전자우편기반의 다형성 바이러스에 대한 거시적 관점에서의 대응기법

김철민[O] 이성욱 홍만표
아주대학교 정보통신전문대학원
ily@ajou.ac.kr[O], suleeip@yahoo.co.kr mphong@ajou.ac.kr

Macroscopic Treatment to Polymorphic E-mail Based Viruses

Cholmin Kim[O] Sung-uck Lee Manpyo Hong
Graduate School of Information and Communication, AJOU university

요 약

최근의 많은 바이러스들이 전달매체(Vector)로서 전자우편을 이용하며 탐지를 어렵게 하기 위해 다형화 기법을 사용한다. 현재의 안티 바이러스(Anti-virus) 제품들은 일반적인 바이러스와 전자우편에 의해 전파되는 바이러스에 대한 탐지 방식을 구분하지 않고 있으므로 전자우편 기반의 바이러스 탐지에 적절치 않다 [1]. 또한 시그너쳐(signature) 기반의 탐지 방식을 이용하므로 다형성 바이러스를 탐지 할 수 없다. 본 논문에서는 전자우편을 통해 전파되는 바이러스를 대상으로 전자우편의 특성을 이용하되 시그너쳐에 의존하지 않음으로써 다형성 바이러스를 탐지 할 수 있는 기법을 소개한다.

1. 서 론

전자우편을 기반으로 하는 바이러스(편의상 이 논문에서는 인터넷 웜을 포함한다.)는 현재의 바이러스 중에 가장 일반적인 형태가 되었다. 유력한 인터넷 보안업체인 CA(Computer Associate)에 따르면 2001년 최악의 피해를 입힌 10개의 바이러스 중 9개가 전자우편을 통해 전파될 수 있는 것으로 나타났다 [2].

그러나 이러한 바이러스에 대응하기 위한 현재의 기법들은 바이러스의 전파경로에 대한 정보를 이용하지 않으므로 적절한 대응을 하지 못하고 있다 [3]. 즉 바이러스가 전자 우편을 이용함으로 인해 발생되는 특징을 간과하게 되는 것이다. 또한 현재의 바이러스들은 다형성 기법을 이용하므로 시그너쳐 탐지에 의존하는 안티 바이러스로는 탐지의 한계가 있다 [4][5].

본 논문에서는 바이러스가 이용하는 전달 매체인 전자우편의 거시적 특성에 착안하여 바이러스가 첨부된 것으로 의심되는 전자 우편을 식별하는 방법으로 전자우편 기반의 다형성 바이러스에 대한 탐지 기법을 제안한다.

2. 관련연구

2.1 거시적 관점에서의 연구

본 논문에서 제안하는 기법은 바이러스를 막기 위한 거시적 관점에서의 연구를 선행 연구로 하고 있다. 거시적 관점에서의 바이러스에 관한 연구는 곧 바이러스가 어떻게 전파되는가에 대한 연구이다 [6].

컴퓨터 바이러스에 대해서도 병리학이 적용될 수 있다. 즉 생물학적인 질병과 같이 컴퓨터 바이러스도 바이러스를 전파시키는 매체(Vector)를 정의하면 각 매체의 종류에 따른 바이러스의 전파특성을 알아낼 수 있는 것이다.

컴퓨터 바이러스 전파에 관한 최초의 저술은 바이러스 전파에 관여하는 여러 변수들을 수학적으로 나타내는 것에서 시작되었다 [7]. 그러나 사람의 행위를 정확하게 수학적으로 나타낼 수는 없으므로 균일성(uniformity)의 가정이 불가피 했고 이로 인해 실제의 바이러스 전파를 나타내지는 못했다.

이후에 바이러스의 확산을 지수적(exponential) 이라고 가정할 때 자원의 공유가 균일하게 일어날 경우의 바이러스 전파에 대한 티펫 이론(Tippet' s theory)이 발표되었다 [8]. 그러나 이것 역시 현실과 맞지 않다. 실제로 바이러스의 전파속도는 시간이 지나면서 점차 감소하게 된다.

최근에는 인공 생명(Artificial Life) 에서의 접근 방법으로 바이러스의 전파를 결정짓는 것은 결국 바이러스가 전파될 환경의 위상이라는 것이 밝혀 졌다 [9][10].

2.2 바이러스의 다형성에 관한 연구

최근의 바이러스들은 점점 더 복잡한 다형성 기법을 이용하여 안티 바이러스 프로그램으로부터 자신을 숨기려 한다. 자주 이용되는 다형성 기법은 암호화와 코드 순서 섞기 이다 [5]. 이러한 다형성 기법들은

전자우편을 통해 전파되는 바이러스에도 그대로 적용된다. 엄밀히 말해 이러한 기법을 이용하는 전자우편기반의 바이러스를 시그너쳐 방식으로 탐지하는 것은 불가능하다 [4]. 따라서 시그너쳐에 의존하지 않는 탐지 방식이 필요하다.

3. 제안된 기법

바이러스의 가장 일반적인 목적 중 하나는 복제된 후손 바이러스들을 최대한 짧은 시간에 최대한 많은 사용자에게 퍼트리는 것이다. 이러한 특성에 따라 전자우편을 기반으로 한 바이러스는 짧은 시간 안에 최대한 많은 전자우편을 생성하여 각각의 수신자에게 바이러스가 첨부된 전자우편이 전달되도록 한다. 이를 위해 바이러스는 현재 감염된 사용자의 전자우편 주소록을 검색하여 다음 수신자들의 주소를 얻는다. 따라서 바이러스의 전파를 관찰하면 하나의 발신자로부터 다수의 수신자로 가는 우편이 연쇄적으로 발생한다는 것을 알 수 있다. 이러한 연쇄적인 우편을 찾아내는 것을 통해 바이러스가 첨부된 것으로 의심되는 우편을 가려낼 수 있다.

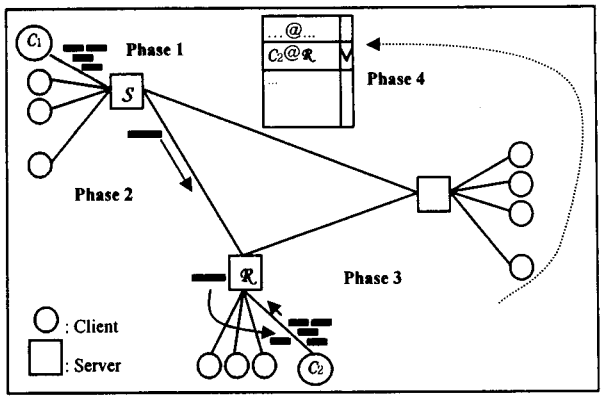
한편 전자우편 서버는 전자우편의 게이트웨이 역할을 하므로 위에서 언급한 연쇄적인 전자 우편의 전달을 찾아내기에 적합한 장소이다. 전자 우편서버를 이용하여 [그림 1]과 같은 4단계로 나누어진 바이러스 탐지가 가능하다.

Phase 1: 전자우편 서버 S가 단일 발신자로부터 다수의 수신자에 대한 다량의 전자우편을 포착한다. 이때 S는 해당 우편의 발신자 리스트를 작성하고 각각의 우편에 의심스러운 것임을 나타내는 태그를 붙여 발송한다.
Phase 2: Phase 1에서 검출된 우편 수신자중 하나의 서버인 $\mathcal{R}$이 태그가 붙은 우편을 전달 받는다. 서버는 S의 주소와 우편 수신자의 주소를 저장한다.
Phase 3: $\mathcal{R}$이 단일 발신자로부터 다량의 우편을 포착하고 이 단일 발신자가 Phase 2에서 저장된 수신자와 같은 경우 이사실을 S 에게 알려 S가 지니고 있는 발신자 리스트에 체크되도록 한다.
Phase 4: S가 지니고 있는 발신자 리스트에 체크된 발신자 수가 일정한 비율을 넘게 되면 S는 Phase1에서 처음 발신자 리스트를 생성시킨 우편이 바이러스를 포함한다고 결정한다.

[그림 1] 4단계 바이러스 탐지 과정

[그림 2]는 제안된 기법에 의한 바이러스 탐지 시나리오를 보여주고 있다. 전자 우편 사용자인 C_1이 바이러스에 감염되었으며 C_1의 주소록에는 C_2가 포함되어 있었다고 가정하자. 이때 [그림 2]는 다음과 같은 단계로 설명할 수 있다.

1. 바이러스는 C_1에서 활동을 시작하여 주소록 검색 후 자신을 복제한 바이러스가 첨부된 전자 우편을 주소록에 있는 사용자들에게 보낸다.
2. (Phase 1) C_1의 전자 우편 서버인 S는 C_1으로부터 동시에 다수의 전자 우편이 발송되었음을 감지 한다. 이때 S는 전자 우편이 목적지로 발송되기 전에 모든 수신자의 리스트를 작성하고 의심스러운 우편임을 알려주는 태그를 각 우편에 삽입한다.



[그림 2] 제안된 기법의 바이러스 탐지 시나리오

3. (Phase 2) C_2의 전자 우편 서버인 $\mathcal{R}$은 태그가 삽입된 전자 우편을 받은 후 이 우편을 보낸 서버 S 와 수신자 C_2의 주소를 저장한다. 전자 우편은 태그가 제거되어 정상적인 형태로 C_2에게 전송된다.
4. C_2로 전파된 바이러스는 C_2를 감염시켜 활동을 시작한다. 바이러스는 C_2의 주소록을 검색 하여 자신을 복제한 바이러스가 첨부된 전자 우편을 주소록에 있는 사용자들에게 보낸다.
5. (Phase 3) $\mathcal{R}$은 C_2로부터 동시에 다수의 전자 우편이 발송되었음을 감지 한다. $\mathcal{R}$은 C_2의 주소를 저장하고 있었으므로 C_2가 의심스러운 메일을 받은 후 다수의 메일을 보냈음을 알 수 있다. $\mathcal{R}$은 이것을 S에게 알리고 S는 수신자 리스트에 해당 수신자를 체크한다
6. C_1으로부터 전자 우편을 받은 다른 서버들도 위의 단계 2 ~ 5까지를 수행한다.
7. (Phase 4) S가 가지고 있는 수신자 리스트에 체크된 수신자의 수가 일정 임계값을 넘을 경우 S는 단계 1 에서 C_1으로부터 발신된 전자 우편이 바이러스를 포함하고 있었다고 판단한다.

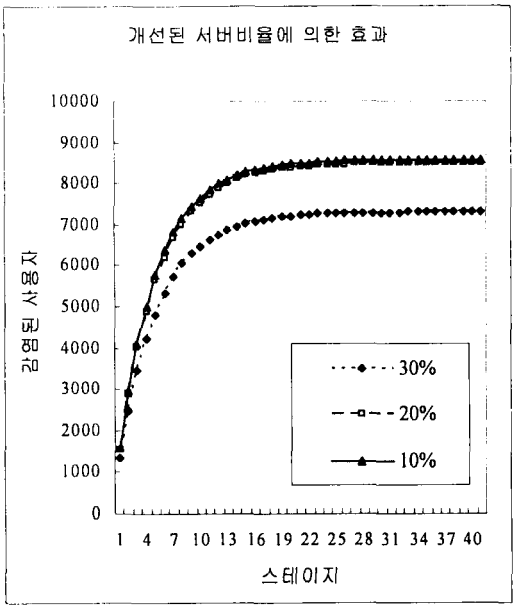
위의 시나리오와 같이 제안된 기법이 현실화 되기 위해서는 전자 우편서버 중 일부를 개선할 필요가 있으며 개선된 서버의 비율에 따른 효과는 다음 장에 설명하고 있다.

단계 7에서 바이러스를 탐지한 서버 S는 처음 바이러스를 발생시킨 것으로 보이는 사용자와 바이러스를 전달 받은 다른 사용자들에게 경고 메시지를 전달함으로써 이후의 바이러스 전파를 지연 시킬 수 있다.

4. 실 험

제안된 기법의 효과를 검증하기 위해 본 연구에서는 간단한 시뮬레이션 환경을 만든 후 가상적으로

바이러스의 전파를 실험하였다. 실험에는 스크립트 형 시뮬레이션 툴인 Simscript II가 사용되었다 [11]. [그림 3]은 제안된 기법을 이용한 전자 우편서버가 설치되었다고 가정할 때의 바이러스 전파 추이를 보이고 있다.



[그림 3] 개선된 서버 비율에 따른 바이러스 전파 추이

이 그래프에 따르면 제안된 기법을 탑재한 개선된 서버 비율이 낮을 때 시간이 지날수록 감염된 사용자의 수가 빨리 증가하는 반면 개선된 서버의 비율이 높으면 감염 속도가 느리다는 것을 알 수 있다. 즉 개선된 서버 비율에 의해 감염 그래프가 오른쪽으로 기울게 되는 것을 알 수 있다. 또한 그래프를 수직 축과 평행하게 보면 개선된 서버비율이 높을수록 평행상태에서 감염된 사용자 수가 적어지는 것을 알 수 있다.

5. 결 론

본 연구에서는 바이러스 전파의 거시적인 특성을 이용하여 전자 우편 기반의 바이러스를 탐지하는 기법을 제시하였다. 제안된 기법은 시그너쳐를 이용하지 않으므로 다형성 바이러스를 탐지 할 수 있다. 또한 제안된 기법을 탑재한 서버들이 일정 비율이상 설치되는 것으로 바이러스의 전파가 느려짐을 시뮬레이션을 통해 보였다. 이후 제안된 기법을 발전시켜 바이러스의 거시적 전파 특성과 미시적 특성을 동시에 이용하여 바이러스를 탐지하는 연구가 가능할 것이다.

참고문헌

[1] Igor Muttik, " STRIPPING DOWN AN AV ENGINE", *Proceeding of the VIRUS BULLETIN CONFERENCE*, pp. 59-68, 2000.

[2] Computer Associate International Inc., " CA Releases Top 10 Virus List for 2001", *http://www3.ca.com/press/pressrelease.asp?id=1856*, 2001.

[3] Cholmin Kim, Seong-uck Lee, Hyeongchol Jung, Yoosuk Jung and Manpyo Hong, " Macroscopic Treatme-nt to E-mail Based Viruses", *Proceeding of the SAM'02*, 2002.

[4] Gabor Szappanos, " ARE THERE ANY POLYMORPHIC MACRO VIRUSES AT ALL? (…AND WHAT TO DO WITH THEM)", *Proceeding of the VIRUS BULLETIN CONFERENCE*, 2002.

[5] Vesselin Bontchev, " MACRO AND SCRIPT VIRUS POLYMORPHISM", *Proceeding of the VIRUS BULLETIN CONFERENCE*, pp. 406-438, 2002.

[6]Frederick B. Cohen, " A Short Course on Computer Viruses 2nd Edition", *John Wiley & Sons*, Inc., pp. 121 - 134, 1994.

[7] Winfried Gleissner, " A mathematical theory for the spread of computer viruses", *Computers & Security*, vol. 8, pp. 35-41, 1989.

[8] Tipet, The Tipet Theory of Computer Virus Propagation, " Foundationware", USA, 1990.

[9] Jeffrey O. Kephart, David M. Chess and Steve R. White, " Computers and Epidemiology", *IEEE Spectrum*, v30 n5, pp. 20-26, 1993.

[10] Jeffrey O. Kephart, " How Topology Affects Population Dynamics", *Proceeding of the Artificial Life III Studies in the Science of Complexity*, Proc. Vol. XVII, pp. 447 - 463, 1993.

[11] Edward C. Russell, " Building Simulation Models with SIMSCRIPT II.5", *CACI Products Company*, 1989.