JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 24, No. 4: 522~528, Apr. 2020

악성코드 확산 모델링에 기반한 확산 예측 도구 개발

신원*

A Spread Prediction Tool based on the Modeling of Malware Epidemics

Weon Shin*

*Professor, Department of Information Security, Tongmyong University, Busan, 48520 Korea

요 약

엄청난 속도로 확산하는 랜섬웨어, 트로이목마, 인터넷 웜과 같은 악성코드는 인터넷의 주요한 위협이 되고 있다. 이러한 악성코드의 행위에 대응하기 위해서는 악성코드의 확산 방식과 영향을 끼치는 영향 요인을 이해하는 것이 필수적이다. 본 논문에서는 악성코드 확산 모델링에 기반을 둔 확산 예측 도구를 개발하였다. 이를 위하여 관련 연구를 살펴보고, 시스템 구성과 구현 방법을 살펴본 후 확산 예측 도구를 이용하여 워머블 악성코드 확산 실험을 수행하였다. 제안 확산 예측 도구를 잘 활용한다면, 최근 악명을 떨치는 워머블 악성코드에 대한 기본 지식만으로도 거시적 관점의 여러 조건에서 확산 형태를 예측하고 다양한 대응 방안을 모색할 수 있게 해준다.

ABSTRACT

Rapidly spreading malware, such as ransomware, trojans and Internet worms, have become one of the new major threats of the Internet recently. In order to resist against their malicious behaviors, it is essential to comprehend how malware propagate and how main factors affect spreads of them. In this paper, we aim to develop a spread prediction tool based on the modeling of malware epidemics. So we surveyed the related studies, and described the system design and implementation. In addition, we experimented on the spread of malware with major factors of malware using the developed spread prediction tool. If you make good use of the proposed prediction tool, it is possible to predict the malware spread at major factors and explore under various responses from a macro perspective with only basic knowledge of the recently wormable malware.

키워드 : 악성코드 확산, 확산 모델링, 확산 예측 도구, 워머블 악성코드

Keywords : Malware epidemics, Spread modeling, Spread prediction tool, Wormable malware

Received 30 December 2019, **Revised** 31 December 2019, **Accepted** 15 February 2020

*** Corresponding Author** Weon Shin(E-mail:shinweon@tu.ac.kr, Tel:+82-51-629-1284)
Professor, Department of Information Security, Tongmyong University, Busan, 48520 Korea

Open Access http://doi.org/10.6109/jkiice.2020.24.4.522 print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. 서 론

현대인들은 정보화 사회의 필수 도구인 컴퓨터를 이용하여 인터넷으로 전 세계를 연결할 뿐 아니라 대용량의 정보를 가공, 처리, 저장하고 있다. 그러나 새로운 컴퓨팅 기술과 다양한 애플리케이션이 보편화되면서 많은 역기능도 함께 증가하고 있는데, 그 중 컴퓨터 바이러스, 트로이목마, 인터넷 웜 등 악성코드는 컴퓨터 시스템 취약점을 이용한 침투, 운영체제 버그와 오류를 이용한 해킹, 좀비 PC를 이용한 분산서비스 거부 공격 등으로 세계적인 피해를 끼치고 있다[1-3]. 특히, 트로이목마, 랜섬웨어 등은 최근 기승을 부리는 악성코드 중 가장 큰 위협으로 조사되고 있으며, IT서비스업체, 정부기관, 금융기관, 통신업체 등을 대상으로 한 사이버범죄에도 활용되고 있다. 그러나 개별 악성코드의 원리나 세부 동작 등에 대한 미시적 관점의 연구는 비교적 많이 진행됐으나, 악성코드의 다양한 확산에 대한 거시적 관점의 연구는 미약한 실정이다.

본 논문에서는 악성코드 확산에 따른 특성을 분석하고 모델을 구성하기 위하여 적용 가능한 악성코드 동작 모델을 조사 및 분석한다. 인터넷상에서 영향 요인을 반영하여 악성코드 확산 예측 도구를 개발하고, 일반 인터넷 웜 및 워머블 악성코드 확산 실험을 수행한다. 이를 통하여 대응 전문가가 악성코드 확산의 양상을 예측하고 관련 대응 방안을 언제까지 어떠한 수준으로 유지해야 하는지에 대한 구체적인 도움을 제공할 수 있다.

Ⅱ. 기존연구

여러 악성코드 중 인터넷 웜의 확산에 대해서는 이미 여러 가지 모델이 제안되어 있으나, 실제 인터넷 환경에서 동작하는 악성코드 확산을 모델링 하기 위해서는 여러 가지 가정과 수정이 필요하다. 본 논문에서는 가장 대표적이며 구현하기 쉬운 모델인 SI 모델과 SIR 모델을 설명하고, 이를 개선한 새로운 SI/SIR 모델을 설명한다.

SI 모델[4, 5]에서 각각의 호스트는 웜에 취약한 S(Susceptible) 상태, 웜에 감염된 I(Infectious) 상태의 2가지 상태를 가진다. 전체 N개의 호스트 중 일부 호스트가 S 상태에서 β의 비율로 웜에 감염되어 I 상태로 변경된다. 이에 대한 미분 방정식은 다음과 같다.

$$\frac{dS(t)}{dt} = -\beta S(t) I(t)$$
$$\frac{dI(t)}{dt} = \beta S(t) I(t)$$

SIR 모델[5, 6]에서는 각 호스트는 취약한 S 상태와 감염된 I 상태, 웜에 감염된 후 제거되는 R(Recovery) 상태의 3가지 상태를 가진다. 전체 N개의 호스트 중 일부 호스트가 S 상태에서 β의 비율로 감염되어 I 상태로 변경되고, 그중 I 상태에서 γ의 비율로 복구되어 R 상태로 변경된다. 이에 대한 미분 방정식은 다음과 같다.

$$\frac{dS(t)}{dt} = -\beta S(t) I(t)$$
$$\frac{dI(t)}{dt} = \beta S(t) I(t) - \gamma I(t)$$
$$\frac{dR(t)}{dt} = \gamma I(t)$$

SI/SIR 모델[7]은 악성코드의 확산이 대응 시점 λ를 기준으로 하여 확산 단계(Spread Period)와 대응 단계(Response Period)로 나뉘어 동작한다. 앞의 모델에서는 확산에 대한 설명은 가능하지만, 치료 또는 대응에 따른 확산의 감소 정도는 알 수 없다. 이러한 문제점을 해결하고 더욱 실제적인 웜의 확산에 대한 확산 단계와 대응 단계를 모두 고려한 미분 방정식은 다음과 같다.

$$\frac{dS(t)}{dt} = -\beta(t) S(t) I(t) - \delta(t) S(t)$$
$$\frac{dI(t)}{dt} = \beta(t) S(t) I(t) - \gamma(t) I(t)$$
$$\frac{dR(t)}{dt} = \gamma(t) I(t) + \delta(t) S(t)$$

$$\begin{cases} \beta(t) \neq 0, \gamma(t) = 0, \delta(t) = 0 & (0 \leq t < \lambda) \\ \beta(t) \neq 0, \gamma(t) \neq 0, \delta(t) \neq 0 & (\lambda \leq t < \infty) \end{cases}$$

단, SI 모델과 SIR 모델에서 β가 고정된 상숫값인데 본 논문에서는 Two-factor Worm Model[8]과 마찬가지로 시간에 따라 변화하는 함수 $\beta(t)$로 나타낸다.

$$\beta(t) = \beta(0)(1 - i(t))^{\phi},\ \beta(0) = \frac{\eta}{\Omega}$$

여기서, $\beta(0)$는 초기 확산율로 β와 같으며, 웜의 단위 시간당 평균 스캐닝 수 η와 웜이 스캐닝할 수 있는 전체 호스트 주소 공간 Ω에 의해 계산된다. ϕ는 감염 호스트 비율에 따라 변화하는 감염률을 반영하는 값이다. 마찬가지로 $\gamma(t)$, $\delta(t)$도 시간에 따라 변화하는 함수로 나

타낸다. 이는 확산이 진행될수록 오버헤드 등에 따른 감소분을 반영한 것이다.

한편, 2019년 5월 CVE-2019-0708로 공개된 보안 취약점 Remote Desktop Services Remote Code Execution Vulnerability는 보안이 취약한 컴퓨터에 원격에서 접속한 후 공격자가 원하는 코드를 실행할 수 있는 취약점인데, Microsoft는 이를 공개하면서 “the vulnerability is wormable”이라고 하였다[9]. 이는 사용자 인증이나 동작 없이 네트워크를 통하여 시스템 사이를 자동으로 확산하게 하는 취약점을 의미한다. 이를 악용하는 대표적인 워머블 악성코드로는 WannaCryptor, Petya와 같은 랜섬웨어가 있다. 워머블 악성코드는 특정 취약점을 이용한 악성코드가 기존 웜과 같은 방식으로 확산하는 새로운 방식의 악성코드로, 인터넷 웜이 될 수도 있고 트로이목마나 랜섬웨어가 될 수도 있다. 기존 자신만의 고유한 동작 방식을 가지지만, 확산 방식은 인터넷 웜의 확산 방식을 그대로 사용하는 것이다. 이를 통하여 사용자의 동작 없이도 스스로 취약점을 이용하여 악성코드를 감염시킨 후 또 다른 취약한 호스트를 스캐닝하고 네트워크를 통하여 다른 호스트로 확산하는 형태를 가진다. 이러한 워머블 악성코드의 확산 방식은 자기 자신을 복제하여 네트워크로 확산하는 인터넷 웜과 동일한 방식이므로 매우 빠른 속도로 확산할 수 있고, 이를 통하여 막대한 피해를 입힐 수 있다. 특히, 확산의 측면에서 한계를 가질 수밖에 없는 트로이목마, 랜섬웨어, 루트킷과 같은 기존 악성코드는 확산의 한계를 뛰어넘는 매우 효과적인 방식이 될 수 있다.

Ⅲ. 악성코드 확산 예측 도구 설계와 구현

본 논문에서는 기존 확산 모델을 기반으로 워머블 악성코드에 적합한 확산 예측 도구를 개발한다. 워머블 악성코드 확산에 필요한 필수 파라미터와 각 모델의 특성에 따른 파라미터 등을 입력하여 정확한 확산 양상을 예측할 수 있고, 다양한 확산 형태를 비교함으로써 악성코드 대응에 대한 적절한 시기와 방법을 적용할 수 있다. 제안 악성코드 확산 예측 도구는 Windows 10 환경에서 Anaconda 3.5 배포판의 Python 3.7을 이용하여 개발하였다. 추가로 사용자 인터페이스를 위하여 PyQt5 라이브러리를, 과학기술용 수치 처리를 위하여 numpy 라이브러리를, 그래프 및 시각화를 위하여 matplotlib 라이브러리를 사용하여 구현하였다.

제안 확산 예측 도구의 시스템 개요는 그림 1과 같이 3부분으로 나누어진다. 입력 부분에서는 사용자 인터페이스를 이용하여 기본 파라미터인 전체 확산 규모, 시간, 최초 감염 호스트 수를 입력한다. 확산 모델을 선택하고, 각 확산 모델에 따른 필수 파라미터를 함께 시각화를 위한 오버랩 여부, 그래프 색깔 등을 선택하여 실행한다. 엔진 부분에서 모델링에 대한 내부 동작을 수행하고 메타데이터를 생성 처리하고 사용자 입력 등에 대한 이벤트 처리 등을 수행한다. 출력 부분에서 시각화에 대한 그래프를 출력하거나 시뮬레이션 처리를 통하여 실시간으로 해당 시점의 상태를 모두 보여준다.

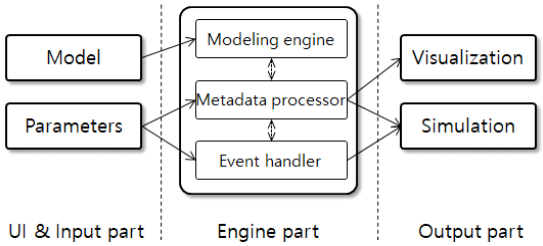

Fig. 1 System overview

표 1은 제안 확산 예측 도구가 지원하는 확산 모델과 각 확산 모델에 필요한 파라미터들을 보여준다. 모델과 무관하게 적용되는 파라미터도 존재하지만, 각 모델에 따라 추가되어야 하는 파라미터도 존재하므로 각 모델의 세부내용을 이해하고 있어야 적절한 값을 선택할 수 있다. 여기서, 확산율 β는 악성코드 특성에 따라 구분되는 고유한 값이지만 그 이외의 값인 감염 호스트 $I(0)$, 복구율 γ, 면역률 δ, 대응 시점 λ, 인터넷 속도 등은 사용자가 직접 입력하는 값이다.

Table. 1 Supported models and its parameters

Model	Default parameters	Required parameters
SI	N, T, $I(0)$	β
SIR	N, T, $I(0)$	β, γ
SI/SIR	N, T, $I(0)$	β, γ, δ, λ

제안 확산 예측 도구의 동작은 그림 2와 같다. 각종 파라미터를 사용자에게 직접 입력받고 선택된 모델에

따라 파라미터를 적절히 사용하여 메타데이터를 생성한 후 사용자에게 시각화하여 보여준다.

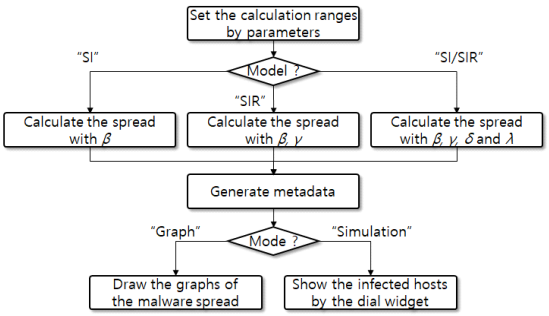

Fig. 2 The proposed prediction tool screen shot

확산 예측 도구의 초기화면은 그림 3과 같다. 확산 모델과 시각화에 필요한 파라미터를 선택할 수 있게 되어 있다. 특히 확산율 β는 가장 중요한 파라미터로 스캐닝 수와 전체 인터넷 주소 공간으로 계산할 수 있다. 스캐닝 수가 정해진 경우와 별도의 계산이 필요한 경우로 나누어 반영하도록 구현하였다.

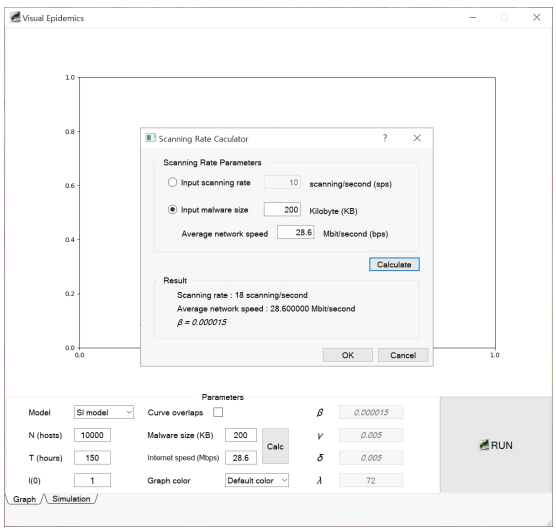

Fig. 3 The proposed prediction tool screen shot

제안 도구에서 여러 확산 모델을 사용하는 이유는 악성코드의 확산과 대응 효과를 모두 설명하기 위함이다. 악성코드 확산을 SI 모델에게만 적용하면 악성코드 확산의 설명만 가능하다. SIR 모델에게만 적용하면 감염된 호스트만 복구될 수 있으므로 감염되기 이전 악성코드 대응의 효과를 분석할 수 없다. SI/SIR 모델은 이러한 문제점을 해결하고 악성코드 확산, 대응과 면역을 좀 더 정확하게 설명하기 위한 모델이라 할 수 있다.

Ⅳ. 악성코드 확산 예측 도구 실험

제안 확산 예측 도구를 이용한 악성코드 확산 실험을 수행하기 위하여 다음과 같은 가정이 필요하다.

<가정>

① 각 호스트는 동일한 악성코드에 중복으로 감염되지 않는다.

② 악성코드는 취약 호스트가 속한 네트워크의 평균 속도에 맞추어 확산한다.

③ 악성코드를 치료하거나 미리 대응하여 면역성을 가지는 호스트의 경우 동일한 악성코드에 다시 감염되지 않는다.

④ 각 호스트의 성능, 스위치 및 라우터에서 일어나는 지연, 네트워크 장비에서 발생하는 패킷 오버헤드 등은 거시적 관점의 악성코드 확산에서는 무시한다.

4.1. 확산 예측 도구의 악성코드 확산 실험

그림 4는 SI 모델에서 감염 가능한 취약 호스트 수 N=10,000, 최초 감염 호스트 1대의 조건으로 평균 인터넷 속도 28.6Mbps 환경에서 200KB 악성코드가 확산하는 실험 결과이다.

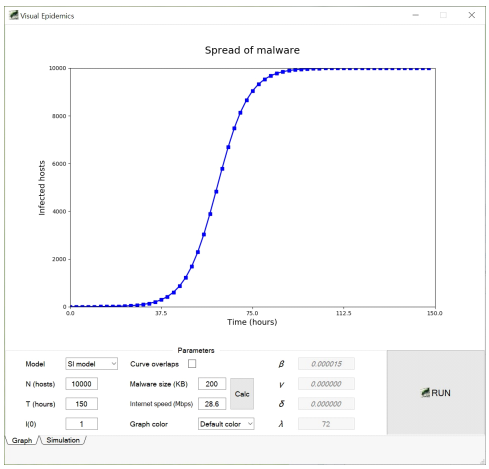

Fig. 4 The spread of malware in terms of time

그림 5는 시뮬레이션 환경으로 같은 조건으로 시간의 추이에 따른 감염 호스트 수를 가상의 네트워크 노드로 표현하여 준다. 그림은 확산 48시간 후 1,301대가 감염되었다는 것을 보여준다.

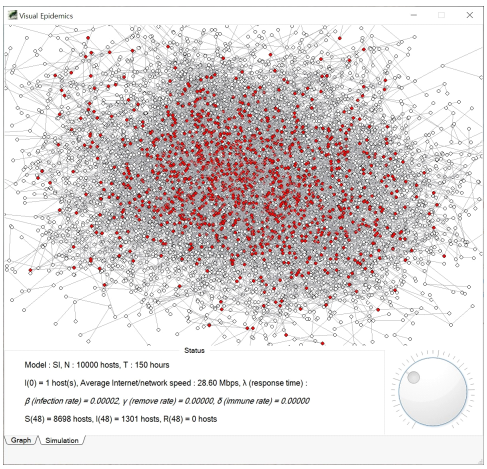

Fig. 5 The spread of malware in terms of node

그림 6은 SI 모델에서 다른 조건은 모두 같지만, 악성코드의 크기가 150KB, 200KB, 250KB인 경우를 가정하여 각각 확산한 결과 그래프를 중첩하여 보여주고 있다. 단, 그림 6부터는 제안 도구의 파라미터 부분은 제외하고 그래프 부분만 발췌하여 보여주도록 한다.

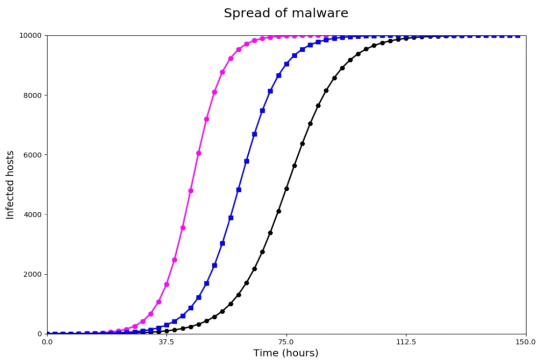

Fig. 6 The spread of 3 types malware at different sizes

그림 7은 SI/SIR 모델에서 감염 가능한 취약 호스트 수 N=10,000, 최초 감염 호스트 1대의 조건으로 평균 인터넷 속도 28.6Mbps 환경에서 악성코드의 크기가 150KB, 200KB, 250KB인 경우를 가정하여 각각 확산한 결과 그래프를 중첩하여 보여주고 있다. 여기서 λ는 대응을 시작한 시점으로 확산 후 72시간부터인데 확산이 줄어드는 것을 확인할 수 있다.

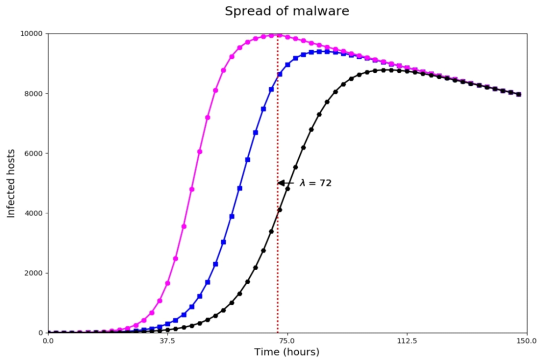

Fig. 7 The spread of 3 types malware at same response period

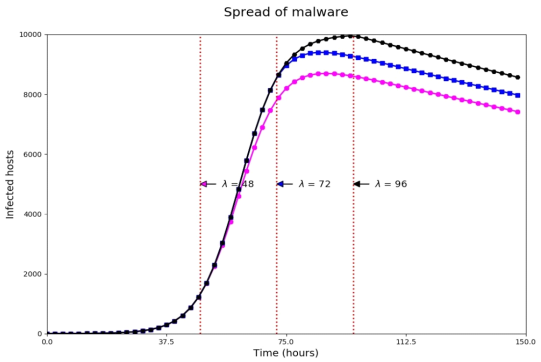

Fig. 8 The spread of 3 types malware at different response periods

그림 8은 SI/SIR 모델에서 다른 조건은 모두 같지만, 악성코드의 크기는 200KB로 고정하고 대응 시점 λ를 48, 72, 96시간인 경우를 가정하여 각각 확산한 결과 그래프를 중첩하여 보여주고 있다. 빨리 대응을 시작할수록 확산이 빠른 속도로 줄어드는 것을 확인할 수 있다.

4.2. 확산 예측 도구의 워머블 악성코드 확산 실험

WannaCry, WannaCrypt 등 여러 이름으로 불리는 WannaCryptor는 대표적인 워머블 악성코드로 전 세계에 막대한 피해를 끼쳤고, 수많은 변형들이 발견된 상태이다. WannaCryptor는 사용자 동작으로 감염되는 기존 랜섬웨어와 달리 Windows의 SMB(Sever Message Block)

취약점을 이용하여 악성코드를 감염시킨 후 접속 가능한 IP 주소를 스캐닝하여 네트워크를 통해 다른 시스템으로 확산한다[10, 11]. 특히, WannaCryptor는 네트워크 속도와 관계없이 초당 10회의 스캐닝을 한다는 것이 알려져 있다[12]. 이를 이용한 확산 그래프는 그림 9와 같다. SI 모델에서 감염가능한 취약 호스트 수 N=10,000, 최초 감염 호스트 1대의 조건으로 평균 인터넷 속도 28.6Mbps 환경에서 확산하는 경우이다.

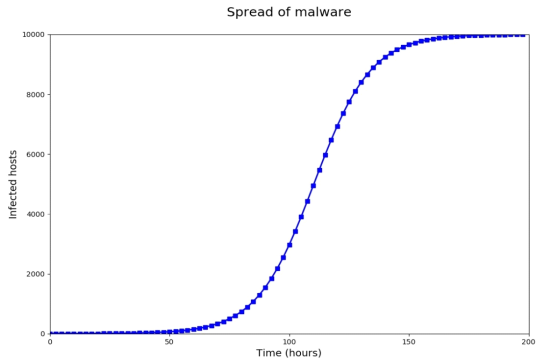

Fig. 9 The estimated spread of WannaCryptor

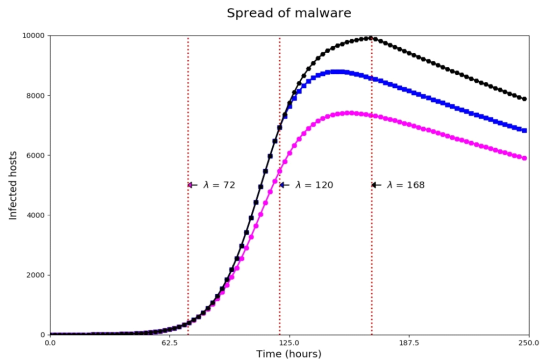

Fig. 10 The estimated spread of 3 types WannaCryptor at different response periods

그림 10은 앞과 같은 조건으로 WannaCryptor가 SI/SIR 모델에서 확산하고, 대응 시점 λ를 72, 120, 168 시간인 경우를 가정하여 각각 확산한 결과 그래프를 중첩하여 보여주고 있다. 168시간 이후는 취약한 호스트 모두가 WannaCryptor에 감염된 상태에서 비로소 대응을 시작하는 것으로 이미 막대한 피해가 발생한 이후이다. 따라서 최소한 그 이전에 이미 대응을 시작하여야 함을 직관적으로 보여주고 있다.

Ⅴ. 결 론

악성코드에 관한 기존연구는 주로 미시적 관점의 연구가 진행되었고 거시적 관점에서 악성코드 확산 연구는 미약한 실정이다. 특히 대응 조직 및 대응 전문가 관점에서 악성코드 특성에 따른 확산을 예측하고 대응 방안의 시행 시점과 효과 분석에 관한 연구도 거의 전혀 없는 실정이다.

본 논문에서는 이러한 문제점을 인식하고 악성코드 확산 모델링에 기반을 둔 확산 예측 도구를 개발하였다. 이를 위하여 관련 연구를 살펴보고 시스템 구성과 구현 방법을 살펴본 후 확산 예측 도구를 이용하여 일반 인터넷 웜 및 워머블 악성코드 WannaCryptor 확산 실험을 수행하였다.

제안 확산 예측 도구는 일반인뿐만 아니라 네트워크, 해킹 및 사이버테러를 다루는 전문 조직 또는 악성코드 대응을 전문으로 하는 조직의 전문가들이 활용할 수 있다. 감염 호스트, 확산율, 면역률, 복구율, 인터넷 속도, 대응 시점 등 악성코드 확산에 영향을 끼칠 수 있는 다양한 요인에 대한 효과를 시뮬레이션하여 정확한 예측을 할 수 있도록 해준다. 또한, 최근 부상하고 있는 워머블 악성코드에 대한 기본 지식만으로도 거시적 관점의 여러 조건에서 확산 형태를 예측하고 다양한 대응 방안을 모색할 수 있게 해준다.

ACKNOWLEDGEMENT

This Research was supported by the Tongmyong University Research Grants 2018 (2018A014)

REFERENCES

[1] Microsoft, Microsoft Security Intelligence Report Vol. 24, [Internet]. Available: https://info.microsoft.com/ww-landing-M365-SIR-v24-Report-eBook.html?lcid=en-us

[2] IBM, IBM X-Force Threat Intelligence Index 2019, [Internet]. Available: https://xforceintelligenceindex.mybluemix.net/

[3] S. Kim, J. Yoo, "A Study on Prediction of Malicious Code Infection Websites Using Markov Chain", *Journal of Security Engineering*, Vol.14, No.1, pp. 9-20, 2017.

[4] H. W. Hethcote, "The Mathematics of Infectious Diseases", *SIAM Review*, vol. 42, No. 4, pp.599-653, 2000.
[5] Y. Wang, S. Wen, Y. Xiang, and W. Zhou, "Modeling the Propagation of Worms in Networks: A Survey", IEEE Communications Surveys and Tutorials, Vol. 16, Issue 2, pp. 942-960, 2014.
[6] J. D. Murray, *Mathematical Biology*, Springer International Publishing, 1993.
[7] W. Shin, "A Study on the Spread and Responses of Mobile Worms by Wireless Network Environments", *Journal of Korea Institute of Information and Communication Engineering*, vol. 10, No. 4, pp.429-440, 2013.
[8] C. C. Zou, W. Gong, D. Towsley. "Code Red Worm Propagation Modeling and Analysis", in Proceedings of the 9th ACM conference on Computer and communications security, pp.138-147, 2002.
[9] Microsoft Security Response Center, Prevent a worm by updating Remote Desktop Services (CVE-2019-0708) [Internet]. Available: https://msrc-blog.microsoft.com/2019/05/14/prevent-a-worm-by-updating-remote-desktop-services-cve-2019-0708/
[10] AhnLab ASEC, "WannaCryptor Ransomware Analysis", AhnLab, Analysis Report, 2017.
[11] KISA KrCERT, WannaCry Analysis Special Report [Internet]. Available:, https://www.krcert.or.kr/data/reportView.do?bulletin_writing_sequence=26747
[12] T. H. Park, K. B. Kim and W. Shin, "A Prediction Model for the Spread of WannaCryptor Ransomware," in *Proceeding of Conference on Information Security and Cryptography-Summer 2019*, Busan, pp. 204-208, 2019.

신원(Weon Shin)
부경대학교 전자계산학과 이학박사
AhnLab Inc. 선임연구원
동명대학교 정보보호학과 교수
※관심분야 : 소프트웨어 보안, 악성코드 확산 대응, 디지털포렌식