情報保護學會論文誌
第19卷 第6號, 2009. 12

정보 자산 보안 위험 추정-정량적, 정성적 방법을 절충한 퍼지 숫자의 활용*

박 노 진,[1†] 이 동 훈[2‡]
[1]단국대학교, [2]고려대학교

Estimating Information Security Risk-Using Fuzzy Number Compromising Quantitative and Qualitative Methods*

Ro-Jin Pak,[1†] Dong-Hoon Lee[2‡]
[1]Dankook University, [2]Korea University

요 약

정보 자산 보안 관련 위험을 추정함에 있어 정성적인 방법과 정량적인 방법이 사용되고 있으나, 두 가지 방법 나름대로 장단점을 갖고 있다. 지나치게 서술적이고 추상적인 정성적 방법과 구체적이지만 자료의 부족으로 인한 정확한 계산이 어려운 정량적 방법의 한계를 어느 정도 극복한 절충된 방법의 개발이 요구된다고 하겠다. 본 논문은 절충의 방법으로서 퍼지 숫자를 이용하는 방법을 제시하고 분석의 예를 보였다. 퍼지 숫자를 이용함으로 자료의 부족함을 전문자의 의견이나 가능한 자료로 대체할 수 있고 위험을 구체적인 수치로 추정할 수 있음을 확인하였다. 가상의 시스템에 대하여 다양한 위협을 가정하여 모의실험을 하였고 시스템에 대한 예상 위험과 비예상 위험을 예측하는 방법을 구현하였다.

ABSTRACT

There have been two methods of estimating computer related security risk such as qualitative and quantitative methods which have distinctive advantages or disadvantages. The former is too narrative and somehow abstract to implement and the latter produces concrete result but needs lots of data, so that it is needed to develop a method overcoming such difficulties. It is advised to mix such two methods in a proper way depending on the conditions of a computer system. In this article, a concept of fuzzy number is employed on the way of mixing the two methods and provide a simple example using fuzzy numbers. Simulation was conducted for an assumed model system and it is demonstrated how to calculated expected and unexpected risk.

Keywords: Fuzzy number, loss, risk

I. 서 론

위험(Risk)은 손실(loss)의 기댓값 혹은 평균 손실로 정의된다[1]. 즉, 위험은 주어진 상황에서 가능한 모든 손실에 대한 개별 확률 값과 손실의 곱을 모두 합친 값이다. 예를 들어, 특정 웹서버에 일정 기간 총 다섯 차례의 침입이 있었고 100만원, 100만원, 200만원, 300만원, 0원의 손실이 발생했다고 하자. 이 경우 위험은

$$100\times 2/5+200\times 1/5+300\times 1/5+0\times 1/5=140$$

접수일(2009년 6월 29일), 수정일(2009년 9월 18일),
게재확정일(2009년 10월 30일)
* 본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었습니다.
† 주저자, rojin@dankook.ac.kr
‡ 교신저자, lee@paper.hankook.ac.kr

(만원)으로 계산된다. 그런데, 위와 같은 확률론적 위험의 정의는 침입의 과정에 따른 손실의 차이를 정확히 반영하지 못하는 문제가 있다. 보안과 관련된 위험은 침입에 대한 결과만큼 그 과정들의 구체적 결과들을 반영해야 하고 따라서 그 과정을 반영하는 다소 변형된 식이 요구 된다. 본 논문에서는 침입의 발생부터 손실이 발생하는 과정을 확률적으로 표현하고 침입으로 인한 훼손의 정도를 의미하는 훼손도라는 개념을 도입하여 위험을 계산하는 방법을 제시하였다. 위험을 계산해내는 과정에서 손실을 직접적으로 구할 수 없는 점을 인정하고 자산과 훼손도를 정성적으로 평가한 후 퍼지 숫자를 이용하여 정량적으로 변환하는 방법을 제안하였다. 논문은 제2장에서 먼저 위험분석에서의 정성적, 정량적 분석의 예와 장단점을 기술하고 퍼지 숫자의 개념을 소개하고 제3장에서 위험에 대한 확률적 표현과 훼손도를 소개하였다. 마지막으로 제4장에서 모의실험을 통해 위험을 계산하는 방법을 구현하였다.

II. 관련 연구

2.1 위험 분석 관련 연구

보안 위험 분석에 대한 연구들이 오랫동안 이루어져왔는데, 이성만 등[2]은 외국의 위험 분석 기법들을 정리하여 국내 기법 개발에 대한 동기를 부여하였다. 임채호 등[3]은 국내 컴퓨터 보안의 문제점을 파악하여 보안 관리자들이 체계적으로 보안 대책을 수립할 수 있는 위험 관리 기술 가이드라인을 제시하였다. 한국전산원은 많은 노력의 결과로 한국형 전산 시스템 보안을 위한 자동화 위험분석 도구인 HAWK를 개발하였다[4]. 한국형 분석도구가 개발된 이후 보안컨설팅에 대한 요구들이 다양해짐에 따라 국내환경에 맞는 위험분석 방법들이 제시되었다[5]. 국내 정보 인프라가 단기간에 거의 완벽한 수준에 이르도록 갖추어지고 새로운 보안 위협이 수시로 보고되는 상황에서 온라인 데이터를 이용한 보안 관리 체계가 제안되었다[6]. 다양한 연구가 이루어지면서 다소 공통된 방법에 대한 요구가 발생했고 이를 위해 표준 또는 지침에 대한 연구가 진행되었다[7]. 우리나라가 IT 강국으로 군림하고 있는 상황에서 위협의 주요 대상은 ISP(Internet Service Provider)이고 문호건 등[8]은 이를 위한 위험분석 시스템 설계 및 구현을 하였다.

아래 대표적인 세 가지 위험 계산의 방법을 점검하고 제 3 장에서 본 논문이 제안하는 계산법을 소개 하겠다

- 한국전산원에서 개발한 HAWK (Hankuk Risk Analysis Watch-out Kit)라는 한국형 보안 관리 도구의 지침서를 보면 '위험은 특정 위협이 취약성을 이용하여 자산을 공격하여 손상을 초래할 수 있는 잠재력이다. 위험은 두 가지 요소의 경험에 의해 특정지어 진다. 발생 가능 확률과 영향이다.'라고 정의되고 위험의 정량적 측도로서 연간손실액[ALE, Annual Loss Estimate=(위협의 기대분포)*(위협 발생 한 건당 손실크기)]을 사용하고 있는데, 현실적으로 연간 손실액 계산을 위한 자료가 부족한 것은 큰 장애물이 된다. 한편, 위험을 계산하는 과정에서 실제 손실의 발생에 앞서 존재하는 시스템에 대한 위협의 수준과 영향, 위협 대상의 상태가 반영 되지 않는다는 것이 보안 측면에서 다소 부족한 면이 있다[5, 9].
- 영국에서 개발된 위험 분석의 선구적인 도구인 CRAMM(the UK Government's preferred Risk Analysis and Management Method)은 '위험은 손실 또는 파손에 대한 기회 혹은 가능성으로 정의된다. 위험은 두 가지 구성 요소의 함수, 즉 원치 않는 사건에 대한 가능성과 그로 인한 효과/충격의 함수로 정의 하고 있으며 위험의 강도는 자산 가치, 위협의 정도, 취약성의 수준에 따라 정성적으로 일곱 층으로 나누어 측정하고 있다. 이 방법은 정량적 분석이 부족하다는 것이 한 가지 약점이라고 하겠다[10].
- 김성원 등[11]은 CRAMM과 같이 보안의 여러 요소를 감안하고 연간 손실액과 같은 정량적 수치로써 위험을 측정할 것이 요구 되는 상황에 맞도록 특정 자산 a에 부과될 위험을 다음과 같이 정의하고 있다;

$$Risk_a = A_{av} \times \sum(T_{ij} \times T_p) \times \sum(V_{ij} \times AP_a), \qquad (1)$$

여기서 A_{av}: 대상자산의 가치; V_{ij}: j번째 대상자산의 i번째 포트가 가지는 취약성 수준; AP_a: 기존통제(1 또는 0); T_{ij}: j번째 대상자산의 i번째 포트에 위협수준; T_p: 위협 T_{ij}의 가능성 (1 또는 0)으로 정의되기에 이른다. 기존 통제 및 가능성 분석 값은 단위 위협 및 취약성에 대하여 통제가 존재하는지의 여부, 위협이 실현 가능한지의 여부를 의미하는 0 혹은 1 값을 가진다. (0: 존재함, 1:존재하지 않음). 또한, 단위 자산에 대한 모든 위험을 더해진 값으로 상위 수준 위험만의 합으로 나누어 전체 자산의 상위수준 위험의 비율을 구하여 위험의 상태를 측정한다. 앞의 두 가지 방법(HAWK, CRAMM)에 비해 침입의 과정을 계산에 담

고 있는 장점이 있다.

위에서 언급한 것과 같이 위험분석에서 정성적 분석과 정량적 분석이 혼재되어 사용되고 있다. 최근 연구를 살펴보면 정보통신단체표준(TTAS)에 의한 위험분석 방법론에 취약성 분석 단계에서의 정량적 평가가 가능한 방법이 제시되었다[12]. 한편, 자산 분석에 이어 자산을 관리하는 조직의 특성을 반영한 보안 요소 가중치를 부여하여 관리 측면의 안정성을 확보하는 방법이 제시되었다[13]. 정성적 판단에 근거하여 모든 자산 중 상위수준의 위험의 비율로 정량화 하는 방법도 제안되었다[11]. 자산을 정성적으로 평가하되 자산 분류를 클래스화하고 객체지향 기법을 이용하여 모델링함으로 시스템의 위험을 평가하기도 한다[14]. 특별히 ISP(internet service provider) 네트워크의 정량적인 위험분석을 시도하였는데 직접적인 자산분석보다는 서비스별 총 가치로부터 자산별 가치를 역 산출하여 네트워크의 위험을 계산할 수도 있다[15].

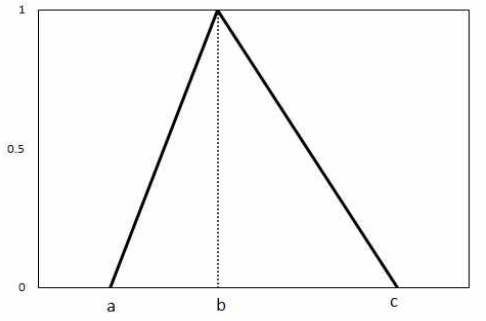
2.2 퍼지 연산의 기초 개념

퍼지집합이란 집합 내의 요소가 애매모호한 경계를 갖고 있어서, 특정 부분집합에 대한 소속과 비소속의 경계가 분명하지 않고 점진적으로 변화되는 요소들의 집합을 의미한다. 퍼지집합이론의 근본 개념은 하나의 요소가 퍼지집합에 부분적으로 소속될 수 있다는 점이다[16,17].

X를 요소 x가 속하는 공간이라고 하자. $A \subseteq X$인 집합 A로부터 [0,1]으로의 함수 $\mu_A(x)$가 정의되어, 만일 $\mu_A(x)=1$ 이면 x는 전적으로 A에 속하고, $\mu_A(x)=0$이면 x가 A에 전혀 속해 있지 않음을 의미하며, $0<\mu_A(x)<1$라면 x가 A에 부분적으로 속해 있다고 할 때, $\mu_A(x)$를 퍼지집합 A에 대한 소속함수라고 한다. $\mu_A(x)$의 값은 x가 A에 속한 소속 정도를 나타낸다. 즉, 소속함수의 값이 1에 가까우면, x가 A에 속한 소속 정도가 높다는 것을 의미한다. 퍼지 소속 함수의 대표적인 함수가 삼각형 소속 함수인데 식 (2)와 같이 정의되고 [그림 1]과 같은 모양을 갖는다.

$$\mu_A(x)=\begin{cases}(x-a)/(b-a), & a \le x \le b \\ (x-c)/(b-c), & b \le x \le c \\ 0 & \text{,그외}\end{cases} \quad (2)$$

또는 간편하게 $\mu_A(x)=(a;\ b;\ c),\ a \le b \le c$로 표시하고 $(a;\ b;\ c)$를 퍼지 숫자라고 한다. a와 c는 요소의 획득 가능 영역의 하한과 상한이고 b는 퍼지 숫자 A의



[그림 1] 삼각형 퍼지 소속함수

평균값이다. b는 소속 함수의 꼭짓점으로서 $\mu_A(x)=1$인 점이다. 따라서 삼각형 퍼지 숫자 $B=(a;b;c)$ 는 a와 c 사이에 존재하며 b에서 소속 가능성이 최대가 되는 퍼지수로 해석될 수 있다. 소속 함수 값은 a에서 b까지는 점차 증가되고 b에서 c까지는 감소된다.

예를 들어, 특정 시스템의 자산이 '약 500만원'이라면 개인에 따라 (480만; 500만; 510만)의 퍼지 숫자로서, 작게는 480만원, 많게는 510만원이며, 평균 500만원이라고 표현할 수 있다.

퍼지 숫자의 곱셈과 덧셈은 순서쌍의 대응하는 항끼리의 곱셈과 덧셈으로 정의된다.

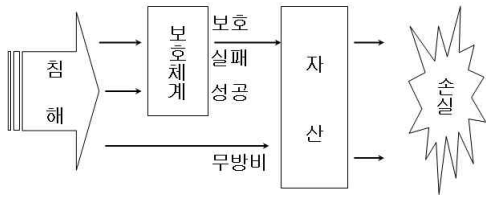
덧셈: $(a1;b1;c1)+(a2;b2;c2)=(a1+a2;b1+b2;c1+c2)$
곱셈: $(a1;b1;c1)\times(a2;b2;c2) \cong (a1\times a2;b1\times b2;c1\times c2)$

본 논문에서는 삼각형 퍼지 숫자를 사용하고자 한다. 많은 연구에서 삼각형의 퍼지 숫자를 사용하는 이유는 의사결정자들이 직관적으로 쉽게 사용할 수 있고 그 효과도 충분이 입증되었기 때문이다[17]. 또한, 본 논문에서 제시할 계산법을 수행하는 과정에서 삼각형 퍼지 소속 함수 보다 다소 진보된 형태의 퍼지 함수를 사용하는 경우 보다 정확한 계산 값을 얻을 수 있겠지만 계산이 종료된 시점의 마지막 퍼지 함수의 형태를 수학적으로 표현하는 것이 사실상 가능하지 않을 수도 있는 현실적인 이유로 인해 삼각형 퍼지 숫자를 사용하겠다.

III. 위험 계산을 위한 제안

3.1 손실 확률의 정의

침해가 발생하여 손실이 발생하는 과정은 [그림 2]



[그림 2] 자산에 대한 침해, 보호 그리고 손실의 관계

에 표현한 바와 같이 '침해 → 보호실패 → 손실발생'의 과정을 거친다고 하겠다.

확률의 정의적 관점에서 위험은 손실액, x, 와 대응하는 확률, $P(\text{손실}=x)$, 의 곱을 합한 결과로써

$$E[\text{손실}]=\sum xP(\text{손실}=x)$$

와 같이 표시할 수 있고 구체적 과정을 포함하는 위험을 계산하기 위해 아래와 같이 몇 차례의 조건부 확률을 반복적으로 사용하여 손실에 대한 확률을 침입시도와 보호실패에 의한 확률로 표시할 수 있겠다.

$$\begin{aligned}&P(\text{손실})\\&=P(\text{손실}\cap\text{침입시도})+P(\text{손실}\cap\text{침입시도없음})\\&=P(\text{손실}\cap\text{침입시도})+0\\&=P(\text{손실}\cap\text{보호성공}\cap\text{침입시도})\\&+P(\text{손실}\cap\text{보호실패}\cap\text{침입시도})\\&=0+P(\text{손실}\cap\text{보호실패}\cap\text{침입시도})\\&=P(\text{손실}|\text{보호실패}\cap\text{침입시도})\\&\times P(\text{보호실패}|\text{침입시도})P(\text{침입시도})\end{aligned}$$

따라서 본 논문은 위험을

$$\begin{aligned}\text{위험}=&\sum_{\text{모든 손실}}(\text{손실})P(\text{손실}|\text{보호실패}\cap\text{침입시도})\\&\times P(\text{보호실패}|\text{침입시도})P(\text{침입시도})\end{aligned}\quad(3)$$

와 같이 정의하고자 한다. 위 식에서 $P(\text{침입시도})$와 $P(\text{보호실패}|\text{침입시도})$는 과거 자료나 기술적인 측면에서 추정이 이루어져야 한다. 손실의 양이 확실치 않은 상황에서 $P(\text{손실}|\text{보호실패}\cap\text{침입시도})$의 계산은 다소 어려움이 있으며 자세한 내용은 다음 소단원에서 설명하겠다.

3.2 훼손도의 도입

보안의 목적이 침해가 시도되어 손실이 발생한 후보다는 발생하기 전에 그 의미가 더 크다고 보았을 때, 소 잃고 외양간 고치는 사후적 손실 분석 보다는 사전적 손실을 따지는 것이 요구된다고 하겠다. 즉, 손실을 결과적 손실과 예측적 손실로 나누어 본다면 보안과 관련된 손실은 할 수 만 있다면 예측적 손실을 구하는 것이 더 의미가 크다고 볼 수 있다. 사실 예측적 손실의 정량적 추론에 있어서 자료의 부족으로 인한 현실적 어려움이 많이 있다고는 하겠으나 한 가지 명확한 것은 미래에 발생할 손실의 크기는 자산의 크기에 비례한다고 할 수 있겠다. 즉, 손실=(비례상수)(자산의 크기)로 정의할 수 있겠다. 따라서 만일 비례상수가 1이라면 손실은 자산 그 자체의 크기가 되고 예컨대 손실에 자산의 완전한 복구비용 까지를 포함한다면 비례상수를 2로 정하면 된다고 하겠다.

이제, 식 (3)로 다시 돌아가면 실제로 계산상 어려움이 예상되는 것은 첫 번째 확률인 P(손실|보호실패∩침입시도) 인데, 특정한 자산에 특정한 침입이 시도되어 보호가 실패했을 때 발생하는 손실의 크기에 따른 확률을 의미한다. 그런데 손실의 확률을 계산하기 위해서는 손실의 크기를 알아야 하는데 이는 최종적으로 알고자 하는 것이므로 실제로는 그 계산 자체에 어려움이 있다. 본 논문에서는 다소 우회적 이지만 침해에 따른 자산의 훼손 정도를 '훼손도'라고 칭하고, 훼손도= (비례상수)·P(손실|보호실패∩침입시도) 로 정의하자. 정의상 훼손도는 확률 보다는 일종의 개연성(likelihood) 혹은 가능성으로 이해할 수 있겠다. 만일 비례상수가 1이면 훼손도는 손실 발생 확률 그 자체가 될 수도 있지만 비례상수가 P(손실|보호실패∩침입시도)의 역수가 되면 훼손도는 1이 되고 이는 침입이 시도되고 보호가 실패하여 자산에 손실 발생할 개연성이 확실히 존재한다고 이해할 수 있겠다. 무론 훼손도가 2라면 자산의 손실에 대한 개연성이 2배로 커지고 본래 상태로 회복시키기 위한 비용까지 손실에 포함하여 복구비용이 손실만큼 필요한 상태에 이르게 될 정도의 경우를 의미한다고 하겠다.

이제, 식 (3)으로부터

$$\begin{aligned}&\text{위험}=\\&\sum_{\text{모든 손실}}(\text{비례상수})(\text{자산})P(\text{손실}|\text{보호실패}\cap\text{침입시도})\\&\times P(\text{보호실패}\,|\text{침입시도})P(\text{침입시도})\\&=\sum_{\text{모든 손실}}(\text{자산})(\text{훼손도})P(\text{보호실패}\,|\text{침입시도})\\&\times P(\text{침입시도})\end{aligned}$$

라고 위험을 재 정의할 수 있겠다.

이제, 기존의 식 (1)과의 비교를 위해 영문을 사용하여 좀 더 구체적으로 위험을 정의하자. 자산 A의 가치를 A_{av}라 하고 i번째 형태의 침입의 j번째 시도의 결과를 T_{ij} 그리고 그에 따른 보안의 결과를 S_{ij} $(i=1,\ldots,a\;\; j=1,\ldots,a_i)$라고 하자.

[표 1] 자산에 대한 정량적, 정성적 정의에 대한 비교

	정량적	정성적
장점	- 위험이 금전적 영향에 따라 우선순위가 지정되며, 자산은 금전적 가치에 따라 우선순위가 지정됩니다. - 결과에 따라 보안 투자 수익(ROSI)에 의한 위험 관리를 촉진합니다. - 결과에 따라 관리 특정 용어(예, 특정 비율로 표시는 금액 가치 및 가능성)로 표시될 수 있습니다. - 조직이 경험을 쌓으면서 데이터 기록을 구축해 나감에 따라 정확성이 증가하는 경향이 있습니다.	- 위험 순위를 이해하고 가시적으로 볼 수 있습니다. - 합의에 도달하기가 더 쉽습니다. - 위협의 빈도를 수량화할 필요가 없습니다. - 반드시 자산의 금전적 가치를 파악할 필요가 없습니다. - 보안 또는 컴퓨터 전문가가 아닌 사람들이 보다 쉽게 참여할 수 있습니다.
단점	- 위험에 지정된 영향의 가치가 참가자의 주관적인 의견에 기반을 둡니다. - 신뢰할 만한 결과와 합의에 도달하는 프로세스에 많은 시간이 소요됩니다. - 계산법이 복잡하고 많은 시간이 소요됩니다. - 결과가 금액 용어로만 표현되어 기술적 지식이 없는 사람들이 해석하기 어려울 수 있습니다. - 프로세스는 전문 기술을 필요로 하여, 참가자가 이를 통해 쉽게 교육 받을 수 없습니다.	- 중요한 위험 간에 차이가 충분하지 않습니다. - 비용 이점 분석의 기초가 없기 때문에 제어 구현에 대한 투자를 정당화하기 어렵습니다. - 생성된 위험 관리 팀의 품질에 따라 결과가 달라집니다.

자산 A에 대한 침입으로 인한 훼손도의 수준 $(v_{ij},\ i=1,...,a;\ j=1,...,a_i)$이 주어지면, 자산의 위험은 다음과 같이 표현할 수 있다. 식 (3)와 위험의 정의에 의하여

$$Risk_A = \sum_{i=1}^{a}\sum_{j=1}^{a_i} A_{av} v_{ij} P(S_{ij}=0 | T_{ij}=1) P(T_{ij}=1)$$

가 된다. 여기서, $S_{ij}=0$는 보호실패, $T_{ij}=1$는 침입 시도를 의미한다. 한 단계 더 나아가 자산에 따른 중요도(가중치)를 정할 수 있다면 자산크기에 중요도의 크기를 곱하여 합한 가중합의 형태로 정의할 수 도 있겠다.

$$\text{위험} = \sum_{\substack{\text{대상자산에 대한}\\ \text{가능한 침입}}} (\text{중요도})(\text{자산크기})(\text{훼손도수준}) \times P(\text{차단실패} \mid \text{침입시도}) P(\text{침입시도})$$

3.3 자산 크기의 추정

자산의 크기는 정성적인 방법과 정량적인 방법으로 정하여 지는데, 두 가지 방법 모두 장단점을 갖고 있다. 과거에는 여러 가지 자료의 수집에 따른 제약으로 정성적 접근법이 보안 위험 관리의 대부분을 차지했으나, 최근 정량적 접근법이 요구되고 있다. 그런데, 정량적 방법을 엄격하게 따른 결과 가시적인 이점은 거의 없으면서 프로젝트를 어렵게 만들고 오랫동안 끌고 간다는 사실을 인정하는 전문가가 늘어남에 따라 정량적 방법이 우월하다는 생각에 변화가 나타나고 있다[18]. [표 1]에 정량적, 정성적 방법을 비교하여 정리하여 보았다. 우리나라의 표준연구소와 유사한 미국의 NIST(National Institute of Standards and Technology)는 Risk Management Guide for Information Technology Systems라는 컴퓨터 보안에 관한 가이드를 제공하고 있는데[23], NIST는 위험 평가 활동을 9단계로 규정하고 그 중 7번째 단계(risk determination)에서 [그림 3]과 같은 예제를 보여 주고 있다. [그림 3]에서 보듯이 먼저 위협이나 그 효과에 대하여 정성적 판단 (low, medium, high)을 하고 (1.0, 0.5, 0.1) 혹은 (10, 50, 100)을 대응시켜 정량화를 시도하고 있다. 이런 식의 시도는 편리하기는 하지만 정량화하는 과정에서 그 값들의 다양한 가능성을 배재시키고 있다.

한편, 본 논문에서는 정성적, 정량적 방법의 절충된 방법을 제안하고 그 방법의 핵심은 정성적 측정치를 퍼지연산을 통해 정량화하는 방법을 사용하는 것이다.

Table 3-6. Risk-Level Matrix

Threat Likelihood	Impact		
	Low (10)	*Medium* (50)	*High* (100)
High (1.0)	Low 10 X 1.0 = 10	Medium 50 X 1.0 = 50	High 100 X 1.0 = 100
Medium (0.5)	Low 10 X 0.5 = 5	Medium 50 X 0.5 = 25	Medium 100 X 0.5 = 50
Low (0.1)	Low 10 X 0.1 = 1	Low 50 X 0.1 = 5	Low 100 X 0.1 = 10

Risk Scale: High (>50 to 100); Medium (>10 to 50); Low (1 to 10)[8]

[그림 3] NIST가 제공하는 위험 산정 예

식 (3)에서 정의한 위험을 삼각형 퍼지 함수를 이용하여 재정의 한다면, 다음과 같다;

$$\text{위험퍼지수} = \sum_{\substack{i= \\ \text{침입형태}}} \sum_{\substack{j= \\ \text{침입순서}}} (A^a;A^b;A^c)(D_{ij}^a;D_{ij}^b;D_{ij}^c)(F_{ij})(I_{ij})$$
$$= (\sum_{ij} A^a D_{ij}^a F_{ij} I_{ij}; \sum_{ij} A^b D_{ij}^b F_{ij} I_{ij}; \sum_{ij} A^c D_{ij}^c F_{ij} I_{ij}),$$

여기서, A, D, F, I는 자산 가치, 훼손도, 보호 실패 확률, 침입확률을 하고 윗 첨자 a, b, c는 퍼지 숫자의 하한, 평균 그리고 상한을 의미한다.

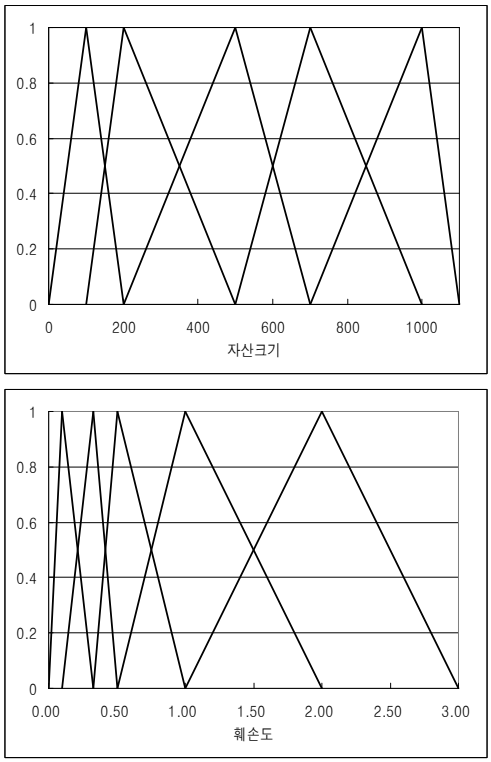
3.4 퍼지 숫자를 이용한 위험 추정의 예

특정 자산에 대한 가치와 그에 대한 훼손도를 예를 들어 [표 2]와 같이 정성적, 정량적으로 평가하였고 하자. 주어진 자산 가치와 훼손도의 범위를 [그림 4]와 같은 삼각형 퍼지 함수로 표현할 수 있겠다.

위험 퍼지수에 제안된 공식을 사용한다는 것은 가능한 침입에 대하여 [표 3]과 같은 계산을 각각 시행하여 합하는 것을 의미한다. 예를 들어, 중간 수준의

[표 2] 자산가치와 훼손도에 대한 퍼지 구간 예시

자산 가치 수준	자산가치의 금액 범위 (단위 만원)	훼손도 수준	훼손 범위
매우 높음	약1000 (700 - 1100)	매우 높음	전파 이상 (1 -3)
높음	약700 (500 - 1000)	높음	거의 전파 (1/2-2)
중간	약500 (200 - 700)	중간	반파 정도 (1/3-1)
낮음	약200 (100 - 500)	낮음	1/3 정도 (1/10-1/2)
매우 낮음	약100 (0 - 200)	매우 낮음	1/10 정도 (0-1/3)

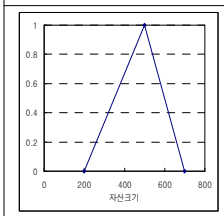
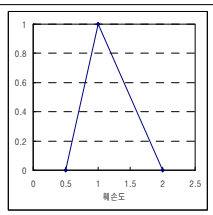
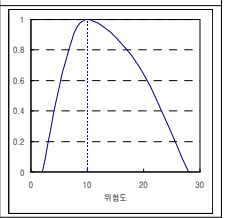


[그림 4] 자산가치와 훼손도에 대한 퍼지 삼각형 소속함수

자산에 대한 높은 수준의 훼손도를 가정하고 보호를 실패할 확률이 1/12, 침입이 일어날 확률이 1/5라고 한다면 위험도는 하한이 2, 평균이 10, 상한이 28인 이차 곡선과 유사한 소속 함수를 갖게 된다.

이제, 마지막으로 퍼지 숫자를 계산한 뒤 정량적 표현을 위해 역퍼지화 (de-fuzzy)를 시행하는데, 그 방법으로 최대법, 최대평균법, 무게중심법이 있다 (자세한 내용은 지면상 생략하겠다). 본 논문에서는 삼각 퍼지 소속 함수의 무게중심을 대표 수치로 삼는 무게 중

[표 3] 위험 퍼지 계산의 예

자산	×	훼손도	×	보호 실패확률	×	침입 확률	=	위험 퍼지수
	×		×	$(\frac{1}{10})$	×	$(\frac{1}{5})$	=	
(200;500;700)	×	(0.5; 1; 2)	×	0.1	×	0.2	≅	(2; 10; 28)

개념: 정보 보안 위험을 예측하기 위하여 손실에 관련된 정성적 정보를 퍼지 숫자를 이용하여 정량적 정보로 변환하여 '위험 = $\sum_{\text{가능한모든손실}}$ (손실의 확률)×(손실)'의 정의에 따라 계산한다.

1. 보안 실패와 침입 시도를 감안한 위험

$$\text{위험} = \sum_{\text{모든 손실}} (\text{손실})P(\text{손실}|\text{보호실패} \cap \text{침입시도}) \times P(\text{보호실패}|\text{침입시도})P(\text{침입시도})$$

2. 자산과 훼손도를 감안한 위험

$$\begin{aligned}\text{위험} &= \sum_{\text{모든 손실}} (\text{비례상수})(\text{자산})P(\text{손실}|\text{보호실패} \cap \text{침입시도}) \times P(\text{보호실패 }|\text{침입시도})P(\text{침입시도}) \\ &= \sum_{\text{모든 손실}} (\text{자산})(\text{훼손도})P(\text{보호실패 }|\text{침입시도}) \times P(\text{침입시도})\end{aligned}$$

3. 정성화된 자산과 훼손도에 대한 퍼지 숫자를 활용한 정량화

$$\begin{aligned}\text{위험퍼지수} &= \sum_{i=\text{침입형태}} \sum_{j=\text{침입순서}} (A^a; A^b; A^c)(D_{ij}^a; D_{ij}^b; D_{ij}^c)(F_{ij})(I_{ij}) \\ &= (\sum_{ij} A^a D_{ij}^a F_{ij} I_{ij}; \sum_{ij} A^b D_{ij}^b F_{ij} I_{ij}; \sum_{ij} A^c D_{ij}^c F_{ij} I_{ij}),\end{aligned}$$

여기서, A, D, F, I는 자산 가치, 훼손도, 보호 실패 확률, 침입확률을 하고 윗 첨자 a, b, c는 퍼지 숫자의 하한, 평균 그리고 상한을 의미한다.

4. 위험퍼지수의 탈퍼지화를 통한 위험계산

심법(center of gravity)을 사용하여 근사적으로 위험도 퍼지수의 실수 값으로 사용하겠다. 퍼지 숫자 $(a;b;c)$의 무게중심 값은

$$g = \int x\mu(x)dx / \int \mu(x)dx$$

로 정의되면 삼각형 소속 함수의 경우는

$$g = c - [(c-a)(c-b)/2]^{1/2} \text{ 또는}$$
$$g = a + [(c-a)(b-a)/2]^{1/2}$$

로 계산된다. 본 논문에서 [표 3]에 제시한 계산의 최종 결과인 위험 퍼지수는 정확히 삼각형 형태를 띠지 않으나 근사적으로 삼각형 퍼지수의 무게 중심값을 사용할 수 있겠다[16,17]. 전체적인 계산과정을 위 글상자에 정리하였다.

IV. 위험 계산의 예

어떤 자산에 대하여 전문가들의 가치 평가가 [표 2]와 같이 '매우 높음', '높음', '중간', '낮음', '매우 낮음'의 다섯 수준으로 평가가 가능하다고 하자. 인터넷 침해 사고 대응지원센터에서 발행한 인터넷 침해사고 동향 및 분석 월보(2008년 1월)에 근거하여 [표 4]와 같이 기관당 년 평균 침해건수와 침해비율을 계산하였다[19]. 훼손도는 2006 CSI/FBI Computer Crime and Security Survey의 결과[1]를 참조하여

[표 4] 인터넷 침해에 관련 통계치

구분	2007 총합	기관 평균	침해비율	훼손도
웜・바이러스	5,996	1.31	0.22	매우 높음
스팸 릴레이	11,668	2.55	0.42	낮음
피싱 경유지	1,095	0.24	0.04	중간
단순 침입시도	4,316	0.94	0.16	중간
기타 해킹	2,360	0.52	0.09	중간
홈페이지변조	2,293	0.50	0.08	낮음
총	27,728	6.06	1	

'웜・바이러스-매우 높음', '스팸릴레이-낮음', '피싱 경유지-중간', '단순침입시도-중간', '기타해킹-높음', '홈페이지변조-낮음'으로 대응하였고 퍼지 구간은 [표 2]와 같이 정하였다고 하자[20].

모의실험은 다음과 같이 실시하였다. 자산 가치가 '매우 높은' 수준부터 '매우 낮은' 수준을 각각 하나의 수준으로 갖는 모두 다섯 개의 자산으로 이루어진 시

1) 2006년 총 손실=$52,494,290; virus contamination(웜·바이러스)=$15,691,460, instant messaging misuse(스팸릴레이)=$291,510, phishing in which your organization was fraudulently represented by sender(피싱 경유지)=$647,510, system penetration by outsider(단순 침입 시도)=$758,000, bots within the organization+sabotage of data or networks(기타 해킹)=$1,183,700, web site defacement(홈페이지 변조)=$162,210.

[표 5] 모의실험의 계산 예

침입	웜 바이러스	스팸 릴레이	피싱 경유지	단순 침입	기타 해킹	홈페이지 변조	자산 하한	자산 중간	자산 상한	위험 하한	위험 중간	위험 상한	무게 중심
훼손도	매우높음	낮음	중간	중간	중간	낮음	700	1000	1100	10.08	899.2	1500.4	824.04
회수	2	1	0	0	1	0							

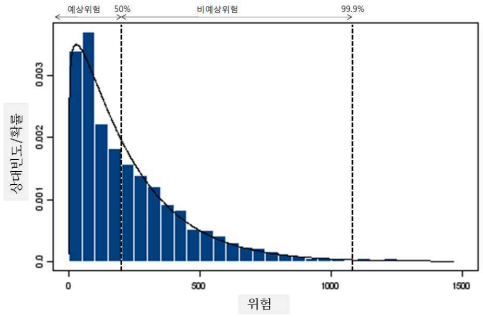
스템이 존재한다고 하자. 각 자산에 대하여 [표 4]에 제시된 여섯 가지 침해 비율에 따라 확률을 갖고 발생한다고 하고 이 과정을 각각의 자산에 대하여 500번 반복 시행하여 [표 3]에 제시된 방법으로 위험을 5자산×500번=총2500번계산하였다. 이제 모의실험을 구체적으로 기술하면 다음과 같다.

단계1: 각 침해에 대한 예상 발생수를 주어진 [표 4]에 있는 기관 평균을 모수로 하는 포아송 분포에서 생성한다. 예를 들어, 첫 번째 모의실험 의 경우에 웜바이러스 2회, 스팸릴레이 1회,…, 홈페이지변조 0회 발생한다고 하자[표 5].

단계2: 단계1에서 침해 조합이 주어지면 자산, 훼손도, 보호 실패 확률(주어진 자료가 이미 발생된 피해관련 자료임으로 보호실패 확률은 1로 함.), 침입 확률([표 4]의 침해 비율)에 따라 [표 3]과 같이 계산한다. 예를 들어, 첫 번째 모의실험의 경우 자산의 가치가 매우 높은 수준이고 훼손도는 앞서 언급한 것 같이 각 침해에 대한 CSI/FBI 평가에 준하여 계산하여 첫 번째 모의실험 자료의 경우 위험은 824.04로써 계산되는데 이는

(웜·바이러스 횟수)(자산 매우 높음)(웜·바이러스 훼손도 매우높음)(보안실패)(웜·바이러스 침해비율)
+(스팸릴레이 횟수)(자산 매우 높음)(스팸릴레이 훼손도 낮음)(보안실패)(스팸릴레이 침해비율)
+(단순침임 횟수)(자산 매우 높음)(단순침입 훼손도 낮음)(보안실패)(단순침입 침해비율)
+(피싱 경유지 횟수)(자산 매우 높음)(피싱 경유지 훼손도 낮음)(보안실패)(피싱 경유지 침해비율)
+(기타 해킹 횟수)(자산 매우 높음)(기타 해킹 훼손도 낮음)(보안실패)(기타 해킹 침해비율)
+(홈페이지 변조 횟수)(자산 매우 높음)(홈페이지 변조 훼손도 낮음)(보안실패)(홈페이지 변조 침해비율)

$$=(2)(700,1000,1100)(1,2,3)(1)(0.22)$$
$$+(1)(700,1000,1100)(1/10,1/3,1/2)(1)(0.42)$$
$$+0+0$$
$$+(1)(700,1000,1100)(1/3,1/2,1)(1)(0.09)$$
$$+0=(10.0,899.2,1500.4)\approx 824.0435491$$



[그림 5] 위험에 대한 히스토그램과 감마확률함수

와 같이 구해진다[표 5].

단계3: '매우 낮음'부터 '매우 높음'까지 각각 하나의 수준을 갖는 다섯 자산들에 대하여 단계2와 같은 계산을 500회씩 2500회의 모의실험을 실시하였다.

그 결과를 히스토그램으로 그리면 [그림 5]와 같고 그 분포는 감마분포($\alpha=1.447, \beta=0.005$)에 대하여 유의수준 0.01에서 부합함을 Kolmogorov-Smirnov 검정(KS통계량=0.0307, p-값=0.0179)을 통해 검정하였다[21]. 경영학의 위험 이론[22]에 근거하여 50 백분위수(누적 비율이 50%가 되는 수)에 해당하는 226.139만원을 예상위험의 상한으로 볼 수 있고 99.9 백분위수 (누적 비율이 99.9%가 되는 수)에 해당하는 1113.031만원을 비예상 위험의 상한으로 볼 수 있겠다. 226.139만원은 자산분류에서 낮은 수준에 해당하고 1113.031만원는 매우 높은 수준의 상한 (1100만원)을 약간 상회하는 값이 되겠다.

V. 결 론

본 논문에서는 침해의 과정과 보안 성공 여부를 모두 포함하는 손실 확률과 훼손도라는 새로운 개념을 도입하고 정성적 판단을 정량적으로 분석하는 과정에서 보다 많은 가능성을 내포하는 퍼지 숫자를 이용한 위험 계산법을 제안하였고 모의실험을 통해 그 활용법을 예시로 보였다. 기존의 분석들에 비하여 위험의 예

측값을 정확한 수치로 계산할 수 있다는 것이 큰 장점이라고 하겠다.

참 고 문 헌

[1] 이재창, 이용구, 수리통계학개론, 경문사, 2008년 11월.

[2] 이성만, 이필중, “해외의 보안위험분석 방법론 현황 및 분석,” 한국통신정보보호학회 종합학술발표회 논문집, pp. 288-302, 1994년 11월.

[3] 임채호, 박태완, 이경석, “전산망보안을 위한 위험관리기술지원서 개발 연구,” 한국통신정보보호학회 종합학술발표회논문집, pp. 316-323, 1994년 11월.

[4] 윤정원, 신순자, 김기수, 이병만, 송관호, “전산 시스템 보안을 위한 자동화 위험분석 도구(HAWK: Hankuk risk Analysis Watch-out Kit)의 개발에 관한 연구,” 한국통신정보보호학회 종합학술발표회논문집, pp. 65-74, 1996년 11월.

[5] 윤정원, 신순자, 이병만, “자동화 위험분석도구의 개발 및 적용과정을 통하여 분석한 국내 정보시스템 보안관리체계의 문제점,” 한국통신정보보호학회 종합학술발표회논문집, pp. 68-77, 1997년 11월.

[6] 주성진, 김종, “온라인 위험 가능성 평가를 통한 지속적인 보안관리 체계,” 한국정보보호학회 종합학술발표회논문집, pp. 544-547, 2002년 11월.

[7] 최상수, 방영환, 최성자, 이강수, “보안관리 및 위험분석을 위한 분류체계, 평가기준 및 평가스케일의 조사연구,” 정보보호학회지, 13(3), pp. 38-49, 2003년 6월.

[8] 문호건, 이종필, “ISP(Internet Service Provider)의 네트워크 보안 위험을 고려한 예상 자산손실 모델링,” 한국정보보호학회 동계학술대회 pp. 121-127, 2003년 12월.

[9] 한국전산원, 정보시스템 보안을 위한 위험 분석 실무 지침서, 2007년 12월.

[10] CCTA, CRAMM User's Guide (Version 2.0), The central computer and telecommunication agency, 1991.

[11] 김성원, 김휘영, 권영찬, 윤호상, 김철호, “네트워크 기반의 실시간 위험관리를 위한 위험분석 및 평가 방법연구,” 한국컴퓨터종합학술대회논문집, pp. 29-34, 2007년 6월.

[12] 박중길, “정량적 방법을 이용한 위험분석 방법론 연구,” 정보처리학회논문지C, 13(7), pp. 851-858, 2006년 12월.

[13] 조경식, “보안 위험분석을 위한 안정성 평가 시스템 설계 및 구현,” 한국컴퓨터정보학회논문지, 12(2), pp. 333-339, 2007년 5월.

[14] 이혁로, 안성진, “객체지향 자산분류모델을 이용한 위험분석에 관한 연구,” 한국인터넷정보학회 논문지, 9(4), pp. 79-84, 2008년 8월.

[15] 문호건, 최진기, 김형순, “ISP(Internet Service Provider)네트워크의 정량적인 위험분석을 위한 시스템 설계 및 구현,” 정보보호학회논문지, 14(2), pp. 101-111, 2004년 4월.

[16] 유동선, 이교원, 기초퍼지이론, 교우사, 1998년 3월.

[17] 이광현, 오길록, 퍼지이론과 응용 I, II, 홍릉과학출판사, 1997년 5월.

[18] Microsoft Corporation, 보안위험관리 가이드, http://www.microsoft.com/koera/technet, 2004.

[19] 한국정보보호진흥원, 인터넷 침해사고 동향 및 분석 월보, 2008년 11월.

[20] Computer Security Institute, “2006 CSI/FBI Computer Crime and Security Survey,” 2006.

[21] 송문섭, 비모수 통계학(S-LINK를 이용한), 자유아카데미, 2003년 8월.

[22] 김진호, 리스크의 이해, 경문사, 2005년 8월.

[23] G. Stoneburner, A. Goguen, and A. Feringa, “Risk Management Guide for Information Technology Systems,” NIST, 2002.

<著者紹介>

박 노 진(Ro Jin Pak) 정회원
1985년 2월: 서강대학교 수학과 졸업
1993년 12월: University of Texas 수학과 박사
1994년 3월 ~ 2001년 2월: 대전대학과 통계학과 조교수
2002년 3월 ~ 현재: 단국대학교 정보통계학과 교수
<관심분야> 데이터 마이닝, 정보보호

이 동 훈(Dong Hoon Lee) 종신회원
1983년 8월: 고려대학교 경제학과 졸업
1987년 12월: University of Oklahoma 전산학과 석사
1992년 5월: University of Oklahoma 전산학과 박사
1993년 3월 ~ 1997년 2월: 고려대학교 전산학과 조교수
1997년 3월 ~ 2001년 2월: 고려대학교 전산학과 부교수
2001년 3월 ~ 현재: 고려대학교 정보보호 대학원 교수
<관심분야> 암호이론, 암호프로토콜, USN 이론, 키 교환, 익명성 연구, PET 기술