



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2009-0022960
(43) 공개일자 2009년03월04일

(51) Int. Cl.

G06Q 50/00 (2006.01)

(21) 출원번호 10-2007-0088684

(22) 출원일자 2007년08월31일

심사청구일자 2007년08월31일

(71) 출원인

고려대학교 산학협력단

서울 성북구 안암동5가1 고려대학교 내

(72) 발명자

김도훈

서울 마포구 창전동 삼성아파트 103동 302호

인호

서울 성북구 안암동5가 고려대학교 자연계 캠퍼스
과학도서관604D

(74) 대리인

현종철

전체 청구항 수 : 총 15 항

(54) 사이버 기상 예측을 위한 악성 코드 분석 장치, 그 방법 및 이를 기록한 기록매체

(57) 요약

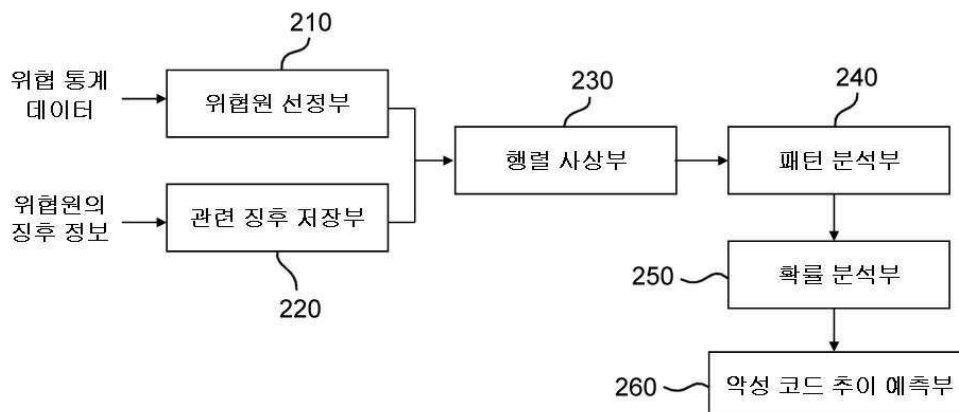
사이버 기상 예측을 위한 악성 코드 분석 장치, 그 방법 및 이를 기록한 기록매체가 개시된다.

본 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석 장치는,

네트워크 기반의 사이버 위협을 발생시키는 위협원의 위협 통계 데이터를 기반으로 악성 코드를 발현하는 위협원을 선정하는 위협원 선정부; 상기 네트워크 기반의 사이버 위협을 발생시키는 위협원에 따라 발생된 위협원의 징후 정보를 저장하는 관련 징후 저장부; 상기 위협원 선정부에서 선정된 위협원과 상기 관련 징후 저장부에 저장된 상기 위협원의 징후 정보를 행렬로 사상하여 혼합 행렬을 생성하는 행렬 사상부; 상기 행렬 사상부에서 생성된 혼합 행렬을 이용하여 상기 위협원과 상기 위협원의 징후 정보의 상관 분석을 수행하고, 상기 위협원의 징후 정보의 발현 주이 패턴을 나타내는 파라미터를 생성하는 패턴 분석부; 상기 위협원의 통계 데이터를 복수 개의 시구간으로 분할하고, 예측하고자 하는 시구간의 위협원의 악성 코드 발현 확률을 산정하는 확률 분석부; 및 상기 확률 분석부에서 산정된 악성 코드 발현 확률에 따라 사이버 기상 예측을 수행하는 악성코드 주이 예측부를 포함한다.

본 발명에 의하면, 보안 위협원의 징후와 보안 위협원 간의 상관관계에 따른 악성 코드 분석 모델을 제공하여 네트워크 기반의 사이버 위협의 발생을 예측할 수 있고, 다양한 보안 사고의 발생을 통계 자료에 근거하여 징후 자료를 습득하고, 이를 기반으로 특정 사이버 위협의 발생 추이를 다양한 시간의 축에 따라 알 수 있으며, 통계치의 훈련기간에 따라 다양한 예측치를 알 수 있기 때문에 특정 위협에 대한 논리적이고 합리적인 훈련기간을 선택할 수 있고, 이에 따라 특정 위협원의 사이버 위협의 예측 정확도를 높일 수 있으며, 보안 시스템의 운영에 있어서 거시적이고 독자적인 정책을 세울 수 있으며, 특히 보안 솔루션에 대한 투자 대비 수익률을 산정하기 위한 중요한 투자 지표를 제공할 수 있는 효과가 있다.

대표도 - 도2



특허청구의 범위

청구항 1

네트워크 기반의 사이버 위협을 발생시키는 위협원의 위협 통계 데이터를 기반으로 악성 코드를 발현하는 위협원을 선정하는 위협원 선정부;

상기 네트워크 기반의 사이버 위협을 발생시키는 위협원에 따라 발생된 위협원의 징후 정보를 저장하는 관련 징후 저장부;

상기 위협원 선정부에서 선정된 위협원과 상기 관련 징후 저장부에 저장된 상기 위협원의 징후 정보를 행렬로 사상하여 혼합 행렬을 생성하는 행렬 사상부;

상기 행렬 사상부에서 생성된 혼합 행렬을 이용하여 상기 위협원과 상기 위협원의 징후 정보의 상관 분석을 수행하고, 상기 위협원의 징후 정보의 발현 추이 패턴을 나타내는 파라미터를 생성하는 패턴 분석부;

상기 위협원의 통계 데이터를 복수 개의 시구간으로 분할하고, 예측하고자 하는 시구간의 위협원의 악성 코드 발현 확률을 산정하는 확률 분석부; 및

상기 확률 분석부에서 산정된 악성 코드 발현 확률에 따라 사이버 기상 예측을 수행하는 악성코드 추이 예측부를 포함하는 사이버 기상 예측을 위한 악성 코드 분석 장치.

청구항 2

제 1 항에 있어서,

상기 확률 분석부는,

상기 위협원의 악성 코드 및 상기 생성된 파라미터를 이용하여 초기 확률 벡터를 생성하고, 상기 생성된 초기 확률 벡터, 상기 분할된 복수 개의 시구간의 징후 정보 간의 상관 분석의 결과를 행렬로 사상한 징후 정보 상관 행렬 및 상기 혼합 행렬을 이용하여 예측하고자 하는 시구간의 상기 위협원의 악성 코드의 발현 확률을 산정하는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 장치.

청구항 3

제 1 항에 있어서,

상기 악성코드 추이 예측부는

상기 예측하고자 하는 시구간의 악성 코드 발현 빈도를 생성하여, 상기 생성된 악성 코드 발현 빈도에 따라 예측하고자 하는 시구간의 사이버 기상 예측을 수행하는 악성 코드 추이 분석 모듈을 포함하는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 장치.

청구항 4

제 1 항에 있어서,

상기 위협원 선정부는

상기 위협원의 발생 빈도 정보를 수집하여 상기 위협원 발생의 확률 분포를 출력하는 은닉열 모듈을 포함하는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 장치.

청구항 5

제 1 항에 있어서,

상기 관련 징후 저장부는 상기 위협원의 징후 정보를 수집하여 상태 전이 확률을 생성하는 관찰열 모듈을 포함하며,

상기 관찰열 모듈에서 생성되는 상태 전이 확률은 각각의 시구간에 따른 위협원의 조건부 확률 값으로 연산되는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 장치.

청구항 6

제 1 항에 있어서,

상기 복수 개의 시구간으로 분할된 상기 위협원의 통계 데이터는

상기 위협원의 발생 빈도를 기반으로 하여 은닉 마코프 모델에 따른 퍼지 로직 추정 모델에 의해 상기 위협원의 발생 빈도로 설정되는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 장치.

청구항 7

제 6 항에 있어서,

상기 복수개로 분할된 상기 위협원의 발생 빈도 H_n 은 하기의 식 (1)에 의해 연산되는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 장치.

$$\begin{aligned} \beta &= H_1, \\ H_1 + \theta &= H_2, \\ H_2 + \theta &= H_3, \\ &\dots\dots \\ H_{n-1} &= \theta, \\ H_n &= a \end{aligned} \quad (1)$$

(상기 식에서, $N_{\max}$ 가 특정 시구간에서 최대로 발생한 위협원의 발생 빈도를 의미하고, $N_{\min}$ 이 상기 특정 시구간에서 최소로 발생한 위협원의 발생 빈도를 의미하며, $a = \text{Log}_{10}N_{\max}$ 에 따라 상기 패턴 분석부의 파라미터 α 가 생성되며, $\beta = \text{Log}_{10}N_{\min}$ 에 따라 상기 패턴 분석부의 파라미터 β 가 생성되며, 상기 패턴 분석부의 파라미터 ε 가 가장 최근에 발생된 위협원의 발생 빈도를 의미하면, 상기 패턴 분석부의 파라미터 γ 는 $\varepsilon - 1$ 로 정의되며, 상기 패턴 분석부의 파라미터 θ 는 $\theta = \frac{\alpha - \beta}{\gamma}$ 로 정의되고, H_n 을 연속된 각각의 시구간의 위협원의 발생 빈도를 의미한다.)

청구항 8

네트워크 기반의 사이버 위협을 발생시키는 위협원의 위협 통계 데이터를 기반으로 악성 코드를 발현하는 위협원을 선정하는 단계;

상기 네트워크 기반의 사이버 위협을 발생시키는 위협원에 따라 발생된 위협원의 징후 정보를 저장하는 단계;

상기 선정된 위협원과 상기 저장된 상기 위협원의 징후 정보를 행렬로 사상하여 혼합 행렬을 생성하는 단계;

상기 생성된 혼합 행렬을 이용하여 상기 위협원과 상기 위협원의 징후 정보의 상관 분석을 수행하고, 상기 위협원의 징후 정보의 발현 추이 패턴을 나타내는 파라미터를 생성하는 단계;

상기 위협원의 통계 데이터를 복수 개의 시구간으로 분할하고, 예측하고자 하는 시구간의 위협원의 악성 코드 발현 확률을 산정하는 단계; 및

상기 산정된 악성 코드 발현 확률에 따라 사이버 기상 예측을 수행하는 단계를 포함하는 사이버 기상 예측을 위한 악성 코드 분석 방법.

청구항 9

제 8 항에 있어서,

상기 악성 코드 발현 확률을 산정하는 단계는

상기 위협원의 악성 코드 및 상기 생성된 파라미터를 이용하여 초기 확률 벡터를 생성하는 단계; 및

상기 생성된 초기 확률 벡터, 상기 분할된 복수 개의 시구간의 징후 정보 간의 상관 분석의 결과를 행렬로 사상한 징후 정보 상관 행렬 및 상기 혼합 행렬을 이용하여 예측하고자 하는 시구간의 상기 위협원의 악성 코드의 발현 확률을 산정하는 단계를 포함하는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 방법.

청구항 10

제 8 항에 있어서,

상기 사이버 기상 예측을 수행하는 단계는

상기 예측하고자 하는 시구간의 악성 코드 발현 빈도를 생성하여, 상기 생성된 악성 코드 발현 빈도에 따라 예측하고자 하는 시구간의 사이버 기상 예측을 수행하는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 방법.

청구항 11

제 8 항에 있어서,

악성 코드를 발현하는 위협원을 선정하는 단계는

상기 위협원의 발생 빈도 정보를 수집하여 상기 위협원 발생의 확률 분포를 출력하는 단계를 포함하는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 방법.

청구항 12

제 8 항에 있어서,

상기 위협원의 징후 정보를 저장하는 단계는

상기 위협원의 징후 정보를 수집하여 상태 전이 확률을 생성하는 단계를 포함하며,

상기 상태 전이 확률은 각각의 시구간에 따른 위협원 정보의 조건부 확률값으로 연산되는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 방법.

청구항 13

제 8 항에 있어서,

상기 위협원은

상기 위협원의 발생 빈도를 기반으로 하여 은닉 마코프 모델에 따른 퍼지 로직 추정 모델에 의해 위협원의 발생 빈도로 설정되는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 방법.

청구항 14

제 13 항에 있어서,

상기 위협원의 발생 빈도 H_n 은 하기의 식 (2)에 의해 연산되는 것을 특징으로 하는 사이버 기상 예측을 위한 악성 코드 분석 방법.

$$\begin{aligned}
 \beta &= H_1, \\
 H_1 + \theta &= H_2, \\
 H_2 + \theta &= H_3, \\
 &\dots\dots \\
 H_{n-1} &= \theta, \\
 H_n &= a
 \end{aligned}
 \tag{2}$$

(상기 식에서, $N_{\max}$ 가 특정 시구간에서 최대 발생한 위협원의 발생 빈도를 의미하고, $N_{\min}$ 이 상기 특정 시구간에서 최소 발생한 위협원의 발생 빈도를 의미하며, $a = \text{Log}_{10}N_{\max}$ 에 따라 상기 패턴 분석부의 파라미터 a 가 생성되며, $\beta = \text{Log}_{10}N_{\min}$ 에 따라 상기 패턴 분석부의 파라미터 β 가 생성되며, 상기 패턴 분석부의 파라미터 ε 가 가장 최근에 발생된 위협원의 발생 빈도를 의미하면, 상기 패턴 분석부의 파라미터 γ 는 $\varepsilon - 1$ 로 정의되며, 상기 패턴 분석부의 파라미터 θ 는 $\theta = \frac{\alpha - \beta}{r}$ 로 정의되고, H_n 을 연속된 각각의 시구간의 위협원의 발생 빈도를 의미한다.)

청구항 15

제 8 항 내지 제 14 항 중 어느 한 항의 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

명세서

발명의 상세한 설명

기술 분야

- <1> 본 발명은 네트워크의 보안 및 투자에 관한 것으로서, 특히 네트워크상에서 발현될 수 있는 위협의 추이를 예측하며, 시계열 분석과 회귀 분석 뿐만 아니라, 은닉 마코프 모델을 이용한 확률 및 통계적 분석을 통하여 확률 예측을 함으로써 오차율을 줄이고, 위협 발현 추이를 용이하게 분석할 수 있는 사이버 기상 예측을 위한 악성 코드 분석 장치, 그 방법 및 이를 기록한 기록매체에 관한 것이다.

배경 기술

- <2> 최근 범세계적으로 사용되고 있는 인터넷과 접속된 시스템의 특성은 개방형 시스템으로서, 보안에 상당한 취약점이 있다. 이는 아주 정밀하게 구축된 보안 시스템이라고 할지라도 해커의 침입을 완전히 차단할 수 없다.
- <3> 따라서, 민감한 데이터는 분리된 시스템에서 작업하는 것이 바람직하고, 인터넷 서버의 외부 이용자는 계속적으로 외부 해커나 악성 코드 등의 침입을 모니터링 하여야 한다.
- <4> 과거에는 단순히 물리적인 보안, 즉 전산실 출입 통제만으로 어느 정도 보안이 되었으나, 통신기술이 발달하면서 LAN을 통하여 어떠한 데이터도 다운로드하여 불법이용이 가능하게 되었다. 다시 말하여, 새로운 기술에 대한 새로운 보안체제로 대처하지 않으면 보안에는 약점을 가지게 된다.
- <5> 항상 새로운 기술에 대하여 보안의 위험성을 연구하고, 체계화하기 위해서는 보안책임자가 있어야 하고, 데이터의 불법유출 및 오용을 방지하기 위하여 데이터의 내용을 잘 알고 있는 구성원으로 이루어진 이용 심의위원회도 필요한 실정이다.
- <6> 정보화 사회가 진전되면 될수록 필수적으로 많은 자료의 공개를 요구하게 마련이다. 이에 따라 정보보호 및 보안을 위해 고도의 보안기술과 많은 경비 및 인력 등이 필요하게 된다. 역으로 고도의 보호 기술과 전문 인력이

있어야 정보의 공개도 용이할 수 있다.

- <7> 그러나, 현실적으로 정보 보안 기술과 전문 인력이 부족하여 누구도 정보의 공개를 쉽게 결정 못하고 있는 실정이다. 특히 인터넷을 이용한 네트워크에 따른 정보 공개는 엄두도 못 내고 있는 것이 현실이다.
- <8> 뿐만 아니라, 정보 공개와 정보 보호는 서로 그 기능이 상반된다. 정보의 이용자는 많은 정보의 공개를 요구하지만, 그에 반하여 정보 제공자는 공개를 하면 할수록 그 위험도는 높아지기 때문에 공개를 기피하는 경향이 있다. 이를 적절한 선에서 조정하려면 정보에 대한 사이버 위협에 대한 보호가 선행되어야 한다.
- <9> 고도 정보화 사회로의 발전을 위해서는 정보 보호 및 보안 기술의 연구 개발이 필요하다. 이를 위해서 정부 차원의 정책적 제도적 지원에 의존하는 것이 현실이다.
- <10> 현재 국내에서 사용되는 인터넷 관련 보안장비는 대부분 외국 제품으로 장비 내부의 핵심 보안 기술은 블랙 박스화하여 공급받고 있는 실정일 뿐만 아니라, 이미 상용화된 기술의 일부만 공개되어 있는 상황이다.
- <11> 특히 통계 데이터의 경우, 자료 유형에 따라 보호방법이 달리 적용되어야 하는데, 원시 데이터는 철저한 내부 통제가 필요하고, 집계데이터와 요약데이터는 일반 컴퓨터 시스템 안전대책과 인터넷 보안 모델에 준하여 정보 보호가 이루어져야 한다. 그리고, 집계 및 표본조사의 모집단으로 활용되는 유효데이터는 별도의 데이터 보호 방법이 필요한 데이터이다.
- <12> 통계 데이터는 통계 조사 응답자의 보호뿐만 아니라 개인 사생활 보호차원에서 당연히 보호되어야 한다. 반면에 통계 이용자는 좀 더 세분화된 자료, 좀 더 많은 자료의 공개를 요구한다. 이와 같이 당연히 보호되어야 하고, 당연히 공개되어야 하는 서로 상반되는 요구사항이 발생된다.
- <13> 그러나, 정보 제공자는 공개를 하면 할수록 그 위험도가 높아지기 때문에 공개를 기피하는 경향이 있다. 이를 위해서는 앞서 상술한 바와 같이 정보에 대한 외부 공격으로부터의 차단이 전제되어야 하는 실정이다.
- <14> 이와 같이, 종래의 보안 통계 보고서를 살펴보면 특정의 보안 위협원의 월별 혹은 년별 보고는 있었으나, 위협원들의 향후 발생 여부에 대해서는 관심을 가지지 않았으며, 보안 위협원 들의 발생 여부가 주기를 가지거나, 공격자들의 행동을 용이하게 파악할 수가 없었다.
- <15> 따라서, 종래의 보안 위협의 예측 알고리즘은 네트워크 환경에 의존하고, 네트워크의 흐름에 따라서만 보안 위협의 발생 여부를 알 수 있어서 특정 보안 위협의 추이를 예측할 수 없고, 보안 시스템의 운영에 대한 정책을 독자적으로 수립할 수 없는 문제점이 있다.

발명의 내용

해결 하고자하는 과제

- <16> 따라서, 본 발명이 해결하고자 하는 첫 번째 과제는 네트워크 기반의 사이버 위협의 발생을 예측할 수 있을 뿐만 아니라, 보안 위협의 각각의 성격에 적합한 훈련에 따른 다양한 예측치를 제공하고, 보안 시스템의 운영에 따른 독자적인 정책을 수립할 수 있는 사이버 기상 예측을 위한 악성 코드 분석 장치를 제공하는 것이다.
- <17> 또한, 본 발명이 해결하고자 하는 두 번째 과제는 상기 사이버 기상 예측을 위한 악성 코드 분석 장치를 이용한 사이버 기상 예측을 위한 악성 코드 분석 방법을 제공하는 것이다.
- <18> 그리고, 본 발명이 해결하고자 하는 세 번째 과제는 상기 악성 코드 분석 방법을 컴퓨터에서 수행할 수 있는 프로그램으로 기록된 기록매체를 제공하는 것이다.

과제 해결수단

- <19> 상기 첫 번째 과제를 해결하기 위하여 본 발명은,
- <20> 네트워크 기반의 사이버 위협을 발생시키는 위협원의 위협 통계 데이터를 기반으로 악성 코드를 발현하는 위협원을 선정하는 위협원 선정부, 상기 네트워크 기반의 사이버 위협을 발생시키는 위협원에 따라 발생된 위협원의 징후 정보를 저장하는 관련 징후 저장부, 상기 위협원 선정부에서 선정된 위협원과 상기 관련 징후 저장부에 저장된 상기 위협원의 징후 정보를 행렬로 사상하여 혼합 행렬을 생성하는 행렬 사상부, 상기 행렬 사상부에서 생성된 혼합 행렬을 이용하여 상기 위협원과 상기 위협원의 징후 정보의 상관 분석을 수행하고, 상기 수행된 상관 분석에 따라 상기 위협원의 발현 주이 패턴을 나타내는 파라미터를 생성하는 패턴 분석부, 상기 위협원의 통계 데이터를 복수 개의 시구간으로 분할하고, 예측하고자 하는 시구간의 위협원의 악성 코드 발현 확률을 산정하는

확률 분석부, 및 상기 확률 분석부에서 산정된 악성 코드 발현 확률에 따라 사이버 기상 예측을 수행하는 악성 코드 추이 예측부를 포함하는 사이버 기상 예측을 위한 악성 코드 분석 장치를 제공한다.

- <21> 한편, 상기 확률 분석부는, 상기 위협원의 악성 코드 및 상기 생성된 파라미터를 이용하여 초기 확률 벡터를 생성하고, 상기 생성된 초기 확률 벡터, 상기 징후 정보의 상관 분석의 결과를 행렬로 사상한 징후 정보 상관 행렬 및 상기 혼합 행렬을 이용하여 외부에서 미리 설정된 다음 시구간의 상기 위협원의 악성 코드의 발현 확률을 예측 산정하는 것을 특징으로 한다.
- <22> 그리고, 상기 악성코드 추이 예측부는 상기 예측하고자 하는 시구간의 악성 코드 발현 빈도를 생성하여, 상기 생성된 악성 코드 발현 빈도에 따라 예측하고자 하는 시구간의 사이버 기상 예측을 수행하는 악성 코드 추이 분석 모듈을 포함하는 것을 특징으로 한다.
- <23> 아울러, 상기 위협원 선정부는 상기 위협원의 발생 빈도 정보를 수집하여 상기 위협원 발생의 확률 분포를 출력하는 은닉열 모듈을 포함하는 것을 특징으로 한다.
- <24> 또한, 상기 관련 징후 저장부는 상기 위협원의 징후 정보를 수집하여 상태 전이 확률을 생성하는 관찰열 모듈을 포함하며, 상기 관찰열 모듈에서 생성되는 상태 전이 확률은 각각의 시구간에 따른 위협원 정보의 조건부 확률 값으로 연산되는 것을 특징으로 한다.
- <25> 그리고, 상기 복수 개의 시구간으로 분할된 상기 위협원의 위협원 정보는 상기 위협원의 발생 빈도를 기반으로 하여 은닉 마코프 모델에 따른 퍼지 로직 추정 모델에 의해 위협원의 발생 빈도로 설정되는 것을 특징으로 한다.

- <26> 여기서, 상기 복수개로 분할된 상기 위협원의 발생 빈도는 $N_{\max}$ 가 특정 시구간에서 최대로 발생한 위협원의 발생 빈도를 의미하고, $N_{\min}$ 이 상기 특정 시구간에서 최소로 발생한 위협원의 발생 빈도를 의미하며, $\alpha = \text{Log}_{10}N_{\max}$ 에 따라 상기 패턴 분석부의 파라미터 α 가 생성되며, $\beta = \text{Log}_{10}N_{\min}$ 에 따라 상기 패턴 분석부의 파라미터 β 가 생성되며, 상기 패턴 분석부의 파라미터 ϵ 가 가장 최근에 발생한 위협원의 발생 빈도를 의미하면, 상기 패턴 분석부의 파라미터 γ 는 $\epsilon - 1$ 로 정의되며, 상기 패턴 분석부의 파라미터 θ 는 $\theta = \frac{\alpha - \beta}{\gamma}$ 로 정의되고, H_n 을 연속된 각각의 시구간의 위협원의 발생 빈도를 의미할 때, 하기의 식 (1)에 의해 연산되는 것을 특징으로 한다.

$$\beta = H_1,$$

$$H_1 + \theta = H_2,$$

$$H_2 + \theta = H_3,$$

.....

$$H_{n-1} = \theta,$$

$$H_n = \alpha$$

(1)

- <27> 상기 두 번째 과제를 해결하기 위하여 본 발명은,

- <28> 네트워크 기반의 사이버 위협을 발생시키는 위협원의 위협 통계 데이터를 기반으로 악성 코드를 발현하는 위협원을 선정하는 단계, 상기 네트워크 기반의 사이버 위협을 발생시키는 위협원에 따라 발생한 위협원의 징후 정보를 저장하는 단계;

- <29> 상기 선정된 위협원과 상기 저장된 상기 위협원의 징후 정보를 행렬로 사상하여 혼합 행렬을 생성하는 단계, 상기 생성된 혼합 행렬을 이용하여 상기 위협원과 상기 위협원의 징후 정보의 상관 분석을 수행하고, 상기 수행된 상관 분석에 따라 상기 위협원의 발현 추이 패턴을 나타내는 파라미터를 생성하는 단계, 상기 위협원의 통계 데이터를 복수 개의 시구간으로 분할하고, 예측하고자 하는 시구간의 위협원의 악성 코드 발현 확률을 산정하는

단계, 및 상기 산정된 악성 코드 발현 확률에 따라 사이버 기상 예측을 수행하는 단계를 포함하는 사이버 기상 예측을 위한 악성 코드 분석 방법을 제공한다.

- <31> 그리고, 상기 악성 코드 발현 확률을 산정하는 단계는 상기 위협원의 악성 코드 및 상기 생성된 파라미터를 이용하여 초기 확률 벡터를 생성하는 단계, 및 상기 생성된 초기 확률 벡터, 상기 징후 정보의 상관 분석의 결과를 행렬로 사상한 징후 정보 상관 행렬 및 상기 혼합 행렬을 이용하여 외부에서 미리 설정된 다음 시구간의 상기 위협원의 악성 코드의 발현 확률을 예측 산정하는 단계를 포함하는 것을 특징으로 한다.
- <32> 아울러, 상기 사이버 기상 예측을 수행하는 단계는 상기 예측하고자 하는 시구간의 악성 코드 발현 빈도를 생성하여, 상기 생성된 악성 코드 발현 빈도에 따라 다음 시구간의 사이버 기상 예측을 수행하는 것을 특징으로 한다.
- <33> 그리고, 악성 코드를 발현하는 위협원을 선정하는 단계는 상기 위협원의 발생 빈도 정보를 수집하여 상기 위협원 발생의 확률 분포를 출력하는 단계를 포함하는 것을 특징으로 한다.
- <34> 아울러, 상기 위협원의 징후 정보를 저장하는 단계는 상기 위협원의 징후 정보를 수집하여 상태 전이 확률을 생성하는 단계를 포함하며, 상기 상태 전이 확률은 각각의 시구간에 따른 위협원 정보의 조건부 확률값으로 연산되는 것을 특징으로 한다.
- <35> 또한, 상기 위협원은 상기 위협원의 발생 빈도를 기반으로 하여 은닉 마코프 모델에 따른 퍼지 로직 추정 모델에 의해 위협원의 발생 빈도로 설정되는 것을 특징으로 한다.

- <36> 여기서, 상기 위협원의 발생 빈도는 $N_{\max}$ 가 특정 시구간에서 최대로 발생한 위협원의 발생 빈도를 의미하고, $N_{\min}$ 이 상기 특정 시구간에서 최소로 발생한 위협원의 발생 빈도를 의미하며, $\alpha = \text{Log}_{10}N_{\max}$ 에 따라 상기 패턴 분석부의 파라미터 α 가 생성되며, $\beta = \text{Log}_{10}N_{\min}$ 에 따라 상기 패턴 분석부의 파라미터 β 가 생성되며, 상기 패턴 분석부의 파라미터 ε 가 가장 최근에 발생된 위협원의 발생 빈도를 의미하면, 상기 패턴 분석부의 파라미터 γ 는 $\varepsilon - 1$ 로 정의되며, 상기 패턴 분석부의 파라미터 θ 는 $\theta = \frac{\alpha - \beta}{\gamma}$ 로 정의되고, H_n 을 연속된 각각의 시구간의 위협원의 발생 빈도를 의미할 때, 하기의 식 (1)에 의해 연산되는 것을 특징으로 한다.

$$\beta = H_1,$$

$$H_1 + \theta = H_2,$$

$$H_2 + \theta = H_3,$$

.....

$$H_{n-1} = \theta,$$

$$H_n = \alpha$$

(1)

- <37>
- <38> 상기 세 번째 과제를 해결하기 위하여 본 발명은,
- <39> 상기 사이버 기상 예측을 위한 악성 코드 분석 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다.

효 과

- <40> 본 발명에 의하면, 보안 위협원의 징후와 보안 위협원 간의 상관관계에 따른 악성 코드 분석 모델을 제공하여 네트워크 기반의 사이버 위협의 발생을 예측할 수 있고, 다양한 보안 사고의 발생을 통계 자료에 근거하여 징후 자료를 습득하고, 이를 기반으로 특정 사이버 위협의 발생 추이를 다양한 시간의 축에 따라 알 수 있으며, 통계치의 훈련기간에 따라 다양한 예측치를 알 수 있기 때문에 특정 위협에 대한 논리적이고 합리적인 훈련기간을

선택할 수 있고, 이에 따라 특정 위협원의 사이버 위협의 예측 정확도를 높일 수 있으며, 보안 시스템의 운영에 있어서 거시적이고 독자적인 정책을 세울 수 있으며, 특히 보안 솔루션에 대한 투자 대비 수익률을 산정하기 위한 중요한 투자 지표를 제공할 수 있는 효과가 있다.

발명의 실시를 위한 구체적인 내용

- <41> 이하, 본 발명의 바람직한 실시예를 첨부도면에 의거하여 상세히 설명하기로 한다.
- <42> 그러나, 다음에 예시하는 본 발명의 실시예는 여러 가지 다른 형태로 변형할 수 있으며, 본 발명의 범위가 다음에 상술하는 실시예에 한정되는 것은 아니다. 본 발명의 실시예는 당 업계에서 평균적인 지식을 가진 자에게 본 발명을 더욱 완전하게 설명하기 위하여 제공된다.
- <43> 도 1은 본 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석의 개괄도이다.
- <44> 도 1을 참조하면, 네트워크 기반의 사이버 위협을 악성 코드를 통하여 발현시키는 위협원의 악성 코드(110)과 과거 위협원의 징후 정보 데이터(120)를 이용하여 상기 위협원과 위협원의 징후 정보의 상관도를 분석하고, 이를 이용하여 실제 위협원을 통한 사이버 위협, 즉 특정 시구간의 악성 코드를 이용한 사이버 위협의 추이 ($H_1, H_2, H_3, \dots, H_n$)를 시간대 별 예측을 수행한다.
- <45> 이는, 특정 위협원의 과거의 추이를 분석하는 것이 아니라, 과거의 위협원과 위협원의 상관관계를 통하여 실제 사이버 위협 즉, 악성 코드의 추이를 시간대 별로 예측을 하고자 하는 것이다.
- <46> 즉, 본 발명은 네트워크 기반의 사이버 위협의 발생을 예측할 수 있을 뿐 아니라 다양한 보안 사고의 발생을 통계 자료에 근거하여 징후 정보를 습득하고, 이를 기반으로 통계 보안 사고의 발생을 통계 자료에 근거하여 징후 정보를 습득하고, 이를 기반으로 통계 데이터의 패턴을 분석한다. 특히 패턴 분석시에 다양한 확률 통계 기법을 이용하여, 은닉 마코프 모델(Hidden Markov Model:HMM)을 이용하여 특정 사이버 위협의 발생 추이를 다양한 시간의 축에 따라 알 수 있고, 뿐만 아니라 통계치의 훈련 기간에 따라 다양한 예측치를 발현하기 때문에, 특정 위협의 적절한 훈련 기간을 선택 가능하게 하며, 이를 바탕으로 각각의 위협의 성격에 맞는 훈련기간을 제공할 수 있다.
- <47> 이를 위하여, 본 발명은 위협원의 징후 정보간의 전이 행렬, 위협원과 위협원의 징후 정보간의 혼합행렬을 구성하여야 하는데, 이를 위해서는 무엇보다도 훈련기간이 중요시된다.
- <48> 따라서, 사이버 위협의 추이를 예측하기 위해 관찰해야 하는 위협원의 징후 정보에 대한 통계 처리 과정과 실제로 추이를 나타내는 위협원에 대한 통계 처리 단계, 위협원의 징후 정보와 위협원 간의 상관 관계를 규명하는 단계, 최종적으로 실제로 기간별 위협의 추이를 예측하는 단계를 포함할 수 있다.
- <49> 도 2는 본 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석 장치를 도시한 것이다.
- <50> 도 2를 참조하면, 본 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석 장치는 위협원 선정부(210), 관련 징후 저장부(220), 행렬 사상부(230), 패턴 분석부(240), 확률 분석부(250) 및 악성 코드 추이 예측부(260)로 이루어질 수 있다.
- <51> 우선, 위협원 선정부(210)는 네트워크 기반의 사이버 위협을 발생시키는 위협원의 위협 통계 데이터를 기반으로 악성 코드를 발현하는 위협원을 선정한다.
- <52> 이는 실제로 분석하고자 하는 악성 코드를 포함하는 통계 자료의 집합을 기반으로 위협원을 선정하는 것으로, 선정된 위협원의 위협 통계 데이터를 기반으로 확률 분포로 출력할 수 있다.
- <53> 더욱 상세하게는, 상기 위협원 선정부(210)는 은닉열 모듈(미도시)을 포함할 수 있으며, 상기 은닉열 모듈은 상기 위협원의 특정 기간 동안의 발생 빈도 정보를 수집하여 상기 위협원 발생의 확률 분포를 출력한다.
- <54> 그리고, 관련 징후 저장부(220)는 네트워크 기반의 사이버 위협을 발생시키는 위협원에 따라 발생된 위협원의 징후 정보를 저장한다.
- <55> 여기서, 상기 관련 징후 저장부(220)는 상기 위협원의 징후 정보를 수집하여 상태 전이 확률을 생성하는 관찰열 모듈(미도시)을 포함할 수 있으며, 상기 관찰열 모듈에서 생성되는 상태 전이 확률은 각각의 시구간에 따른 위협원 정보의 조건부 확률값으로 연산될 수 있다.
- <56> 또한, 행렬 사상부(230)는 상기 위협원 선정부(210)에서 선정된 위협원과 상기 관련 징후 저장부(220)에 저장된

위협원의 징후 정보를 행렬로 사상하여 혼합 행렬을 생성한다.

- <57> 여기서, 혼합 행렬은 상기 위협원과 위협원의 징후 정보의 상호 상관도 분석을 수행하기 위해 생성되는 행렬이다.
- <58> 상기 행렬 사상부(230)에서 혼합 행렬이 생성되면, 패턴 분석부(240)에서 상기 행렬 사상부(230)에서 생성된 혼합 행렬을 이용하여 상기 위협원과 상기 위협원의 징후 정보의 상관 분석을 수행하고, 상기 수행된 상관 분석에 따라 위협원의 발현 추이 패턴을 나타내는 파라미터(parameter)를 생성한다.
- <59> 상기와 같이 위협원의 발현 추이 패턴을 나타내는 파라미터를 생성하면, 생성된 파라미터를 기반으로 현재의 위협원의 발현을 예측할 수 있게 된다.
- <60> 상기 패턴 분석부(240)에서 위협원의 징후 정보의 발현 추이 패턴을 나타내는 파라미터를 생성하면, 확률 분석부(250)에서 상기 위협원의 통계 데이터를 복수 개의 시구간으로 분할하고, 예측하고자 하는 시구간에서 위협원의 악성 코드 발현 확률을 산정할 수 있다.
- <61> 상세하게는, 상기 확률 분석부(250)는 상기 위협원의 악성 코드 및 상기 생성된 파라미터를 이용하여 초기 확률 벡터를 생성하고, 상기 생성된 초기 확률 벡터, 상기 분할된 복수 개의 시구간의 징후 정보 간의 상관 분석의 결과를 행렬로 사상한 징후 정보 상관 행렬 및 상기 생성된 혼합 행렬을 이용하여 예측하고자 하는 시구간에서 상기 위협원의 악성 코드 발현 확률을 예측 산정하게 된다.
- <62> 한편, 복수 개의 시구간으로 분할된 상기 위협원의 위협원 정보는 위협원의 발생 빈도를 기반으로 하여 은닉 마코프 모델에 따른 퍼지 로직 추정 모델에 의해 위협원의 발생 빈도로 설정될 수 있다.
- <63> 은닉 마코프 모델에 따른 퍼지 로직 추정 모델은 하기의 표 1로 정리할 수 있다.

표 1

N_{max} : 특정 시구간에서 최대 발생한 위협원의 발생 빈도수
N_{min} : 특정 시구간에서 최소 발생한 위협원의 발생 빈도수
$\alpha = \text{Log}_{10} N_{max}, \beta = \text{Log}_{10} N_{min},$
$\gamma: \varepsilon - 1, (\varepsilon \text{ 은 가장 최근에 발생한 위협원의 발생 빈도수}),$
$\theta = \frac{ \alpha - \beta }{\gamma}, \beta = H_1, H_1 + \theta = H_2, H_2 + \theta = H_3,$
$H_{n-1} = \theta, H_n = \alpha$
(H_n 은 특정 시구간에서의 위협원의 발생 빈도수)

<64>

<65> 여기서, 복수 개로 분할된 위협원의 발생 빈도 H_n 은,

<66> N_{max} 가 특정 시구간에서 최대 발생한 위협원의 발생 빈도를 의미하고, N_{min} 이 상기 특정 시구간에서 최소 발생한 위협원의 발생 빈도를 의미하며, $\alpha = \text{Log}_{10} N_{max}$ 에 따라 상기 패턴 분석부의 파라미터 α 가 생성되며, $\beta = \text{Log}_{10} N_{min}$ 에 따라 상기 패턴 분석부의 파라미터 β 가 생성되며, 상기 패턴 분석부의 파라미터 ε 가 가장 최근에 발생한 위협원의 발생 빈도를 의미하면, 상기 패턴 분석부의 파라미터 γ 는 $\varepsilon - 1$ 로 정의되며, 상기 패턴 분석부의 파라미터 θ 는 $\theta = \frac{\alpha - \beta}{\gamma}$ 로 정의되고, H_n 을 연속된 각각의 시구간의 위협원의 발생 빈도를 의미할 때, 하기의 수학적 식 1에 의해 연산될 수 있다.

수학식 1

$$\beta = H_1,$$

$$H_1 + \theta = H_2,$$

$$H_2 + \theta = H_3,$$

.....

$$H_{n-1} = \theta,$$

$$H_n = a$$

<67>

<68>

도 3은 멀티 포트 스캐닝 액티베이션을 통하여 사이버 위협을 발생시키는 일례를 도시한 것이다.

<69>

봇(Bots)은 좀비(zombie) 시스템의 설립을 통하여 대규모의 호스트로 부터 특정 이득을 얻기 위하여 멀티 포트 스캐닝을 시도한 후에 서비스 거부(Denial of Service:DoS)를 발생시킨다.

<70>

도 3을 참조하면, 아고봇(Agobot)과 마이봇(Mybot)에 의해 수행된 멀티 포트 스캐닝의 일례를 도시하고 있다.

<71>

이와 같은 봇(Bots)의 발생 경향을 본 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석 장치에 의하여 적 절한 시기에 미리 분석하여 관리자/클라이언트 서버로 구성된 중요한 정보를 포함하고 있는 보안 정보를 효율적 으로 유지하고, 보안을 수행할 수 있게 된다.

<72>

도 4는 본 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석의 다이어그램을 도시한 것이다.

<73>

도 4를 참조하면, 은닉열 모듈(410)에서는 과거의 위협 통계 데이터를 미리 설정한 시구간으로 나누어, 각 시구 간별 위협원의 발생 빈도를 수집하여 상기 위협원 발생의 확률 분포로 출력한 수치를 나타내고 있다.

<74>

여기서, 은닉열 모듈(410)에서 출력하는 위협원 발생의 확률 분포는 $b_i(x)$ 로 표현될 수 있으며, 이는 위협원 x 의 특정 시구간 i에 해당되는 발현횟수로 나타낼 수 있다. 특히 특정 시구간에서의 위협원의 상태를 H_n 으로 정 의한다.

<75>

그리고, 관찰열 모듈(420)은 은닉열 모듈(410)의 위협원의 징후 통계 정보를 수집하여 상태 전이 확률(state transition probability)(a_{ij})을 구성한 것으로 $a_{ij}=p(H_j|H_i)$ 로 표현할 수 있다.

<76>

이때, a_{ij} 는 이전의 시구간에서의 위협원의 발생 빈도(H_j)가 연속된 다음 시구간에서의 위협원의 발생 빈도 (H_i)에 미칠 확률로서 특정 기간 동안 순차적으로 누적되어 최종적으로 상태 전이 확률로 표현하는 것이다.

<77>

이를 기반으로 악성 코드 추이 분석 모듈(430)에서 위협원의 징후 정보와 위협원 간의 상관관계를 규명하여 기 간별 위협의 추이를 예측하게 된다.

<78>

도 5a 내지 도 5d는 특정 시구간에서의 위협원의 징후 정보에 따라 위협원의 징후 정보의 발현 추이 패턴을 도 시한 것이다.

<79>

예를 들어, 위협원(135 포트와 445 포트)의 징후 정보를 포트 스캐닝 횟수로 산정하고, 포트 스캐닝된 상기 위 험원의 징후 정보의 상관 관계를 분석하여 파라미터 값을 생성하기 위하여 x,y를 135 포트와 445 포트의 문자열 값으로 사용한다.

<80>

그러면, 두 포트 간의 상관분석은 (x,y)에 대하여 n개의 측정값 $(x_1,y_1), (x_2,y_2), \dots, (x_n,y_n)$ 이 주어졌을 때, 하기의 수학식 2에 의하여 x,y의 상관계수를 구한다.

수학식 2

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}$$

<81>

<82> 여기서, x_i 와 y_i 는 두 포트의 특정 시구간 i 에서의 징후 정보의 측정값을 나타내고, $\bar{x}$ 와 $\bar{y}$ 는 x_i 와 y_i 의 평균, σ_x 와 σ_y 는 x_i 와 y_i 의 표준 편차를 의미 한다.

<83> 이와 같이, 상기와 같은 정의 및 조건에 의하여 포트 스캐닝 데이터를 분류한다.

<84> 하기의 도 5a 내지 도 5d에 도시된 $N_{PS}(t)$ 는 t 시점에서의 포트 스캐닝 횟수를 의미하며, S_n 은 상관 분석의 결과 해당하는 구간의 상태 정보를 정의한 것이고, O_n 은 징후 정보를 도시한 것으로 현재의 징후 정보 상태와 이전 징후 정보 상태를 표현하기 위하여 정의한 것이다.

<85> 우선 도 5a는 135 포트와 445 포트의 징후 정보의 추이가 동시에 증가하는 징후 정보 증가 패턴을 도시한 것이다.

<86> 이는 상기 수학식 2에 의해 연산된 상관계수 r 이 하기의 수학식 3에 따른 조건을 만족하는 경우이다.

수학식 3

$$0.5 < \text{상관계수}(r) \leq 1,$$

$$135\text{포트} : N_{PS}(t_{n-2}) < N_{PS}(t_{n-1}) < N_{PS}(t_n),$$

$$445\text{포트} : N_{PS}(t_{n-2}) < N_{PS}(t_{n-1}) > N_{PS}(t_n)$$

<87>

<88> 그리고, 도 5b는 135 포트와 445 포트의 징후 정보의 추이가 동시에 감소하는 징후 정보 감소 패턴을 도시한 것이다.

<89> 이는 상기 수학식 2에 의해 연산된 상관계수 r 이 하기의 수학식 4에 따른 조건을 만족하는 경우이다.

수학식 4

$$0.5 < \text{상관계수}(r) \leq 1,$$

$$135\text{포트} : N_{PS}(t_n) < N_{PS}(t_{n-1}) < N_{PS}(t_{n-2}),$$

$$445\text{포트} : N_{PS}(t_n) < N_{PS}(t_{n-1}) < N_{PS}(t_{n-2})$$

<90>

<91> 그리고, 도 5c는 135 포트와 445 포트의 징후 정보의 추이가 상호 교차하는 징후 정보 교차 패턴을 도시한 것이다.

<92> 이는 상기 수학식 2에 의해 연산된 상관계수 r 이 하기의 수학식 5에 따른 조건을 만족하는 경우이다.

수학식 5

$$-1 \leq \text{상관계수}(r) < -0.5$$

<93>

<94> 그리고, 도 5d는 135 포트와 445 포트의 징후 정보의 추이가 동시에 평형을 이루는 징후 정보 평형 패턴을 도시한 것이다.

<95> 이는 상기 수학식 2에 의해 연산된 상관계수 r 이 하기의 수학식 6에 따른 조건을 만족하는 경우이다.

수학식 6

$$-0.5 \leq \text{상관계수}(r) \leq 0.5$$

도 6은 본 발명에 따른 위협원의 각각의 시구간 설정시 위협의 증가가 지수 증가를 따른다고 가정할 경우, 퍼지 로직 추정 방법을 사용한 것이다. 퍼지 로직 추정 방법은 일반적으로 특정 데이터의 구간 설정을 위해 사용되는 방법 중의 하나로써, 위협원의 범위를 일괄적으로 나누는 것이 아니라, 지수적 성질을 가미하고, 모호한 경계를 명확하게 나눌 수 있도록 경계선(610)을 설정하기 위해 사용되는 것으로 상기 표 1의 산출 근거를 따른다.

상기 표 1에 따라서, 궁극적으로 H_n 을 나눌 수 있는 경계를 설정할 수 있으므로, 그 나누는 범위에 따라서 위협원의 각각의 시구간별 정보가 다르게 해석되게 된다.

도 7은 본 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석 방법의 흐름도를 도시한 것이다.

도 7을 참조하면 우선, 네트워크 기반의 사이버 위협을 발생시키는 위협원의 위협 통계 데이터를 기반으로 악성 코드를 발현하는 위협원을 선정한다(710 과정).

이는 실제로 분석하고자 하는 악성 코드를 포함하는 통계 자료의 집합을 기반으로 위협원을 선정하는 것으로, 선정된 위협원의 위협 통계 데이터를 기반으로 확률 분포로 출력할 수 있다.

상세하게는, 상기 위협원의 특정 기간 동안의 발생 빈도 정보를 수집하여 상기 위협원 발생의 확률 분포를 출력할 수 있다.

그 다음, 상기 네트워크 기반의 사이버 위협을 발생시키는 위협원에 따라 발생된 위협원의 징후 정보를 저장한다(720 과정).

상세하게는, 상기 위협원의 징후 정보를 수집하여 상태 전이 확률을 생성하는 단계를 포함할 수 있으며, 여기서 상기 상태 전이 확률은 각각의 시구간에 따른 위협원 정보의 조건부 확률값으로 연산될 수 있다.

그 다음, 상기 선정된 위협원과 상기 저장된 상기 위협원의 징후 정보를 행렬로 사상하여 혼합 행렬을 생성한다(730 과정).

여기서, 혼합 행렬은 상기 위협원과 위협원의 징후 정보의 상호 상관도 분석을 수행하기 위해 생성되는 행렬이다.

상기 선정된 위협원과 위협원의 징후 정보의 상호 상관도 분석을 수행하기 위해 생성하는 혼합 행렬은 하기의 표 2와 같이 표시될 수 있다.

표 2

	위협원의 징후정보			
위협원	C_{11}	C_{12}	$\dots$	C_{1n}
	C_{21}	C_{22}	$\dots$	C_{2n}
	$\vdots$	$\vdots$	C_{ij}	$\dots$
	C_{m1}	C_{m2}	$\dots$	C_{mn}

이와 같이 생성되는 혼합 행렬은 위협원과 상기 위협원의 징후 정보간의 상관관계 분석 결과 생성되는 행렬로써, 상기 위협원의 사이버 기상 예측 결과를 산출하는데 중요한 역할을 하게 된다.

그 다음, 혼합 행렬이 생성되면, 상기 생성된 혼합 행렬을 이용하여 상기 위협원과 상기 위협원의 징후 정보의 상관 분석을 수행하고, 상기 수행된 상관 분석에 따라 위협원의 징후 정보의 발현 추이 패턴을 나타내는 파라미터(parameter)를 생성한다(740 과정).

상기와 같이 위협원의 발현 추이 패턴을 나타내는 파라미터를 생성하면, 생성된 파라미터를 기반으로 현재의 위협원의 발현을 예측할 수 있게 된다.

<112> 그 다음, 상기 위협원의 징후 정보의 발현 추이 패턴을 나타내는 파라미터를 생성하면, 상기 위협원의 통계 데이터를 복수 개의 시구간으로 분할하고, 예측하고자 하는 시구간에서 위협원의 악성 코드 발현 확률을 산정할 수 있다(750 과정).

<113> 상세하게는, 상기 위협원의 악성 코드 및 상기 생성된 파라미터를 이용하여 초기 확률 벡터를 생성하고, 상기 생성된 초기 확률 벡터, 상기 분할된 복수 개의 시구간의 징후 정보 간의 상관 분석의 결과를 행렬로 사상한 징후 정보 상관 행렬 및 상기 생성된 혼합 행렬을 이용하여 예측하고자 하는 시구간에서 상기 위협원의 악성 코드 발현 확률을 예측 산정하게 된다.

<114> 이는, 위협원의 악성 코드 및 상기 생성된 파라미터를 이용하여 초기 확률 벡터를 생성하기 위해서, 구하고자 하는 위협원의 최근 발생 데이터를 조사할 수 있다.

<115> 초기 확률을 구하기 위한 최근의 위협 발생 데이터는 사용자의 설정에 따라 다양한 시간 간격으로 설정할 수 있음은 물론이다.

<116> 여기서, 최근의 위협원의 위협 발생 데이터를 분석하여 하기의 수학식 7과 같이 연산하여 수학식 8을 만족하는 각 위협원의 발생 빈도 및 초기 확률 벡터를 생성한다.

수학식 7

$$I(H_1, H_2, \dots, H_n) = I\left(\frac{\alpha}{F}, \frac{\beta}{F}, \dots, \frac{\delta}{F}\right)$$

<117>

<118> 단, I는 초기 확률 벡터를 나타내며, $\alpha, \beta, \gamma, \delta$ 는 각 시구간별 위협원의 상태의 악성코드 발현 빈도 ($H_1, H_2, H_3, \dots, H_n$)이고, F는 하기의 수학식 7에 의해 연산되는 악성코드 발현 횟수의 총합이다.

수학식 8

$$F = \sum_{i=1}^n f_i = \alpha + \beta + \dots + \delta$$

<119>

<120> 이를 기반으로 하여, 초기 확률 벡터 I와 분할된 복수 개의 시구간의 징후 정보 간의 상관 분석의 결과를 행렬로 사상한 징후 정보 상관 행렬(O_{ij}) 및 상기 생성된 혼합 행렬 (C_{ij})의 곱으로 예측하고자 하는 위협원의 악성 코드 발현 확률 P를 산정한다.

<121> 이는 하기의 수학식 9와 같이 표현할 수 있다.

수학식 9

$$I(H_1, H_2, \dots, H_n) \times \begin{bmatrix} O_{11} & O_{12} & \dots & O_{1n} \\ O_{21} & O_{22} & \dots & O_{2n} \\ \dots & \dots & O_{ij} & \dots \\ O_{m1} & O_{m2} & \dots & O_{mn} \end{bmatrix} \\ \times \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \dots & \dots & C_{ij} & \dots \\ C_{m1} & C_{m2} & \dots & C_{mn} \end{bmatrix} \\ = P(H_1, H_2, \dots, H_n)$$

<122>

<123> 한편, 복수 개의 시구간으로 분할된 상기 위협원의 위협원 정보는 위협원의상기 위협원의 위협원 정보는 위협원의 발생 빈도를 기반으로 하여 은닉 마코프 모델에 따른 퍼지 로직 추정 모델에 의해 위협원의 발생 빈도로 설정될 수 있다.

<124> 한편, 은닉 마코프 모델에 따른 퍼지 로직 추정 모델은 상기의 표 1로 정리할 수 있다.

<125> 여기서, 복수 개로 분할된 위협원의 발생 빈도는 $N_{\max}$ 가 특정 시구간에서 최대로 발생한 위협원의 발생 빈도를 의미하고, $N_{\min}$ 이 상기 특정 시구간에서 최소로 발생한 위협원의 발생 빈도를 의미하며, $\alpha = \text{Log}_{10}N_{\max}$ 에 따라 상기 패턴 분석부의 파라미터 α 가 생성되며, $\beta = \text{Log}_{10}N_{\min}$ 에 따라 상기 패턴 분석부의 파라미터 β 가 생성되며, 상기 패턴 분석부의 파라미터 ε 가 가장 최근에 발생한 위협원의 발생 빈도를 의미하면, 상기 패턴 분석부의 파라미터 γ 는 $\varepsilon - 1$ 로 정의되며, 상기 패턴 분석부의 파라미터 θ 는 $\theta = \frac{\alpha - \beta}{\gamma}$ 로 정의되고, H_n 을 연속된 각각의 시구간의 위협원의 발생 빈도를 의미할 때, 상기의 수학식 1에 의해 연산될 수 있다.

<126> 마지막으로, 상기 산정된 악성 코드 발현 확률에 따라 사이버 기상 예측을 수행하게 된다(760 과정).

<127> 도 8은 실제 발생한 사이버 위협과 본 발명과 종래 기술의 사이버 기상 예측효율의 비교도이다.

<128> 도 8을 참조하면, 위협원인 135 포트와 445 포트와 관련이 있는 위협원(AGOBOT와 MYBOT)의 상관관계를 규명한 것으로 실제 발생한 사이버 위협의 횟수(800)와 본원 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석 방법(810)에 의해 예측된 사이버 위협의 횟수가 가장 근사하고, 사이버 위협의 흐름이 가장 유사함을 알 수 있다.

<129> 그리고, 시계열 분석을 통한 예측방법론인 이동평균법(MA)(820)와 지수이동평균법(EMA)(830)과도 비교해 보면 실제 발생한 위협 횟수를 나타내는 데이터(800)와 본원 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석 방법(810)이 가장 유사값을 나타냄을 알 수 있다.

<130> 도 8은 본원 발명의 9개월간의 사이버 기상 예측을 위한 과거 통계 데이터를 이용한 것으로 예측을 위한 훈련기간을 더 증가시킬 경우, 더욱 우수한 결과를 얻어낼 수 있다.

<131> 또한 본 발명에 따른 상기의 각 단계는 일반적인 프로그래밍 방법을 이용하여 소프트웨어적으로 또는 하드웨어적으로 다양하게 구현할 수 있다는 것은 이 기술분야에 통상의 기술을 가진 자라면 용이하게 할 수 있는 것이다.

<132> 그리고, 본 발명의 일부 단계들은, 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 모두 포함한다. 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, CD-RW, 자기 테이프, 플로피디스크, HDD, 광 디스크, 광자기 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어, 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드로 저장되고 실행될 수 있다.

<133> 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명에 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로, 본 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 상기의 설명에 포함된 예들은 본 발명에 대한 이해를 위해 도입된 것이며, 이 예들은 본 발명의 사상과 범위를 한정하지 아니한다. 상기의 예들 외에도 본 발명에 따른 다양한 실시 태양이 가능하다는 것은, 본 발명이 속한 기술 분야에서 통상의 지식을 가진 자에게는 자명할 것이다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면의 간단한 설명

<134> 도 1은 본 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석의 개괄도이다.

<135> 도 2는 본 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석 장치를 도시한 것이다.

<136> 도 3은 멀티 포트 스캐닝 액티베이션을 통하여 사이버 위협을 발생시키는 일례를 도시한 것이다.

<137> 도 4는 본 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석의 다이어그램을 도시한 것이다.

<138> 도 5a 내지 도 5d는 특정 시구간에서의 위협원의 징후 정보에 따라 위협원의 징후 정보의 발현 추이 패턴을 도시한 것이다.

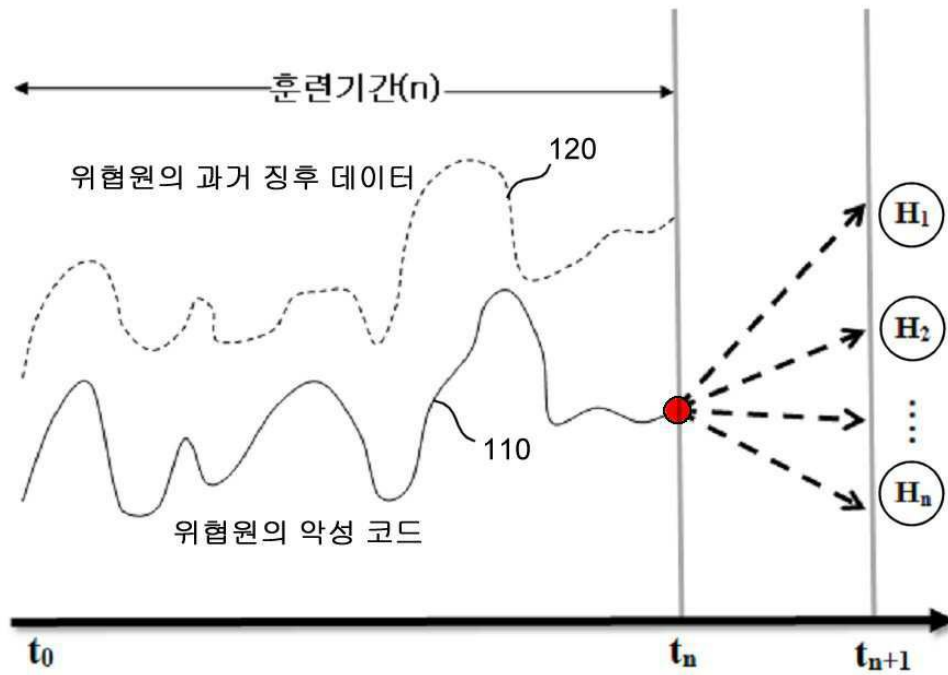
<139> 도 6은 본 발명에 따른 위협원의 각각의 시구간 설정시 위협의 증가가 지수 증가를 따를 경우의 퍼지 로직 추정 방법을 사용한 것이다.

<140> 도 7은 본 발명에 따른 사이버 기상 예측을 위한 악성 코드 분석 방법의 흐름도를 도시한 것이다.

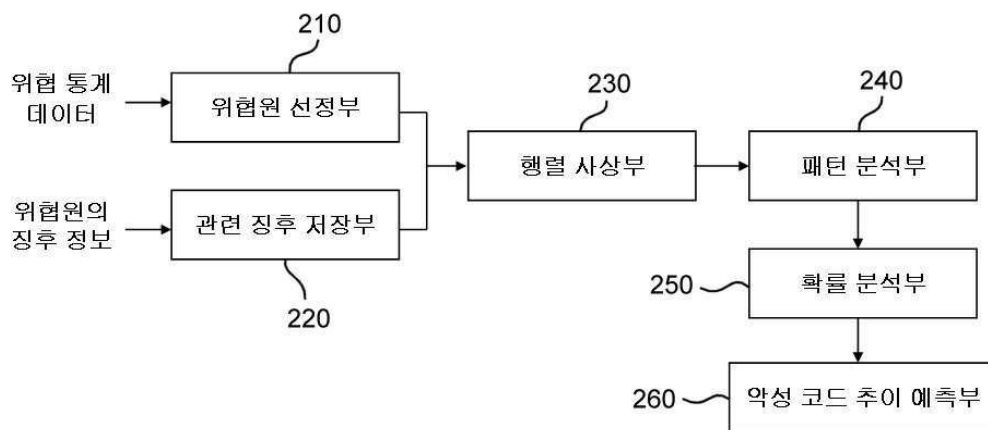
<141> 도 8은 실제 발생한 사이버 위협과 본 발명과 종래 기술의 사이버 기상 예측효율의 비교 그래프이다.

도면

도면1



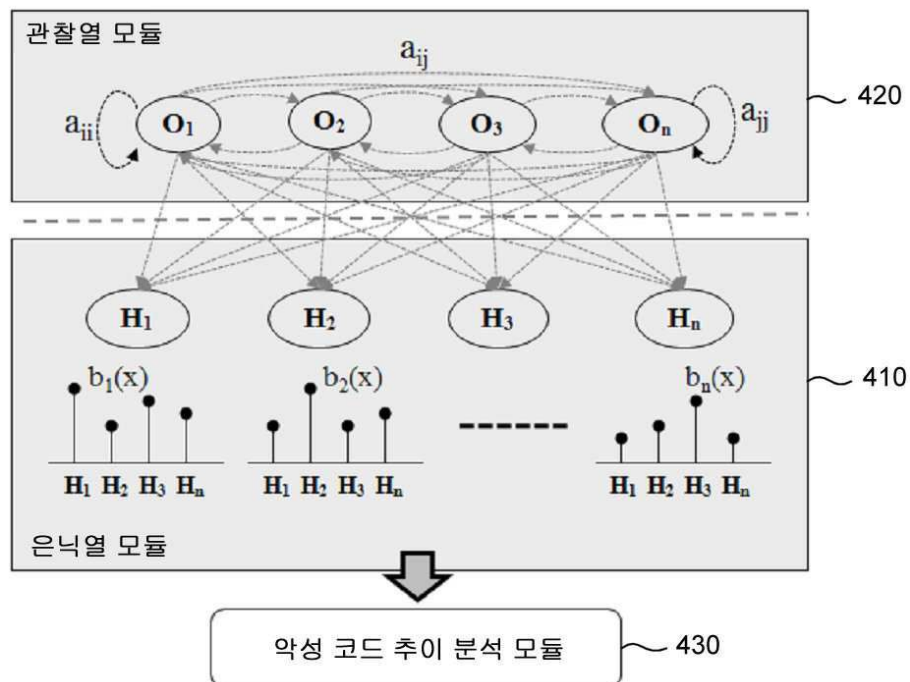
도면2



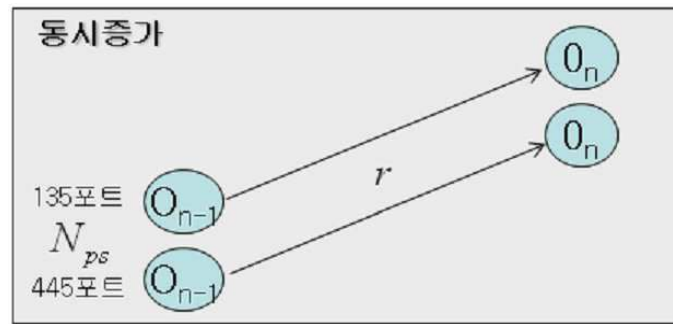
도면3

Proto	Local Address	Foreign Address	State
TCP	192.168.0.2:1026	192.168.0.29:135	SYN_SENT
TCP	192.168.0.2:1028	192.168.0.32:445	SYN_SENT
TCP	192.168.0.2:1030	192.168.0.37:135	SYN_SENT
TCP	192.168.0.2:1031	192.168.0.29:135	SYN_SENT
TCP	192.168.0.2:1077	192.168.0.143:135	SYN_SENT
TCP	192.168.0.2:1078	192.168.0.143:135	SYN_SENT
TCP	192.168.0.2:1080	192.168.0.28:445	SYN_SENT
TCP	192.168.0.2:1081	192.168.0.143:135	SYN_SENT
TCP	192.168.0.2:1139	192.168.0.165:445	SYN_SENT
TCP	192.168.0.2:1261	192.168.0.47:445	SYN_SENT
TCP	192.168.0.2:1267	192.168.0.28:445	SYN_SENT
TCP	192.168.0.2:1269	192.168.0.165:445	SYN_SENT
TCP	192.168.0.2:1275	192.168.0.37:135	SYN_SENT
TCP	192.168.0.2:1283	192.168.0.27:135	SYN_SENT
TCP	192.168.0.2:1322	192.168.0.143:445	SYN_SENT
TCP	192.168.0.2:1333	192.168.0.196:445	SYN_SENT
TCP	192.168.0.2:1393	192.168.0.177:135	SYN_SENT
TCP	192.168.0.2:1543	192.168.0.7:135	SYN_SENT
TCP	192.168.0.2:1553	192.168.0.196:135	SYN_SENT
TCP	192.168.0.2:1711	192.168.0.7:135	SYN_SENT

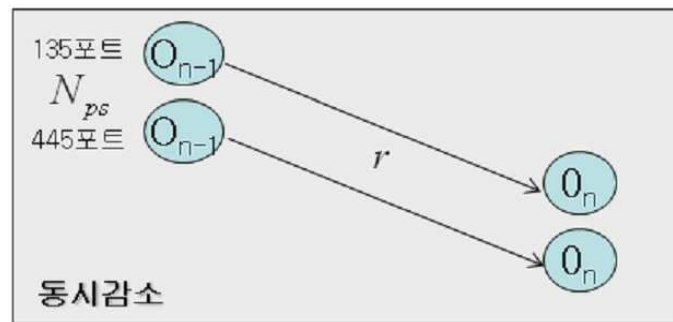
도면4



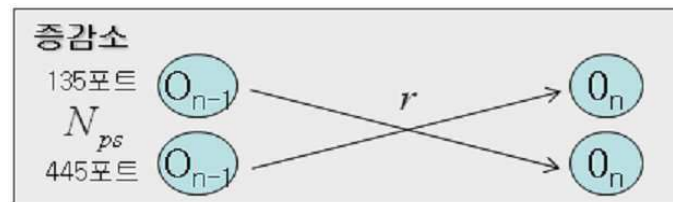
도면5a



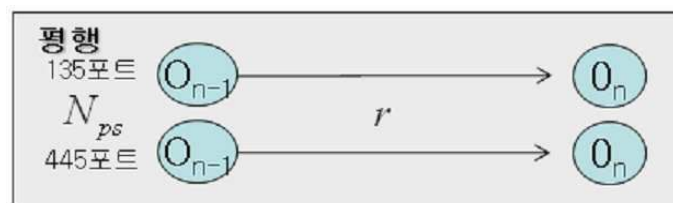
도면5b



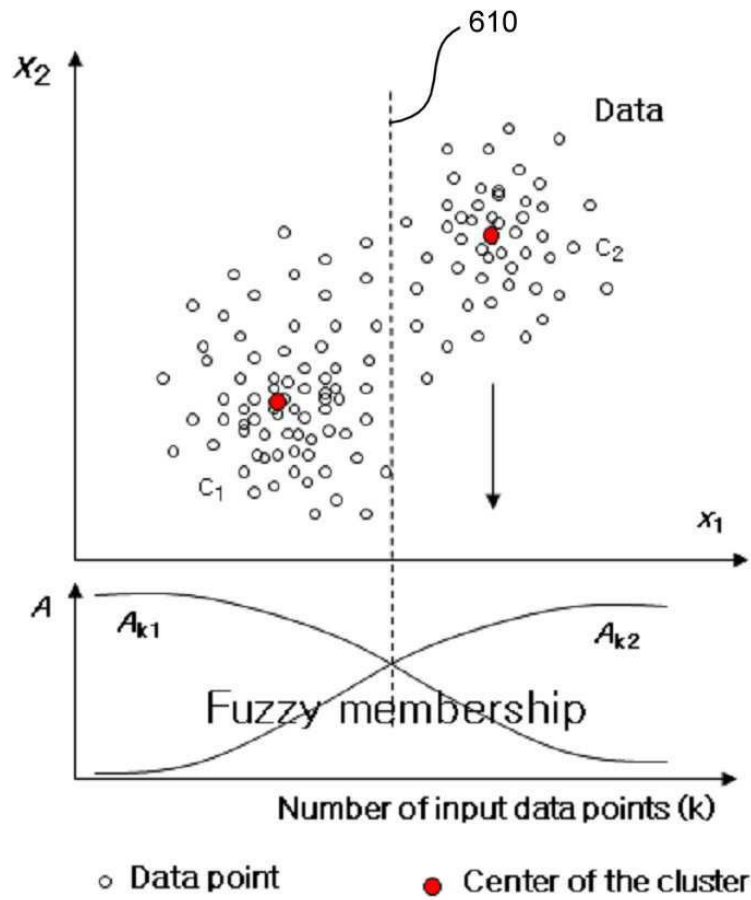
도면5c



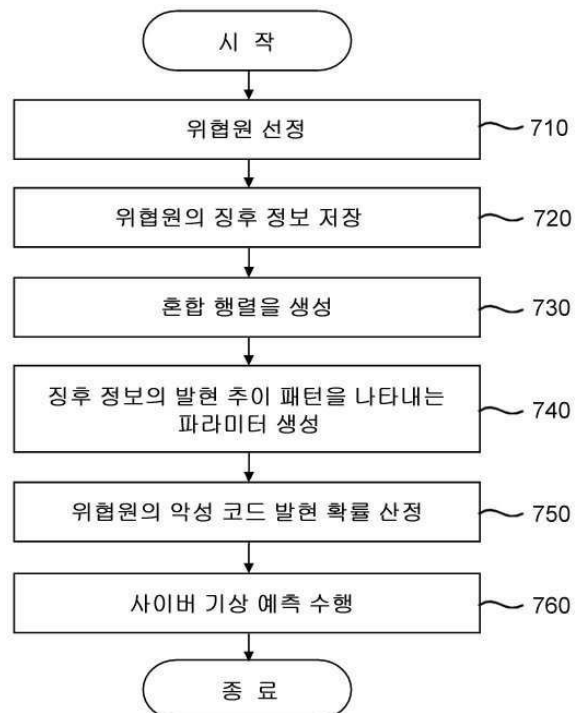
도면5d



도면6



도면7



도면8

