마코브 체인을 이용한 Mass SQL Injection 웜 확산 예측에 관한 연구

박원형* · 김영진** · 이동휘* · 김귀남*

요 약

최근 웜에 의한 사이버 위협이 증가함에 따라 웜의 확산 특성을 분석하기 위한 전파 모델이 연구되고 있다. 대표적인 예로 수학적 모델링 기법인 Epidemic(SI), KM(Kermack-MeKendrick), Two-Factor, AAWP(Analytical Active Worm Propagation)등의 모델 기법들이 제시되었다. 하지만, 기존 모델 방법들은 대부분 코드레드와 같은 네트워크를 대상으로 하는 랜덤 스캐닝 기법에 대해서만 모델링이 가능하다. 또한 거시적인 분석만 가능하고 특정 위협에 대해 예측하는데 한계점을 가지고 있다. 따라서 본 논문에서는 과거의 위협 발생 데이터를 근거로 하여 Mass SQL Injection 같은 사이버위협에 적용 가능한 마코브 체인(markov chain) 기반 예측 방법을 제시한다. 이를 통하여 각 위협별 발생 확률 및 발생빈도를 예측할 수 있다.

A Study on Prediction of Mass SQL Injection Worm Propagation Using The Markov Chain

Won Hyung Park* · Young Jin Kim** · Dong Hwi Lee* · KuiNam J Kim*

ABSTRACT

Recently, Worm epidemic models have been developed in response to the cyber threats posed by worms in order to analyze their propagation and predict their spread. Some of the most important ones involve mathematical model techniques such as Epidemic(SI), KM (Kermack-MeKendrick), Two-Factor and AAWP(Analytical Active Worm Propagation). However, most models have several inherent limitations. For instance, they target worms that employ random scanning in the network such as CodeRed worm and it was able to be applied to the specified threats. Therefore, we propose the probabilistic of worm propagation based on the Markov Chain, which can be applied to cyber threats such as Mass SQL Injection worm. Using the proposed method in this paper, we can predict the occurrence probability and occurrence frequency for each threats in the entire system.

Key words : Markov Chain, Cyber Threat, Worm Propagation

* 경기대학교 정보보호학과
** 고려대학교 정보경영공학전문대학원

1. 서 론

인터넷의 개방성과 정보 시스템의 복잡화로 인하여 사이버공격에 이용될 수 있는 다양한 취약점(Vulnerability)들이 발견되고 있다. 이러한 취약점들을 이용하는 악성코드의 경우 불특정 다수의 시스템을 대상으로 전파되며 피해를 입히기 때문에 공격의 대상이 되는 특정 시스템뿐만 아니라 해당 네트워크 도메인 전체가 원활한 서비스 제공에 큰 장애를 겪게 된다. 이에 해당하는 대표적인 사례로는 2001년의 MS사의 IIS 웹서버 소프트웨어의 취약점을 이용한 CodeRed 웜과 2003년 1월 25일 10분 내에 전 세계 90% 이상의 컴퓨터를 감염시켜 버린 SQL Overflow(Slammer)[1]등의 웜이 있다. 최근에는 SQL Injection 취약점을 이용한 웹 서버 악성코드 삽입 등 다양한 사이버 공격에 이용될 수 있는 봇(Bot) 계열 악성코드들이 이슈화 되고 있다. 이처럼 최근 악성코드들의 성향은 한정된 시간 내에 최다한 많은 호스트들을 감염시킬 수 있도록 설계되고 있는 것이 특징이다. 결국 악성코드의 성격에 따라 달라질 수는 있으나 특별한 대응조치를 하지 않는다면 악성코드에 감염이 되는 호스트들의 수는 시간의 흐름에 따라 지속 증가할 것이다. 만약 이러한 사이버위협 예측에 관한 연구가 없는 상태에서 사이버공격으로 인해 국가 핵심기반시설의 정보통신망이 마비된다면 국가안보에 엄청난 파급 효과를 낳게 될 것이다. 현재 사이버위협 예측모델에 관한 연구가 미흡한 실정이며 사이버위협 징후를 사전 탐지하여 공학적이고 학술적인 기준에서 분석이 요구된다. 사이버위협 요소들을 명확히 분류하고 그 특성을 세부적으로 분석하여 객관적이고 효율적인 통계적 예측기준을 마련하는 것이 시급하다.

본 논문은 최근 발생하고 있는 Mass SQL Injection 웜 발생 이벤트 데이터를 가지고 웜 확산 빈도를 예측할 수 있는 마코브 체인 기반의 사이버위협 예측기술에 대해 연구 한다.

2. 관련 연구

2.1 웜 확산 예측 모델

본 절에서는 논문의 연구배경으로 기존 웜 확산 모델과 문제점에 대해 알아본다. 웜 확산 모델에서 초기의 수학적인 모델 연구는 바이러스에 대한 모델 기법에서 시작되었는데, 데이터의 흐름 또는 정보의 흐름이 웜 확산에 많은 영향을 미친다는 가정에서 이를 바이러스 모델에 적용하였다. Fred Cohen [2]등은 각 호스트 간의 데이터 흐름, 또는 통신 메커니즘에 의해 바이러스가 전파된다는 가정을 하였다.

Epidemic(SI) 모델에 적용되는 파라미터는 아래 <표 2-1>와 식 (2-1), 식 (2-2), 식 (2-3)를 사용한다[3].

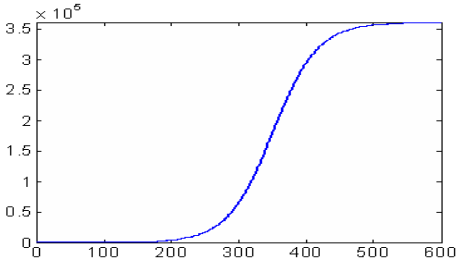
〈표 2-1〉 Epidemic 모델 파라미터

파라미터	설 명
I(t)	감염된 호스트 수
S(t)	감염될 수 있는 취약한 호스트 수
N	시스템 내에 있는 호스트 총 수
β	두 호스트 간의 감염율
α	평균 감염시도율

$$N = S(t) + I(t) \tag{2-1}$$

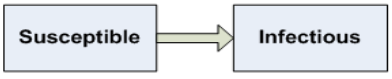
$$\frac{dI(t)}{dt} = \beta I(t)S(t) = \beta I(t)[N - I(t)] \tag{2-2}$$

$$\alpha = \beta N \tag{2-3}$$



(그림 2.1) Epidemic 모델 시뮬레이션

이 연구를 시작으로 Gleissner 등은 이를 확장하여 다중사용자 시스템에서 웜 전파 모델을 제시하였고, 웜은 기하급수적인 전파를 보인다는 것을 증명하였다[4]. 여기서부터 출발한 Simple Epidemic Model은 (그림 2.2)처럼 모든 상태를 감염된 상태(Infectious)와 취약한 상태(Susceptible) 두 가지로 표현하는 수학적 모델로서 SI(Susceptible-Infectious) 모델이라고도 하며 기준시간, 취약한 호스트 수, 네트워크 내의 총 호스트 수 등을 파라미터로 사용한다.

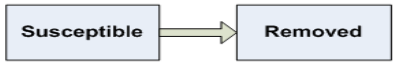


(그림 2.2) SI 모델

하지만 이 모델링 기법은 한번 감염된 호스트가 영원히 감염된 상태로 남게 되어 취약점이 제거된 상태(removed)가 표현되지 않는다는 단점이 있다. General Epidemic Model(Kermack-MeKendrick Epidemic Model)[5]은 (그림 2.3)처럼 이러한 SI 모델링 기법의 단점을 보완, 취약점이 제거된 상태(Removed)를 표현할 수 있는 모델링 기법으로서 SIR(Susceptible-Infectious-Removed)이라고 불린다.

(그림 2.3) General Epidemic(KM) 모델

하지만, 이 모델링 기법은 (그림 2.4)와 같이 취약한 호스트가 웜에 감염되기 전에 취약점이 조치된 상태의 표현이 불가능한 문제점이 있다. 그리고 기존의 SIR 모델에 포함되지 않았던 요소 2가지를 추가하여 Two-Factor 모델[6]을 제안 하였는데, 이는 CodeRed 사고 이후 새롭게 제안된 모델이며 인간의 대응책과 감소되는 감염율을 고려하였다.

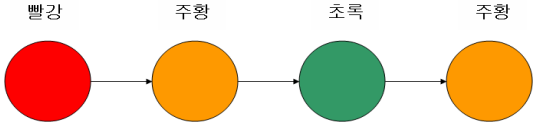


(그림 2.4) 웜 감염 전, 취약점 조치 상태

웜의 관점에서 인간의 대응책은 어떤 호스트들을 웜 확산 활동을 감소시킨다. 감염된 호스트들은 감염된 호스트 및 여전히 취약한 호스트들을 모두 포함한다. 이후 Zesheng Chen, Lixin Gao이 특정 네트워크에서 확산되는 CodeRed 웜에 대한 확산 예측 모델 AAWP(Analytical Active Worm Propagation)를 연구하였다[7]. 하지만 해킹과 웜이 결합된 새로운 형태의 사이버위협인 Mass SQL Injection 웜은 기존 일반 웜 확산 특징과는 달라 기존 웜 확산 모델에 적용하기 어렵다.

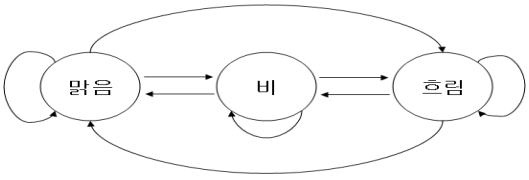
2.2 마코브 체인

본 절에서는 논문에서 제시하는 사이버위협 예측 모델을 설계하기 위해 적용되는 마코브 체인에 대해 설명한다. (그림 2.5)와 같이 각각의 상태는 전적으로 이전 상태에 의해서 결정된다. 만약 아래의 그림처럼 현재 신호등이 초록색이었다면 그 다음은 주황색으로 변할 것이다. 이러한 시스템을 결정 시스템이라 한다[8].



(그림 2.5) 결정시스템(Deterministic System)

그런데 신호등 예와 달리 날씨 예를 들어보면 내일의 날씨가 전적으로 오늘의 날씨에만 의존하여 변한다고 말할 수 없기 때문에 이 시스템은 결정 시스템이 아니다. 다음 (그림 2.6)에서 보는 것처럼 오늘 날씨가 맑았다고 해서 내일 날씨가 맑다고 말할 수 없는 것이다. 확률적으로 맑을 수도 있고 흐릴 수도 있고 비가 올 수도 있다.



(그림 2.6) 비결정시스템(Non-Deterministic System)

이러한 문제를 풀기 위해 다음과 같은 가정을 한다. "모델의 상태를 오로지 이전 상태들에만 의존한다." 이 가정을 마코브 가정이라 한다. 마코브 가정을 하게 되면 복잡한 문제를 굉장히 단순화 시킬 수 있다는 장점이 있으나 과도하게 단순화하기 때문에 중요한 정보를 잃을 수 있다는 단점이 있다. 마코브 체인은 다음의 세 가지로 설명될 수 있는 모든 시스템을 말한다.

- 상태 집합(set of state) : 전체시스템이 가지고 있는 위협들이 가질 수 있는 상태들의 집합을 나타낸다. 위협 상태의 집합을 T라 정의하면 식 (2-4)와 같다.

$$T = T_1, T_2, T_3, \cdots, T_n \qquad (2\text{-}4)$$

위협 상태 집합은 하나의 위협이 가질 수 있는 값들의 범위(임계값)을 나타내거나, 여러 위협 상태들의 쌍(조합)이 될 수 있다.

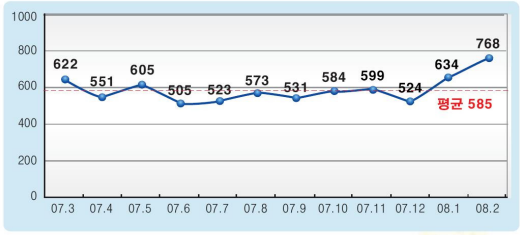
- π 벡터(초기 확률) : 정의된 위협 상태들이 초기 상태에 가질 수 있는 위협 발생 확률을 나타낸다. 초기 위협 확률의 총 합은 1이 되어 식 (2-5)을 만족 해야 한다.

$$\sum_{i=1}^{N} P_{Ti} = 1(\text{T는 위협 상태}) \qquad (2\text{-}5)$$

- 상태 전이행렬(state transition matrix) : 정의된 위협 상태들 간의 전이 확률을 나타낸다. 과거의 관찰된 자료, 즉 분석된 각 위협별 발생 빈도수와 정의된 위협 상태 집합과의 매핑(mapping)을 통해 전이 행렬을 구한다[9].

3. Mass SQL Injection 웜

3.1 사이버위협 동향



(그림 3.1) 월별 사이버 침해사고 발생 추이

최근 웹 서핑 중 사용자 모르게 악성코드에 감염되는 피해가 많이 발생 하고 있다. 2008년 3월 국가사이버안전센터 사이버 시큐리티 보고서[10]에 따르면 2월은 해킹을 통한 자료훼손 및 유출 사고는 감소하였으나, 악성코드 감염사고가 전월 대비 증가하여 전체 사고가 1월에 이어 증가 추세를 보였다. 침해사고는 지난달보다 21.1%가 증가하여 2007년 12월 이후 증가 추세인데 주로 악성코드 감염과 홈페이지관련 사고 증가가 원인이다.

3.2 Mass SQL Injection 개요

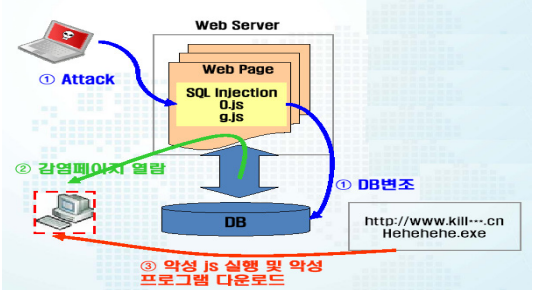
2008년 4월 초부터 전 세계 130만개 이상의 웹 사이트에 악성코드를 유포하는 SQL Injection 공격코드가 숨어 있는 것을 확인하였다. 언론에 알려진 것은 4월이지만, 2008년 1월 아파치 보안 모듈인 Mod Security 프로젝트의 블로그[11]에 공개되면서 알려지게 되었다.

MMass SQL Injection은 기존 SQL Injection 취약점에서 확장된 개념으로 한 번의 공격으로 대량의 DB값이 변조하여 홈페이지에 악성코드를 삽입하는 공격을 의미한다. Mass SQL Injection의 두

가지 특징은 일부분을 HEX 인코딩하거나, 전체 HEX 인코딩 하는 방식을 사용하고 있다. DB값 변조시 악성 스크립트를 삽입하여, 사용자들이 변조된 사이트를 방문하여 감염되거나 봇(Bot)이 설치되어 서비스거부공격에 이용이 가능하다. 악성 스크립트에 사용되는 형식은 js, swf, exe 확장자를 가진 파일을 주로 사용한다. Mass SQL Injection 웜 공격 대상은 MS-SQL를 사용하는 DB 서버, 공격자는 ASP가 가동중인 IIS 웹서버를 주요 공격 대상으로 한다.

〈표 3-1〉Mass SQL Injection 웜 감염 경로

1. Google, Naver, Yahoo 등 검색사이트를 통해 URL 정보를 수집(파라미터를 변수로 받은 URL검색)
   URL에 HEX 인코딩된 공격, DB 테이블 문자열 칼럼값 변조
2. 사용자가 감염된 웹페이지 게시판 열람 후 악성 스크립트 다운로드
3. 악성 스크립트 *.js 실행 및 악성코드 재감염
   (봇넷 감염 및 개인정보 유출 추정)



(그림 3.2) Mass SQL Injection 감염 경로

3.3 Mass SQL Injection 상세 분석

Mass SQL-Injection 웜 공격 등 악성코드 유포지로 악용하기 위한 홈페이지 대상 해킹시도가 증가하고 있다. 해커(중국 추정)는 IFRAME 태그를 이용한 악성프로그램의 유포가 탐지·차단되자 DB 프로시저 명령을 이용한 공격코드를 탐지회피를 위해 16진수 값으로 위장하고 홈페이지 게시글에 악성코드를 삽입하는 새로운 형태의 SQL Injection 공격 웜을 유포시켰다. 아래 SQL 구문은 sysobjects 테이블의 xtype = U(User) 필드에서 모든 row를 가져오는 것이다. 각 오브젝트에 s.cawjb.com/s.js 사이트 주소 코드를 추가하도록 업데이트 명령을 실행시키는 구문으로, 이 공격을 받은 웹 사이트는 IIS와 MS-SQL 서버가 설치된 경우이다. 특히, 필터링을 우회하기 위해 CAST나 CONVERT 명령어를 쓴다.

(그림 3.3) SQL-Injection 공격 코드(decode)

4. Mass SQL Injection 웜 예측 모델

4.1 마코브 체인 기반의 웜 예측 모델

본 논문에서는 독립 또는 단일 시스템의 피해로 인하여 그 시스템과 관련된 전체 시스템에 영향을 주는 정도를 확률적으로 예측할 수 있는 사이버위협 예측 모델을 구현 하였다. 제안된 예측모델은 시간의 흐름 또는 위협 종류에 따라 사이버위협으로 인한 피해 확산 속도가 달라질 수 있다. 본 논문에서 제안된 위협발생 예측 값은 아래 식 (4-1)을 통해 구할 수 있다.

$$T = \sum_{i=1}^{n} P(S_i)\, M(S_i)$$

$$= \sum_{i=1}^{n} P(\frac{\alpha}{F}\frac{\beta}{F}\frac{\gamma}{F}\cdots\frac{\delta}{F})(\frac{\sum_{i=1}^{n} S_i}{n}) \quad (4\text{-}1)$$

- T = 위협발생 예측값 • S = 위협 상태
- n = 위협 발생 상태집합이 개수
- $P(S_i) = P(S_1 S_2 S_3 S_4 \cdots S_n)$

$= P(\frac{\alpha}{F}\frac{\beta}{F}\cdots\frac{\delta}{F})$(각 위협 상태의 발생 확률) (4-2)

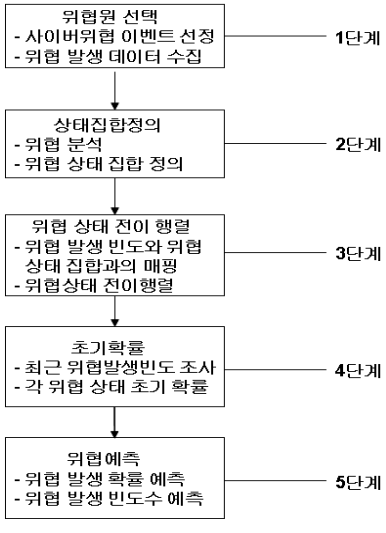
단, $F = \sum_{i=1}^{n} f_i = \alpha+\beta+\gamma+\cdots+\delta$

$\sum_{i=1}^{n} P(S_i) = 1$(각 위협상태의 발생확률의 합은 1)

• $M(S_i) = \frac{\sum_{i=1}^{n} S_i}{n}$(각 위협 상태의 평균값) (4-3)

4.2 마코브 체인 기반의 웜 예측 모델

본 절에서는 마코브 체인에 기반한 예측 모델에 대하여 단계별 프로세스를 정의한다. 본 논문에서 제안한 마코브 프로세스에 기반한 사이버위협 모델은 (그림 4.1)과 같은 과정을 거쳐 생성되며, 크게 5단계(위협원 선택, 상태집합의 정의, 위협상태 전이행렬, 초기확률, 위협예측)로 이루어진다.



(그림 4.1) 사이버위협 예측 모델 생성 절차

1단계인 '위협원 선택'에서는 보안장비에서 탐지된 이벤트를 대상으로 주간, 일간별로 의미있는 데이터를 수집하여 통계화 한다. 2단계인 '상태집합의 정의'단계에서는 조직이나 기관이 가지고 있는 위협들이 취할 수 있는 상태를 정의한다. 상태(state)란 주요 정보통신 기반 시설이 가지고 있는 위협의 상태를 말하며, 상태집합은 하나의 위협 상태가 가질 수 있는 값들의 범위를 나타내거나, 여러 위협 상태들의 쌍(조합)이 될 수 있다. 3단계인 '위협상태 전이행렬'에서는 상태집합에서 정의된 위협 상태와 위협 발생 빈도 데이터를 이용하여 위협 상태들 간의 전이 행렬을 구한다. 4단계인 '초기확률(π 벡터)'에서는 정의된 각 위협 상태가 초기 상태에 발생할 수 있는 확률을 구한다. 마지막 5단계 '위협 예측' 단계에서는 전 단계에서 구한 위협 상태 전이행렬과 초기 확률 값을 통해 앞으로 발생할 위협 발생 확률이나 빈도수를 예측할 수 있다.

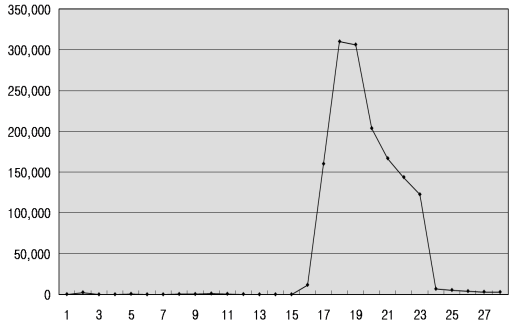
5. Mass SQL Injection 적용 및 결과

5.1 Mass SQL Injection 위협원 선택

본 논문의 적용 사례는 K기관에서 2008년 10월부터 11월까지 보고된 Mass SQL Injecton 공격 통계 데이터를 이용하였다. 우선 위협원 단계에서 위협 발생데이터를 수집하고 수집된 데이터를 분석하여 위협 유형을 분류하고 위협 순위를 산정한다.

- 웜 전파 이벤트(Mass SQL-Injection 감염장비에서 특정 홈페이지 감염시도) : 악성코드 유포 경유 홈페이지에서 접속한 감염된 시스템이 다시 SQL Injection취약점이 있는 홈페이지로 악성코드를 전파하는 웜 확산 공격으로 '통신망에 대한 불법적인 공격' 위협이다.

웜 전파 위협 유형의 일별 발생 수는 다음 (그림 5.1)과 같다.



(그림 5.1) 웜 확산 일별 수 및 통계 추이

5.2 Mass SQL Injection 상태집합 정의

본 위협은 웜 확산(악성 프로그램을 이용한 통신망에 대한 불법적인 침입)에 대한 상태 집합을 정의한다. 웜 확산에 대한 일별 발생 빈도수는 <표 5-1>과 같으며 식 (5-1)와 같은 임계값의 범위로 상태 집합(S) 식 (5-2)을 정의한다.

본 논문에서 임계값은 위협의 단위 시간(일별) 발생 빈도를 몇 개의 구간으로 나누어 표현한다.

• S의 위협 기준 :

S1 : 정상 S2 : 관심 S3 : 주의 S4 : 경계 S5 : 심각

위협기준은 K기관의 사이버위협 발령기준을 적용하여 신뢰도를 높혔다.

• S의 임계값의 범위 :

S1 : 0~1,000, S2 : 1,001~10,000,
S3 : 10,001~100,000, S4 : 100,001~200,000
S5 : 200,001~ (5-1)

임계값의 범위는 K기관에서 사용하는 일일 임계값 기준을 적용하여 신뢰도를 높혔다.

• S = {S1, S2, S3, S4, S5} (5-2)

식 (5-1), 식 (5-2)에서 볼 수 있듯이, 위협이 독립적으로 발생하는 경우에는 상태집합이 단지 해당 위협의 임계값 범위에 의해 결정된다.

5.3 Mass SQL Injection 위협상태 전이행렬

위 단계에서 정의한 내용을 바탕으로 1주 부터 4주까지 28일 간의 위협 발생수를 정의된 상태집합(S)과 매핑하여 상태를 열거한다.

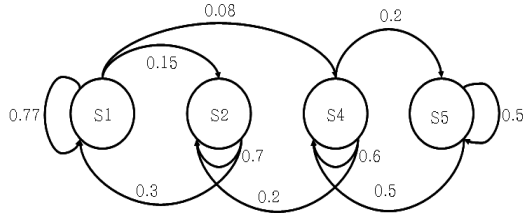
$$\begin{bmatrix} S_1 & S_2 & S_1 & S_1 & S_1 & S_1 & S_1 \\ S_1 & S_1 & S_2 & S_1 & S_1 & S_1 & S_1 \\ S_1 & S_3 & S_4 & S_4 & S_5 & S_5 & S_4 \\ S_4 & S_4 & S_2 & S_2 & S_2 & S_2 & S_2 \end{bmatrix}$$

열거된 상태들로부터 각 상태(S1, S2, S3, S4, S5)에서 다른 상태로의 전이 횟수를 구하고 이를 바탕으로 상태전이행렬 식 (5-3)을 구한다.

$$\begin{pmatrix} 10 & 2 & 1 & 0 & 0 \\ 2 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0.77 & 0.15 & 0.08 & 0 & 0 \\ 0.30 & 0.70 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0.2 & 0 & 0.6 & 0.2 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix} \quad (5\text{-}3)$$

상태 전이 행렬 식 (5-3)로부터, 각 위협에서 다른 위협으로의 전이 확률 값의 합이 1이 되어 식 (4-2)을 만족한다. 또한 위 위협 상태 전이 확률을 상태 다이어그램으로 나타내면 (그림 5.2)와 같다.



(그림 5.2) 위협에 대한 상태 다이어그램

5.4 Mass SQL Injection 초기 확률

본 논문에서는 웜 확산에 대한 초기 확률을 구하기 위해 최근 5일 동안 발생한 빈도수를 이용한

다. 앞서 제시한 식 (4-2)를 이용하여 최근 5일 동안 발생한 빈도수와 이에 따른 초기 확률 값을 구하면 식 (5-4)와 같다.

- 빈도수 : 6,741, 5,329, 3,951, 3,135, 2,911
  $= S_2, S_2, S_2, S_2, S_2$ (5-4)
- 초기 확률 : $P(S_1, S_2, S_3, S_4, S_5) = P(0\,1\,0\,0\,0)$

5.5 Mass SQL Injection 위협예측

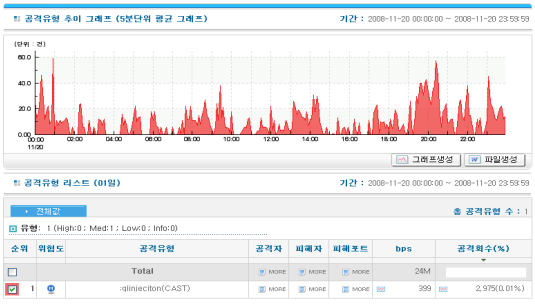
위협 상태 전이 행렬과 초기 확률를 이용하여 다음에 발생하게 될 위협 발생 확률을 예측하고 또한 위협 발생 빈도수를 예측할 수 있다. 즉, 위협 발생 확률은 식 (5-5)과 같이 계산할 수 있다.

$$(0\,1\,0\,0\,0)\begin{pmatrix} 0.77 & 0.15 & 0.08 & 0 & 0 \\ 0.30 & 0.70 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0.2 & 0 & 0.6 & 0.2 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix} = (0.3\,0.7\,0\,0\,0) \quad (5\text{-}5)$$

결과 식 (5-5)부터 웜 확산 예측 확률은 다음 날 S1 상태로 0.3, S2 상태로 0.7확률로 발생할 것이라고 예측할 수 있다. S2상태 즉, 1,001과 10,000사이의 발생빈도를 가질 것이라는 것을 알 수 있다. 좀 더 구체적인 다음 날의 위협발생 빈도를 구하기 위해서 다시 식 (4-3)을 이용한다. 이를 위해 위협 발생 확률 식 (5-5)와 각 임계값의 평균값을 이용한다. 본 적용사례에서는 평균값을 구하기 위해 최근 5일 동안 발생한 이벤트의 발생빈도 평균을 이용하였다. 앞서 정의한 임계값의 범위에 따라 평균값을 구하면 다음과 같다.

- 평균값 = $\frac{6,741+5,329+3,951+3,135+2,911}{5} = 4,413$
- M(S1) = 0, M(S2) = 4,413, M(S3) = 0, M(S4) = 0 M(S5) = 0 또한 각 위협상태에 대한 발생확률은 (5-4)으로부터 얻을 수 있다.
- P(S1) = 0.3, P(S2) = 0.7, P(S3) = 0, P(S4) = 0, P(S5) = 0 따라서 위협 발생 빈도는 식 (4-1)에서 n = 5일 때, 즉 위협 상태의 개수가 4개일 때 이므로 아래와 같이 구한다.
- 예상 위협 발생 빈도 = 0.7 × 4,413 ≒ 3,089

위 결과로부터 다음날 발생빈도수는 약 3,089으로 예측할 수 있다. 아래 (그림 5.7)은 실제 다음날 이벤트 발생수가 2,975건 이었다.



(그림 5.3) 실제 다음날 웜 확산 이벤트 건수

두 사이의 차이는 114건, 실제 이벤트와의 편차는 -60, 표준편차는 59로 확인 하였다. 결과적으로 마코브 체인기반 위협 예측값이 표준편차 범위 및 임계값 S2 범위안에 있는 것을 확인 하였다.

6. 결 론

본 논문에서는 기존 바이러스나 웜뿐만 아니라 여러 가지 위협의 종류들로부터 사이버공격을 받았을 때 그 피해 정도를 예측할 수 있는 마코브 체인 기반의 사이버위협 예측 모델을 제안하였다. 또한 실제 사이버위협 이벤트인 Mass SQL Injection 웜 데이터를 이용하여 제안한 모델을 적용하여 신뢰성 있는 결과를 얻었다.

하지만, 제안하는 모델의 적용 및 결과를 통해 좀 더 정확한 예측 값을 얻기 위해서는 위협 상태 집합을 정의하기 위해 적절한 범위의 임계값을 정해야 한다. 임계값의 범위를 어떻게 정의하느냐에 따라 예측값이 실제 발생하는 위협과 어느 정도 가깝게

예측할 수 있는지를 결정하게 되고 임계값을 세분화하면 할수록 실제 값과 가까운 값을 예측할 수 있다. 또한 본 논문에서는 관측된 일별 데이터를 이용하였는데, 시간별, 또는 단위별 데이터가 확보되고 이용한다면 좀 더 정확한 예측치를 얻을 수 있다. 추가로 정확한 웜 확산 예측을 위해 인간의 대응활동인 악성코드 차단설정, 안티바이러스백신 업데이트, 침입탐지패턴 주입으로 인한 실시간 탐지 등 통계적 확률만 가지고 표현할 수 없는 요소까지 고려해야 정확한 데이터 값을 구할 수 있을 것이다.

참 고 문 헌

[1] David Moore, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, Nicholas Weaver, The Spread of the Sapphire/Slammer Worm, 2003.

[2] F. B Cohen, “A Formal Definition of Computer Worms and Some Related Results”, Computers & Security, pp. 641-652, 1992.

[3] 전영태, “AAWP와 LAAWP를 확장한 웜 전파 모델링 기법 연구”, 고려대학교 정보경영공학전문대학원, 2007.

[4] Dr. Winfried Gleissner, “A Mathematical Theory for the Spread of Computer Viruses”, Computers and Security, Vol. 8, 1989, pp. 35-41, 1989.

[5] D. J. Deley and J. Gani, “Epidemic Modeling : An Introduction”, Cambridge university Press, 1999.

[6] D. Moore, C. Shannon, G. M. Voelker, and S. Savage, “Internet Quarantine : Requirements for Containing Self-Propagating Code”, INFOCOM, 2003.

[7] Z. Chen, L. Gao, and K. Kwiat, “Modeling the Spread of Active Worms”, INFOCOM2003, 2003.

[8] http://www.comp.leeds.ac.uk/roger/HiddenMarkovModels/html_dev/main.html.

[9] 국가사이버안전센터 사이버시큐리티 3월호, 2008.

[10] http://www.modsecurity.org/blog/archives/2008/01/sql_injection_a.html.

박 원 형

2002년 서울산업대학교 산업정보시스템공학과(공학사)

2005년 서울산업대학교 정보산업공학과(공학석사)

2006년～현재 경기대학교 정보보호학과 박사과정

김 영 진

2007년～현재 고려대학교 정보경영공학전문대학원 박사과정

이 동 휘

2000년 경기대학교 전자계산학과(이학사)

2003년 경기대학교 정보보호기술공학과(공학석사)

2007년 경기대학교 정보보호학과(정보보호학 박사)

현재 경기대학교 산업기술보호 특화센터 연구교수

김 귀 남

미국 캔자스대학(공학사)

미국 콜로라도주립대학(공학석사)

미국 콜로라도주립대학(공학박사)

현재 경기대학교 정보보호학과 교수

현재 경기대학교 산업기술보호특화센터장