(19)대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) 。Int. Cl.
G06F 17/00 (2006.01)

(45) 공고일자 2007년06월19일
(11) 등록번호 10-0729151
(24) 등록일자 2007년06월11일

(21) 출원번호 10-2006-0066832
(22) 출원일자 2006년07월18일
심사청구일자 2006년07월18일

(65) 공개번호
(43) 공개일자

(73) 특허권자 고려대학교 산학협력단
서울 성북구 안암동5가1 고려대학교 내

(72) 발명자 김영갑
서울 동대문구 제기2동 174-32 16/8

백두권
서울 용산구 이촌1동 301-162 현대아파트 31동 306호

인호
서울 중구 필동2가 84-19 우림빌라 라-301호

(74) 대리인 현종철

(56) 선행기술조사문헌
KR1020040011858 A
KR1020040011866 A
KR1020040012285 A

심사관 : 김수섭

전체 청구항 수 : 총 8 항

(54) 마코프 프로세스 기반의 피해 산정 방법, 그 기록 매체 및그 장치

(57) 요약

마코프 프로세스 기반의 피해 산정 방법, 그 기록 매체 및 그 장치가 개시된다.

본 발명은 장기간의 위협 발생 데이터를 수집 분석하여 주요 정보 통신 기반 시설에 발생 가능한 위협 상태 집합을 정의하는 단계, 상기 위협 발생 데이터의 각 위협별 발생 빈도수와 상기 위협 상태 집합 사이의 매핑을 이용하여 위협 상태 전이 행렬을 산출하는 단계, 상기 장기간의 위협 발생 데이터보다 최근의 위협 발생 데이터를 이용하여 상기 위협 상태 집합의 초기 발생 확률을 산출하는 단계 및 상기 위협 상태 전이 행렬과 상기 초기 발생 확률을 이용하여 상기 주요 정보 통신 기반 시설에서 발생 가능한 위협들의 발생 빈도 예측 정보를 생성하는 단계를 포함한다.

본 발명에 의하면, 전체 시스템에서 위협들의 발생 확률 및 발생 빈도를 마코프 프로세스를 이용하여 설명하고 하나의 특정 위협이 아니라 전체 시스템이 가지고 있는 여러 가지 위협들에 대해 적용할 수 있으며, 위협이 발생 하였을 때 생성되는 위협 영역에 대하여 위협들 사이의 관계를 파악할 수 있고, 공분산과 상관계수를 통해 위협들 사이의 상관관계 정도를 파악할 수 있는 효과가 있다.

대표도

도 3

특허청구의 범위

청구항 1.

장기간의 위협 발생 데이터를 수집 분석하여 주요 정보 통신 기반 시설에 발생 가능한 위협 상태 집합을 정의하는 단계;

상기 위협 발생 데이터의 각 위협별 발생 빈도수와 상기 위협 상태 집합 사이의 매핑을 이용하여 위협 상태 전이 행렬을 산출하는 단계;

상기 장기간의 위협 발생 데이터보다 최근의 위협 발생 데이터를 이용하여 상기 위협 상태 집합의 초기 발생 확률을 산출하는 단계; 및

상기 위협 상태 전이 행렬과 상기 초기 발생 확률을 이용하여 상기 주요 정보 통신 기반 시설에서 발생 가능한 위협들의 발생 빈도 예측 정보를 생성하는 단계를 포함하는 마코프 프로세스 기반의 피해 산정 방법.

청구항 2.

제 1 항에 있어서,

상기 위협 상태 집합을 정의하는 단계는

과거 장기간에 걸쳐 기록된 위협 발생 데이터를 수집하는 단계;

상기 위협 발생 데이터로부터 위협을 유형별로 분류하고 각 위협의 주기를 산출하고, 상기 각 위협의 주기를 이용하여 위협 순위를 결정하여 상기 위협 순위에 따른 대표 위협들을 결정하는 단계; 및

상기 대표 위협들의 발생 빈도에 대한 범위값들로 구성된 위협 상태 집합을 생성하는 단계를 포함하는 것을 특징으로 하는 마코프 프로세스 기반의 피해 산정 방법.

청구항 3.

제 2 항에 있어서,

상기 위협 상태 집합은

하나의 위협이 독립적으로 발생하는 경우 상기 하나의 위협이 가질 수 있는 발생 빈도에 대한 범위값들로 구성되고, 복수의 위협들이 상호 연관되어 발생하는 경우 상기 복수의 위협들의 발생 빈도에 대한 범위값들을 조합한 형태로 구성되는 것을 특징으로 하는 마코프 프로세스 기반의 피해 산정 방법.

청구항 4.

제 1 항에 있어서,

상기 위협 상태 전이 행렬을 산출하는 단계는

상기 위협 발생 데이터의 각 위협별 발생 빈도수와 상기 위협 상태 집합을 서로 매핑하여 위협 상태들을 열거하는 단계;

상기 열거된 위협 상태들 중 하나의 위협 상태에서 다른 위협 상태로 전이하는 횟수를 산출하고, 상기 전이하는 횟수를 이용하여 위협 상태 전이 행렬을 산출하는 단계를 포함하는 것을 특징으로 하는 마코프 프로세스 기반의 피해 산정 방법.

청구항 5.

제 1 항에 있어서,

상기 초기 발생 확률을 산출하는 단계는

상기 장기간의 위협 발생 데이터보다 최근의 위협 발생 데이터로부터 각 위협 상태별 최근 발생 빈도를 산출하는 단계; 및

상기 각 위협 상태별 최근 발생 빈도로부터 초기 발생 확률의 총 합이 1이 되도록하는 상기 위협 상태 집합의 초기 발생 확률을 산출하는 단계를 포함하는 것을 특징으로 하는 마코프 프로세스 기반의 피해 산정 방법.

청구항 6.

제 1 항에 있어서,

상기 위협들의 발생 빈도 예측 정보를 생성하는 단계는

상기 위협 상태 전이 행렬과 상기 초기 발생 확률을 곱하여 피해 파급 확률을 연산하는 단계; 및

상기 피해 파급 확률 및 상기 각 위협 상태의 평균값을 곱하여 발생 가능한 위협들의 발생 빈도 예측 정보를 생성하는 단계를 포함하는 것을 특징으로 하는 마코프 프로세스 기반의 피해 산정 방법.

청구항 7.

제 1 항 내지 제 6 항 중 어느 한 항의 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

청구항 8.

장기간의 위협 발생 데이터를 수집 분석하여 주요 정보 통신 기반 시설에 발생 가능한 위협 상태 집합을 정의하는 위협 정의부;

상기 위협 발생 데이터의 각 위협별 발생 빈도수와 상기 위협 상태 집합 사이의 매핑을 이용하여 위협 상태 전이 행렬을 산출하는 행렬 산출부;

상기 장기간의 위협 발생 데이터보다 최근의 위협 발생 데이터를 이용하여 상기 위협 상태 집합의 초기 발생 확률을 산출하는 초기값 산출부; 및

상기 위협 상태 전이 행렬과 상기 초기 발생 확률을 이용하여 상기 주요 정보 통신 기반 시설에서 발생 가능한 위협들의 발생 빈도 예측 정보를 생성하는 위협 예측부를 포함하는 마코프 프로세스 기반의 피해 산정 장치.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 네트워크 보안에 관한 것으로, 특히, 마코프 프로세스 기반의 피해 산정 방법, 그 기록 매체 및 그 장치에 관한 것이다.

최근 급속한 인터넷 기술의 발전으로 기업이나 기관에서의 업무 처리는 인터넷 기반 기술에 의존하고 있다. 또한 주요정보통신 시설의 네트워크 의존도와 결합도가 증가함에 따라 시스템내의 취약성을 대상으로 침해 행위와 같은 사이버 보안 사고의 수가 크게 증가하고 있다. 이에 따라 개인정보는 물론 컴퓨터 자원들의 침해와 관련된 위험 분석(risk analysis) 및 피해 파급(damage propagation)에 관한 연구가 요구된다.

정확한 위험 분석은 적절한 보안 대응책 선정을 가능하게 하고 결과적으로 위험 발생 가능성을 크게 감소시켜 차후에 실제적으로 발생될 수 있는 보안 사고의 피해규모를 크게 감소시킨다. 이와 같은 위험 분석은 보안 사고를 사전에 예방하기 위해 의미가 크고, 실제적으로 많은 연구가 진행되어 왔다.

전파 모델 (Epidemic Model)은 특정 모집단 내에서 발병원이 발생 한다 가정하고 시간변화에 따라 감염된 개체와 감염에 노출된 개체, 치유되는 개체들 간의 개체수 관계를 설명하기 위한 모델이다. 이를 통하여 실제 네트워크상의 웜이 전파되는 확산력을 설명할 수 있다.

기본적으로 네트워크상에서 발생 가능한 전파 특성을 지니는 위협들의 피해 확산력을 설명하기 위하여 생물학적 고전모델인, 고전단순전파모델 (Classical Simple Epidemic Model)을 사용하고 있다. 그 중에서 가장 범용적으로 사용되는 모델이 바로 Kermack-Mckendrick 모델 (또는 전통적 SIR 전파모델)이다. 이 모델에서는 웜에 의해 변화하는 호스트의 상태를 S (Susceptible), I (Infectious), R(Removed) 세 가지로 나누고 있어 SIR 모델이라고도 부른다. 웜에 노출된 호스트는 최초에 웜에 취약한 상태(S)를 띠고 웜에 걸리게 되면 감염된 상태(I)로 변한다. 그리고 마지막으로 감염된 호스트가 웜을 치료 하거나 웜의 기능을 완전 상실케 되는 상태(R)로 전환된다.

SIRS 모델은 종래의 SIR모델을 개량한 전파 모델이다. 즉, 일반적인 경우는 시간에 따른 감염된 호스트 I(t), 감염에 노출된 호스트 S(t), 그리고 제거되는 호스트 R(t)를 생각하였는데, SIRS 모델에서는 S → I → R 단계 이후에 다시 해당 호스트가 감염될 수 있다는 사실을 가정하였다. 즉, 감염에 노출된 호스트에서 감염이 발생되면서 일정 시간 후 제거과정을 통해 완전 제거가 아닌, 다시 감염될 수 있는 호스트의 상태를 추가한 것이다.

마지막으로, 종래의 SIR, SIRS 모델에 포함되지 않았던 요소 2가지를 추가하여 TWO Factor Model을 제안하였다. 이는 코드레드 사고 이후 새롭게 제안된 모델이다. 이 모델은 인간의 대응책과 감소되는 감염율을 고려하였다. 웜의 관점에서, 인간의 대응책은 어떤 호스트들을 웜 확산 유통으로부터 제거한다. 제거되는 호스트들은 감염된 호스트 및 여전히 취약한 호스트들을 모두 포함한다.

이에 반해, 피해 파급에 대한 개념은 아직 정립되지 않았으며, 종래의 피해 파급 모델은 현재의 IT 환경에 대한 사고 유형들의 특성을 적절히 반영하기에 부적절하다.

따라서, 종래의 피해 산정 방법들은 바이러스(virus)나 웜(worm) 같은 특정 위협(threats)에 대한 피해 파급 모델을 제시하여 전체 시스템이 가지고 있는 다양한 위협에 대해 적용하기 어렵고, 위협이 발생 하였을 때 생성되는 위협 영역에 대하여 위협들 사이의 관계 또는 시간적인 요소에 의한 분석이 어려운 문제점이 있다.

발명이 이루고자 하는 기술적 과제

따라서, 본 발명이 이루고자 하는 첫번째 기술적 과제는 전체 시스템에서 위협들의 발생 확률 및 발생 빈도를 마코프 프로세스를 이용하여 설명하고 하나의 특정 위협이 아니라 전체 시스템이 가지고 있는 여러 가지 위협들에 대해 적용할 수 있으며, 공분산과 상관계수를 통해 위협들 사이의 상관관계 정도를 파악할 수 있는 마코프 프로세스 기반의 피해 산정 방법을 제공하는 데 있다.

본 발명이 이루고자 하는 두번째 기술적 과제는 상기의 마코프 프로세스 기반의 피해 산정 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공하는데 있다.

본 발명이 이루고자 하는 세번째 기술적 과제는 상기의 마코프 프로세스 기반의 피해 산정 방법이 적용된 마코프 프로세스 기반의 피해 산정 장치를 제공하는데 있다.

발명의 구성

상기의 첫번째 기술적 과제를 이루기 위하여, 본 발명은 장기간의 위협 발생 데이터를 수집 분석하여 주요 정보 통신 기반 시설에 발생 가능한 위협 상태 집합을 정의하는 단계, 상기 위협 발생 데이터의 각 위협별 발생 빈도수와 상기 위협 상태 집합 사이의 매핑을 이용하여 위협 상태 전이 행렬을 산출하는 단계, 상기 장기간의 위협 발생 데이터보다 최근의 위협 발생 데이터를 이용하여 상기 위협 상태 집합의 초기 발생 확률을 산출하는 단계 및 상기 위협 상태 전이 행렬과 상기 초기 발생 확률을 이용하여 상기 주요 정보 통신 기반 시설에서 발생 가능한 위협들의 발생 빈도 예측 정보를 생성하는 단계를 포함하는 마코프 프로세스 기반의 피해 산정 방법을 제공한다.

상기의 두번째 기술적 과제를 이루기 위하여, 본 발명은 상기의 마코프 프로세스 기반의 피해 산정 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다.

상기의 세번째 기술적 과제를 이루기 위하여, 본 발명은 장기간의 위협 발생 데이터를 수집 분석하여 주요 정보 통신 기반 시설에 발생 가능한 위협 상태 집합을 정의하는 위협 정의부, 상기 위협 발생 데이터의 각 위협별 발생 빈도수와 상기 위협 상태 집합 사이의 매핑을 이용하여 위협 상태 전이 행렬을 산출하는 행렬 산출부, 상기 장기간의 위협 발생 데이터보다 최근의 위협 발생 데이터를 이용하여 상기 위협 상태 집합의 초기 발생 확률을 산출하는 초기값 산출부 및 상기 위협 상태 전이 행렬과 상기 초기 발생 확률을 이용하여 상기 주요 정보 통신 기반 시설에서 발생 가능한 위협들의 발생 빈도 예측 정보를 생성하는 위협 예측부를 포함하는 마코프 프로세스 기반의 피해 산정 장치를 제공한다.

피해 파급 영역(damage propagation area)은 위협의 종류나 위협들 사이의 연관 관계에 따라 달라지므로 피해 파급 모델은 이러한 요구사항을 만족해야 한다.

따라서 본 발명은 과거의 위협 발생 데이터를 근거로 하여 정보 시스템이 가지고 있는 위협들 사이의 상호 관계 및 시간의 흐름에 따라 피해 파급을 종합적으로 예측 및 분석할 수 있는 마코프 프로세스(markov porcess) 기반의 확률적 피해 파급 모델을 제공한다.

본 발명에서 제공하는 피해 파급 모델을 통하여 주요 위협들에 대해 사전에 위협 발생 확률 및 위협 발생 빈도를 예측하고 피해 파급 정도에 따른 위협 대책 수립에 활용할 수 있다. 또한, 각 위협간의 상관관계를 분석하는데도 적용 가능하다.

먼저, 피해 파급 모델을 설계하기 위해 적용되는 마코프 프로세스에 대해 설명한다. 마코프 프로세스는 상태간 전이가 오로지 이전 n개의 상태에 의존하여 이루어지는 프로세스를 말한다. 이 때 이 모델을 n차원 모델이라 하는데 n은 다음 상태를 결정하는데 영향을 미치는 상태의 개수를 말한다. 과거의 상태를 기억하지 않는다는 점에서 비기억 프로세스(memoryless process)라고도 한다.

마코프 프로세스를 $X(t)$라 하면 임의의 시간 $t_1<t_2<\ldots<t_k<t_{k+1}$에 대해 $X(t)$가 이산값이면

$P[a<X(t_{k+1})=x_{k+1}|X(t_k)=x_k,\ldots,X(t_1)=x_1]$ $P[a<X(t_{k+1})\le b|X(t_k)=x_k,\cdots,X(t_1)=x$

$=P[X(t_{k+1})=x_{k+1}|X(t_k)=x_k]$이고 $X(t)$가 연속값이면 $=P[a<X(t_{k+1})\le b|X(t_k)=x_k]$ 로 마코프 성질이 기술된다.

위의 식에서 t_k는 현재, t_{k+1}은 미래, 그리고 $t_1, \ldots, t_{k-1}$은 과거의 시점이다. 마코프 프로세서의 값이 이산값이면 마코프 체인(markov chain)이라고 한다. 마코프 체인은 t가 이산적이냐 연속적이냐에 따라 이산시간 마코프 프로세스(discrete-time markov provess), 연속시간 마코프 프로세스(continuous-time markov chain)로 나뉜다.

마코프 프로세스는 다음의 세 가지로 설명될 수 있는 모든 시스템을 말한다

먼저, 상태 집합(set of state)은 프로세스로부터 모든 가능한 상태들의 집합이다. 다음, π 벡터(초기 확률)는 시스템의 초기화 확률 벡터이고, 상태 전이 행렬(state transition matrix)은 각 상태간 전이 확률을 나타낸다.

본 발명은 마코프 프로세스의 세 가지 구성요소를 정의함으로써 피해 파급 모델을 설계하고 적용한다.

마코프 프로세스 기반의 확률적 피해 파급 모델은, 독립 또는 단일 시스템의 피해로 인하여 그 시스템과 관련된 전체 시스템에 영향을 주는 정도를 확률적 접근방법으로 예측하는 것을 말한다. 전체 시스템이 가지고 있는 피해 파급 모형과 피해 파급 속도는 시간의 흐름, 또는 위협의 종류에 따라 달라질 수 있다. 마코프 프로세스에 기반한 피해 파급 모델은 전체 시스템에서 주요 위협들에 대한 위협 전이 확률 및 위협 발생확률을 마코프 프로세스를 이용하여 설명한다.

이하에서는 도면을 참조하여 본 발명의 바람직한 실시예를 설명하기로 한다. 그러나, 다음에 예시하는 본 발명의 실시예는 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 다음에 상술하는 실시예에 한정되는 것은 아니다.

도 1은 본 발명이 적용되는 다양한 피해 파급 영역을 도시한 것이다.

도 1은 위협의 종류 (T_1, T_2, T_3,...,T_8)에 따른 피해 파급 모형 나타낸다. 도 1에서 볼 수 있듯이, 피해 파급 모형은 위협의 종류, 위협들 사이의 관계 및 위협 행위에 대한 대응책에 의해 다양한 형태를 이룰 수 있다. 본 발명에 따른 마코프 프로세스에 기반한 피해 파급 모델은 다양한 위협들에 적용 가능할 뿐만 아니라 이들 사이의 상호 관계 정도(밀접도)를 바탕으로 그림 1과 같은 다양한 형태의 피해 파급 영역에 대하여 적용할 수 있다. 이를 위해서 확률적 피해 파급 모델은 몇 가지 요구사항을 만족해야 한다. 첫째, 시간에 따른 피해 파급 영역을 고려해야 한다. 즉, 기관이나 사용자들은 시간의 흐름에 따라 위협에 대한 패치 (patch) 또는 업그레이드(upgrade)를 취할 수 있기 때문에 이에 의하여 피해 파급 영역이 바뀔 수 있다. 둘째, 피해 파급 속도는 위협 및 취약점의 종류에 따라 달라질 수 있으므로, 피해 파급 모델은 다양한 종류의 위협에 적용될 수 있도록 설계해야 한다.

위에서 언급한 요구사항을 만족하는 마코프 프로세스에 기반한 피해 파급 모델을 설계하기 위해서는 몇 가지 가정이 필요하다. 첫째, 위협과 관련된 과거의 데이터가 충분하고 신뢰할 만하다. 마코프 프로세스를 이용하여 모델을 설계하기 위해서는 신뢰할 만한 과거 데이터이 수집이 무엇보다도 중요하다. 둘째, 정보 자산(asset)의 피해 파급은 이산 시간(discrete time)에 의해 확산된다. 즉, 이산 시간 또는 특정 이벤트에 의해 확산된다.

이것은 정보 통신 시설의 자산이 완전히 피해를 받았을 경우에만 이것과 관련된 자산에게 피해를 전파한다는 것을 의미한다. 셋째, 전체 시스템은 하나 이상의 정의된 위협을 가지고 있다. 즉, 피해 파급 모델은 정의된 위협에 대한 전체 시스템의 피해 산정 및 피해 파급 효과를 구할 수 있어야 한다. 넷째, 기관이나 사용자들은 시간의 흐름에 따라 위협에 대한 대책을 수립, 적용할 수 있다. 이에 따라 피해 파급 영역은 증가 또는 감소 될 수 있다. 마지막으로, 마코프 프로세스 전이 확률은 전체 시스템의 위협 상태에 적용되는 확률이다.

따라서 하나 이상의 위협 상태에 대해 적용할 수 있다. 마코프 프로세스에 기반한 피해 파급 모델은 앞서 설명한 마코프 프로세스의 요소, 즉, 상태 집합, π 벡터(초기 확률), 전이확률을 정의함으로써 이루어진다.

본 발명에서는 피해 파급에 대한 전체 시스템의 피해 손실을 정량적으로 표현하기 위해서 수학식 1을 이용한다.

수학식 1

$$RISK(s, t) = Loss(s, t) \times Prob.(s, t)$$

수학식 1에서 RISK는 자산의 취약한 부분에 위협 요소가 발생하여 자산의 손실, 손상이 일어났을 때의 손실액(피해 산정액)을 의미하고, 공간(s, scope)과 시간(t, time) 개념을 함께 고려하였다. 공간은 피해 파급 시스템의 크기 즉, 특정 이벤

트 후의 피해 영역의 확산 및 감소를 연산해 주기 위한 영역 개념이다. 시간 t는 시간 경과에 따른 피해 파급의 정도를 고려하기 위한 요소로서 일정한 시간 또는 특정 이벤트에 의한 피해 영역의 모양 및 크기가 변경되는 것을 적용하기 위한 것이다. 즉, 피해 파급 손실액은 자산이 특정한 위협의 발생에 의해 자산의 하락하는 정도(Loss)와 위협이 발생할 확률(Prob.)에 의해 연산된다.

도 2는 본 발명에 따른 마코프 프로세스 기반의 피해 산정 장치의 블럭도이다.

위협 정의부(210)는 장기간의 위협 발생 데이터를 수집 분석하여 주요 정보 통신 기반 시설에 발생 가능한 위협 상태 집합을 정의한다.

행렬 산출부(220)는 장기간의 위협 발생 데이터의 각 위협별 발생 빈도수와 위협 상태 집합 사이의 매핑을 이용하여 위협 상태 전이 행렬을 산출한다.

초기값 산출부(230)는 장기간의 위협 발생 데이터보다 최근의 위협 발생 데이터를 이용하여 위협 상태 집합의 초기 발생 확률을 산출한다.

위협 예측부(240)는 위협 상태 전이 행렬과 초기 발생 확률을 이용하여 주요 정보 통신 기반 시설에서 발생 가능한 위협들의 발생 빈도 예측 정보를 생성한다.

바람직하게는, 위협 정의부(210) 및 초기값 산출부(230)는 대량의 위협 발생 데이터를 저장하는 데이터 베이스와 연결될 수 있다. 바람직하게는, 위협 정의부(210), 행렬 산출부(220), 초기값 산출부(230) 및 위협 예측부(240)는 컴퓨터 상에서 실행가능한 프로그램 루틴으로 제작될 수 있다. 바람직하게는, 위협 예측부(240)는 위협들의 발생 빈도 예측 정보를 시각적으로 표시하기 위한 디스플레이 장치와 연결될 수 있다.

도 3은 본 발명에 따른 마코프 프로세스 기반의 피해 산정 방법의 흐름도이다.

본 발명에 따른 마코프 프로세스에 기반한 확률적 피해 파급 모델은 도 3과 같은 과정을 거쳐 생성되며, 크게 상태 집합 정의(310 과정), 위협 상태 전이 행렬 산출(320 과정), 초기 확률 산출(330 과정), 위협예측(340 과정) 과정으로 이루어진다.

먼저, 상태 집합 정의(310 과정) 과정에서는 조직이나 기관이 가지고 있는 위협들이 취할 수 있는 상태를 정의한다. 상태(state)란, 주요 정보통신 기반 시설이 가지고 있는 위협의 상태를 말하며, 상태 집합은 하나의 위협 상태가 가질 수 있는 값들의 범위를 나타내거나, 여러 위협 상태들의 쌍(조합)이 될 수 있다.

위협 상태 집합이 정의되면, 위협 상태 전이 행렬 산출(320 과정) 과정에서는 상태 집합에서 정의된 위협 상태와 위협 발생 빈도 데이터를 이용하여 위협 상태들 간의 전이 행렬을 구한다.

다음, 초기 확률 산출(330 과정) 과정에서는 정의된 각 위협 상태가 초기 상태에 발생할 수 있는 확률을 구한다.

마지막으로, 위협예측(340 과정) 과정에서는 전 단계에서 구한 위협 상태 전이행렬과 초기 확률 값을 통해 앞으로 발생할 위협 발생 확률이나 빈도수를 예측할 수 있다.

도 4는 도 3의 상태 집합 정의 과정(310 과정)의 상세 흐름도이다.

상태 집합 정의 과정(310 과정)에서는 주요 정보통신 기반 시설이 가지고 있는 위협의 종류를 조사하고 위협 발생 데이터를 수집, 분석하여 전체 시스템이 가질 수 있는 위협 상태 집합을 정의한다.

상태 집합의 정의는 도 4와 같은 세부 과정를 따른다.

먼저, 위협 발생 데이터 수집 과정(411 과정)에서는 조직이나 기관이 가지고 있는 과거의 위협 발생 데이터를 수집하는 단계이다. 마코프 프로세스 기반의 피해 파급 모델은 이러한 과거의 데이터가 다른 요소들보다 높은 비중을 차지하기 때문에 신뢰성 있고 대용량의 데이터 수집이 무엇보다 중요하다. 예를 들어, 한국정보보호진흥원(Korea Information Security Agency, KISA)에서 2001년 1월부터 2005년 6월까지 보고된 해킹바이러스 통계 및 분석 월보를 분석, 이용하여 데이터에 대한 신뢰를 높일 수 있다.

데이터 수집이 완료된 후, 위협 분석 과정(412 과정)에서는 자산의 가치에 악영향을 줄 수 있는 잠재적인 위협을 파악하고 발생 가능성 등을 파악하는 단계로 피해를 산출하는데 있어서 중요한 과정이다. 위협 분석은 크게 위협 파악과 위협 순위 산정으로 구성된다. 위협 파악은 위협을 유형별로 분류, 조사하고 각 위협의 주기를 산출한다. 예를 들어, 위협 유형은 한국정보보호진흥원의 보고서를 토대로 하여, 크게 해킹, 바이러스, 스캔탐지로 구분할 수 있다. 위협 순위는 파악된 위협 주기를 바탕으로 위협의 심각성에 따라 가장 고려해야 될 위협 순위를 결정한다. 이는 조직에 대해 가장 많은 영향을 줄 수 있는 위협의 중요도를 정하여 조직에 피해를 줄 수 있는 위협을 우선적으로 고려하여 보다 효과적인 대책을 마련하기 위해 필요하다. 이를 통해 마코프 프로세스 기반의 피해 파급 모델에 우선적으로 적용해야 할 위협을 결정한다. 예를 들어, 한국정보보호진흥원 보고된 해킹바이러스 통계 및 분석 월보를 이용하여 각 위협 유형별 위협 주기를 분석하여 대표적인 위협을 선택할 수 있다.

마지막으로, 위협 상태 집합 정의 과정(413 과정)에서는 위협 주기를 분석하여 위협 발생수의 적절한 임계값(threshold)을 결정하여 위협 상태를 정의한다. S가 위협 상태의 집합이라면 다음과 같이 정의할 수 있다.

$T=\{T_1, T_2, \cdots, T_n\}$는 임계값의 집합, T_i은 해킹, 바이러스, 웜 등과 같은 특정 위협, $S=\{S_1, S_2, \cdots, S_i, \cdots, S_n\}$는 위협 상태 집합, $S_i=(T_\alpha, T_\beta, \cdots, T_\gamma)$은 서로 다른 위협들의 임계값 쌍, α, β, γ 는 서로 다른 위협을 나타낸다.

위협 상태 집합은 '위협이 독립적으로 발생하는가?' 또는 '위협이 위협들 사이에 서로 연관성을 가지고 발생하는가?'에 따라 정의하는 방법이 다르다. 전자의 경우, 위협 상태 집합은 하나의 위협이 가질 수 있는 임계값 범위가 집합이 된다. 반면, 후자의 경우에는 여러 위협들의 임계값 범위의 조합으로 상태 집합이 정의된다. 따라서, 각 위협 주기의 임계값을 얼마나 세분화하여 설정하느냐에 따라 위협 상태의 수와 복잡도가 다르게 되며, 또한 전이 행렬이 가지는 요소(element)의 수 및 복잡도도 달라진다. 즉, 각 위협들의 임계값 범위를 세분화 할수록 생성되는 위협 상태 집합이 커지며, 이에 따라 복잡도도 증가하게 된다.

도 5는 도 3의 위협 상태 전이 행렬 산출 과정(320 과정)의 상세 흐름도이다.

위협 상태 전이 행렬 산출 과정(320 과정)은 상태 집합 정의 과정(310 과정)에서 정의된 위협 상태들 간의 전이 확률을 구하는 단계이다. 분석된 각 위협별 발생 빈도수와 전 단계에서 정의된 상태 집합과의 매핑을 통해 위협 상태 전이 행렬을 구한다.

위협 상태 전이 행렬을 구하기 위해서는 도 5와 같이 2가지의 세부 과정를 수행한다.

먼저, 각 위협별 발생 빈도 데이터를 위협 상태 집합과 매핑(mapping)하여 상태들을 열거한다(521 과정). 다음, 열거된 상태들을 분석하여 하나의 위협 상태에서 다른 상태로의 전이 횟수를 구하고 이를 이용하여 전이 행렬을 구한다(522 과정). 위협 발생 빈도와 위협 상태 집합과의 매핑은 다음의 수학식 2와 같다.

수학식 2

위협 상태: $S \rightarrow 2^T$, 위협 발생 빈도의 위협 상태로의 매핑

위협 상태 전이 행렬 산출 과정(320 과정)은 상태 집합 정의 과정(310 과정)과 마찬가지로, 위협들이 서로 독립적으로 발생하여 각 위협별 상태 전이 행렬을 구하는 경우와, 여러 위협들이 서로 연관되어 만들어지는 경우로 나눌 수 있다. 위협들이 서로 독립적으로 발생하여 만들어지는 위협 상태 전이 행렬은 복잡도가 작아 위협 발생 빈도와 정의된 상태 집합과의 매핑을 통해 쉽게 생성할 수 있다. 반면에 여러 위협들이 서로 연관되어 만들어지는 경우의 전이 행렬은 연관된 위협의 종류가 많을수록, 또한 각 위협별 정의된 상태의 개수에 따라 전이 행렬의 크기 및 복잡도가 커지게 된다. 복잡도를 적절하게 만들기 위해서는 상태 집합 정의 단계에서 각 위협별 적절한 임계값 범위를 설정하여 적당한 개수의 위협 상태를 정의해야 한다.

위협 상태 전이 행렬단계에서 생성되는 상태 전이 행렬 P를 나타내면 수학식 3과 같고 수학식 4를 만족한다.

수학식 3

$$P=\begin{pmatrix} P_{11} & P_{12} & P_{13} & \cdots & \cdots & P_{1n} \\ P_{21} & P_{22} & P_{23} & \cdots & \cdots & P_{2n} \\ P_{31} & P_{32} & P_{33} & \cdots & \cdots & P_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & P_{ij} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ P_{n1} & P_{n2} & P_{n3} & \cdots & \cdots & P_{nn} \end{pmatrix}$$

이때, $P_{ij} = P(S_j \text{ at } t+1 \mid S_i \text{ at } t)$ (t 는 시간)이다.

수학식 4

$$\sum_{j=1}^{n} P_{1j} = 1,\ \sum_{j=1}^{n} P_{2j} = 1,\ \sum_{j=1}^{n} P_{3j} = 1,\ \cdots,\ \sum_{j=1}^{n} P_{nj} = 1$$

이때, $P_{ij} \geq 0$, $\sum_{j=1}^{n} P_{ij} = 1$, i = 1, 2, ... , n이다.

각 열(row)은 하나의 위협 상태에서 다른 위협 상태로의 확률을 나타내고, 각 행의 합은 1이 되어야 한다.

도 6은 도 3의 초기 확률 산출 과정(330 과정)의 상세 흐름도이다.

초기 확률(π 벡터) 산출 과정(330 과정)은 전체 시스템에서 정의된 각 위협 상태가 초기 상태에 가질 수 있는 위협 발생 확률로써 도 6와 같은 세부 과정를 따른다.

먼저, 위협 상태의 초기 확률값을 구하기 위해서, 우선 구하고자 하는 위협의 최근 발생 데이터를 조사한다(631 과정). 초기 확률을 구하기 위한 최근의 위협 발생 데이터는 3개월, 6개월, 9개월, 1년 등의 시간 단위로 이용한다. 최근의 위협 발생 데이터를 분석하여 수학식 5와 같이 연산하며 수학식 6을 만족하는 각 위협 상태별 발생 빈도 및 초기 확률을 연산한다(632 과정).

수학식 5

$$P(S_1\ S_2 \ldots S_k \ldots\ S_n) = P\left(\frac{\alpha}{F}\ \ \frac{\beta}{F} \ldots \frac{\gamma}{F} \ldots \frac{\delta}{F}\right)$$

단, α, β, γ, δ 는 각 상태 (S_1, S_2, S_k, S_n)에서의 발생 횟수이다.

수학식 6

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

또한, 초기 위협 확률의 총 합은 1이 되어야 하므로 수학식 7을 만족한다.

수학식 7

$$\sum_{i=1}^{N} P(S_i) = 1$$

이때, S는 위협 상태를 나타낸다.

도 7은 도 3의 위협 예측 과정(340 과정)의 상세 흐름도이다.

위협 예측 과정(741 과정)은 전 단계에서 생성된 위협 상태 전이행렬과 초기 확률 값을 이용하여 앞으로 발생할 위협 발생 확률이나 빈도수를 예측한다.

위협 발생 빈도를 예측하기 위해서 임계값(범위를 갖는 값)의 평균값 또는 중간값(median)을 대표값으로 이용하게 된다.

먼저, 다음의 수학식 8과 같이, 위협 상태 전이 행렬과 초기 확률을 이용하여 위협 발생 확률을 구한다(741 과정). 또한 수학식 9와 같이 여러 위협 상태 중 특정 위협 상태에 대한 발생 확률을 구할 수 있다.

수학식 8

$$(P(S_1)\ P(S_2)...P(S_k)...\ P(S_n))\begin{pmatrix} P_{11} & P_{12} & P_{13} & \cdots & \cdots & P_{1n} \\ P_{21} & P_{22} & P_{23} & \cdots & \cdots & P_{2n} \\ P_{31} & P_{32} & P_{33} & \cdots & \cdots & P_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & P_{ij} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ P_{n1} & P_{n2} & P_{n3} & \cdots & \cdots & P_{nn} \end{pmatrix}$$

$=(P(S_1)'P(S_2)'..P(S_k)'..P(S_n)')$

수학식 9

$$P(S_k)' = \sum_{i=1}^{n} P(S_i)P_i$$

S는 위협 상태, k는 특정 위협 상태를 나타내고, n은 위협 발생 상태 집합이 개수, $P(S_i)$는 각 위협 상태의 초기 발생 확률, $P(S_i)'$는 각 위협 상태의 다음 발생 확률을 나타낸다.

위협 발생 확률과 각 상태의 평균값을 이용하여 수학식 10과와 같은 예상 위협 발생 빈도를 구할 수 있다(742 과정).

수학식 10

$$\text{예상 위협 발생 빈도} = \sum_{i=1}^{n} P(S_i)M(S_i)$$

단, n은 위협 상태 집합의 개수, $P(S_i)$는 각 위협 상태의 발생 확률, $M(S_i)$는 각 위협 상태의 평균값을 나타낸다.

본 발명의 적용 사례로서 한국정보보호진흥원에서 2001년 1월부터 2005년 6월까지 보고된 해킹바이러스 통계 및 분석 월보에 보고된 데이터를 이용할 수 있다. 우선 상태 집합의 정의 단계에서 위협 발생 데이터를 수집하고 수집된 데이터를 분석하여 위협 유형을 분류하고 위협 순위를 산정한다. 위협 유형은 크게, 해킹, 바이러스, 스캔탐지로 나누었고 각 유형별 대표적인 위협과 월별 발생수는 표1, 2, 3과 같다.

T_1 및 표 1은 해킹 (악성 프로그램을 이용한 통신망에 대한 불법적인 침입)을 나타낸다. 해킹 위협 유형 중 T_1은 Netbus, Subseven처럼 악성 프로그램을 이용한 공격과 함께 시스템의 취약점을 자동으로 공격하는 프로그램을 이용한 '통신망에 대한 불법적인 침입' 위협이다. 이 위협은 사용자의 시스템에 설치되어 백도어를 오픈하여 정보를 유출하거나 시스템의 정상적인 동작을 방해하는 위협으로써 발생하기 쉬운 해킹 위협이다.

[표 1]

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	총계
2001년	85	125	70	89	85	64	65	495	268	77	51	97	1,571
2002년	401	119	82	59	286	417	313	298	210	465	472	990	4,112
2003년	1148	557	1132	934	306	450	185	544	119	137	129	96	5,837
2004년	154	148	118	1066	493	181	72	22	16	24	125	90	2,509

2005년	29	20	15	3	15	36							118
평균	363.4	193.8	283.4	430.2	237.0	229.6	158.7	339.7	153.2	175.7	194.2	318.2	

T_2및 표 2는 바이러스 (인터넷 웜)를 나타낸다. T_2는 '인터넷 웜'으로써 바이러스 위협 유형 중 하나로, 독립적으로 자기 복제를 실행하여 번식하는 빠른 전파력을 가진 컴퓨터 프로그램 또는 실행 가능한 프로그램으로 인한 위협이다.

[표 2]

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	총계
2001년	1	1529	2429	625	684	520	6106	5965	10772	4795	4068	3024	40,518
2002년	2005	1384	1306	3165	2760	1774	1706	1458	1610	3566	3028	1684	25,446
2003년	1361	1320	2537	2350	3704	1854	1185	9748	19682	3999	11658	8949	68,347
2004년	4824	5750	9820	4233	19728	22767	15228	8132	3153	2658	2319	2117	100,727
2005년	1832	1205	1049	648	1302	1040							7,076
평균	2,004	2,237	3,428	2,204	5,635	5,591	6,056	6,325	8,804	3,754	5,268	3,943	

T_3및 표 3은 스캔탐지 (네트워크 도청 및 감청)를 나타낸다. T_3는 스캔탐지 위협 유형의 한 종류로, 대상 시스템의 운영체제, 설정 등을 알아보기 위하여 스캔하는 등의 행위로 주로 해킹의 사전 단계로 이용되는 '네트워크 도청 및 감청' 위협이다.

[표 3]

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	총계
2002년	1665	1256	2080	2110	1776	1418	1177	1216	1601	2483	1596	1597	20335
2003년	648	593	656	402	345	1489	469	1168	3120	3560	2201	315	14966
2004년	2004	3389	10631	18546	11618	833	1591	1964	901	1794	2628	1381	57217
평균	1,439	1,746	4,455	7,019	4,579	1,246	1,079	1,449	1,874	2,612	2,141	1,097	

위의 위협들과 발생 빈도를 분석한 다음 적절한 임계값을 설정하여 위협 상태를 정의한다. 상태 집합을 정의하기 전에, 각 위협이 서로 독립적으로 발생하는 경우와 서로 연관성을 가지고 발생하는 경우로 나누어 설명한다.

도 8a 및 도 8b는 본 발명의 적용 사례에서 위협 상태를 도시한 것이다.

다음은 각 위협이 서로 독립적으로 발생하여 다른 위협들과 연관 관계가 없는 경우의 적용 사례를 보여준다. 따라서 주요 정보통신 기반 시스템이 가지고 있는 각 위협별 서로 다른 위협 상태 전이 행렬을 가지게 된다. 단, 위협이 발생할 때마다 각 위협은 동일한 조건에서 발생한다고 가정한다. 예를 들어, 위협에 대한 보안 대응책, 시스템 자원 및 동일한 환경을 가지고 있다.

우선, 위협 T_1(악성 프로그램을 이용한 통신망에 대한 불법적인 침입)에 대한 상태 집합을 정의한다. T_1에 대한 월별 발생 빈도수는 표 1과 같으며 다음과 같은 임계값의 범위로 상태 집합(S)을 정의한다. S의 임계값의 범위: S_1 : 0~300 , S_2 : 301~600, S_3 : 601~900, S_4 : 901~1200 일때, S = {S_1, S_2, S_3, S_4} 이다.

본 발명에서 임계값은 위협의 단위 시간(월별) 발생 빈도를 몇 개의 구간으로 나누어 표현한다.

위협이 독립적으로 발생하는 경우에는 상태 집합이 단지 해당 위협의 임계값 범위에 의해 결정된다. 다음으로, 일 예로 2001년 1월부터 2005년 6월까지 월별 위협 발생수를 정의된 상태 집합(S)과 매핑하여 상태를 다음과 같이 열거한다.

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_2, S_1, S_1, S_1, S_1, S_2, S_2, S_1, S_1, S_2, S_2, S_4, S_4, S_2, S_4, S_4, S_2, S_2, S_1, S_2, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_4, S_2, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1

열거된 상태들로부터 각 상태(S_1, S_2, S_3, S_4)에서 다른 상태로의 전이 횟수를 구하고 이를 바탕으로 다음의 수학식 11과 같은 상태 전이행렬을 구한다.

수학식 11

$$\begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \end{array}\begin{array}{c} \begin{array}{cccc} S_1 & S_2 & S_3 & S_4 \end{array} \\ \begin{pmatrix} 31 & 5 & 0 & 1 \\ 6 & 3 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 2 \end{pmatrix} \end{array} \quad \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \end{array}\begin{array}{c} \begin{array}{cccc} S_1 & S_2 & S_3 & S_4 \end{array} \\ \begin{pmatrix} 0.84 & 0.13 & 0 & 0.03 \\ 0.55 & 0.27 & 0 & 0.18 \\ 0 & 0 & 0 & 0 \\ 0 & 0.60 & 0 & 0.40 \end{pmatrix} \end{array}$$

수학식 11로부터, 각 위협에서 다른 위협으로의 전이 확률 값의 합이 1이 된다. 위의 위협 상태 전이 확률을 상태 다이어그램으로 나타내면 도 8a와 같다.

이상에서는 T_1에 대한 전이 행렬을 구하는 과정을 보였으며, 다른 위협들도 위와 같은 과정을 거쳐 각 위협별 상태 전이 행렬을 구할 수 있다.

예를 들어, 위협 T_1에 대한 초기 확률을 구하기 위해 최근 6개월(표 1의 2005년 1월~6월) 동안 발생한 빈도수를 이용할 수 있다. 최근 6개월 동안 발생한 빈도수와 이에 따른 초기 확률 값을 구하면 다음과 같다. 즉, 빈도수 : 29, 20, 15, 3, 15, 36 = S_1, S_1, S_1, S_1, S_1, S_1 으로서, 초기 확률 : $P(S_1\ S_2\ S_3\ S_4) = P(1\ 0\ 0\ 0)$이다.

위협 상태 전이 행렬과 초기 확률을 이용하여 다음에 발생하게 될 위협 발생 확률을 예측하고 또한 위협 발생 빈도수를 예측할 수 있다. 즉, 위협 발생 확률은 수학식 12와 같이 연산할 수 있다.

수학식 12

$$(1\,0\,0\,0)\begin{pmatrix} 0.84 & 0.13 & 0 & 0.03 \\ 0.55 & 0.27 & 0 & 0.18 \\ 0 & 0 & 0 & 0 \\ 0 & 0.60 & 0 & 0.40 \end{pmatrix} = (0.84\,0.13\,0\,0.03)$$

수학식 12로부터 T_1은 다음 달 S_1 상태로 0.84, S_2 상태로 0.13, S_4 상태로 0.03의 확률로 발생할 것이라고 예측할 수 있다. 즉, 0과 300 사이의 발생빈도를 가질 것이라는 것을 알 수 있다.

좀더 구체적인 다음 달의 위협 발생 빈도를 구하기 위해서 수학식 10을 이용한다. 이를 위해 위협 발생 확률과 각 임계값의 평균값을 이용한다. 본 적용사례에서는 평균값을 구하기 위해 전달의 위협 발생빈도를 이용하였다. 앞서 정의한 임계값의 범위에 따라 평균값을 구하면 다음과 같다. 즉, $M(S_1)=36$, $M(S_2)=0$, $M(S_3)=0$, $M(S_4)=0$ 이다. 또한, 각 위협상태에 대한 발생확률을 얻을 수 있다. 즉, $P(S_1)=0.84$, $P(S_2)=0.13$, $P(S_3)=0$, $P(S_4)=0.03$ 이다.

따라서 위협 발생 빈도는 수학식 7에서 n=4 일 때, 즉 위협 상태의 개수가 4개일 때 이므로 다음과 같이 구한다. 즉, 예상 위협 발생 빈도 = $\sum_{i=1}^{4} P(S_i)M(S_i)$ = $0.84 \times 36 \fallingdotseq 30$ 이다.

위 결과로부터 다음 달 T_1이 발생할 빈도수는 약 30으로 예측할 수 있다. 또 다른 위협 T_2와 T_3의 경우도 위와 같은 방식으로 각각의 위협 발생 확률 및 빈도를 구할 수 있다.

각각의 상태가 서로 독립적이지 않고 서로 연관되어 발생되는 경우에 이들 위협들 사이의 관계를 반영하기 위해서 위협 상태들의 쌍(조합)으로 처리한다. 본 절에서는 이에 대한 예를 들기 위해, T_1(악성 프로그램을 이용한 통신망에 대한 불법적인 침입)과 T_2(인터넷 웜)에 관련된 위협 상태 전이 행렬을 구한다. 5.1절의 예와 마찬가지로, 각 위협이 발생할 때마다, 전체 시스템은 모든 동일한 조건을 가진다고 가정한다. 즉, 동일한 시스템 자원 및 환경을 가지고 있다.

T_1, T_2에 대한 월별 발생 빈도수는 표 1, 2와 같으며 각 위협에 대하여 아래와 같은 임계값의 범위를 정의한다. 즉, T_1의 임계값은 각각 H_1: 0~400, H_2: 401~800, H_3: 801~1200 으로 정의하고, T_2의 임계값은 각각 W_1: 0~4000, W_2: 4001~8000, W_3: 8001 이상으로 정의 할 수 있다.

위협 상태 집합은 T_1과 T_2의 임계값의 쌍을 조합하여 총 9개의 상태가 정의된다. 즉, 위협 상태의 집합은 S = { S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9}가 된다.

단, $S_1 = (H_1, W_1)$, $S_2 = (H_1, W_2)$, $S_3 = (H_1, W_3)$, $S_4 = (H_2, W_1)$, $S_5 = (H_2, W_2)$, $S_6 = (H_2, W_3)$, $S_7 = (H_3, W_1)$, $S_8 = (H_3, W_2)$, $S_9 = (H_3, W_3)$ 이다.

위협 상태 전이 행렬을 구하기 위해, 일 예로 2001년 1월부터 2005년 6월까지 T_1, T_2의 월별 위협 발생 데이터를 이용하여 각 달의 T_1, T_2 의 위협 발생수 쌍 (T_1, T_2)을 정리한다. 즉, (85, 1), (125, 1529), (70, 2429), (89, 625), (85, 684), (64, 520), (65, 6106), (495, 5965), (268, 10772), (77, 4795), (51, 4068), (97, 3024), ... 중간 생략 ... , (29, 1832), (20, 1205), (15, 1049), (3, 648), (15, 1302), (36, 1040) 과 같이 나타낼 수 있다.

위의 (T_1, T_2) 쌍을 각 위협의 임계값(T_1, T_2의 임계값) 범위에 맞추어 (H_i, W_j) 쌍을 구한다. 즉, (H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_2), (H_2, W_2), (H_1, W_3), (H_1, W_2), (H_1, W_2), (H_1, W_1), ... 중간 생략 ... ,(H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1) 과 같이 구한다.

각 (H_i, W_j) 쌍을 정의된 위협상태 집합과 매핑하여 상태를 열거한다.

S_1, S_1, S_1, S_1, S_1, S_1, S_2, S_5, S_3, S_2, S_2, S_1, S_4, S_1, S_1, S_1, S_1, S_4, S_1, S_1, S_1, S_4, S_4, S_7, S_7, S_4, S_7, S_7, S_1, S_4, S_1, S_6, S_3, S_1, S_3, S_3, S_2, S_2, S_4, S_2, S_6, S_3, S_3, S_3, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1

열거된 위협 상태들로부터 각 상태(S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9)에서 다른 상태로의 전이 횟수를 구하고 이를 바탕으로 전이행렬을 구하면 수학식 13과 같다.

수학식 13

$$\begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \\ S_7 \\ S_8 \\ S_9 \end{array} \begin{array}{c} \begin{array}{ccccccccc} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 & S_7 & S_8 & S_9 \end{array} \\ \begin{pmatrix} 19 & 1 & 1 & 4 & 0 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \quad \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \\ S_7 \\ S_8 \\ S_9 \end{array} \begin{array}{c} \begin{array}{ccccccccc} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 & S_7 & S_8 & S_9 \end{array} \\ \begin{pmatrix} 0.73 & 0.04 & 0.04 & 0.15 & 0 & 0.04 & 0 & 0 & 0 \\ 0.17 & 0.32 & 0 & 0.17 & 0.17 & 0.17 & 0 & 0 & 0 \\ 0.29 & 0.29 & 0.42 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.50 & 0.17 & 0 & 0.17 & 0 & 0 & 0.16 & 0 & 0 \\ 0 & 0 & 1.00 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.00 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0 & 0.25 & 0 & 0 & 0.50 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

수학식 13으로부터, 각 상태에서 다른 상태로의 전이 확률 값의 합이 1이 됨을 알 수 있다. 또한 위협상태 전이확률을 상태 다이어그램으로 나타내면 도 8b와 같다.

T_1,T_2 에 대한 위협 상태의 초기 확률을 구하기 위해 최근 1년(표 1, 2의 2004년 7월~2005년 6월) 동안 발생한 빈도수를 이용한다. 최근 1년 동안 발생한 T_1, T_2의 빈도수 쌍과 이에 따른 초기 확률 값을 연산한다.

빈도수는 (72, 15228), (22, 8132), (16, 3153), (24, 2658), (125, 2319), (90, 2117), (29, 1832), (20, 1205), (15, 1049), (3, 648), (15, 1302), (36, 1040) = S_3, S_3, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1 와 같고, 초기 확률은 $P(S_1\ S_2\ S_3\ S_4\ S_5\ S_6\ S_7\ S_8\ S_9)$ = P(0.83 0 0.17 0 0 0 0 0 0) 가 된다.

위협 상태 전이 행렬과 초기 확률을 이용하여 다음에 발생하게 될 위협 발생 확률을 예측하고 또한 위협 발생 빈도수를 예측할 수 있다.

즉, $(0.83\,0\,0.17\,0\,0\,0\,0\,0\,0)\begin{pmatrix} 0.73 & 0.04 & 0.04 & 0.15 & 0 & 0.04 & 0 & 0 & 0 \\ 0.17 & 0.32 & 0 & 0.17 & 0.17 & 0.17 & 0 & 0 & 0 \\ 0.29 & 0.29 & 0.42 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.50 & 0.17 & 0 & 0.17 & 0 & 0 & 0.16 & 0 & 0 \\ 0 & 0 & 1.00 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.00 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0 & 0.25 & 0 & 0 & 0.50 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

= (0.66 0.08 0.10 0.13 0 0.03 0 0 0) 가 된다.

위의 결과부터 다음 달 각 위협 상태(S_1~S_9)는 0.66, 0.08, 0.10, 0.13, 0, 0.03, 0, 0, 0의 확률로 발생할 것이라고 예측할 수 있다. 이는 S_1의 상태, 즉, (H_1, W_1)의 값이 갖는 발생빈도를 가질 확률이 가장 크다는 것을 말한다. 따라서 T_1은 0에서 400 사이의 값을, T_2는 0에서 4000 사이의 값이 발생할 것이라는 것을 예측할 수 있다. 다음 달의 T_1, T_2의 위협 발생 빈도를 예측하기 위해 앞서 구한 위협 발생 확률과 각 위협들의 임계값 평균값(M)을 이용한다.

본 적용사례에서도 평균값을 구하기 위해 전달의 위협 발생빈도를 이용할 수 있다. 우선 상태 정의 단계에서 구분한 T_1, T_2의 임계값 범위에 따라 평균값을 구하면 다음과 같다.

T_1의 평균값($M(H_i)$)은 $M(H_1)$=36, $M(H_2)$=0, $M(H_3)$=0 이고, T_2의 평균값($M(W_i)$)은 $M(W_1)$=1040, $M(W_2)$=0, $M(W_3)$=0 이다.

각 위협의 발생 빈도를 구하기 전에, 각 위협별 임계값에 대한 발생 확률을 구해야 하는데, 위협 발생 확률 값을 이용하여 구할 수 있다. 위협 T_1, T_2의 각 임계값 발생 확률은 아래와 같다. 즉, T_1의 임계값 확률($P(H_i)$)은 H_1 : 0.66 + 0.08 + 0.10 = 0.84, H_2 : 0.13 + 0 + 0.03 = 0.16, H_3 : 0 이고, T_2의 임계값 확률($P(W_i)$)은 W_1 : 0.66 + 0.13 + 0 = 0.79, W_2 : 0.08 + 0 + 0 = 0.08, W_3 : 0.10 + 0.03 + 0 = 0.13 이다.

마지막으로, 각 위협별 예상 발생 빈도를 구할 수 있다.

예상 위협 발생 빈도 = $\sum_{i=1}^{n} P(S_i)M(S_i)$ = 예상 위협 발생 확률(임계값 확률) × 임계값의 평균값이다.

즉, T_1의 예상 발생 빈도 = $\sum_{i=1}^{3} P(H_i)M(H_i)$ = 0.84×36 ≒ 30 이고, T_2의 예상 발생 빈도 = $\sum_{i=1}^{3} P(W_i)M(W_i)$ = 0.79×1040 ≒ 821 이다.

위 결과로부터 다음 달 T_1이 발생할 빈도수는 약 30, T_2가 발생할 빈도는 821로 예측할 수 있다.

정보 기반 시스템에는 다양한 위협들이 존재하는데 이들 사이의 명확한 관계성 분석을 하기 위해, 먼저 위협들 사이에 어느 정도의 연관성이 있는지를 분석할 필요가 있다. 이러한 분석을 통해 서로 연관된 위협을 탐색할 수 있으며, 연관된 위협들 사이의 관계성 분석을 위한 기초가 된다. 본 발명에서는 이를 위해 공분산 (covariance, Cov)과 상관계수 (corrleation coefficient)를 이용한 위협간의 연관 관계 정도를 분석한다.

공분산은 변수들 간의 상관관계 정도를 나타내며 수학식 14와 같이 정의된다.

수학식 14

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

단, E(X)는 변수 X의 기대값(expectation)이다. Cov(X, Y)= 0 일 경우에는, 임의의 변수 X, Y 값이 아무런 관련이 없음을 나타낸다.

상관계수 ρ(X, Y)는 관련성의 밀접성을 평가하는 지표로 수학식 15와 같이 정의되고 수학식 16을 만족한다.

수학식 15

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

단, $Var(X)=E(X^2)-E(X)^2$, $\sigma_X=\sqrt{Var[X]}$, $\sigma_Y=\sqrt{Var[Y]}$이다.

수학식 16

$$-1 \le \rho(X,Y) \le 1$$

$$\rho(X,Y)=\begin{cases}-1, & \text{if } X=-aY(a>0)\\ 0, & \text{if } X \text{ and } Y \text{ are uncorrelated}\\ +1, & \text{if } X=aY(a>0)\end{cases}$$

이때, 으로서, 특히 상관계수의 값이 0일 경우에는 변수간의 관련성이 0인 경우이다. 상관계수의 값이 -1값에 가까울수록 X가 증가 할수록 Y는 감소하고, 1에 가까울수록 X가 증가 할수록 Y는 감소하는 관계를 가지고 있다.

적용사례의 세 가지 위협 T_1, T_2, T_3 의 상관관계 정도에 대하여 살펴보면 수학식 15에 의해 다음과 값을 얻을 수 있다. 즉, ρ(T1, T2)= -0.0401 , ρ(T1, T3)= 0.19979 , ρ(T2, T3)= 0.25718 이다. 위의 결과로 두 위협 T_2, T_3는 관련성의 밀접성이 0.2571 값으로 세 위협 사이에 가장 밀접한 관련이 있는 것으로 분석할 수 있다. 즉, 위협 T_2(바이러스 : 인터넷 웜)의 발생이 위협 T_3(스캔탐지 : 네트워크 도청 및 감청)의 발생에 어느 정도의 영향을 미친다고 분석할 수 있다. 또한, 적용사례에 보인 두 위협 T_1, T_2는 관련성의 밀접성이 -0.040 로 T_1, T_2 사이의 관련성이 거의 없다는 것을 알 수 있다. 즉, 위협 T_1(해킹: 악성 프로그램을 이용한 통신망에 대한 불법적인 침입)의 발생이 위협 T_2(바이러스 : 인터넷 웜)의 발생에 거의 영향을 주지 않는다고 분석할 수 있다.

본 발명에서는 세 가지 위협에 대해서만 연관 관계 정도를 보였지만, 이 위협 이외의 다양한 위협에도 공분산과 상관관계를 이용한 연관 관계 정도를 분석할 수 있다. 이를 이용하여 수많은 위협들 중에서 어떤 위협에 대해 적용할 것인지를 결정하여 피해 산정을 연산하기 위한 성능을 높일 수 있다. 또한 이를 바탕으로 각 위협들 사이의 명확한 연관 관계 분석에 이용할 수 있다.

도 9 내지 도 13b는 본 발명에 따른 시뮬레이션 결과 그래프이다.

본 발명에 따른 마코브 프로세스에 기반한 확률적 피해 파급 모델의 검증을 위해 5가지의 시나리오를 구성하여 모델을 검증할 수 있다. 확률적 피해 파급 모델의 시나리오는 각 위협이 독립적으로 발생하는 경우와 위협들이 서로 연관되어 발생하는 경우를 모두 포함한다.

시뮬레이션을 통한 모델 검증을 위해, 우선 KISA에서 발행한 해킹바이러스 통계 및 분석 월보의 2001.1 ~ 2004.6 의 자료를 이용하여 피해 파급 모델을 생성하고, 이 모델을 바탕으로 이후 2004.7 ~ 2005.6의 위협 발생 빈도를 예측하여 실제 발생빈도와 비교 분석한다. 즉, 2001.1 ~ 2004.6의 자료를 이용해 위협 상태 집합, 위협 전이 확률, 초기 확률 값을 정의하고 정의된 모델을 통해 앞으로 일어날 발생 빈도를 예측(2004.7~ 2005.6)하여, 예측된 발생 빈도와 실제 일어난 빈도와의 비교 분석한다. 시나리오는 크게 5개로 작성되였다.

시나리오 1에서는 대표값의 적용 범위를 달리 함으로써 위협 발생 빈도에 영향을 미치는 정도를 분석하였다. 즉, 대표값의 적용 범위를 최근1달, 2달, 6달의 값을 이용함으로써 위협 발생 빈도에 어느 정도 영향을 주는지 실험하였다. 시뮬레이션 조건은 다음과 같다.

대표값을 연산하기 위해 적용 범위를 최근 1달, 최근 2달의 평균, 최근 6달의 평균 발생 빈도로 나누고, 초기 확률을 연산하기 위해 최근 6 개월의 발생 빈도 사용한다. 매달마다 초기값이 변경된다. 위협 상태 전이 행렬이 6 개월마다 갱신된다.

위의 조건을 바탕으로 시뮬레이션 한 결과, 도 9와 같은 결과를 얻었다. 결과를 분석해 보면, 대표값으로 최근의 데이터를 사용할수록 실제 발생 빈도(그래프의 KISA)와 가까워진다는 것을 확인할 수 있다. 즉, 최근 1개월(1 Month)의 빈도를 대표값으로 사용함으로써 최근 2달의 평균 값(2 Months), 또는 최근 6달의 평균값을 대표값으로 이용하는 경우(6 Months)보다 실제 발생 빈도와 가까운 좋은 결과를 얻을 수 있다. 이는 마코브 체인의 정의에서도 알 수 있듯이 앞으로 발생할 사건이 최근의 상태, 즉 최근의 발생빈도 데이터를 이용함으로써 좀 더 정확한 예측을 할 수 있다는 것을 말한다.

시나리오 2에서는 초기 확률을 구하기 위한 적용 범위를 달리 함으로써 위협 발생 빈도에 영향을 미치는 정도를 분석하였다. 즉 초기 확률 값을 이용하기 위해 최근 3개월, 6개월, 1년의 값을 사용하였다. 시뮬레이션 조건은 다음과 같다.

대표값을 연산하기 위해 전달(최근 1달)의 발생 빈도 사용하고, 대표값은 매달마다 갱신된다. 초기 확률 값을 구하기 위해 위협 발생빈도를 최근 3개월, 최근 6개월, 최근 1년의 발생 빈도 값으로 구분한다. 위협 상태 전이 행렬이 6 개월마다 갱신된다.

위의 조건을 바탕으로 시뮬레이션 한 결과 도 10과 같은 결과를 얻었다. 도 10을 바탕으로 시나리오 2의 시뮬레이션 결과 값을 살펴보면, 초기 확률 값을 구하기 위해 최근 3개월의 값을 이용하는 것(3 Months)이, 최근 6개월(6 Months)이나 1년(1 Year)의 발생 빈도를 이용하는 것보다 실제 발생 빈도(그래프의 KISA)에 가까운 좋은 결과를 얻을 수 있다. 즉, 시나리오 1와 마찬가지로 최근의 데이터를 이용함으로써 앞으로 발생할 빈도 데이터를 좀 더 정확히 예측할 수 있다.

시나리오 3에서는 위협 상태 전이 행렬의 변경 주기를 달리 함으로써 위협 발생 빈도에 영향을 미치는 정도를 분석하였다. 즉, 위협 상태 전이 행렬을 만들기 위해 사용되는 위협 발생빈도를 1개월, 3개월 6개월 단위로 나누어 실험하였다. 시뮬레이션 하기 위한 조건은 다음과 같다. 대표값을 연산하기 위해 전달(최근 1달)의 발생 빈도 사용하고, 대표값은 매달마다 갱신된다. 초기 확률을 연산하기 위해 최근 6 개월의 발생 빈도 사용, 매달마다 초기값이 변경된다. 위협 상태 전이 행렬을 만들기 위해 위협 발생빈도를 1개월, 3개월, 6개월 단위로 나누어 변경된다.

위의 조건을 바탕으로 시뮬레이션 한 결과 도 11과 같은 결과를 얻었다. 위의 시뮬레이션 결과 값을 살펴보면, 위협 상태 전이 행렬의 변경 주기가 서로 다름에도 불구하고 예상 발생 빈도수는 모든 경우가 거의 일치하는 것을 볼 수 있다. 이는 위협 상태 전이 행렬을 생성할 때 작은 값들의 변경이 전체 행렬에 작게 영향을 주기 때문이다. 이를 통해 전이 행렬의 잦은 변경은 발생 확률의 예측에는 별로 영향을 끼치지 않음을 확인할 수 있다.

시나리오 1~3의 경우에는 위협 상태를 임계값의 범위에 따라 4개로 정의하였다. 시나리오 4에서는 임계값의 범위를 6개로, 즉 위협 상태를 6개로 늘여 위협 상태가 4개인 경우와 비교 분석하였다. 즉, 임계값의 범위(위협 상태)를 2가지 경우로 나누어 시뮬레이션 하였다. 시뮬레이션 조건은 다음과 같다.

즉, 위협 상태 4개 := S1: 0~300, S2: 301~600, S3: 601~900, S4: 901~1200 이고, 위협 상태 6개 := S1: 0~200, S2: 201~400, S3: 401~600, S4: 601~800, S5: 801~1000, S6: 1001~1200 이다. 대표값을 연산하기 위해 전달(최근 1달)의 발생 빈도 사용하고, 대표값은 매달마다 갱신된다. 초기 확률을 연산하기 위해 최근 6 개월의 발생 빈도 사용, 매달마다 초기값이 변경된다. 위협 상태 전이 행렬이 6 개월마다 갱신된다.

위 조건을 가지고 시뮬레이션 한 결과, 도 12와 같은 결과를 얻었다. 시나리오 4의 시뮬레이션 결과 값을 살펴보면, 위협 상태의 수가 6개인 경우가 위협 상태의 수가 4개인 경우보다 2004년의 7월, 8월의 빈도 값만을 제외하고 거의 차이가 없음을 보여주고 있지만, 이는 발생 빈도 데이터의 부족 결과로 판단된다. 실제로 데이터의 양이 많을 경우에는 좀더 큰 영향을 준다고 말할 수 있다. 즉, 발생 빈도 임계값의 범위가 작게 함으로써 위협 상태의 수가 증가되고, 이를 통해 좀더 정확한 값을 예측할 수 있다. 그러나, 위협 상태의 수가 증가할수록 연산의 복잡도는 증가하게 되므로, 위협의 특성 및 정보 시스템의 환경에 따라 적정한 수의 위협 상태가 결정되어져야 한다.

시나리오 5에서는 앞서 적용 사례로 보여준 두 개의 위협, 즉 T_1(악성 프로그램을 이용한 통신망에 대한 불법적인 침입), T_2(인터넷 웜)가 서로 연관 지어 발생하는 경우의 위협 발생 빈도 예측을 실험하였다. 시뮬레이션을 위한 조건은 아래와 같다. 즉, T_1의 임계값은 각각 H_1 : 0~400, H_2 : 401~800, H_3 : 801~1200 이고, T_2의 임계값은 각각 W_1 : 0~3000, W_2 : 3001~6000, W_3 : 6001 이상 이다. 대표값을 연산하기 위해 전달(최근 1달)의 발생 빈도 사용하고, 대표값은 매달마다 갱신된다. 초기 확률을 연산하기 위해 최근 6 개월의 발생 빈도 사용, 매달마다 초기값이 변경된다. 위협 상태 전이 행렬이 6 개월마다 갱신된다.

위 조건을 가지고 시뮬레이션 한 결과 도 13a 및 도 13b와 같은 결과를 얻었다. 시나리오 5을 시뮬레이션 한 결과, T_1에 대해서, 시나리오 1~4에서 독립적으로 발생하는 경우와는 수치적으로 좀 차이가 난다. 이는 임계값의 범위가 독립적으로 발생하는 경우와 서로 연관 지어 발생하는 경우에서 서로 다르기 때문이다. 즉, 독립적으로 발생하는 경우에는 임계값의 범위를 4개로 나누었지만, 연관 지어 발생하는 경우(즉, 시나리오 6의 경우)에는 3개로 나누어 범위가 넓어졌다. 시나리오 4에서 보았듯이, 임계값의 범위에 따라 차이가 나는 것을 확인할 수 있다. 그러나 그래프의 유형은 앞선 시나리오 1~4와 거의 비슷함을 알 수 있다. T_2의 경우에는, 최초 1, 2달의 경우를 제외하고 실제 발생 빈도(도 13a 및 도 13b의 KISA)와 거의 비슷하게 발생함을 확인할 수 있다.

5개의 시나리오에 기반한 시뮬레이션 결과, T_1의 경우는 특정 해킹 툴의 배포, T_2의 경우는 특정 웜 또는 바이러스에 의한 공격에 의해 위협 발생 빈도가 크게 달라지는 경우를 제외하고는 본 발명에서 제시하는 마코브 프로세스에 기반한 확률적 피해 파급 모델을 통해 예측되는 발생 빈도 값이 실제 값과 근사하다는 것을 확인할 수 있다.

본 발명에서는 바이러스나 웜 뿐만 아니라 여러 가지 위협의 종류들로부터 공격을 받았을 때 그 피해 파급 정도를 예측할 수 있는 마코프 프로세스 기반의 확률적 피해 파급 모델을 제공한다. 제공된 모델은 전체 시스템에서 위협들의 발생 확률 및 발생 빈도를 마코프 프로세스를 이용하여 설명함으로써 하나의 특정 위협이 아니라 전체 시스템이 가지고 있는 여러 가지 위협들에 대해 적용할 수 있다. 더불어 공분산과 상관계수를 통해 위협들 사이의 상관관계 정도를 파악할 수 있다.

바람직하게는, 본 발명의 마코프 프로세스 기반의 피해 산정 방법을 컴퓨터에서 실행시키기 위한 프로그램을 컴퓨터로 읽을 수 있는 기록매체에 기록하여 제공할 수 있다.

본 발명은 소프트웨어를 통해 실행될 수 있다. 소프트웨어로 실행될 때, 본 발명의 구성 수단들은 필요한 작업을 실행하는 코드 세그먼트들이다. 프로그램 또는 코드 세그먼트들은 프로세서 판독 가능 매체에 저장되거나 전송 매체 또는 통신망에서 반송파와 결합된 컴퓨터 데이터 신호에 의하여 전송될 수 있다.

컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 테이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 장치의 예로는 ROM, RAM, CD-ROM, DVD±ROM, DVD-RAM, 자기 테이프, 플로피 디스크, 하드 디스크(hard disk), 광데이터 저장장치 등이 있다. 또한, 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 장치에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.

본 발명은 도면에 도시된 일 실시예를 참고로 하여 설명하였으나 이는 예시적인 것에 불과하며 당해 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 실시예의 변형이 가능하다는 점을 이해할 것이다. 그러나, 이와 같은 변형은 본 발명의 기술적 보호범위내에 있다고 보아야 한다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위의 기술적 사상에 의해서 정해져야 할 것이다.

발명의 효과

상술한 바와 같이, 본 발명에 의하면, 바이러스나 웜 뿐만 아니라 여러 가지 위협의 종류들로부터 공격을 받았을 때 그 피해 파급 정도를 예측할 수 있는 마코프 프로세스 기반의 확률적 피해 파급 모델을 제공함으로써, 전체 시스템에서 위협들의 발생 확률 및 발생 빈도를 마코프 프로세스를 이용하여 설명하고 하나의 특정 위협이 아니라 전체 시스템이 가지고 있는 여러 가지 위협들에 대해 적용할 수 있으며, 위협이 발생 하였을 때 생성되는 위협 영역에 대하여 위협들 사이의 관계를 파악할 수 있고, 공분산과 상관계수를 통해 위협들 사이의 상관관계 정도를 파악할 수 있는 효과가 있다.

도면의 간단한 설명

도 1은 본 발명이 적용되는 다양한 피해 파급 영역을 도시한 것이다.

도 2는 본 발명에 따른 마코프 프로세스 기반의 피해 산정 장치의 블럭도이다.

도 3은 본 발명에 따른 마코프 프로세스 기반의 피해 산정 방법의 흐름도이다.

도 4는 도 3의 상태 집합 정의 과정의 상세 흐름도이다.

도 5는 도 3의 위협 상태 전이 행렬 산출 과정의 상세 흐름도이다.

도 6은 도 3의 초기 확률 산출 과정의 상세 흐름도이다.

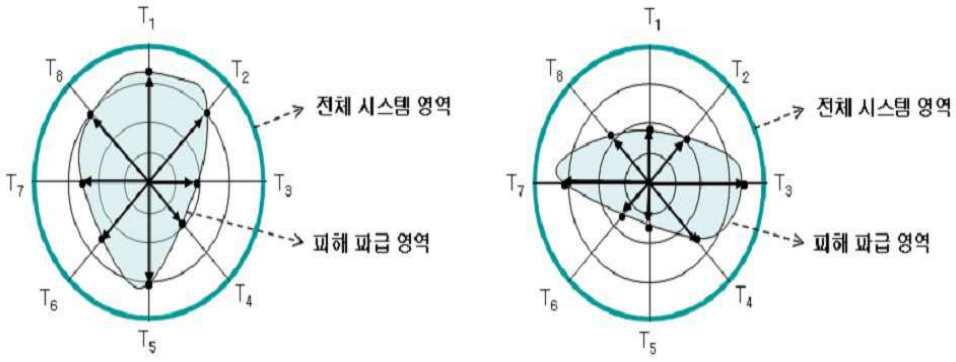
도 7은 도 3의 위협 예측 과정의 상세 흐름도이다.

도 8a 및 도 8b는 본 발명의 적용 사례에서 위협 상태를 도시한 것이다.

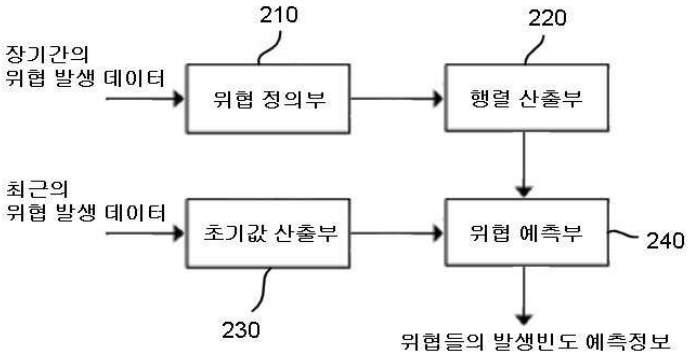
도 9 내지 도 13b는 본 발명에 따른 시뮬레이션 결과 그래프이다.

도면

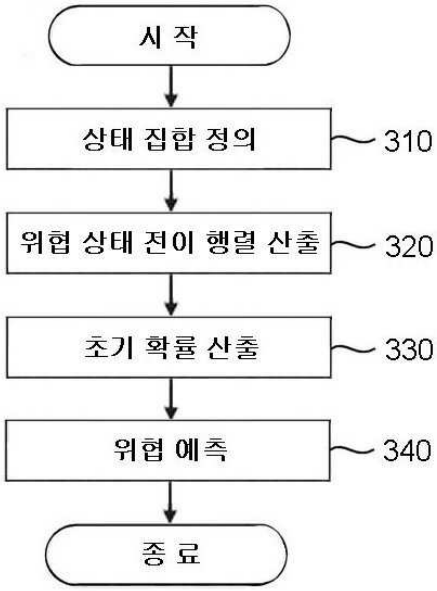
도면1



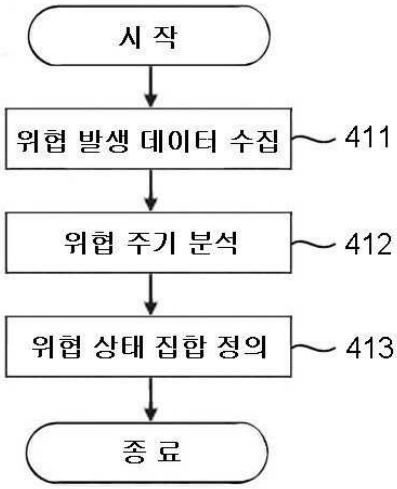
도면2


210
220
장기간의
위협 발생 데이터
위협 정의부
행렬 산출부
최근의
위협 발생 데이터
초기값 산출부
위협 예측부
240
230
위협들의 발생빈도 예측정보

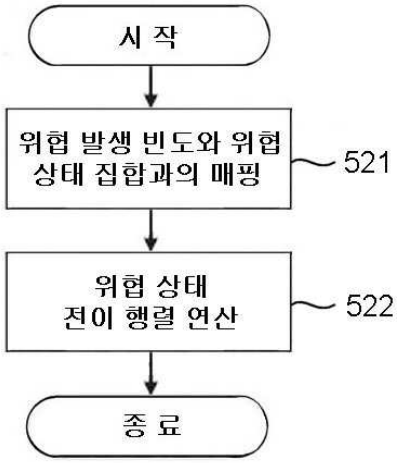
도면3


시 작
상태 집합 정의
310
위협 상태 전이 행렬 산출
320
초기 확률 산출
330
위협 예측
340
종 료

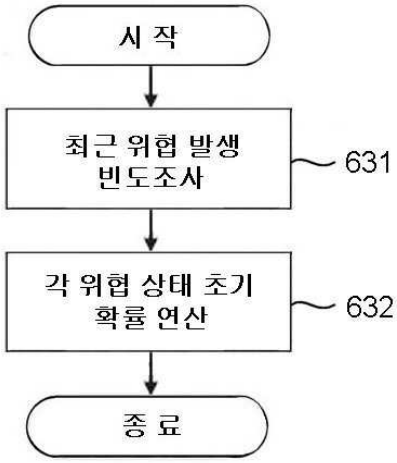
도면4


시 작
위협 발생 데이터 수집
411
위협 주기 분석
412
위협 상태 집합 정의
413
종 료

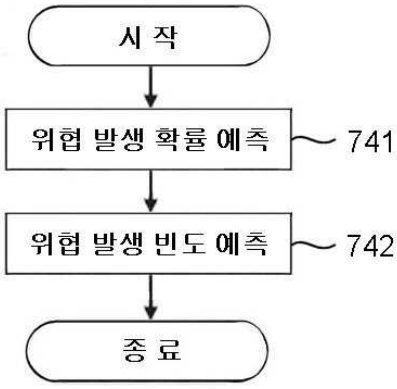
도면5


시 작
위협 발생 빈도와 위협 상태 집합과의 매핑
521
위협 상태 전이 행렬 연산
522
종 료

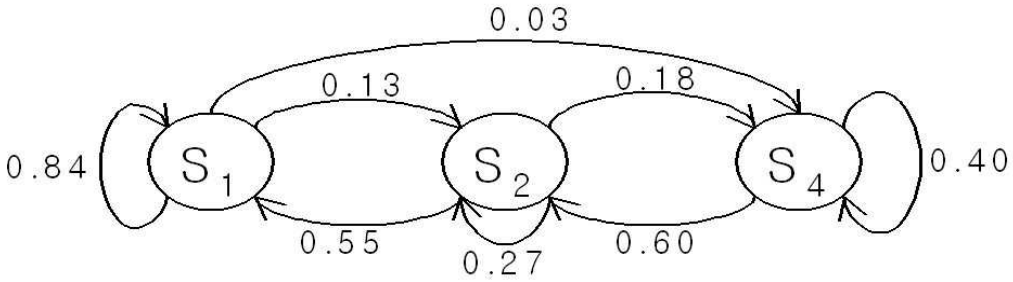
도면6


시 작
최근 위협 발생 빈도조사
631
각 위협 상태 초기 확률 연산
632
종 료

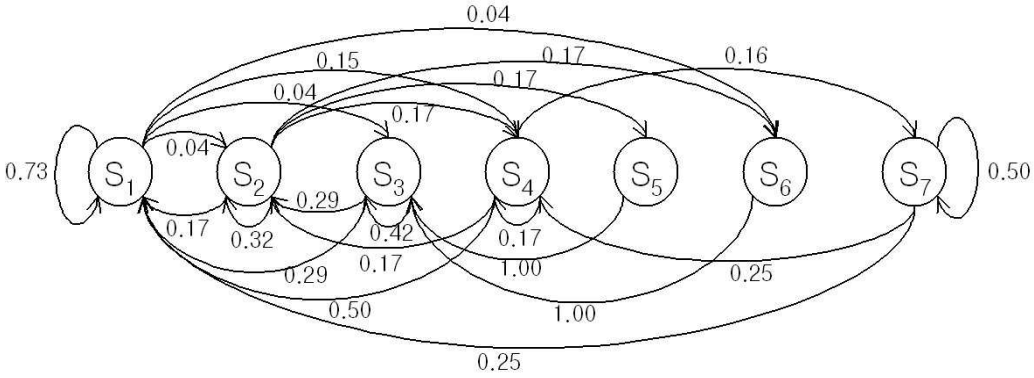
도면7


시 작
위협 발생 확률 예측
741
위협 발생 빈도 예측
742
종 료

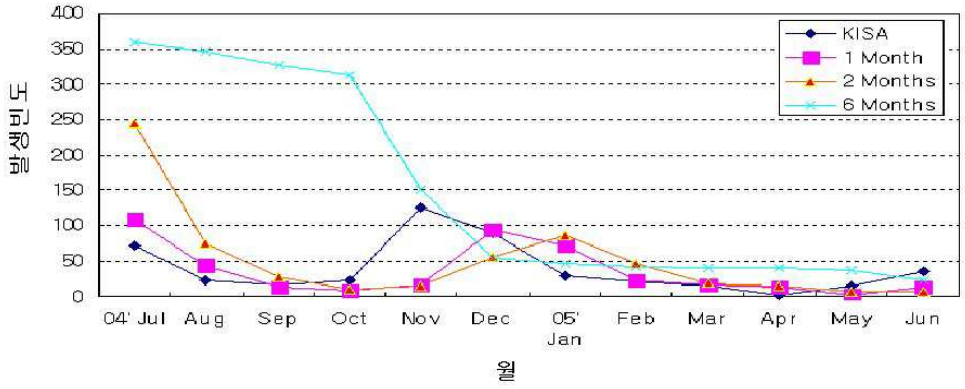
도면8a



도면8b

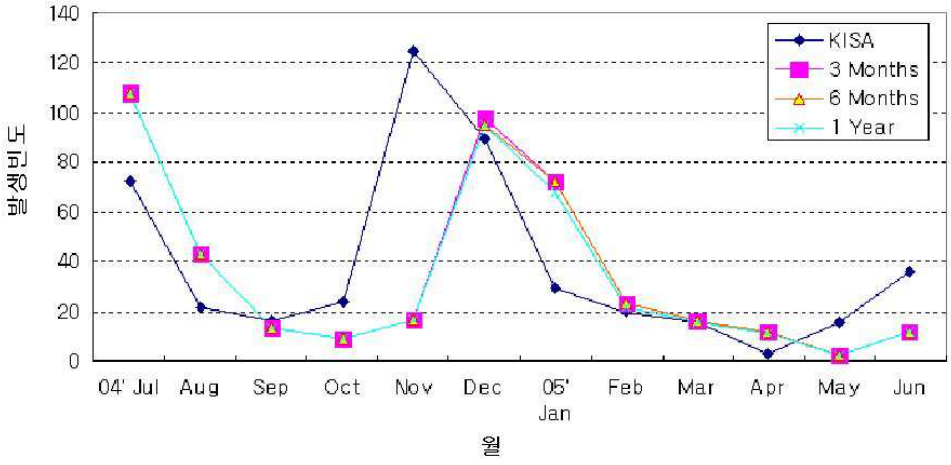


도면9



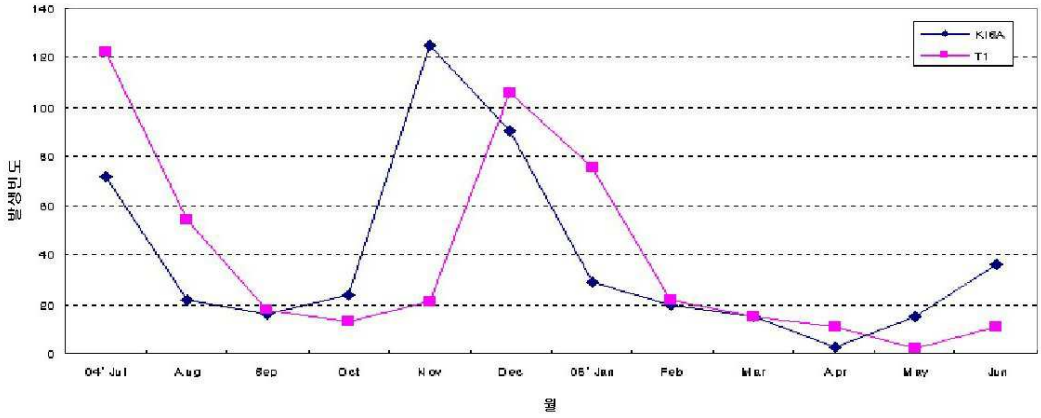
도면10

도면11

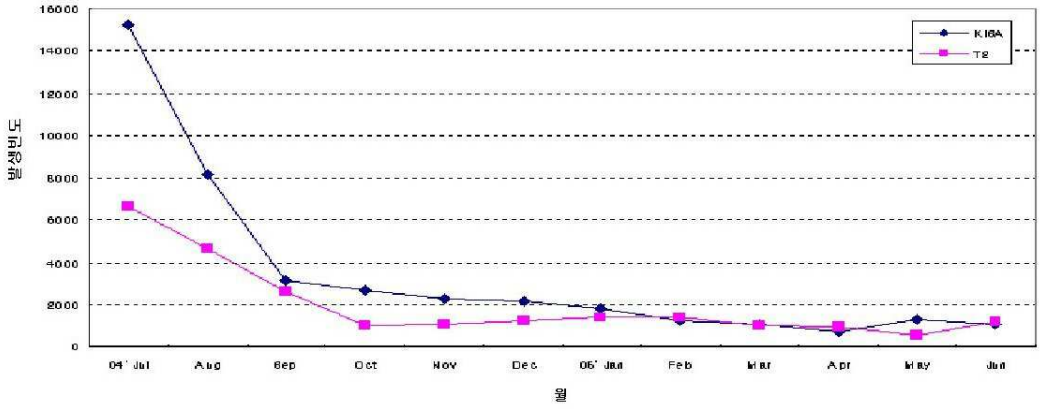


도면12

도면13a



도면13b


16000
14000
12000
10000
8000
6000
4000
2000
0
04' Jul
Aug
Sep
Oct
Nov
Dec
05' Jan
Feb
Mar
Apr
May
Jun
월
KIGAM
T2