확률 기반 보안 투자 수익율 모델을 위한 방법론

김도훈○ 김능회 인호
고려대학교
karmy○@korea.ac.kr, nunghoi@korea.ac.kr, hohin@korea.ac.kr

Methodology for Probability based Return on Security Investment Model

요 약

최근 들어 보안에 대한 피해의 급증으로 많은 기업들이 정보시스템에 막대한 기업 자원을 투입하고 있다. 뿐만 아니라, 다양한 ROI 평가 방법을 통해 투자대비 최대 수익률을 이끌어 내기 위해 끊임없이 시도하고 있다. 이는 많은 기업들이 정보시스템에 대한 투자를 유보하거나 정확한 평가 및 방법을 내리고 싶어한다는 증거이다. 이러한 기업의 추세는 최근 보안 분야에 있어 합리적인 보안 투자 방침을 세우는데 좋은 지침이 되어 지고 있다. 그러나, 지금의 대규모 정보 시스템 구축 및 웹기반의 인프라 환경에서는 적절한 보안 투자를 한다는 것은 쉽지 않다. 이러한 근본적인 이유 중 하나는 수익률을 측정하는 방법의 부재에서 찾을 수 있다. 따라서, 본 논문에서는 보안 분야에서 쓰이는 기존의 ROSI(Return On Security Investment)모델을 소개하고, 투자의 위험부담을 줄이기 위한 확률 기반의 개선된 ROSI(Probability based ROSI = PROSI) 방법론을 제안하고자 한다.

1. 서 론

최근 들어 전 세계적으로 해커, 컴퓨터 바이러스, 사이버테러가 급증하고 있으며, 그 유형도 다양해지고 있다. 이러한 이유로 많은 기업들이 정보시스템에 막대한 기업 자원을 투입하고 있다. 그러나 대부분의 경우 보안의식의 부재, 비용상의 문제점, 현실성등 다양한 이유로 인하여 보안에 대한 투자가 효익을 간과한 채 투자에 힘써온 것이 사실이다. 특히, 2003년 1월 25일 인터넷 대란[1] 이후에도 각 부처 및 민간에서 보안 분야에 관련한 다양한 연구가 진행되고 있으나, 현실성 있는 투자 및 수익률을 기대하기는 쉽지 않았다. 대부분의 정보시스템은 투자의 규모가 큰 반면 그 적용 범위와 요구사항이 복잡하여 투자의 결과가 실패로 돌아갈 가능성이 매우 크다. 이러한 근본적인 이유중 하나는 수익률을 측정하는 방법의 부재에서 찾을 수 있다.

그렇다면 어느 정도의 보안투자가 합리적일까? 항상 많은 돈과 시간의 투자만이 올바른 선택일까? 결국 이러한 의문은 대부분의 정책결정자들이 보안에 투자를 했을 때, 최대 수익률을 이끌어 내기 위해 무엇을 고려해야 할지를 고민하게 만드는 대목이다. 따라서 다음과 같은 조건을 생각해 볼 필요가 있다.

- 기업 운영시 부족한 보안 비용액은 얼마나 되는가?
- 생산력 증가 측면에 있어 보안의 부재는 얼마나 중요한 부분을 차지하는가?
- 심각한 보안의 허점을 야기 시키는 문제점은 무엇인가?
- 비용 효과가 가장 큰 솔루션은 무엇인가?
- 생산력 증가 측면에 있어 보안문제의 해결은 얼마나 중요한 것인가?

이 같은 질문에 대해 본질적으로 해결해야 할 다양한 문제점을 논해야 할 필요가 있다. 즉, 특정 보안 자산에 대해 피해를 입을 확률요인을 분석하고 그 값을 측정하여 초기 뿐 아니라 전반적인 보안비용을 추정하기 위해 다양한 분석 기법이 필요하다는 것이다. 다시말해, 대부분의 경영자들은 주로 정보시스템을 평가하기 위해 투자수익율(Return on Investment - 이하 ROI라 칭함) [2]을 사용하고 있다. 특히 보안의 관점에서 만들어진 ROI 기법을 보안 투자 수익율(Return on Security Investment - 이하 ROSI라 칭함) [3]이라 하며 이를 통해 보안 자산에 효과적으로 투자해야할 비용이 산정 될 수 있다는 것이다. 특히, 본 논문에서는 이러한 기존의 ROSI 모델을 개량하여 각종 위협에 노출된 보안자산에 시간적인 개념과 위협의 발생확률을 산정하여 미래의 투자액을 가늠해 볼 수 있는 확률기반의 ROSI를 제안 하고자 한다. 뿐만

아니라, 이러한 보안 자산간의 상관관계를 분석을 통해 향후 자산의 위협 노출 정도도 알 수가 있을 것이다.

2. 관련연구

2.1 Markov Chain

본 논문에서는 확률기반의 ROSI를 연산하기 위해 다음과 같은 확률 모델을 제안 하고자 한다. 본 논문에서 제시하는 확률기반의 ROSI를 설계하기 위해 적용되는 마르코프 체인에 대해 설명하고자 한다[5]. 먼저 상태의 변화에 대해 알아보면, "모델의 상태는 오로지 이전 상태들에만 의존한다." 이 가정을 마르코프 가정(markov assumption)이라 한다. 이는 복잡한 문제를 단순화 시켜 분석을 용이하게 할 수 있다. 또한, 마르코프 프로세스는 상태간 전이가 오로지 이전 n개의 상태에 의존하여 이루어지는 프로세스를 말한다. 이때, 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. 마르코프 프로세스를 $X(t)$라 하면 임의의 시간 $t_1 < t_2 < \cdots < t_k < t_{k+1}$ 에 대해 $X(t)$가 이산값이면,

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

$$= P[X(t_{k+1}) = x_{k+1} | X(t_k) = x_k]$$

이고 $X(t)$가 연속값이면

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

$$= P[a < X(t_{k+1}) \le b | X(t_k) = x_k$$

로 마르코프 성질이 기술된다. 위의 식에서 t_k는 현재, t_{k+1}은 미래, 그리고 $t_1, \cdots, t_{k-1}$은 과거의 시점이다. 마르코프 프로세서의 값이 이산값이면 마르코프 체인(markov chain)이라고 한다. 마르코프 체인은 t가 이산적이냐 연속적이냐에 따라 이산시간 마르코프 프로세스(discrete-time markov process), 연속시간 마르코프 프로세스(continuous-time markov chain)로 나뉜다. (표. 1)의 상태간 전이 확률은 위에서 제시한 날씨 모델을 위한 것으로 날씨의 상태간 전이하는 확률을 보여주고 있다.

		오늘날씨		
		맑음	구름	비
어제 날씨	맑음	0.5	0.25	0.25
	구름	0.375	0.125	0.5
	비	0.125	0.625	0.25

(표. 1) 상태간 전이 확률

즉, 마르코프 체인은 다음의 세 가지로 설명될 수 있는 모든 시스템을 말한다.

- 상태 집합(state) : 가능한 상태들
  (예) 맑음, 구름, 비
- π 벡터 : 시스템의 초기화 확률 벡터
- 상태 전이 행렬 : 각 상태간 전이 확률

3. 본 론

3.1 보안 측면을 고려한 ROSI

기업의 경쟁력을 알아보는 지표 중의 하나가 투자 대비 이익률이다. 투자 대비 이익률(ROI)은 기업이 어느 정도의 돈을 투자하여 얼마만큼의 이익을 올리는 지를 알아보는 지표이다. ROI 추정은, 주어진 제안서를 위한 비즈니스 사례를 계발하기 위해, 다른 접근방법과 함께 사용된다. 어떤 기업에 대한 전반적인 ROI는, 그 기업이 얼마나 잘 관리되고 있는지를 평가하는 방법으로 사용되기도 한다. 따라서 ROI를 계산하기 위한 일반식은 다음과 같다.

$$ROI = \frac{\text{기대수익} - \text{투자비용}}{\text{투자비용}} \quad (1)$$

즉, 이를 통하여 회사의 순이익을 측정할 수 있으며, 때때로 수익자산과 균등해 지기도 한다. 때문에 이러한 ROI를 극대화시키기 위해 다음과 같은 세 가지 조건을 고려해야 한다.

- 최소 비용
- 최대 수익
- 단위시간당 빠른 수익률

여기서 기본적인 ROI의 성질을 가지고 보안 측면을 고려한 ROSI에 대해 알아보기로 한다. ROSI 역시 ROI의 기본 모델에 충실하기 때문에 다음과 같이 표현 할 수 있다.

$$ROSI = \frac{(\text{위험노출} \times \text{위험감소율}) - \text{보안제품비용}}{\text{보안제품비용}} \quad (2)$$

이때, 위험노출(Risk Exposure)을 정량화하는 과정이 중요한데 다음과 같은 조건에 의해서 정의 할 수 있다.

- 단일손실노출(Single Loss Exposure-SLE)
- 연간발생율(Annual rate of occurrence-ARO)
- 연간손실노출(Annual Loss Exposure-ALE)

특히 SLE 또는 ARO는 측정을 위한 일반적인 방법은 없기 때문에, 실생활에서 보고되는 피해 사례 보고서를 통해 통계적 평균값으로 산출해야 한다. 이를 구성하는데 보험 산정 자료, 학술연구, 또는 단독 조사에 의해 만

들어 질 수 있다.

$$\text{위험노출} = ALE = SLE \times ARO \quad (3)$$

그러나 실제로 보안사고(SLE)에 의한 실 손실액을 구한다는 것은 쉽지가 않다. 즉, 네트워크 기반의 시스템 구조에서는 그 피해액을 산정한다는 것은 사실상 어렵기 때문이다. 뿐만 아니라 보안 사고는 즉시 그 피해현상이 하루만에 발생한다고 보기 힘들다. 그러나 미국에서는 최근 들어 FBI(미 연방 수사국)와 CSI(컴퓨터 보안회)에서 신뢰성이 높은 자료[4]들이 보고됨으로써 ROSI 연산에 있어 신뢰성을 높일 수 있다. 즉, 실제로 투자 규모 및 위협 발생에 대한 통계자료는 곧 ROSI를 계산하기 위한 중요한 요소(factor)로 작용한다는 것을 알 수 있다.

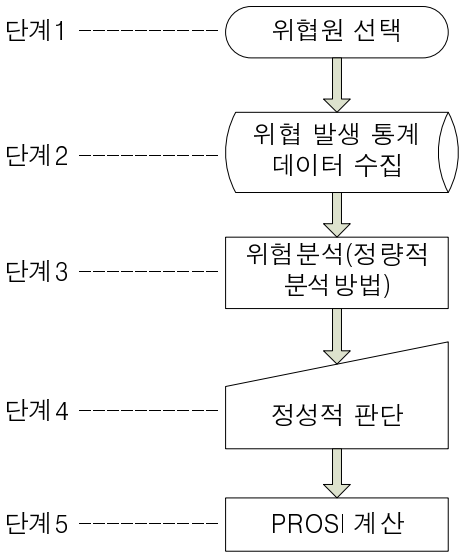
이 처럼 ROSI를 계산하기 위해 필요한 각각의 요소들에 대해서 알아보았다. 그러나 이러한 보안 측면에서의 ROSI 계산법은 과거 통계 데이터에 의존하고 있으며, 자산간의 확률적 상관관계 분석 즉, 정량적 분석이 실질적으로 부족하여 실제 ROSI값을 산정하기가 쉽지 않다. 따라서 본 논문에서는 자산간의 확률적 의존도를 알아 보기 위해 마르코프 가정(Markov Assumption)을 도입하여 자산간의 확률적 발생 의존도를 알아보고자 한다.

3.2 확률기반 ROSI 제안(PROSI)

본 논문에서는 위에서 설명한 마르코프 체인의 세 가지 구성요소를 정의함으로써 확률기반의 ROSI 모델(Probability based ROSI = PROSI)을 정의하였다. 이는 기존의 ROSI 모델에서의 '위험 노출' 부분을 마르코프 체인으로 재디자인(위험노출체인)하여 시간에 따른 위협원의 전이 확률을 분석하여 실제로 위험에 노출되어질 확률을 산정할 수 있다. 다음은 PROSI의 식을 유도해 낸 것이다.

$$PROSI = \frac{(\text{위험노출체인}(\$) \times \text{위험감소율}) - \text{보안제품비용}}{\text{보안제품비용}} \quad (4)$$

이는, 보안 자산의 위험 수위를 결정하고 특히 보안 자산의 위험노출의 연결 고리를 각종 보안사고와 연결지어 설명이 가능하다. (그림. 1)과 같이 PROSI를 산정하기 위한 프로세스이다.



(그림. 1) PROSI 프로세스

단계1. 위협원 설정. 보안 사고와 관련 있는 특정 위협원 2개 이상을 선택한다.

단계2. 위협발생 통계 데이터 수집. 상기 위협원을 잘 묘사하고 있는 통계 데이터를 수집하여 실제로 위험 분석을 위한 준비를 하는 단계라 할 수 있다.

단계3. 위험분석. 위험 분석 단계에서는 마르코프 체인으로 위협 발생을 실제로 예측하는 연산을 수행 하게 된다. 이는 krCERT의 보안 통계 월보[6]를 인용하여 다음과 같은 확률식과 행렬식으로 만들어 낸다.

$$P(S_1 \ S_2 \ \dots \ S_k \ \dots \ S_n) \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \dots & \dots & P_{ij} & \dots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix} = P'(S_1 \ S_2 \ \dots \ S_k \ \dots \ S_n) \quad (5)$$

$P(S_n)$ = 과거위협발생확률
P_{mn} = 위협전이확률행렬
$P'(S_n)$ = 위협원의 발생확률

$$P'(S_k) = \sum_{i=1}^{n} P(S_i) P_{ik} \quad (6)$$

$P'(S_k)$ = 위협원의 특정 발생확률

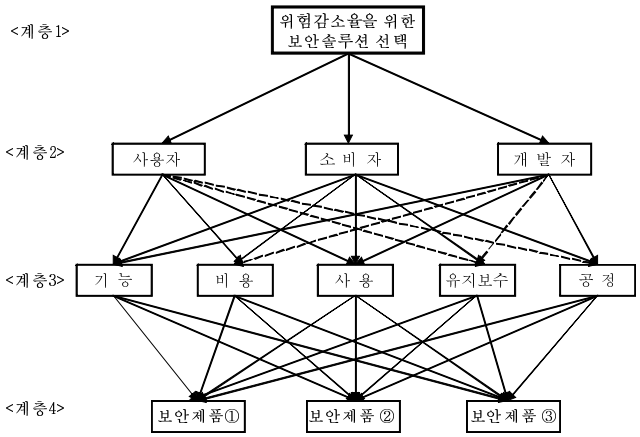
$$EF = \sum_{i=1}^{n} P(S_i) M(S_i) \quad (7)$$

$P'(S_n)$ = 위협원의 발생건수예측값

위 식을 통해 위험 노출 값을 산정 할 수 있다. 즉, 향후 수개월 후의 위협 발생건수가 예측이 되며, 이는 실제로 위험 노출값으로 인식하고 이를 비용으로 산정하는 단계를 거치게 된다.

단계4. 정성적 판단. 본 단계에서는 실질적인 보안 솔루션의 위험 감소율을 결정하는 단계로서, 전문가들의 판단과 실제적인 성능 평가를 위해 계층 분석적 의사결정법(AHP) [7]를 이용하여 최적의 보안 솔루션을 선택하게 된다. 이때, 선택되어질 확률이 위험 감소율로서 다음

(그림. 2)와 같은 단계별 프로세스를 거쳐서 측정된다.



(그림. 2) 보안제품의 선택을 위한 계층분석적 의사결정

단계5. PROSI 계산. 마지막 단계로서 위에 식(4)에서와 같이 선택 되어진 보안 솔루션의 비용과 함께 PROSI를 계산하는 단계가 되겠다.

4. 결 론

이 처럼 마르코프 체인을 이용하여 위험 노출 정도를 산정하고 계층적 의사결정법에 의해 위험 감소율을 측정함으로써 PROSI 모델을 계산할 수 있었다. 이는 기존의 통계적 방법에 의해서 산정된 값들 보다 정량적으로 해석할 수 있는 초석을 마련하였고, 확률분석을 통해 ROSI 값을 예측할 수도 있었다. 이는 단순히 ROSI값을 산정하는 것과는 달리 예측이라는 측면에서 향후 보안 자산의 가치가 보안 위협의 추이에 따라 동적으로 변하여도 능동적으로 대처 할 수 있음을 시사한다. 즉, 기존의 ROI 산정 기간은 짧게는 분기별, 장기적으로는 연간으로 산정하는 것이 일반적이나, 본 모델을 이용하면 월별로 분석이 가능하여 기존 모델 보다는 효율적인 투자 정책의 제안이 가능해 졌다.

향후 본 모델의 적용 가능성을 살피기 위해 실제 사례를 들어 실험을 하고, 실험의 검증을 위해 몬테카를로 시뮬레이션을 통해 검증을 하고자 한다. 또한, 정확도 및 다양한 적용 방안을 고려하기 위해 보안 위협의 관찰 정보를 동적으로 감시할 수 있는 은닉 마르코프 모델[8]을 이용한 PROSI 기법을 연구할 예정이며, 더불어 합리적인 보안 제품을 선택하기 위한 AHP 기법 또한, 개량한 모델을 추후 도입하고자 한다.

참고 문헌

[1] David Moore, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, Nicholas Weaver, “The Spread of the Sapphire/Slammer Worm”, 2003. 3.
http://www.caida.org/publications/papers/2003/sapphire/sapphire.html

[2] Greg McLean and Jason Brown, “Determining the ROI in IT security.”, 2003. 4.
http://www.cica.ca/index.cfm/ci_id/14138/la_id/1.htm

[3] Nicolas Sklavos, Panagiotis Souras, “Economic Models and Approaches in Information Security for Computer Networks”, *International Journal of Network Security*, Vol.2, No.1, PP.14--20, Jan. 2006.

[4] Jennifer Cincu, Robert Richardson, “Virus Attacks Named Leading Culprit of Financial Loss by U.S. Companies in 2006 CSI/FBI Computer Crime and Security Survey”, July 13, 2006.
http://www.gocsi.com/press/20060712.jhtml

[5] Wai-Ki Ching, Michael K. Ng, “Markov Chains Models, Algorithms and Applications.” Springer Science-Business Media(2006).

[6] Korea Information Security Agency (KISA) : Statistics and Analysis on Hacking and Virus.
http://www.krcert.or.kr

[7] 김도훈, 이택, 인호, “요구사항 협의모델을 위한 계층분석적 의사결정 방법, *KCSE(소프트웨어공학 학회)*, 8권1호, 2006. 2.

[8] A. Arnes, F. Valeur, G. Vigna, and R. Kemmerer, “Using Hidden Markov Models to Evaluate the Risks of Intrusions: System Architecture and Model Validation”, *Proceedings of the International Symposium on Recent Advances in Intrusion Detection (RAID),* Hamburg, Germany, September 2006.