인공신경망 회로를 이용한 RFID 기반 유비쿼터스 화물 추적시스템 동작 시 EDI 데이터 보안 대책에 관한 연구

박필구*, 유혁*
*고려대학교 컴퓨터정보통신대학원
e-mail : genius8989@korea.ac.kr, hxy@korea.ac.kr

A Study of Security Method of EDI Data in Progress of Ubiquitous Cargo Tracing System based on RFID Technology by using a Artificial Neural Network

Pil-Goo Park*, Prof. Chuck Yoo *
*Graduate School of Computer & Information Communication, Korea University

요 약

RFID 를 이용한 화물추적시스템은 물류분야의 특성상 서로 다른 소속의 이 기종 간의 데이터의 인터페이스로 화물의 흐름을 체계화한다. 국내뿐 아니라 국제적으로도 여러 종류의 데이터를 인터페이스하고 있으며, 이 데이터들은 EDI 표준을 이용하여 다양한 환경의 시스템으로 인터페이스 되어 적용되고 있다. 하나의 물류흐름을 만들기 위하여 RFID 를 이용한 데이터의 인터페이스가 이루어지다 보니 다양한 보안상의 문제를 유발시키고 있는 실정이다. 본 논문에서는 인공신경망 회로를 이용하여 이 기종 간의 EDI 데이터 인터페이스 시 발생할 수 있는 보안상의 취약점을 미리 파악하여 적절한 조치를 취할 수 있도록 방향을 제시하였다.

1. 서론

많은 분야에서 유비쿼터스의 개념이 도입되며 대한민국의 특수성을 이용한 네트워크 인프라를 바탕으로 여러 분야에서 급속도로 발전을 해오고 있다. 그 중에서도 현재 기 개발 되어있으며 활발하게 움직이고 있는 분야가 물류 분야이다. 물류 분야에서도 화물추적을 위한 시스템은 RFID 기반 유비쿼터스 시스템과 더욱 가까이 와 있는 것이 사실이다. 이런 기반기술을 100% 이용하기 위해서는 믿을 수 있는 데이터 인터페이스가 절실히 요구되고 있다. 실시간에 가까운 데이터를 요구하는 화물추적시스템에서는 데이터가 곧 경쟁력이라고 말 할 수 있을 정도로 정확한 정보를 요구한다. 또한 여러 지역에서 이 기종간의 데이터를 통합 함으로 인해 원치 않은 곳에서 데이터 보안상의 문제점들이 나타나게 된다.

기존의 환경에서 RFID 를 이용한 화물 추적 시스템을 원활히 사용하기 위해서는 크게 두 가지 문제점이 해결 되어야 한다. 첫 번째로는 기업간 이 기종간의 시스템 환경에서 데이터를 표준화하는 문제이다. 표준화된 데이터 형식을 사용하지 않음으로 인해 기업 확장이나 지점 확장에 저해 요인이 될 수 있다., 또한 기업간 데이터 교환 시 장애 요인으로 작용하고 있다. 두 번째로는 기업간의 데이터 신뢰성 확보이다. 다양한 경로의 데이터를 수신하다 보니, 데이터 중에는 실제로 시스템에 영향을 줄 수 있는 오류를 내포한 데이터도 발생을 하고, 해커들의 의해 데이터가 변조가 일어나는 경우도 있다. 이와 같이 다양한 데이터 보안상의 문제를 내포하고 있다.

본 논문에서는 상기에 언급된 문제점을 바탕으로 기업간 국내 지점간의 RFID 를 이용하여 화물 추적 시스템 구축 시 표준화를 위하여 국제 전자 문서 양식인 EDI(Electronic Data Interchange)를 이용하였다. 더불어 데이터 신뢰성 확보를 위환 통합된 시스템 구조를 제안하고, 데이터 전송 시 발생 될 수 있는 다양한 데이터 보안 상의 문제점을 시스템에서 분류하여 대처 할 수 있는 방안을 실험을 통해 고찰하였다.

본 논문의 구성은 다음과 같다. 2 장에서는 관련분야의 개념과 연구분야에 대해서 살펴보고 3 장에서는 제안된 시스템 통합구조와 전송되어지는 데이터 포맷 그리고 제안된 시스템 내부의 작동원리와 검증과정을 기술한다. 4 장에서는 실험과 분석을 통해 문제점을 고찰한다. 마지막으로 5 장에서는 본 논문의 결론을 맺고, 향후 연구과제를 제시한다.

2. 관련연구

본 논문에서 사용된 RFID 기반 화물 추적시스템에서의 EDI 데이터 보안 관한 연구는 RFID, EDI 와 데이터 보안에 관한 개별 연구가 많이 진행되고 있으나, 실제 특정 시스템에 적용된 통합된 연구는 그 사례가 많지 않다. 개별 정보에 관한 내용을 살펴보면 다음과 같다.

2.1 EDI(Electronic Data Interchange)

전자상거래의 한 형태이며, 기업간 거래에 관한 데이터와 문서를 표준화하여 컴퓨터 통신망으로 거래 당사자가 직접 전송·수신하는 정보전달 시스템이다. 전자문서교환에서 사용하는 국제적인 통신표준은 X.435, ANSI 의 X12 등 3~4 가지 중 X.435 가 표준으로 사용되었다가, 현재 국제연합이 중심이 되어 만든 UN/EDIFACT 의 표준을 따르고 있다. 본 논문에서도 UN/EDIFACT 의 표준 문서인 CODECO(Container Gate-In/Gate-Out)문서를 이용하게 된다.

2.2 비정상탐지(Anomaly Detection)

일종의 비정상적인 접근 행위를 나타내는 표현으로 시스템 내에서 인지하지 못하는 행위가 나타났을 때 탐지하는 방법을 의미한다. 탐지하는 방법으로는 통계적인 방법, 비정상적인 행위 측정 방법 등 여러 가지 방법들이 있다. 그 대표적인 방법이 인공 신경망 기법이다. 반복되는 학습을 통해 자주 발생되는 문제에 대해서는 가중치를 부여하고, 패턴을 인식함으로써 비정상을 탐지하는 방법이다.

3. 제안 시스템 통합 구조

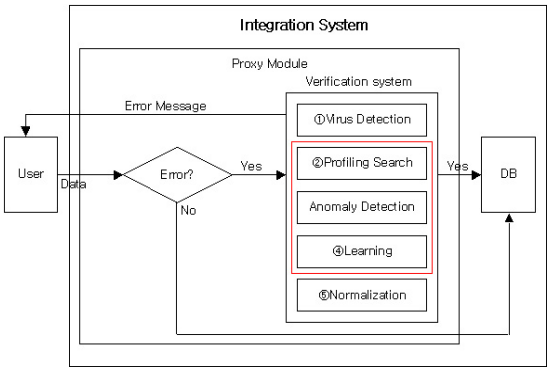
3.1 통합시스템 개요

정보의 데이터 타입, 허용 가능 문자 셋, 변수 중복의 여부, 숫자의 범위 등 기본 적인 검증에서부터 악의적인 해킹이나 바이러스에 대한 전문적인 검증에 이르기까지 다양한 접근의 검증이 필요하다. 본 논문에서는 부적절한 정보가 전달되는 것을 검증하기 위하여 데이터의 무결성을 인공신경망의 패턴인식으로 분석하고 이를 프로파일 기법에 적용하려 한다.

프로파일 기법은 기본적인 정보가 정확히 전달되어 이상을 탐지하는 기법으로 이상 상태 탐지를 위한 방법으로는 매우 적합한 방법이다. 더불어 프로파일러를 시스템 전송 전의 프락시 형태로 구성하여 프락시를 경유하여 전송되는 입력 패턴을 학습하여, 반복 학습을 통한 정확한 데이터 검증에 사용되도록 하였다. 이 시스템을 적용하게 되면 공개되지 않은 어플리케이션에 대한 필터링이 가능하다는 장점도 내포하고 있다. 설계 될 시스템의 전체적인 구조는 다음과 같다.

3.2 통합시스템 동작방식

(그림 1)에서와 같이 User(Product & Other Company & Container and RFID Reader)를 통해 수집된 정보는 통합 시스템(Integration System)이 구축되어 있는 대상 서버로 데이터를 전송한다. 전송된 데이터는 일차적으로 프락시 모듈을 통해 최소한의 정보 검증을 한다. 최소한의 정보 검증을 통해 에러가 없는 것으로 검증을 마치면 데이터 정보 조합을 위해 해당 데이터베이스로 전송 되고 이상이 발견된 데이터는 검증시스템(Verification system)에 전달이 되어 검증을 시작한다. 검증 시스템에서는 바이러스 탐지(Virus Detection), 프로파일 검색(Profiling Search), 비정상 탐지(Anomaly Detection), 학습(Learning), 정규화(Normalization) 5 단계의 검증 과정을 거쳐 검증을 하게 된다. 우선적으로 데이터는 바이러스 탐지 과정을 통해 해당 데이터가 악의적인 정보인지를 판단하게 된다. 악의적인 정보로 판별이 되면 검증 시스템에서는 더 이상의 검증을 하지 않고 에러 메시지를 발생 시키게 된다.



(그림 1) 통합 시스템

반대로 정상적인 데이터로 판단이 되면 다음단계를 진행하게 된다. 기존의 데이터를 이용하여 프로파일된 정보가 해당 데이터가 가진 오류사항이 없는지 검증하게 된다. 해당 프로파일 된 정보는 레코드로 구성된 테이블 형태로 정보를 가지고 있으며, 수행 속도를 고려하여 해당 데이터는 인덱싱이 되어 있는 정보로 이루어져 있다. 또한 해당 테이블의 정보는 국내뿐이 아닌 다양한 언어를 지원하기 위해 UTF8 형태의 유니코드를 지원하는 형태로 이루어져있다.

```
UNH+071008161349+CODECO:D:95B:UN'
BGM+34::UN:CONTAINER GATE-IN/GATE-OUT REPORT+1+9+NA'
TDT+20+TST+++LTD:172:20'
NAD+MS+DNATC:ZZZ'
NAD+MR+DNAL:ZZZ'
EQD+CN+TEST4001718+4310:102:5++2+5'
RFF+BN:AOD'
TMD+12'
DTM+7:20071008180 5:203'
LOC+5+KRZZZ:139:5:INCHON'
LOC+60+KRINC:139:5:ODCY'
LOC+9+KRINC:139:5:INCHON'
LOC+165+ZZZ:139:5:ODCY'
SEL+056043+CA'
TDT+1++3++LTD:172:20+++6168:146'
TDT+20+SC10119+++DNA:172:20'
NAD+MS+DNATC:ZZZ'
NAD+MR+DNAL:ZZZ'
EQD+CN+TEST3005508+4510:102:5++3+4'
RFF+BM:XAMOBSV01192001'
TMD+13'
DTM+7:20071008133 5:203'
LOC+5+KRPTK:139:5:WAREHOUSE1'
LOC+60+KRPTK:139:5:PCT'
LOC+11+KRINC:139:5:INCHON'
LOC+165+ZZZ:139:5:PCT'
SEL+11111+CA'
TDT+1++3++LTD:172:20+++8605:146'
NAD+CA+CONTAINER SHIP'
CNT+16:5'
UNT+31+071008161349'
```

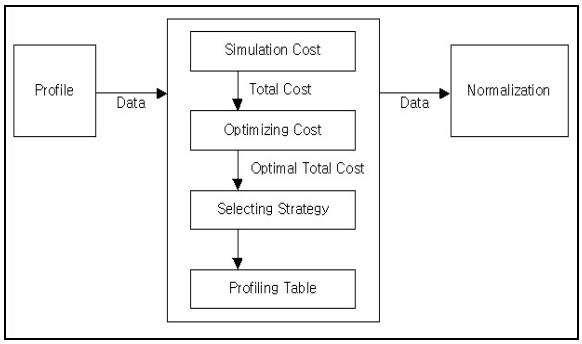
(그림 2) Container Gate-In/Gate-Out EDI Data

마지막으로 데이터 정규화(Normalization) 과정을 거치게 된다. 정규화 과정은 EDI 포맷 형식으로 수신된 데이터로 EDI 데이터의 단위 그룹 정보 형태인 세그먼트(Segment)의 맞는 데이터를 추출하게 된다. 여기서 사용되는 EDI 포맷은 CODECO(Container Gate-In / Gate-Out Report) 문서를 사용하게 되고 버전(Version)

은 D95B 표준은 UN 표준에 맞추어서 사용한다. 화물 추적 시스템의 특성상 국내 화물 추적 시스템 만이 아닌 해외 화물 추적 시스템까지 포괄하기 위해 UN 표준을 사용한다. 전송되어 온 EDI 데이터는 아래 (그림 2)와 같은 형태로 이루어져 있다.

3.3 가중치와 프로파일을 이용한 보안 기법

(그림 1) 통합 시스템에서의 검증 시스템 중에서 ④학습(Learning) 부분은 생명체의 신경망을 이용하여 만든 인공 신경망을 이용하여 데이터의 대한 검증을 실행하고 계속된 학습을 통해 미래에 발생할 수 있는 문제에 대해 예측한다.



(그림 3) 가중치와 프로파일 모델

학습을 통해 작성된 데이터들은 분석을 통해 현재 시스템에 적용된 수행탐지 기법뿐 만이 아닌 인식하지 못했던 새로운 침입 탐지 패턴까지 발견하여 보다 향상된 침입 탐지 시스템을 구현한다.

새로 발견된 문제에 대해서는 반복적인 학습 과정을 통해 그 결과에 따라 가중치를 부여하여 학습하게 함으로써 새로운 패턴의 내용을 프로파일 테이블에 계속적으로 저장한다. 프로파일 된 정보가 저장될 테이블 구조는 다음과 같다. 프로파일 테이블은 크게 3가지로 구성하였다. 첫 번째로 프로파일 들의 마스터 정보를 저장 할 (A) Profiling Master 테이블, 두 번째로 EDI 전송 시 발생되는 세그먼트 단위 에러 정보를 저장 할 (B) EDI Segment Error Detail 테이블, 마지막으로 악의적인 목적으로 공격을 수행하는 주소를 파악 할 수 있는 (C) Attack IP 테이블로 구성 하였다.

인공신경망 모델에서 발견된 패턴 중 계속 반복적으로 발생하는 것에 대해서는 누적 시키지 않고 해당 패턴에 대한 시간을 체크하고 해당패턴에 대한 가중치를 증가 시키는 과정을 걷히게 된다. 또한 반복 학습을 통해 해당 프로파일 데이터들은 공격 형태 뿐만이 아닌 데이터 마이닝을 위한 자료로도 사용하게 된다. 이렇게 본 논문에서는 반복적인 과정을 통해 EDI와 인공 신경망 모듈간의 조화를 이루면서 데이터 보안상의 문제를 해결하게 된다.

4. 시스템 실험 및 분석

4.1 실험 개요

지금까지 RFID 기반 유비쿼터스 화물 추적 시스템에서의 기반 기술 및 연구에 관한 많은 자료들을 살펴 보며 시스템 운영 상에서 일어날 수 있는 데이터 상실과 데이터 보안에 대한 연구를 진행하였다. 또한 그에 따른 대책에 대해서도 고찰해 보았다. 이번 장에서는 실제 환경에서 테스트를 통해 제안된 시스템의 효용성을 확인하려 한다. 실험내용과 환경은 다음과 같다.

위에서 언급한 기술을 바탕으로 RFID 를 통해 취득한 정보를 EDI 형태로 정보를 전달 하여, 미들웨어에서 각 지역의 정보를 안전하게 수신 할 수 있는지에 대한 기술을 검증하기 위한 실험이다. 실험 시에는 실제 지역별 화물 운송 시 화주의 화물이 시작되는 지점인 DOOR 지역, CY(Container Yard), CFS 간의 이동 시 데이터 전송은 안정적으로 유지되어있는 상태로 실험을 실시하여 전송이 원활히 이루어 질 수 있도록 하였다. 개별 컨테이너의 대한 정보는 국제간 회사간 이 기종 간의 시스템 통합을 위하여 EDI 를 이용한 데이터 통신을 하도록 구성하였다. 문서양식은 CODECO EDI 표준을 이용하였으며, 버전은 UN 버전을 이용하여 데이터 교환을 하였다.

본 논문에서의 실험환경은 MS 2003 Sever 상에서 구현하였으며, 미들웨어로는 Oracle AS 10g 버전을 사용하였다. 실험대상으로 삼은 데이터의 양은 약 15 만 여 건의 데이터를 표준으로 삼아 4 개의 다른 이기종간의 데이터를 받을 수 있도록 하였다. 첫 번째로 시스템을 제안하기 이전과 이후의 환경에서의 데이터의 정확성과 효율성을 확인하였다. 두 번째로 제안된 시스템으로 정상적인 요청과 비정상적인 요청에 대한 정상 처리 여부를 확인 하였다.

4.2 실험 분석

실험에 사용된 데이터들은 일정 기간 동안 수집된 데이터들을 이용하여 실험을 진행하였다. 총 표본 데이터의 수는 152,198 Row 의 데이터들을 대상으로 하여 해당 데이터들의 보안 문제들을 체크 하였다. 또한 발생한 보안문제에 대한 표본 데이터들을 일정한 기간으로 나누어 (그림 1)의 통합 시스템을 적용전과 적용 후로 나누어 시스템을 비교 분석 하였다.

<표 1> 비정상 데이터 수

공격유형	1 차 표본 데이터	2 차 표본 데이터
표본데이터 양	77,788	74,829
Data 타입조작	108	93
Data 길이조작	57	67
Terminate 조작	5	7
비정상 명령어 실행 공격	25	14
시스템명령 실행공격	13	3
사용자 오류	167	209
보안 문제 총 발생 건수	375	393
발생 비율(%)	0.48%	0.53%

실험을 수행하기 이전 일부 프로파일 들이 존재하

였으며, 작업 수행 도중 데이터의 상태에 따라 미리 분석된 프로파일 들을 참조하는 경우도 발생하였다. 그 중에는 정상이지만 프로파일을 참조하는 경우도 있었으며, 실제로 프로파일이 적용이 되어 정상데이터로 바뀌는 경우도 있었다. 반대로 일반적이지 않은 문제가 발생되는 부분에 대해서는 관리자가 개입하여 수정 작업을 수행하여, 프로파일 화 하지 않았다. 이는 참조 할 수 있는 프로파일 수를 최소화 하여 수행 속도의 문제를 사전의 예방하는 역할을 하였다. 시스템 분석 작업을 수행한 결과 아래<표 1>과 같은 보안상의 문제에 대한 결과를 얻을 수가 있었다.

일정 기간 동안 비슷한 양의 데이터를 표본으로 삼아 실험한 결과 기간에 따른 보안상의 문제는 같지는 않았지만, 대체적으로 비슷한 발생 빈도를 보였다.

4.3 시스템 비교

위의 <표 1>의 실험 결과를 토대로 시스템 적용 이전과 이후의 데이터의 변화를 측정해보았다. 또한 1 차 표본데이터와 2 차 표본 데이터를 비교하여 또 다른 문제가 없는지 확인하였다.

<표 2> 표본데이터와 System 적용 후 비교

공격유형	표본데이터	시스템 적용
표본데이터 량	74,829	74,829
Data 타입조작	93	5
Data 길이조작	67	3
Terminate 조작	7	2
비정상 명령어 실행 공격	14	1
시스템명령 실행공격	3	4
사용자 오류	209	1
보안 문제 총 발생 건수	393	16
발생 비율(%)	0.53%	0.02%

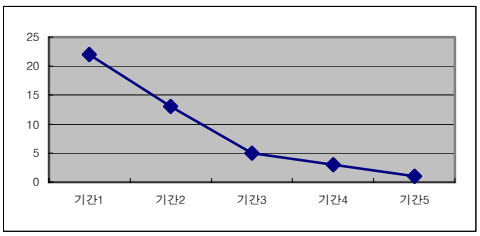
<표 3> 1 차 표본데이터와 2 차 표본데이터 비교

	1 차 표본데이터	2 차 표본데이터
표본데이터 량	77,788	74,829
Data 오류건수	375	16
Data 오류 율(%)	0.126%	0.02%
False(+)건수	0	45

전체적으로 프로파일 적용 후 데이터의 문제는 현저히 줄어드는 것으로 확인하였다. 수행 속도 측면에서도 차이를 느끼지 못할 정도의 결과를 보였다.

하지만 실험 진행 과정에서 프로파일 매치 시 정상적인 것(Positive)을 비정상적인 것으로 판단하는 오류(False)인 양성오류(False Positive)가 발생하였다. 본 논문에서는 편의상 False Positive 를 False(+)로 표현 하도록 하겠다.

<표 3>에서와 같이 시스템 적용 이전의 데이터에서는 False(+) 발생건수가 없었으나, 시스템 적용 이후에는 False(+) 발생건수가 늘어나는 것을 확인 할 수 있었다. False(+) 문제의 발생으로 원인을 파악하고 대책을 마련하기 위해 (그림 5)와 같이 일정 기간별 분석 작업을 수행하였다.



(그림 5) False(+) 발생 건 수

분석 결과 False(+)는 시스템 발생 초기에는 발생하였으나, 관리자가 개입하여 일정 프로파일에 대해서 예외 사항을 추가하여주고, 관리한 결과 시스템이 안정화된 이후에는 거의 발생하지 않음을 보여주고 있다. 이는 초기 시스템에서 발생할 수 있는 문제점으로 시간이 지난 후에는 거의 발생하지 않을 것으로 보여진다.

5. 결론 및 향후 연구과제

앞에서 살펴본 것과 같이 기존의 시스템에서는 이기종간의 데이터 표준화가 힘들고, 데이터 인터페이스 시 발생되는 데이터 보안문제에 대해서는 탐지가 힘들다는 단점이 있었다.

하지만 본 논문에서 제안한 시스템에서는 표준화가 가능하고 새로운 문제에 대한 탐지가 가능하다는 것을 실험을 통해서 증명하였다. 동작 방식을 살펴보면 발생된 문제에 대한 프로파일들을 식별하고 데이터의 타입, 길이, 사용자의 오류로 인한 이상현상들의 특징을 인식할 수 있기 때문에 이상 현상에 대한 탐지가 가능하도록 하였다. 또한 대부분의 데이터들은 약속된 장소로부터 데이터를 전송 받기 때문에 인공신경망과 프로파일을 이용한 탐지가 효과적이라고 할 수 있다.

이를 증명한 실험에서는 많은 데이터의 검증을 통해 이상현상 탐지가 효과적임을 증명하였으며, 패턴 매치 시 단점으로 지적되는 False(+)에 대한 문제도 실험을 통해 조절이 가능함을 확인하였다.

향후 연구 과제로는 EDI 와 RFID 이용 시 발생할 수 있는 데이터 보안상의 문제를 일반화된 보안 모듈로 구현함으로써, 많은 분야에서 사용이 가능하도록 관련 분야 연구를 계속 수행해야 한다.

참고문헌

[1] RFID handbook :Fundamentals and Applications in Contactless Smart Cards and Identification by Klaus Finkenzeller

[2] Machine Learning Book by Tom M. Mitchell

[3] James A. Freeman and David M. Skapura, Neural Networks : Algorithms, Applications, and Programming Techniques, Addison Wesley, 1991.

[4] UN/EDIFACT Message CODECO Release: D95B