기업 보안을 위한 융합보안 컴플라이언스 관리 모델에 관한 연구

김민수*

요 약

최근 지속적으로 발생하는 보안위협은 기업의 비즈니스 연속성을 저해할 뿐만 아니라 사회적·국가적 차원에서도 그 심각성이 높아지고 있는 실정이다. 이러한 보안위협은 기업과 국가 간 경쟁력이 심화 되면서 기업의 지적재산권 침해가 지속적으로 증가함에 따라, 기업들은 다양한 IT 컴플라이언스(compliance) 법제들에 대하여 의무적 준수와 더불어 엄격한 법적 책임을 부담하여야 한다. 따라서 본 연구에서는 기업의 능동적인 IT 컴플라이언스 활용을 위해 머신러닝(machine learning) 기술을 이용한 융합보안 컴플라이언스 관리 모델을 제안하고자 한다.

A Study on A Model of Convergence Security Compliance Management for Business Security

Minsu Kim*

ABSTRACT

Recently, increasing security threats are not only interfering with business continuity of companies but they are al so causing serious problems on social and national levels. As violation of intellectual property rights increases due to growing competition between different companies and countries, companies are now required to follow various IT co mpliance regulations, under relevant legal obligations. This study proposed a model of convergence security complian ce management by using machine learning, in order to help companies actively utilize IT compliance.

Key words : IT Compliance, Machine Learning, Machine Learning Algorism, Business Information Security, Convergence Security

접수일(2016년 9월 25일), 수정일(1차: 2016년 9월 29일), 게재확정일(2016년 9월 30일)

* 경기대학교 융합보안학과

1. 서 론

지식정보사회에서 기업의 비즈니스 프로세스는 IT 인프라(Information Technology Infra) 환경을 기반으로 디지털화가 가속화 되었고, 이는 기업과 산업의 경쟁력 강화를 위한 필수 요건으로써 국가차원에서 그 중요성을 강조하고 있다.

이에 따라 각 기업은 IT 인프라를 바탕으로 경쟁력 있는 비즈니스 프로세스를 통해 수많은 정보를 생산, 활용 및 저장, 관리 업무를 운영하게 된다. 하지만 비즈니스 프로세스 중에서 경쟁력 있는 중요 기술 및 정보의 경우 잠재적 위험에 노출될 가능성이 높기 때문에, 기업은 정보보호에 대한 인식이 높아지고 있는 실정이다.

이러한 위험요소는 기업과 국가 간 경쟁력이 심화되면서 기업의 지적재산권 침해가 지속적으로 증가함에 따라, 최근의 비즈니스 환경은 기업의 사회적 책임과 의무가 요구되어지고 있다. 이를 위해 기업들은 다양한 IT 컴플라이언스(compliance) 법제들에 대하여 의무적 준수와 더불어 엄격한 법적 책임을 부담하여야 한다. 이로 인해 기업은 비즈니스 환경에서 다양한 위험요소들에 대하여 법적, 기술적 영역에서 능동적 대응과 더불어 지속적 관리를 수행하여야 하는 전사적 정보보호 패러다임으로 전환되고 있다. 즉, 기업은 다양한 보안위협으로부터 비즈니스 연속성을 유지하기 위해 IT 컴플라이언스를 활용하여, 효과적인 위험관리와 기업의 가치를 높일 수 있는 도구로써 적극적인 활용이 병행되어야 한다.

따라서 본 연구에서는 능동적인 IT 컴플라이언스 활용을 위해 머신러닝(machine learning) 기술을 이용하여 각 기업의 정보보호 정책 및 보안 업무 규정을 분석하고, 기업에 필요한 보안 업무를 생성한 후 IT 컴플라이언스와 매핑하여 효율적인 보안 업무를 수행할 수 있는 융합보안 컴플라이언스 관리 모델을 제안하고자 한다.

2. 관련 연구

2.1 IT 컴플라이언스

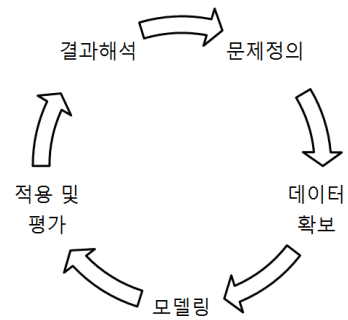
컴플라이언스(compliance)의 사전적 의미는 법규준수, 준법감시, 내부통제 등의 의미로 “사업 추진 과정에서 기업이 자발적으로 관련 법규를 준수하도록 하기 위한 일련의 시스템”이며, 컴플라이언스 프로그램(compliance program)이라는 표현을 사용하기도 한다[1].

그렇다면 기업에서의 컴플라이언스는 협의적 개념에서 대개 ‘주식회사에서의 법령준수체제’의 의미로 통용되지만, 광의적 개념에서는 법적 영역 이외에도 ‘사회적 가치’ 혹은 ‘사회규범’의 의미도 포함되어야 있다고 볼 수 있으며 위험관리의 한 수단으로써 그 역할을 수행할 수 있다[2][3][4]. IT 컴플라이언스는 기업의 비즈니스 프로세스를 수행하기 위해 IT 인프라 환경에서 새로운 규제와 환경 변화에 대한 능동적 대처를 지원하기 위한 IT 솔루션 및 프로세스 영역을 통합한 것이다. 그리고 IT 컴플라이언스의 범위를 살펴보면 정보에 대한 기밀성, 무결성, 가용성을 보장할 적정한 통제기능인 ‘정보 관리’, 개인정보 취급에 대한 규정인 ‘개인정보보호’, 기업의 주요 인프라를 효과적인 보호 및 대응에 대한 ‘정보보안’으로 구분할 수 있다.

2.2 Machine Learning

머신러닝이란 수집된 다양한 데이터 분석을 위해 알고리즘(algorism)을 바탕으로 학습(learn)을 통해 주어진 일에 대한 해결책 제시를 자동화하는 것을 의미하며[5], 최근에는 빅 데이터 기술 분야와의 융합으로 머신러닝기법이 고도화 되고 있다.

머신러닝을 통한 문제 해결 과정은 (그림 1)과 같이 5단계로 정의될 수 있다.



(그림 1) 머신러닝을 활용한 문제해결 과정[6]

머신러닝을 활용한 문제해결 과정은 적용 대상에 대한 문제를 분석 방법론을 활용하여 명확히 정의하여 의미 있는 데이터를 확보하고, 머신러닝 알고리즘 등을 활용하여 확보된 데이터에 대한 모델링을 수행하게 된다. 그 결과로 얻어진 모델을 실제 문제 해결에 적용 및 평가하고 마지막으로 적용 결과를 해석하여 문제에 대한 해결책과, 새롭게 추가되는 문제를 정의하고 문제 해결 과정을 반복하여 최종적으로 개선된 결과를 도출하는 일련의 과정을 수행하게 된다[6]. 확보된 데이터에 대한 모델링을 위해 사용되는 알고리즘은 다음과 같다.

2.3 Machine Learning Algorism

나이브 베이스(Naïve Bayes Algorithm)는 머신러닝에 사용되어지는 알고리즘으로 항목별 문서 분류와 사용 횟수를 특정 단어로부터 기록(log)하여 분류항목 간의 확률적 차이를 이용하여 높은 확률의 항목을 선택하는 방법으로 다음과 같은 수식을 사용하게 된다[7].

$$p(c|x) = \frac{p(x|c)p(c)}{p(x)} \quad (1)$$

정의된 입력 데이터(x), 분류항목(c)에서 베이스 규칙(Bayes′rule)의 조건부 확률(conditional probability)로 두개의 항목에 데이터를 분류하기 위한 방법으로 p1(x,y) → p2(x,y)이면 첫 번째 항목, p1(x,y) ← p2(x,y)이면 두 번째 항목에 속한다.

이 두 가지 규칙을 베이스 규칙에 적용하면 다음과 같은 수식을 사용하게 된다[8].

$$p(x_i|x,y) = \frac{p(x,y|c_i)p(c_i)}{p(x,y)} \quad (2)$$

조건부 확률의 두가지 규칙을 비교하여 x, y로 확인된 데이터를 기준으로 p(c1|x,y) → p(c2|x,y)이면 c1, p(c1|x,y) ← p(c2|x,y)이면 c2에 속한다.

또한 얼굴인식 등에 사용되는 패턴인식 알고리즘인 Knn(K-nearest neighbor) 알고리즘은 기존 데이터와 새로운 데이터를 비교하여 유사 거리 데이터를 분류하는 방법으로 유클리드 거리(Euclidean distance) 측정 방법을 사용하며 다음과 같은 수식을 사용하게 된다[9].

$$d = \sqrt{(xA_0 - xB_0)^2 + (xA_1 - xB_1)^2} \quad (3)$$

두 가지 데이터 xA 그룹과 xB 그룹 간의 정수형 수치 값을 가지고 거리 계산을 한다. 이때 데이터 정규화가 필요한데 0에서 1이나 1에서 1 등으로 정의하며 다음과 같은 수식을 사용하게 된다[10].

$$V = (old\,V - \min V)/(\max V - \min V) \quad (4)$$

정규화(V)는 집합 내 가장 작은 값(min V)와 가장 큰 값(max V)를 사용하게 된다.

그리고 의학 관련 연구, 마케팅 등의 분야에 사용되는 일반적인 의사결정 트리(Decision Tree Algorithm)은 부모 노드(parent node)에서 데이터 집합이 분류될 때 까지 반복하는 형태로, 양자화 방법(Quantitative) 방법을 적용하여 데이터 분할을 결정하고 정보이득(Information Gain)을 통해 노드들을 분할하게 된다.

분류항목에 대한 선택확률을 계산하기 위해 다음과 같은 수식을 사용하게 된다[11].

$$L = \log_2 p(x_i) \quad (5)$$

정보이득에 대한 측정은 다음 수식과 같이 엔트로피(Entropie) 수식을 사용하게 된다.

$$H = -\sum_{i=1}^{n}(x_i)\log_2 p(x_i) \quad (6)$$

마지막으로 연관규칙(association rules)과 같이 데이터 간의 관계에 있어 규칙을 찾는 알고리즘으로 Apriori 알고리즘을 사용하게 되며, 자주 발생하는 데이터 간의 집합이나, 데이터 간의 관계에 따른 연관규칙에 따라 관계 여부를 판단하게 된다[12].

데이터 집합 비율 G와 신뢰도 S는 연관규칙으로 정의되어 군집화 되어, 규칙에 만족하는 데이터들을 찾는데 이용되며, 다음과 같은 수식을 사용하게 된다[13].

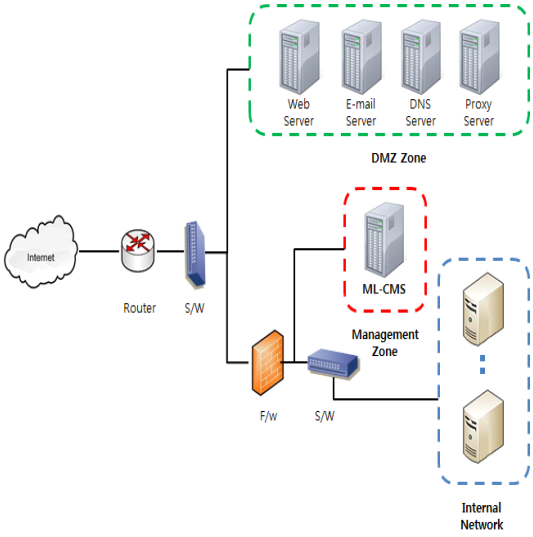
$$G = \frac{DG_i}{DG_{total}} \; {}^{IF A \rightarrow B} \; S = G(A, B / G(A) \quad (7)$$

3. 제안하는 방법

본 논문은 제한적인 연구로써 웹서비스를 제공하는 중소기업을 대상으로 해당되는 기업 보안 업무와 정보보호 통제항목을 매핑한 융합보안 컴플라이언스 관리 모델을 위한 기술적 방법론을 제안한다.

3.1 제안된 시스템 구성도

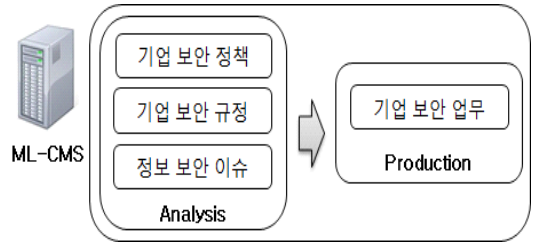
(그림 2)는 웹 서비스를 제공하는 중소기업의 모델을 바탕으로 머신러닝 기반 보안관리 시스템을 적용한 Test Bed 구성도를 나타낸 것이다. 외부 광역통신망을 통해 연결된 내부 네트워크 영역을 살펴보면, 웹서비스를 제공하는 DMZ(Demilitarized Zone) 구간은 일반적으로 웹(web), 이메일(e-mail), DNS(Domain Name Server), 프록시(proxy) 서버로 구성되어 진다. 그리고 기업의 내부망(internal network)을 보호하기 위해 방화벽(firewall)을 내부망 앞에 설치하고 컴플라이언스 관리 서버(ML-CMS, Machine Learning -Compliance Management Server)를 연결하게 된다.



(그림 2) 제안 시스템 구성도

3.2 컴플라이언스 관리 서버

컴플라이언스 관리 서버(ML-CMS)는 기업의 보안 정책 및 규정을 바탕으로 수행되어지는 보안 업무에 대한 요소들은 (그림 3)과 같이, 연관규칙(association rules)을 통해 기업보안정책, 기업 보안 규정, 정보보안 이슈에 대한 군집화를 통해 규칙에 만족하는 데이터들을 생산하게 된다.



(그림 3) 기업 보안 업무 생산

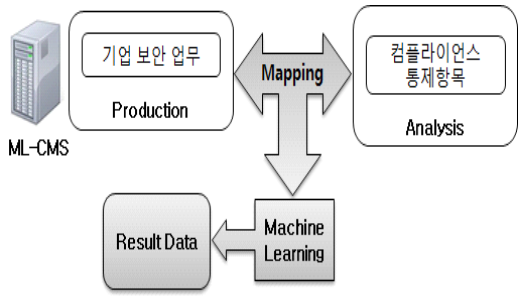
연관규칙에 의해 생산된 기업 보안 업무 데이터는 <표 1>과 같이 컴플라이언스 관리적, 물리적, 기술적 분야 통제항목과의 자동매핑(automatic mapping)을 수행하게 된다. 통제항목 수는 중소기업의 웹(Web) 서비스를 기준에서 필요한 항목인 관리적 분야 25개, 물리적 분야 6개, 기술적 분야 154개의 통제항목을 조정하였고, 이와 같이 해당 분야에서 우선순위로 적용되어야 할 통제항목을 기업 보안 정책 및 규정을 통해 데이터를 도출하여야 한다.

<표 1> 컴플라이언스 통제항목 분류[14]

구 분	통제항목수		비고
관리적 분야	25		정책 등
물리적 분야	6		접근통제
기술적 분야	유닉스	47	계정,DB 관리 등
	윈도우즈	49	
	보안장비	16	계정,패치 관리 등
	네트워크장비	14	
	웹(Web)	28	취약점

(그림 4)는 머신러닝 기술이 적용되어 도출된 데이

터를 기준으로 기업의 보안 업무를 수행할 수 있도록 한 관리 모델이다. 결과 데이터의 도출은 나이브 베이스 알고리즘(Naïve Bayes Algorithm)을 바탕으로, 기업 보안 업무와 컴플라이언스 통제항목의 조건부 확률(conditional probability)을 통해 베이스 규칙에 적용하여 그 결과 데이터를 도출하게 된다.



(그림 4) 매핑을 통한 머신러닝

4. 기존 모델과의 비교 검증

지금까지 살펴본 머신러닝 기반의 컴플라이언스 관리 모델과 기존의 컴플라이언스 관리 시스템과의 비교 검증을 실시하였다.

<표 2>를 살펴보면, 기존연구 보안 모델의 경우 기업의 보안환경보다는 통제항목에 대한 보안업무 수행을 우선시 하는 경향을 보이고 있다. 보안 위협에 대한 일련의 행위를 분석 및 대응하기 위해 증적 관리는 필수 요소로써 그 기능을 수행하고 있다.

하지만, 기존 모델은 보안 관리자의 수동적 관리 및 주관적 판단에 의해 이루어지기 때문에 실시간 대응 측면과 보안 요소 적용의 적절성 측면에 효율성을 극대화 하기 어렵다. 세부적인 사항으로 보안 환경 변화 적용의 경우 기존의 시그니처 기반과 같이 새로운 변화에 능동적적용이 어려우며, 이러한 환경 변화에 따른 보안 요소의 구성 또한 수동적으로 생성할 수 밖에 없다. 또한 새로운 보안 요소에 대한 업데이트 및 취약점 점검 항목에 대한 등급 변화도 수동적으로 진행되어 진다.

<표 2> 기존 모델과의 비교 검증

구 분	기존 연구 보안 모델	제안 연구 보안 모델
통제항목별 수행	가능	가능
증적 관리 기능	가능	가능
보안 환경 변화 적용	수동적 적용	능동적 적용
보안 요소 구성	수동적 생성	능동적 생성
보안 요소 업데이트	수동적 업데이트	능동적 업데이트

5. 결 론

기업은 비즈니스 연속성과 경쟁력 강화를 위해 IT 인프라를 바탕으로 수많은 정보를 생산, 활용 및 저장, 관리 업무를 운영하게 된다. 하지만 기업과 국가간 경쟁력이 심화 되면서 기업의 지적재산권 침해가 지속적으로 증가함에 따라, 최근의 비즈니스 환경은 기업의 사회적 책임과 의무가 요구되어지고 있으며 이를 위해 기업들은 다양한 IT 컴플라이언스(compliance) 법제들에 대하여 의무적 준수 및 엄격한 법적 책임을 부담하여야 한다.

이를 위해 본 연구에서는 능동적인 IT 컴플라이언스 활용을 위해 머신러닝(machine learning) 기술을 이용하여 각 기업의 정보보호 정책 및 보안 업무 규정을 분석하고, 기업에 필요한 보안 업무를 생성한 후 IT 컴플라이언스와 매핑하여 효율적인 보안 업무를 수행할 수 있는 융합보안 컴플라이언스 관리 모델을 제안하였다. 제안된 관리 모델은 머신러닝 알고리즘을 이용하여 도출된 데이터를 기준으로 기업의 보안 업무를 수행할 수 있도록 한 모델로써, 기존 관리 모델과의 비교에서 살펴본 '보안환경 변화 적용', '보안 요소 구성', '보안 요소 업데이트' 항목에서 능동적으로 대처하면서 실시간 대응 측면과 보안 요소 적용의 적절성 면에서 그 효율성을 극대화 할 수 있을 것이라

생각된다. 이에, 향후 융합보안환경에서의 컴플라이언스 관리 모델을 구축하고 관리하기 위한 제도적 마련과 더불어 정보보안 분야에도 인공지능(AI, Artificial Intelligence)을 활용한 능동형 모델 개발을 위한 심도 있는 논의와 연구가 필요하다.

참고문헌

[1] https://ko.wikipedia.org/wiki/%EC%BB%B4%ED%94%8C%EB%9D%BC%EC%9D%B4%EC%96%B8%EC%8A%A4.

[2] 落合誠一, “企業コンプライァンス確立の意義,” ジュリスト1438号, p.12, 2012.

[3] 松木和道, “企業コンプライァンスの現實,” ジュリスト1438号, p.18, 2012.

[4] 육태우, “기업 컴플라이언스의 발전 및 집행주체”, 강원법학, Vol.47, p.311, 2016.

[5] 최도현 외, “빅데이터 환경에서 사용자 거래 성향 분석을 위한 머신러닝 응용 기법”, 한국정보통신학회논문지, Vol.19, No.10, p.2233, 2016.

[6] 조성준 외, “머신러닝(인공지능)의 산업 응용”, Industrial Engineering Magazine, Vol.23, No.2, pp.34-38, 2016.

[7] McCallum, Andrew, and Kamal Nigam, “ comparison of event models for naive bayes text classification” AAAI-98 workshop on learning for text categorization, 1998.

[7] Winkler, Robert L, “Introduction to Bayesian inference and decision” Vol. 15, No. 4, pp. 938-939, 1973.

[9] Altman, Naomi S. “n introduction to kernel and nearest-neighbor nonparametric regression” The American Statistician, Vol. 46, No. 3, pp. 175-185, 1992.

[10] Hastie, Trevor, and Rolbert Tibshirani. “Discriminant adaptive nearest neighbor classif-ication” Pattern Analysis and Machine Intelli-gence, IEEE Transactions on, Vol. 18, No. 6, pp. 607-616, 1996.

[11] Choi-Jonghu, et al. “pplication of Data Mning Decision Tree” Statistic Korea, Journal of The Korean Official Statistics, Vol. 4, No. 1, pp. 61-83, 1999.

[12] Agrawal, Rakesh, and Ramakrishnan Srikant. “Fast algorithms for mining association rules” Proc. 20th int.conf. very large data bases, VLDB, Vol. 1215, pp. 487-499, 1994.

[13] Perego, Raffaele, Salvatore Orlando, and P. Palmerini. “nhancing the apriori algorithm for frequent set counting” Data Warehousing and Knowledge Discovery, Springer Berlin Heidelberg, pp. 71-82, 2001.

[14] 주요정보통신기반시설 취약점 분석·평가 기준, 행정자치부.

[저자소개]

김 민 수 (Minsu Kim)

2004년 컴퓨터공학사
2012년 경호안전학석사
2015년 산업보안학박사
현 재 경기대학교 융합보안학과 초빙교수

email : fortcom@hanmail.net