데이터마이닝을 이용한 DDoS 예측 모델링

김종민* · 정병수**

요 약

최근 인터넷 등 정보통신 기술의 발달로 인해 언제 어디서나 인터넷을 이용할 수 있는 환경이 구축 되었으며, 이로 인한 사이버위협은 다양한 경로를 통해 시도되고 있다. 본 연구에서는 사이버위협 중 지속적으로 증가 추세인 DDoS 예측 모델링하기 위해 이벤트 데이터를 근거로 하여 통계적 기법을 통해 DDoS 위험지수 예측식을 도출하였고, 도출된 위험지수를 정량화하였다. 제시된 위험지수를 활용하여 DDoS 위협에 대해 사전 대응정책을 세움으로써 피해를 최소화 시킬 수 있는 객관적이고 효율적인 예측 모델이 될 것으로 기대한다.

DDoS Prediction Modeling Using Data Mining

Jong-Min Kim* · Byung-soo Jung**

ABSTRACT

With the development of information and communication technologies like internet, the environment where people are able to access internet at any time and at any place has been established. As a result, cyber threats have been tried through various routes. Of cyber threats, DDoS is on the constant rise. For DDoS prediction modeling, this study drew a DDoS security index prediction formula on the basis of event data by using a statistical technique, and quantified the drawn security index. It is expected that by using the proposed security index and coming up with a countermeasure against DDoS threats, it is possible to minimize damage and thereby the prediction model will become objective and efficient.

Key words : DDoS, Security Threats, Malicious Code, Prediction Modeling, data mining

접수일(2016년 3월 28일), 게재확정일(2016년 3월 30일)

* 경기대학교 융합보안학과
** 세한대학교 경찰행정학과(교신저자)

1. 서 론

DoS(Denial of Service)공격은 시스템이나 네트워크의 구조적인 취약점을 공격하여 시스템의 성능을 저하시키거나 마비시키는 공격방식이다. DoS공격은 공격 방법이 비교적 간단하고 손쉽게 이루어질 수 있다는 측면에서 상당한 위험요소를 내포하고 있지만, 한 대의 서버를 마비시키기위한 공격 트래픽을 전송하는 것은 공격자 한명이 하기에는 매우 어려운 일이며, 다수 공격자를 확보하는 것도 어려운 일이다[1].

하지만 수많은 컴퓨터들이 네트워크를 이루어 운영되는 현재의 컴퓨터 통신환경은 공격자의 지시를 따르는 악성코드를 유포시켜 DoS공격에 충분한 공격자를 확보하여 공격 가능하다[2,3]. 과거 연구들은 DoS 공격을 방어하기 위한 공격자의 공격기법에 국한되어 연구되어 왔으며, 방어 기법에 적용하기는 한계점을 가지고 있다. 따라서 본 연구에서는 방어 기법에 적용할 수 있는 DDoS 예측식을 도출하기 위해 위협발생 데이터를 근거로 하여 위험지수 예측모델을 제시하고 위협에 대해 보완할 수 있는 기법을 제안하였다.

2. 관련연구

2.1 DDoS

DDoS 공격은 공격자가 일반 사용자들의 PC에 악성코드를 심어 감염시킨 후 감염 된 PC를 이용하여 특정웹 사이트나 서버에 대량의 트래픽을 송신하는 공격방법으로서 공격자가 감염자의 PC를 이용하기 때문에 공격자 스스로가 외부에 노출되지 않으면서도 대량의 데이터를 보낼 수 있다는 특징이 있다. 현재에 들어서는 C&C(Command and Control)서버 등을 이용하여 손쉽게 공격대상을 변경 가능하고 공격 방식 또한 여러 가지 형태로 바꾸는 형태로 계속적으로 진화해 나가고 있다[4].

2.2 CVSS

CVSS(Common Vulnerability Scoring System)는 미국의 주요기반시석자문회의(NIAC : National Infrastructure Advisory Council)의 지원하에 개발되었다. 주요기반시설자문회의는 기국 국토안보부(Department of Homeland Security) 산하의 자문기구로써 국가 핵심기반시설(의료, 운송, 에너지 등), 국가중요정보시스템, 사이버공간에 대한 전반적인 보안관련 사항을 다루고 국토안보부장관을 통해 대통령에게 보안정책을 건의하는 기구이다. CVSS는 1-10까지의 숫자를 사용한 새로운 등급 시스템으로 기업들이 IT 시스템과 관련된 정보를 추가하고, 자체 시스템 환경에 대해서도 특정 위험을 예측할 수 있도록 해주는 것이다. 또한 패치에 대한 우선순위도 선정할 수 있다. 이밖에도 기업이 위험 밸런스에 대한 자체 환경 측정을 추가할 수 있을 뿐 아니라 취약점으로 인해 위험이 발생할 수 있는 공격코드와 보안패치 가용성 등의 요소들도 고려할 수 있도록 되어 있다[5].

2.3 회귀분석

회귀분석은 데이터 분석에서 가장 널리 쓰이는 통계적 분석방법의 하나로서, 종속변수와 독립변수들간의 관계를 밝히는 분석방법으로 선형모형을 통해서 모형을 적합하고, 적합된 모형을 통해서 관련된 현상을 연구할 수 있다. 회귀분석에서 독립변수가 여러개인 경우를 다중회귀분석이라고 한다. 이런 다중회귀분석은 시간의 흐름에 따른 변화를 적용하여 예측기법에 많이 사용되고 있다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_n X_n \qquad (1)$$

여기서 β_0는 절편을 나타내는 회귀계수이고, β_0, … β_k는 종속변수와 독립변수간의 기울기를 나타내는 회귀계수로 다른 독립변수들의 고정되었을 경우 X_1이 Y에 미치는 영향을 나타낸다고 할 수 있다[6][7][8].

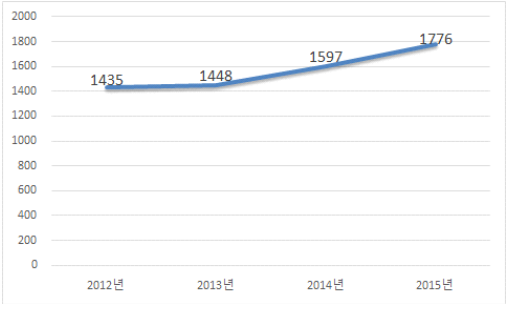
3. DDoS 공격 분석 및 예측 모델 설계

위협 대응에 가장 중요한 핵심은 위협이 가시화되

기 전에 조기에 잠재 위협을 예측하여 사전에 예방 정책을 수립하여 대응할 수 있는 모델이 필요하다. 제안한 모델을 설계하기 위해 DDoS 발생 데이터들을 각각 요소별(Access Vector, Access Complexity, Authentication, Confidentiality Impact, Integrity Impact, Availability Impact)로 나누어 분석하였다.

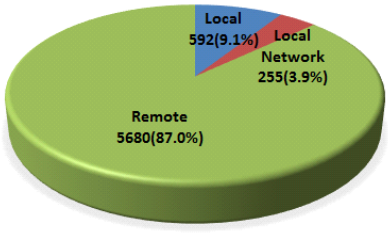
3.1 DDoS 공격 분석

(그림 1)은 본 논문에서 사용될 2009년 1월부터 2015 12월까지의 DDoS 공격 데이터이며, 매년 지속적으로 증가하는 것을 볼 수 있다.



(그림 1) DDoS 발생현황[9]

3.1.1 공격 수행 위치(Access Vector/AV)

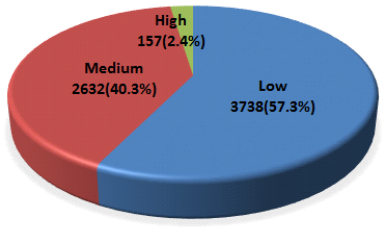


(그림 2) 공격 수행 위치별 DDoS 빈도수

(그림 2)는 공격 수행 위치별 DDoS 빈도수를 나타낸 것으로, Local, Local Network의 공격이 이루어지지 않고 네트워크의 접근으로 원격 공격을 하는 Remote(Network)의 상태가 5,680(87.0%)건으로 가장 많이 나타났으며, 그 다음으로 방화벽, USB DMA 공격, 로컬 권한 상승 등 주변장치를 공격하는 Local의 상태가 592(9.1%)건으로, IP 서브넷, 블루투스 등 인접 네트워크를 통해 취약점을 공격하는 Local Network의 상태 255(3.9%)건의 순서로 나타났다.

공격 수행 위치는 공격자가 멀리 위치할수록 큰 취약점의 상태를 가지는데 Local Network < Local < Remote(Network)와 같이 나타났다.

3.1.2 공격 복잡도(Access Complexity/AC)

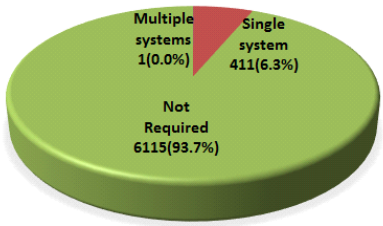


(그림 3) 공격 복잡도별 DDoS 빈도수

(그림 3)은 공격 복잡도별 DDoS 빈도수를 나타낸 것으로, 기술적인 접근 조건은 필요하지만 특별한 접근 조건들이 없어 쉽게 접근이 가능한 Low의 상태가 3,738(57.3%)건으로 가장 많이 나타났으며, 다소 전문성이 있어야 접근 가능한 Medium의 상태가 2,632(40.3%)건, 그 다음으로 시스템에 접근할 때 특별한 접근 조건을 요구하는 High의 상태가 157(2.4%)건의 순서로 나타났다.

공격 복잡도는 공격자가 취약점 공격을 위해 접근할 때 특별한 접근 조건들이 미비할 때 큰 취약점의 상태를 가지며, High < Medium < Low 와 같이 나타났다.

3.1.3 인증(Authentication)여부

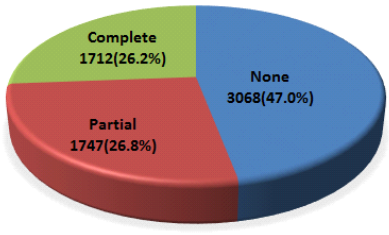


(그림 4) 인증여부별 DDoS 빈도수

(그림 4)는 인증여부별 DDoS 빈도수를 나타낸 것으로, 공격자가 취약점 공격을 할 때 시스템이 인증을 요구하지 않는 Not Required의 상태가 16,115(93.7%)건으로 가장 많이 나타났으며, 그 다음으로 Single system의 상태가 411(6.3%)건, 공격자가 취약점 공격을 할 때, 두 개 이상의 인증을 필요로 하는 Multiple systems의 상태가 1(0%)건의 순서로 나타났다.

인증여부는 공격자가 취약점 공격을 할 때 인증의 수가 적을 때 큰 취약점의 상태를 가지며, Multiple systems < Single system < Not Required와 같이 나타났다.

3.1.4 기밀성 공격영향(Confidentiality Impact)

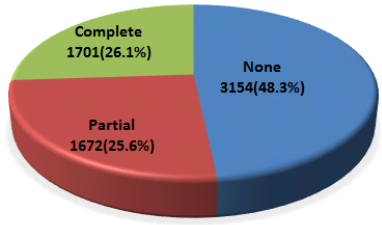


(그림 5) 기밀성 공격영향별 DDoS 빈도수

(그림 5)는 기밀성 공격영향별 DDoS 빈도수를 나타낸 것으로, 공격자가 취약점 공격을 성공하여 시스템에 대해 미치는 영향이 없는 None의 상태가 3,068(47.0%)건으로 가장 많이 나타났으며, 그 다음으로 시스템 파일에 한정적으로 접근만 가능한 Partial의 상태가 1,747(26.8%)건, 취약점 공격 성공하여 시스템에 대해 모든 파일의 정보를 읽을 수 있는 Complete의 상태가 1,712(26.2%)건의 순서로 나타났다.

기밀성 공격영향은 공격자가 취약점 공격을 성공하였을 때 공격이 기밀성에 미치는 영향이 증가 할수록 큰 취약점의 상태를 가지며, Complete < Partial < None와 같이 나타났다.

3.1.5 무결성 공격영향(Integrity Impact)

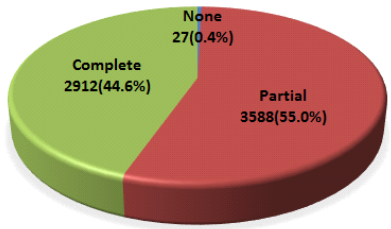


(그림 6) 무결성 공격영향별 DDoS 빈도수

(그림 6)은 무결성 공격영향별 DDoS 빈도수를 나타낸 것으로, 공격자가 취약점 공격을 성공하여 시스템에 대해 미치는 영향이 없는 None의 상태가 3,154(48.3%)건으로 가장 많이 나타났으며, 그 다음으로 시스템 파일에 한정적으로 정보 수정이 가능한 Partial의 상태가 1,672(25.6%)건, 취약점 공격 성공하여 시스템에 대해 보호기능을 완전 손실시키는 Complete의 상태가 1,701(26.1%)건의 순서로 나타났다.

무결성 공격영향은 공격자가 취약점 공격을 성공하였을 때 공격이 무결성에 미치는 영향이 증가 할수록 큰 취약점의 상태를 가지며, Partial < Complete < None와 같이 나타났다.

3.1.6 가용성 공격영향(Availability Impact)



(그림 7) 가용성 공격영향별 DDoS 빈도수

(그림 7)은 가용성 공격영향별 DDoS 빈도수를 나타낸 것으로, 공격자가 취약점 공격을 성공하여 시스템에 대해 정보 자원의 접근을 방해하거나 성능을 감

소시키는 Partial의 상태가 3,588(55.0%)건으로 가장 많이 나타났으며, 그 다음으로 취약점 공격 성공하여 시스템에 네트워크 대역폭, 디스크 공간 소모 등 모든 자원을 중단시키는 Complete의 상태가 2,912(44.6%)건, 시스템에 대해 미치는 영향이 없는 None의 상태가27(0.4%)건의 순서로 나타났다.

가용성 공격영향은 공격자가 취약점 공격을 성공하였을 때 공격이 가용성에 미치는 영향이 증가 할수록 큰 취약점의 상태를 가지며, None < Complete < Partial와 같이 나타났다.

4. 통계적 기법을 이용한 DDoS 예측

본 논문에서는 제안한 모델에 적용할 데이터를 정량화하기 위해 요소별(Access Vector, Access Complexity, Authentication, Confidentiality Impact, Integrity Impact, Availability Impact) 상관분석과 회귀분석을 실시하여 예측식을 제시하였고, 회귀분석에서 사용된 예측 모델은 선형모형, 2차 모형, 3차 모형, 지수모형, 로지스틱 5가지이며, 예측 모델의 식은 <표 1>과 같다.

<표 1> 예측 모델식

Model	수 식
선형모형	$E(Y_t) = A_1 + A_2 t$
2차 모형	$E(Y_t) = A_0 + A_1 t + A_2 t$
3차 모형	$E(Y_t) = A_0 + A_1 t + A_2 t^2 + A_3 t^3$
지수모형	$E(Y_t) = A_0 e^{A_1 t}$
로지스틱	$E(Y_t) = (\frac{1}{u} + A_0 A_1^t)^{-1}$

4.1 요소별 영향도

DDoS 예측 모델을 구성하기 위하여 DDoS 공격과 요소와의 관련성을 회귀분석을 통해 영향정도(R Square)를 산출하였고 각 회귀식의 영향력의 크기를 표준화하여 표준화된 결정계수(Adjusted R Square)를 산출하였다. 또한, 표준화된 결정계수들을 상대적 영향력으로 백분율로 환산하였다.

그 결과 Confidentiality Impact가 가장 높은 설명력을 나타냈으며, Integrity Impact, Availability Impact, Access Vector, Authentication, Access Complexity의 순으로 나타났다.

<표 2> DDoS에 대한 요소별 영향력

구분	R2	Adjusted R2	상대적 영향력 (%)	영향 순위
Access Vector	0.091	0.091	4.970	4
Access Complexity	0.033	0.033	1.802	6
Authentication	0.052	0.051	2.785	5
Confidentiality Impact	0.57	0.569	31.076	1
Integrity Impact	0.559	0.559	30.530	2
Availability Impact	0.528	0.528	28.837	3

4.2 DDoS 위험지수 예측

6개의 변수에서 구해진 예측 식 $(Y_0,\ Y_1,\ Y_2,\ Y_3,\ Y_4,\ Y_5)$과 요소별 상대적 영향력을 곱하여 DDoS 위험지수를 예측한다.

<표 3>은 요소별(Access Vector, Access Complexity, Authentication, Confidentiality Impact, Integrity Impact, Availability Impact) 추정 회귀식을 정리한 것이다.

<표 3> 요소별 추정 회귀식

구분	회귀식
Access Vector	$Y_0 = 0.000 \times AV^1 + 0.143 \times AV^2 + 0.028 \times AV^3 + 4.526$
Access Complexity	$Y_1 = 3.269 \times AC - 0.57 \times AC^2 + 1.880$
Authentication	$Y_2 = 1.165 \times A + 0.551 \times A^2 + 2.512$
Confidentiality Impact	$Y_3 = 0.000 \times CI + 0.058 \times CI^2 + 0.137 \times CI^3 + 4.749$
Integrity Impact	$Y_4 = 0.000 \times II - 0.312 \times II^2 + 0.251 \times II^3 + 5.125$
Availability Impact	$Y_5 = 0.000 \times AI + 0.412 \times AI^2 + 0.024 \times AI^3 + 3.936$

요소에 의한 DDoS 위험지수를 예측하기 위하여 각 요소에 따른 회귀식에 앞에서 구해진 영향도를 곱하여 합한 값을 DDoS 위험지수 예측식으로 하고 이를 수식으로 나타내면 (식 2)과 같다.

$$\begin{aligned} DDoS\ \text{위험지수}(DD_1) = {} & (Y_0 \times 0.04970) \\ & + (Y_1 \times 0.01802) \\ & + (Y_2 \times 0.02785) \\ & + (Y_3 \times 0.31076) \\ & + (Y_4 \times 0.30530) \\ & + (Y_5 \times 0.28837) \end{aligned} \quad (2)$$

여기서 DD_1은 6가지 요소에 의한 예측식, Y_0은 Access Complexity, Y_1은 Access Complexity, Y_2는 Authentication, Y_3은 Confidentiality Impact, Y_4는 Integrity Impact, Y_5는 Availability Impact이다.

회귀식의 합에 따라 DDoS의 위험지수에 따라 정상, 관심, 주의, 경계, 심각의 5등급으로 구분하여 위험지수로 나타내었다.

<표 4>은 각 함수에 제3장에서 분석된 요소별 DDoS에 대한 위협수준이 높은 순으로 각각의 수치를 회귀식에 대입하여 위험지수를 추출하였다.

<표 4> 요소별 위협수준에 의한 위험지수

구분	심각	경계	주의	관심	정상
Access Vector	0.246	0.243	0.239	0.236	0.232
Access Complexity	0.180	0.155	0.131	0.107	0.083
Authentication	0.180	0.189	0.166	0.142	0.118
Confidentiality Impact	1.658	1.627	1.597	1.567	1.536
Integrity Impact	1.509	1.518	1.527	1.537	1.546
Availability Impact	1.512	1.449	1.386	1.324	1.261

<표 3>은 <표 4>에 의해 얻어진 위험지수가 요소별 어떠한 상태에서 예측 값을 가지는지를 나타낸 것으로 위험지수가 예측되면 이 자료를 통해 어떠한 상태에서 예측되었는지 분석하고 위협에 대한 정책을 설정할 수 있다.

<표 5> 위험지수에 대한 요소별 상태

위험지수	번호	AV	AC	A	CI	II	AI
4.9	1	R	H	S	P	N	N
4.95	1	R	L	NR	N	C	N
4.96	1	R	H	S	N	N	P
	2	R	L	NR	N	P	N
… 중간생략 …							
5.34	1	L	L	NR	C	N	C
5.36	1	R	L	NR	C	N	C

본 연구에서는 Access Vector, Access Complexity, Authentication, Confidentiality Impact, Integrity Impact, Availability Impact의 요소들을 (식 2)에서 계산된 값이 5.1 이상일 때는 DDoS공격 위험지수가 심각의 상태인 것으로 하고, 4.8미만일 때는 이를 정상인 상태로 하여 <표 6>과 같이 심각, 경계, 주의, 관심, 정상의 5가지 범위로 나누어 지수화 하였다.

<표 6> DDoS공격 위험지수

위험수준	정상	관심	주의	경계	심각
예측지수	4.8 미만	4.9 미만	5.0 미만	5.1 미만	5.1 이상

5. 결 론

DDoS 공격이 발생하면 탐지하고 대응까지 시간이 비교적 많이 소요된다. 과거에는 공격기법에 국한되어 연구가 되어 왔으며, 이런 연구가 방어기법에 적용하기에는 제한적이였다. 이를 보완하기 위해 발생 전 잠재 위협을 예측하여 사전에 정책을 수립하고 대응할 수 있는 예측에 관한 연구가 현재 활발히 진행되고 있으며, 예측 기법들의 장점을 활용한 시스템들이 점차 늘어나고 있다.

본 논문에서는 DDoS 공격에 적용 가능한 위험지수의 예측식을 도출하고 미래에 발생할 수 있는 위협에 대해 위험지수를 예측함으로써 예측된 위험지수에 대한 요소(Access Vector, Access Complexity, Authentication, Confidentiality Impact, Integrity Impact, Availability Impact) 들이 어떠한 상태에서 나타나는지를 알 수 있었다. 이 자료를 토대로 미래에 발생할 수 있는 위협에 대해 예측하여 분석함으로써 위협에 대한 예방 정책 수립하여 피해를 최소화 할 수 있을 것이라 기대할 수 있다.

참고문헌

[1] 서진원, 곽진, “다단계 방어기법을 활용한 DDoS 방어시스템 설계”, 정보보호학회논문지 제22권 제3호, pp. 679-689, 2012.

[2] 전용희, 장종수, 오진태, “DDoS 공격 및 대응기법 분류”, 정보보호학회논문지, 제19권 제3호, pp. 46-57, 2009.

[3] 최양서, 오진태, 장종수, 류재철, “분산서비스 거부(DDoS) 공격 통합 대응체계 연구”, 정보보호학회논문지, 제19권 제5호, pp. 11-20, 2009.

[4] 홍성혁, “DDoS 공격에 대한 분석 및 대응방안”, 디지털융복합연구, 제12권 제1호, pp. 423-429, 2014

[5] 백종욱, “연관성 분석에 의한 사이버위협 경보모델”, 경기대학교 박사학위논문, 2006.

[6] Özbayoğlu G, Özbayoğlu ME (2006) A new approach for the prediction of ash fusion temperatures: a case study using Turkish lignites. Fuel 85:545 - 552.

[7] Mehmet Bilgili, "Prediction of soil temperature using regression and artificial neural network models", Meteorology and Atmospheric Physics Vol. 110, Numbers.1-2, 2010, pp.59-70

[8] K. T. Chau, J. E. Chan, "Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island", Landslides Vol. 2, No. 4, 2005, pp.280-290.

[9] http://www.cvedetails.com/browse-by-date.php

[저 자 소 개]

김 종 민 (Jong-Min Kim)

2010년 2월 체육학사
2012년 2월 경기대학교
경호안전학석사
2015년 2월 경기대학교
산업보안학박사

email : dyuo1004@gmail.com

정 병 수 (Byung-Soo Jung)

2005년 02월 행정학사
2007년 08월 경찰학 석사
2011년 02월 범죄학 박사

email : 2079bs@daum.net