프라인 처리 모듈인 C4.5 training 모듈과 3개의 온라인 처리 모듈인 MIB update detection, MIB data collection, flooding attack detection 모듈로 구성된다(그림 1 참조). 1) C4.5 training 모듈에서는 다양한 트래픽 공격을 임의로 발생시켜 C4.5 기반의 학습을 실시한다; 2) MIB update detection 모듈은 ifInOctets MIB을 수집하여 탐지 시스템의 동작 시점을 결정하고 MIB data collection 모듈을 실행 시킨다; 3) MIB data collection 모듈은 MIB 정보를 타깃 시스템으로부터 C4.5 training 모듈에서 결정된 MIB 정보만을 선택한다; 4) 수집된 정보는 flooding attack detection 모듈로 전달되어 공격유무와 공격유형을 판단한다(그림 2 참조).

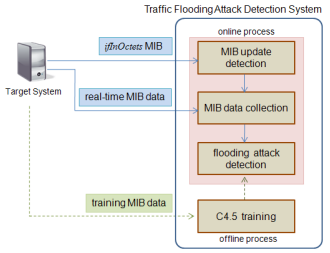

그림 1. 트래픽 폭주 공격 탐지 시스템의 전체 구조도

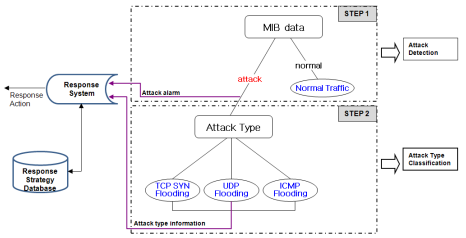

그림 2. C4.5 기반 탐지 모듈의 계층적 탐지 절차

제안된 flooding attack detection 모듈의 각 계층별 기능은 다음과 같다. 첫 번째 계층은 정상트래픽과 공격트래픽을 분류하는 계층으로써 공격트래픽이 탐지되면 침입대응 시스템에 침입 사실을 실시간으로 보고한다. 두 번째 계층은 트래픽 폭주 공격으로 판단된 공격트래픽을 TCP-SYN flooding, UDP flooding, ICMP flooding으로 각각 분류하고 침입 대응 시스템에 공격유형에 대한 추가적인 정보를 제공한다. 트래픽 폭주 공격을 유형별로 분류함으로써 공격이 발생한 프로토콜에 대해서만 서비스를 제한하고 관리할 수 있으므로 보다 안정적인 네트워크 환경과 원활한 자원관리를 지원할 수 있다.

3. 실험 및 결과 분석

본 논문에서는 java 기반의 machine learning tool인 weka[9]를 사용하였으며, RFC1213[10]에서 정의한 mib-2 그룹의 MIB 객체들 중, 실제 트래픽 폭주 공격에 반응하는 13개의 MIB 객체들만을 선정하였다. 트래픽 폭주 공격의 대표적 공격 툴인 Stacheldraht[11]를 이용하여 TCP-SYN flooding 공격, UDP flooding 공격, ICMP flooding 공격 등을 타깃 시스템에 실시하였다. 본 논문의 실험에서 사용된 MIB 객체들을 [표 1]에 정리하였다.

[표 1] 탐지 시스템에서 사용된 MIB 객체들[10]

mib-2 group	SNMP MIB objects	MIB object description
ip	ip.*ipInReceives*	인터페이스로부터 받은 ip 데이터그램의 총 개수
	ip.*ipInDelivers*	수신된 ip 데이터그래 중 상위계층으로 전달된 데이터그램의 총 개수
	ip.*ipOutRequests*	상위계층에서 송신 요청한 ip 데이터그램의 개수
	ip.*ipOutDiscards*	정상적으로 송신 요청된 ip 데이터그램 중 buffer overflow등에 의해 drop되는 ip 데이터그램의 개수
tcp	tcp.*tcpAttemptFails*	비정상적으로 tcp connection이 종료된 횟수

	tcp.*tcpOutRsts*	RST flag를 포함하여 송신된 tcp 세그먼트의 개수
udp	udp.*udpInErrors*	udp 패킷 구성오류에 의해 전달되지 못한 데이터그램의 개수
icmp	icmp.*icmpInMsgs*	수신된 icmp 메시지의 총 개수
	icmp.*icmpInDestUnreachs*	수신된 icmp destination unreachable 메시지의 총 개수
	icmp.*icmpInEchos*	수신된 icmp echo 메시지의 총 개수
	icmp.*icmpOutDestUnreachs*	송신된 icmp destination unreachable 메시지의 개수
	icmp.*icmpOutEchoReps*	송신된 icmp echo response 메시지의 개수
	icmp.*icmpOutMsgs*	송신된 icmp 메시지의 총 개수

첫 번째 실험은 공격트래픽의 신속한 탐지와 이에 대한 규칙들을 의미론적으로 분석하는 실험으로써, 정상트래픽 1000개와 공격트래픽은 유형별로 500개씩 랜덤하게 추출하여 학습하였고, 학습에 참여하지 않은 정상트래픽 1000개와 공격유형별 트래픽 500개씩을 테스트 하였다. 성능 측정을 위하여 침입 탐지율(detection rate), false positive rate(FPR) 및 false negative rate(FNR)[1]를 성능지표로 사용했으며 실험결과는 [표 2]에 정리하였다. [표 2]의 실험 결과에 의하면, 만족스러운 침입 탐지율과 안전한 FNR을 보여줌을 확인할 수 있다.

[표 2] 트래픽 폭주 공격 탐지의 성능 측정 표

침입 탐지율	FPR	FNR
99.13	0.3	0.87

그림 3은 정상트래픽과 공격트래픽의 분류를 위한 의사결정나무이다. 의사결정나무의 구성 시, 시스템에서 정의된 13개의 MIB 객체 중 4개의 객체만으로도 정상트래픽과 공격트래픽이 정확하게 분류됨을 알 수 있다. 여기서 leaf node안의 수치는 학습데이터에 의해 해당 클래스로 분류된 트래픽의 개수와 오분류된 트래픽의 개수를 의미한다.

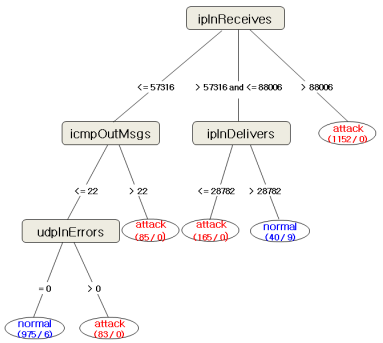

그림 3. 정상 및 공격트래픽 분류를 위한 의사결정나무

그림 3으로부터 6개의 규칙을 얻을 수 있다([표 3] 참조).

[표 3] 정상/공격트래픽을 분류하는 규칙들

규칙 번호	규칙 내용
1	IF ipInReceives > 88006 THEN attack
2	IF 57316 < ipInReceives <= 88006 AND ipInDelivers > 28782 THEN normal
3	IF 57316 < ipInReceives <= 88006 AND ipInDelivers <= 28782 THEN attack
4	IF ipInReceives <= 57316 AND icmpOutMsgs > 22 THEN attack
5	IF ipInReceives <= 57316 AND icmpOutMsgs <= 22 AND udpInErrors > 0 THEN attack
6	IF ipInReceives <= 57316 AND icmpOutMsgs <= 22 AND udpInErrors = 0 THEN normal

위 [표 3]의 규칙들을 의미론적으로 분석한 결과를 [표 4]에 정리하였다.

[표 4] 정상/공격트래픽 분류 규칙들의 의미론적 해석

규칙 번호	규칙에 대한 의미론적 해석
1	트래픽 폭주 공격은 대량의 패킷을 전송하기 때문에 공격이 발생하면 대량의 데이터그램이 발생한다. 따라서 ipInReceives 값이 임계값 88006보다 큰 1152개가 공격트래픽으로 분류된다.
2	전송받은 IP 패킷이 IP 계층에서 상위계층으로의 전달률이 50%(ipInDelivers/ipInReceives) 이상일 경우 대부분 정상 트래픽으로 분류된다.
3	트래픽 폭주 공격이 발생할 경우 전송받은 데이터그램이 상위계층으로 전달되지 못하는 현상이 빈번히 발생한다. 본 규칙에 의하면 전송받은 데이터그램의 전달율이 50%에도 (ipInDelivers/ipInReceives) 미치지 못하기 때문에 공격 트래픽으로 분류된다.
4	전송받은 데이터그램의 개수는 적지만, icmp 응답메시지인 icmpOutMsgs 값이 대량으로 발생되었기 때문에 트래픽 폭주 공격으로 분류되며, 보다 구체적으로 ICMP flooding 공격이다.
5	udpInErrors가 발생하였기 때문에 UDP flooding 공격으로 판단된다. udpInErrors는 패킷 구성의 오류로 인해서 발생되는 값이기 때문에 현재의 operating system을 고려한다면 정상적인 트래픽은 패킷 구성의 오류를 발생시키지 않는다. 따라서 udpInErrors의 발생 유무를 통해 공격을 판단할 수 있다.
6	전송받은 데이터그램의 개수가 적고, ICMP flooding 공격일 경우 icmp 응답메시지가 대량으로 발생하는데 이 값이 22 이하이기 때문에 공격으로 판단되지 않으며, udpInErrors의 반응이 없기에 정상적인 트래픽으로 분류된다.

위 [표 4]의 규칙들을 종합적으로 분석해보면 데이터그램의 개수를 반영하는 ipInReceives의 범위에 따라 세부적인 추가 조건으로 공격 유무를 판단할 수 있다. 특히 대부분의 트래픽 폭주 공격은 대량의 패킷을 전송하기 때문에 ipInReceives 값이 88006 이상일 경우 정확히 공격으로 판단됨을 확인하였다.

두 번째 실험은 DDoS의 대표적 공격유형별로 분류하고 이에 대한 규칙들을 의미론적으로 분석하는 실험으로써 공격유형별로 랜덤하게 500개씩 학습하였으며, 학습에 참여하지 않은 공격유형별 트래픽 500개로 테스트 하였다. 1500개의 공격트래픽 중 TCP-SYN flooding 공격트래픽 11개, UDP flooding 공격트래픽 2개가 정상트래픽으로 분류되었기에 실제 분류 테스트에 참여한 공격트래픽은 총 1487개이다. 성능 측정을 위하여 분류 정확도(classification accuracy)[1]를 성능지표로 사용했으며 실험결과는 아래의 [표 5]에 정리하였다.

[표 5] 트래픽 폭주 공격유형별 분류 정확도

TCP-SYN flooding	UDP flooding	ICMP flooding	전체 분류정확도
100.0	100.0	100.0	100.0

그림 4는 공격유형별 분류를 위한 의사결정나무로써 단지 2개의 속성만으로 공격유형을 정확하게 분류함을 보여준다. 그림 4로부터 3개의 규칙을 얻을 수 있다([표 6] 참조).

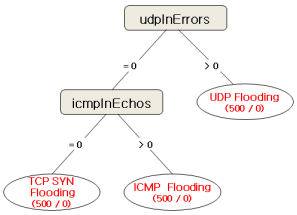

그림 4. 공격유형별 분류를 위한 의사결정나무

[표 6] 공격유형별 분류를 위한 규칙들

규칙 번호	규칙 내용
1	IF udpInErrors > 0 THEN UDP flooding
2	IF udpInErrors = 0 AND icmpInEchos > 0 THEN ICMP flooding
3	IF udpInErrors = 0 AND icmpInEchos = 0 THEN TCP-SYN flooding

위 [표 6]의 공격유형별 분류를 위한 규칙들을 의미론적으로 해석한 결과를 아래의 [표 7]에 정리하였다.

[표 7] 공격유형별 분류를 위한 규칙들의 의미론적 해석

규칙 번호	규칙에 대한 의미론적 해석
1	udpInErrors는 udp 패킷 구성오류에 의해 증가되는 값으로 UDP flooding 공격에만 반응한다. 따라서 UDP flooding 공격을 정확히 분류한다.
2	icmpInEchos는 수신된 icmp 메시지의 개수를 의미하는 값으로 TCP-SYN flooding, UDP flooding 공격에 영향을 받지 않기 때문에 ICMP flooding 공격을 정확히 분류한다.
3	TCP에 종속적인 MIB을 사용하지 않고도 UDP flooding, ICMP flooding 공격이 아닌 공격트래픽을 정확히 TCP-SYN flooding 공격으로 분류한다.

위 [표 7]의 규칙들은 DDoS의 대표적 공격유형인 TCP-SYN flooding, UDP flooding, ICMP flooding을 단지 2개의 MIB 속성만으로도 신속하고 정확하게 분류함으로써 네트워크의 부하를 최소화하고, 시스템을 효율적으로 운용할 수 있음을 보여준다.

4. 결 론

본 논문에서는 1) 데이터마이닝의 대표적인 분류 모델인 C4.5 알고리즘을 기반으로 SNMP MIB 정보를 사용하여 트래픽 폭주공격을 탐지하고 각 공격유형별 분류를 수행하는 시스템을 설계 및 구현하였다; 2) C4.5에서 추가적으로 제공하는 동작원리에 관한 규칙들을 상세히 분석함으로써 공격탐지 및 공격유형별 분류에 관한 시스템의 의미론적 해석을 시도하였다; 3) C4.5는 주어진 SNMP MIB 속성들의 정보이익 값을 이용하여 예측모형을 구축하는 알고리즘으로, 특징선택 및 축소의 효과를 부수적으로 얻을 수 있다. 따라서 시스템의 운용 시, 제안된 모델은 전체 13개의 MIB 정보 중, 5개의 MIB 정보만을 사용하여 보다 신속하고, 정확하며, 또한 가벼운 공격탐지 및 공격유형별 분류를 수행함으로써 네트워크 시스템의 자원관리와 효율적인 시스템 운영에 기여하였다.

특히, 본 논문에서 시도되었던 의미론적 해석은 이제까지 시도되지 않았던 참신한 연구방향으로 향후 기대되는 공헌 가능성이 매우 크다고 할 수 있다. 그러나 의사결정나무계통의 알고리즘은 그 성격상 예측모형 알고리즘으로써, 완벽한 의미론적 해석을 제공치는 못한다. 따라서, 향후 연구 과제로는 공격 및 유형별 분류에 관한 메카니즘에 내재되어 있는 유용한 지식의 발견과 분석에 관한 연구의 도구로써, 연관관계규칙기법(association rule mining)을 사용하여 보다 심층적인 연구를 수행하고자 한다.

참고문헌

[1] J. Yu, H. Lee, M. Kim, and D. Park, “Traffic flooding attack detection with SNMP MIB using SVM”, Computer Communications, In Press.

[2] J. Li and C. Manikopoulos, “Early statistical anomaly intrusion detection of DOS attacks using MIB traffic parameters”, Information Assurance Workshop, IEEE, pp. 53-59, 2003.

[3] R. Puttini, M. Hanashiro, F. Miziara, R. Sousa, L. García-Villalba, and C. Barenco, “On the anomaly intrusion-detection in mobile adhoc network environments”, Proc. of PWC 2006, LNCS 4217, pp. 182-193, 2006.

[4] K. Ramah, H. Ayari, and F. Kamoun, “Traffic anomaly detection and characterization in the Tunisian national university network”, Proc. of Networking 2006, LNCS 3979, pp. 136-147, 2006.

[5] M. Shyu, S. Chen, K. Sarinnapakorn, and L. Chang, “A novel anomaly detection scheme based on principal component classifier,” Proc. of the IEEE Foundations and New Directions of Data Mining Workshop, pp. 172-179, Melbourne, Florida, USA, 2003.

[6] 정광본, 최미정, 김명섭, 원영준, 홍원기, “ML 알고리즘을 적용한 인터넷 애플리케이션 트래픽 분류”, KNOM Review, Vol. 10, No. 2, pp. 39-52, 2007.

[7] S. Ruggieri, “Efficient C4.5”, IEEE Transactions on Knowledge and Data Engineering, Vol. 14, No 2, pp. 438-444, 2002.

[8] J. Han and M. Kamber, Data Mining: Concept and Techniques, Morgan Kaufmann Publishers, 2nd Ed., pp. 291-310, 2007.

[9] Machine Learning Lab in The University of Waikato, http://www.cs.waikato.ac.nz/ml.

[10] IETF RFC 1213, “Management information base for network management of TCP/IP-based internets: MIB-II”, http://www.rfc-editor.org/rfc/rfc1213.txt.

[11] “Distributed denial of service (DDoS) attacks tools”, http://staff.washington.edu/dittrich/misc/ddos/.