박사학위논문

# 망 분리 환경에서 웹 스크래핑을 이용한 웹브라우징 기법

A Web-browsing Technique
using Web Scraping
in Network Separation Environments

제주대학교 대학원

융합정보보안학협동과정

정 원 치

2022년 2월

# 망 분리 환경에서 웹 스크래핑을 이용한 웹브라우징 기법

지도교수 박 남 제

정 원 치

이 논문을 융합정보보안학협동과정 박사학위 논문으로 제출함.

2021년 12월

정원치의 융합정보보안학협동과정 박사학위 논문을 인준함.

| | | | |
|---|---|---|---|
| 심사위원장 | 변 | 영 | 철 ㊞ |
| 위 원 | 이 | 은 | 주 ㊞ |
| 위 원 | 주 | 연 | 수 ㊞ |
| 위 원 | 김 | 인 | 중 ㊞ |
| 위 원 | 박 | 남 | 제 ㊞ |

제주대학교 대학원

2021년 12월

# A Web-browsing Technique using Web Scraping in Network Separation Environments

Won-chi Jung
(Supervised by professor Namje Park)

A thesis submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Convergence Information Security

2021. 12.

This thesis has been examined and approved.

Yung-Cheol, Byun

Thesis director, Namje Park, Prof. of Elementary Computer Education

Eunju, Lee

Yeon-soo, Joo

InJung, Kim

Namje, Park

2021. 12.

Department of Convergence Information Security
GRADUATE SCHOOL
JEJU NATIONAL UNIVERSITY

# 목 차

# 표 목 차

# 그 림 목 차

# 요 약

해킹의 위협이 증가할수록 강력한 대응책이 필요하다. 우리나라는 정부 공공기관을 상대로 일어난 일련의 정보보안 사고로 인하여, 망 분리라는 강력한 대응책이 정책적으로 국가 공공기관과 금융권에 우선 적용되었다.

망 분리는 업무를 수행하는 내부망과 인터넷을 통해 정보를 탐색하는 외부망으로 나누어 외부의 공격자로부터 침입을 원천차단하는 강력한 방식이고, 정보보안 방어전략 측면에서 큰 효과를 발휘하고 있다.

하지만 4차 산업혁명 시대의 고도화 된 기술은 대량의 데이터 공유와 초연결 기술에 집중되어 있고, 코로나 19로 인하여 많은 기관은 방역수칙에 따라 재택근무를 수행하고 있다. 이런 환경적 시대적 변화에 맞는 망 분리 정책이 필요하다는 문제 인식을 가지고, 어떻게 하면 망 분리 환경의 보안 수준을 유지하면서 효율성을 높이는 방안을 찾는 것을 목적으로 한다.

우선 망 분리의 현행 제도와 망 분리의 방식을 파악하여 망 분리 다양한 형태의 장·단점을 파악하여, 망 분리의 목적을 파악한다. 망 분리 구축이후에도 발생할 수 있는 다양한 위협요소와 이를 보완하기 위한 추가 보안 장비들과 특성을 파악하고, 최신 IT와 보안 트랜드를 분석하여 제안하는 모델을 설정한다.

본 논문은 망 분리가 적용 된 환경에서, 내부망 PC를 사용하는 사용자가 인터넷 정보에 접근하는 방법을 제안하고 있다. 내부망에서 접근하기 원하는 인터넷 정보를 외부망으로 전달하고, 웹 스크래핑 기술을 이용하여, 브라우징 된 웹 콘텐츠를 정적인 형태로 고정하여 내부망 PC에 다시 브라우징 하는 메커니즘이다.

차단이 목적인 망 분리 환경에서 내부망과 외부망의 연결을 위한 네트워크 구성 모델을 설계하고 이를 실제 구현하여 실험을 통해 결과를 확인한다. 내부망과 외부망에 데이터를 전달 하는 것에 대한 안전성 확인을 위해 MITRE와 OWASP에서 정의한 일반적인 웹에서 발생 가능한 취약점을 분석하고, 실험 가능한 악성 유형을 추출하여, 실제 구현 될 모델의 비 악성화의 실험 타겟을 설정하였다.

제안하는 메커니즘이 설정된 악성코드를 비 악성화 처리 됨을 상세 분석을 통해 증명하였으며, 현행 망 분리 구조의 제도 내에서 내부망과 외부망의 데이터 연동이 가능하다는 것을 확인하였다.

망 분리 정책을 안정적으로 관리하기 위해서는 강력한 네트워크 정책이나, 보안장비도 중요하지만, 그에 못지않게 그 환경에서 업무를 수행하는 사용자의 보안 의식도 중요하다. 사용자에게 불편함을 강제하는 제도는 예외 사항이나 편법을 만들게 될 것이다. 이를 보완하기 위해 망 분리의 효율성을 높이는 노력도 필요하다. 내부망에서 인터넷의 정보를 활용할 수 있는 방안을 통해 망 분리의 도입이나 제도를 정착하기 위한 도움이 될 것이다.

주제어 : 망 분리, 망 분리 정책, 웹 스크래핑, 비 악성화 웹, 망 연계 장치

# Ⅰ. 서 론

## 1.1. 연구 배경

시대적으로 가장 위험한 사이버 위협과 이에 대응하기 위한 대비책은 상관관계를 가진다. 한국에서는 국가기관을 향한 해킹 공격이 발생한 바 있으며, 이런 일련 공격을 통해서 국가·공공기관의 중요자료가 유출되는 사례가 발생하였으며, 또한 이런 사건이 증가하고 있다. 이런 사이버 위험에 대처하기 위해서, 우리나라는 네트워크망 분리(Network Separation) 정책을 채택하게 된다.

초기, 망 분리 정책 수립과정에서 물리적인 분리를 고수했으나, 다수의 지점을 보유한 기관에 적용하기 위한 현실적인 해결방안과 비용적인 문제 해결을 위해 논리적 망 분리도 허용되었으며, 적용 범위도, 2013년 정보통신 서비스 사업자 또는 인터넷 사업자에게 망 분리 의무를 부과하였고, 2014년 금융권(은행, 보험, 증권, 카드사)을 포함하였다. 그리고 현재는 대부분 정부, 공공기관 그리고 금융권 등에 망 분리 의무를 부과하고 있고, 이미 대다수 기관에 망 분리가 적용되었다.

보안 강화의 목적으로, 망 분리는 정부 주도적으로 확산되고 있다. 망 분리 정책은 외부망과 내부망의 네트워크를 분리 운영하여, 외부로부터 공격에 대해 내부 정보와 데이터를 방어하는 효과적인 방법이다. 하지만, 보안을 강화한 만큼 망 분리 적용 환경에서 업무 효율성은 떨어지기 때문에, 시대의 환경적인 변화에 대응하려는 방안이 요구되고 있다[1].

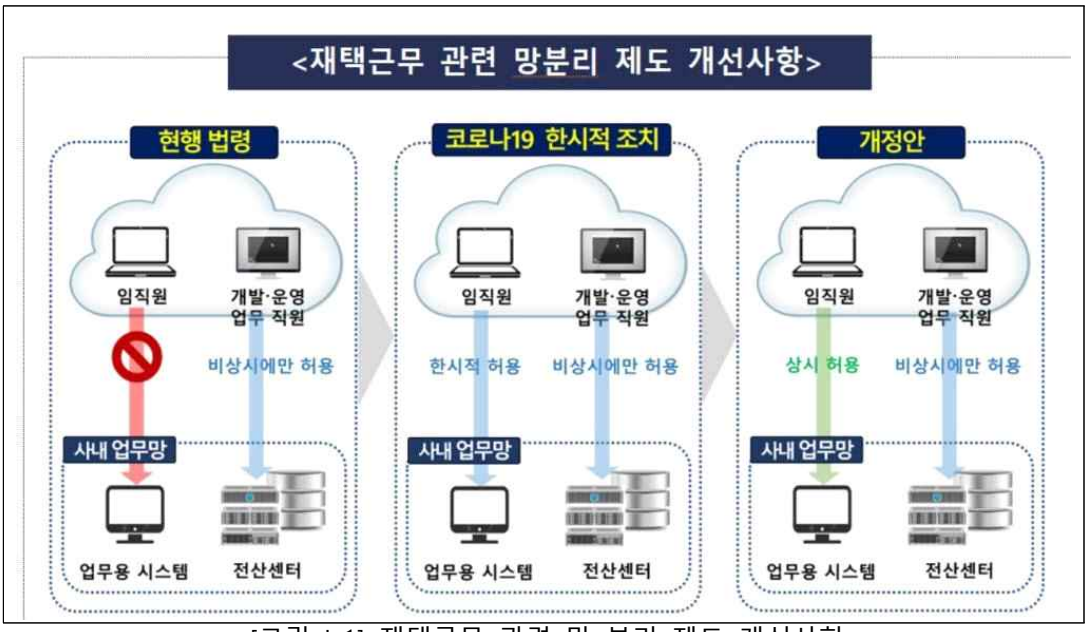
망 분리 정책이 마주하는 환경적 변화는 첫 번째, 근무 환경의 변화이다.

2020년부터 2021년까지, 코로나19의 확산을 방지하기 위해, 사회적 거리두기를 실시하고 있으며, 사회적 거리두기 단계에 따라 단계별로 재택근무를 권고 혹은

의무가 되었다. 중앙재난안전대책본부에서 제정된 “방역수칙 단계별 「사회적 거리두기」 조치내용”을 보면. 공공기관이나 기업은 단계에 따라, 밀집도를 최소화하기 위해 유연·재택근무 등을 전체 인원의 1/3이나 1/2을 근무 형태 변경을 통해 밀집도를 낮추게 하는 방안이 있다. 이를 준수하기 위해 재택근무가 보편화되고 있다. 하지만, 재택근무자는 재택에서 내부망에 있는 업무 시스템과 데이터에 접근할 수 있어야 하지만, 망 분리 규제를 받는 공공기관과 금융기관은 현재의 제도의 틀 안에서 재택근무를 시행하기 위해서는 별도의 방법이나 규제 완화 등의 방안이 요구된다.

현행법에서는 장애나 재해 발생 등 비상 상황 시 신속한 조치를 위해 전산센터에 대해서만 예외적으로 원격접속을 허용하고 있으나, 금감원은 코로나19로 인한 원격접속을 이용한 재택근무를 예외적으로 허용하였다.

보안 전문가들은 금융사들이 충분한 준비기간 없이 급히 재택근무로 전환하면서 보안 조치가 미흡할 것에 대한 우려를 제기하고 있다. 또한, 코로나 팬데믹 사태가 장기화함에 따라 재택근무의 일상화를 고려한 제도 개선이 필요하다는 목소리도 높아지고 있다. 금감원은 감독규정을 개정해 앞으로 금융회사 임직원의 상시 원격접속을 [그림 I-1]과 같이 허용하기로 했다.




[그림 I-1] 재택근무 관련 망 분리 제도 개선사항
(출처 : 금융감독원 망 분리 제도 개선 보도자료, 2020.9.17.)

예외적인 재택근무를 허가하게 하기 위해서는 기존과는 다른 접속을 허용해야 한다. 별도의 접속 환경을 허용하는 것은 기존 망 분리 체계에서의 방어전략과는 별도의 연동·보안 기술이 요구된다.

재택근무 환경을 구축하기 위해서, 자택에서 회사까지 인터넷을 이용하여 통신하게 된다. 이 과정에서 통신 내용에 대한 정보보호를 위해 구간 암호화 등의 방안이 요구된다. 통신 과정의 보안성 강화를 위해 VPN(Virtual Private Networks)이나 전용망(Private Network)을 적용하여 중간자 공격(Man In The Middle Attack, MITM)으로부터 통신 과정을 보호한다. 재택근무 PC(Personal Computer) 환경 보호를 위해 DRM(Digital Rights Management)이나 DLP(Data Loss Prevention) 기술을 적용하여 문서나 중요정보 유출을 차단하며, 악성코드를 방어하기 위한 백신 프로그램, 화면캡처 및 스크린 워터마크(Watermark) 등이 필요하다[2].

금융당국은 원격접속을 통한 재택근무 방식은 각 금융사 사정에 따라 자율적으로 선택할 수 있도록 정책을 임시로 완화하였다. 금융사들을 적용한 망 분리 방식은 대부분 물리적 망 분리로 구조로 사용자의 업무 PC에 두 개의 네트워크 통신선이 연결되는 방식이다. 두 개의 통신선 중 한 개는 내부업무를 위한 내부망 통신선이며, 다른 하나는 외부업무 즉, 인터넷을 사용하기 위한 통신선 이다. 이런 구조를 가진 기관에서는 재택근무를 위해서는 추가적인 보안 조치들이 필요하다. 가정에서 이미 사용하고 있는 PC는 인터넷을 사용하면서, 이미 악성코드에 감염되거나 해커에 의해 제어권을 빼앗긴 상태인지 확인 할 수 있는 방법이 없다. 그러므로 회사에서 노트북을 제공하거나, 내부 서버에 긴급하게 VDI를 구성하는 등의 보완조치가 필요하다. 하지만, 논리적 망분리를 선택하여 VDI를 선택한 기관은 재택근무로 변환이 비교적 간단하다. 이미 내부 VDI 자원 접근이 가능한 상황이기 때문에, 재택근무자의 PC와 내부 VDI 구간 암호화를 보완하고, 인증을 강화하는 조치로 간단히 해결할 수 있다.

위의 사례처럼, 망 분리라는 강력하고 단단한 정책이 코로나19로 인해 유연하게 변화할 필요가 생겼을 때, 지속적으로 예외를 허용하여야 하고 예외로 인하여 보안 정책이 약화 될 수 있는 상황에 있다.

또 다른 예로, 핀테크 산업의 예를 들 수 있다. 핀테크는 금융과 기술이 합성되어 또 다른 가치를 창출할 것으로 기대되는 산업군이다. 하지만, 핀테크 업체들은 전자금융 감독규정에 따라 금융업무를 다룬다는 이유로 일반 금융회사와 똑같이 '망 분리' 규제를 적용받는다. 핀테크 업계는 내부망과 외부망을 분리하는 현재의 규제가 개발 효율성을 떨어뜨리고 보안 측면에서도 오히려 취약하다고 주장하고 있다[3]. 이유는 정보기술(IT)·통신·유통 등 다양한 업계의 이종(異種) 데이터 간 결합을 통한 혁신을 기대했기 때문이다. 21년 5월 이뤄진 마이데이터 사업자 사전 수요조사에 참여한 116개 업체 가운데 비금융회사와 핀테크 업체는 61개로 전체의 52.5%를 차지했다. 핀테크 관련 스타트업은 핀테크 창업에 엄격한 망 분리 규제가 걸림돌이 된다고 한목소리로 지적하고 있다. 망 분리 규제 내에서는 스타트업이 자주 이용하는 오픈소스나 클라우드 서비스를 이용할 수가 없다. 또한 데이터 분석·개발을 위해 반출입하는 데이터와 소스코드를 일일이 허가받은 뒤 내부망으로 옮겨야 하므로 소요되는 시간과 비용이 상당하다. 이는 당연한 정보보안의 수칙일 수 있으나, 업무의 효율성 측면에서, 많은 문제를 발생시킬 수 있다. 민관협력 네트워크인 스타트업 얼라이언스에 따르면 망 분리 규제는 개발자의 생산성을 50% 감소시키고 인건비는 30% 증가시킬 뿐만 아니라 망을 구축하는 데 드는 비용은 25인 사업자 기준 5억 원 이상으로 큰 지출을 요구한다.

다른 환경의 변화는 클라우드 전환의 가속화이다. 소프트웨어 솔루션, 정보시스템 그리고 DB(Data Base) 같은 서버나 장비를 자체적인 전산실이나 IDC(Internet Data Center)의 코로케이션(Colocation) 서버를 이용하여 운영하는 방식을 온프레미스(On-Premise) 방식이라고 한다.

망 분리 정책이 수립된 시점에서는 공공기관의 정보시스템이 외부(클라우드 포함)에 위탁되는 것을 허용하지 않았으며, 망 분리 전략 역시, 온프레미스 환경을 기본으로 발전되어왔다. 내부와 외부의 구분 선을 긋고 외부는 인터넷이나 다른 네트워크를 허용하고, 내부는 내부 시스템 간 소통을 허용하는 정책이기 때문이다. 하지만 클라우드 환경에서는 내부와 외부의 경계선을 논리적으로는 그을 수는 있지만, 모든 물리적인 시스템의 위치는 외부이기 때문에, 어떻게 정책을 적용해야 하는지, 어떤 방어전략을 세워야 하는지 불명확해진다. 이처럼 보안 위협이 정의되지 않기 때문에 보호 대책이 수립되기 어렵다.

이런 상황에서 행정안전부는 2025년까지 모든 행정기관과 공공기관의 정보시스템이 통합 관리되는 클라우드 기반으로 전환될 것이라고 정책을 발표하였다. 공공기관의 서버나 시스템의 내용 연수가 5년 이상임을 고려하고, 예산 수립 기간이 내년도 예산을 올해 수립한다는 것을 생각하면, 상당히 급진적인 발표이다. 행정안전부는 공공기관이 운영하는 정보시스템 전부(10,009개)를 클라우드로 이전·전환하여 통합할 것이라고 발표하였다. 이를 통해서 정부가 달성해야 할 목표는 우선 보안 강화와 안정적인 운영임을 명시하였다. 행정·공공기관이 자체적으로 운영하는 정보시스템의 비율은 약 83%로 대부분 기관은 설비와 규모가 미흡하고, 전문 인력이 부족하여 보안에 취약하다. 또한, 정보자원의 절반 이상이 내용연수(6년 이상)을 초과하여 사용되기 때문에 시스템 효율과 안정성에 문제가 있다고 판단하고 있다. 이 정책은 [그림 I-2]와 같이 전자정부법 제2조에 따른 행정·공공기관 2,105개를 대상으로 한다.

| 대분류 | 중분류 | 기관 수(안) | 대상기관 명 |
|---|---|---|---|
| 행정기관 | 중앙행정기관 | 54 | 국세청, 관세청, 경찰청, 여성가족부 등 |
| | 소속기관 | 504 | 감사교육원, 강원지방기상청, 경찰대학 등 |
| | 광역지자체 | 17 | 세종특별자치시, 서울특별시, 강원도 등 |
| | 기초지자체 | 228 | 서울특별시_종로구, 부산광역시_동구 등 |
| 공공기관 | 공공기관(공공기관운영법) | 340 | 한국기술교육대학교, 한국개발연구원 등 |
| | 지방공사·공단(지방공기업법) | 151 | 인천도시공사, 서울시설공단 등 |
| | 지방 출자·출연기관 | 742 | 서울의료원, 서울연구원 등 |
| | 국·공립대학교 | 54 | 충남대학교, 서울대학교, 제주대학교 등 |
| 교육기관 | 시·도교육청 등 | 18 | 서울특별시교육청, 부산광역시교육청 등 |
| 합계 | | 2,105개 기관 | |

[그림 I-2] 정보자원 통합 및 클라우드 전환 대상[4]
(출처:행정·공공기관 정보자원 통합 및 클라우드 전환을 위한 정책협의회)

공공기관 클라우드 전환을 위해서 행정안전부에서 제시한 계획은 우선, 공공·민간 클라우드센터를 본격적으로 활용하도록 권고하고 있다.

「행정기관 및 공공기관 정보자원 통합기준」에 따라서 국가안보 등 주요한 정보를 처리할 때에 공공클라우드 센터를 운영 하도록 정의되어 있다. 이 외에 공공기관의 정보시스템은 [표 I-1]와 같이 클라우드 안정성에 관한 보안 인증(클라우드 보안 인증제)을 받은 민간 클라우드를 이용하게 되어있다[4].

[표 I-1] 클라우드 보안 인증제 인증기준 및 통제항목[5]

| 구분 | 내용 | 통제항목 |
|---|---|---|
| IaaS 인증 | 관리적·물리적·기술적 보호조치 및 공공기관용 추가 보호조치 | 117개 항목 |
| SaaS 표준 | 관리적·기술적 및 공공기관용 추가 보호조치 | 78개 항목 |
| SaaS 간편 | 관리적·기술적 및 공공기관용 추가 보호조치 | 30개 항목 |
| DaaS 인증 | 관리적·물리적·기술적 및 공공기관용 추가 보호조치 | 110개 항목 |

공공기관 업무를 위해 제공하는 클라우드 대상에 관한 정의는 「클라우드컴퓨팅 발전 및 이용자 보호에 관한 법률 시행령」에 정의되어 있으며, 다음과 같다.

“ 1. 서버, 저장장치, 네트워크 등을 제공하는 서비스
2. 응용프로그램 등 소프트웨어를 제공하는 서비스
3. 응용프로그램 등 소프트웨어의 개발・배포・운영・관리 등을 위한 환경을 제공하는 서비스
4. 그 밖에 제1호부터 제3호까지의 서비스를 둘 이상 복합하는 서비스 “

위에서 살펴본 것과 같이, 공공기관과 금융기관은 공적인 데이터를 다루기 때문에 다른 기업들과는 가치체계가 특수하다. 따라서 정보보안의 가장 강력한 방법의 하나인 망 분리를 강제적으로 적용하고 있다. 하지만 현행 망 분리 제도는 4차 산업혁명과 디지털 전환 혁신 시대에 맞지 않는 부분이 다소 존재한다.

보안을 위한 강한 규제는 업무 효율성을 저하하고, 다양한 복합적인 정보를 이용하기 어렵게 만든다. 따라서 망 분리가 적용된 환경에서 혁신적인 업무 향상을 기대할 수 없다. 실제로 코로나19 같은 사회 시스템의 변화를 가져오는 사건에 의해 클라우드로 급격한 전환이 요구되는 환경에서 유연성의 부족으로 많은 문제가 발생하고 있다[6].

## 1.2. 연구의 문제설정과 목적

본 연구는 망 분리를 기본 원칙인 “업무망(내부망)에서는 인터넷을 사용할 수 없다”에 대안을 제시하기 위해서 다음과 같은 문제를 설정하였다.

□ 내부망에서 인터넷을 차단하는 목적은 무엇인가?
- 현재 망 분리 제도와 규제를 이해하여 인터넷을 차단함으로써 얻는 보안상의 이점을 확인한다.

□ 망 분리가 적용된 내부망-외부망 방식은 안전한 것인가?
- 웹(인터넷)의 취약점이나 위협 요소 중 망 분리 적용을 통해 무력화되는 것과 별도의 보호 대책이 필요한 위협요인을 분류한다.

□ 내부망에서 안전하게 인터넷의 정보를 이용할 방법은 있는가?
- 내부망에서 정보 이용에 대한 절차를 확인하고 이를 효율적으로 개선한 방안에 대해 고찰한다.

□ 웹 스크래핑(Web Scraping) 기술을 이용하여 서로 다른 PC 환경에서 브라우징이 가능한가?
- 웹 브라우저의 기술을 이해하고, 이를 활용한 웹 스크래핑 기술을 통해 본 연구의 목적에 맞게 활용할 수 있는지 실험한다.

본 논문은 「효율적인 데이터 전송을 위한 망 분리 메커니즘」을 제안한다. 현행 망 분리의 비효율성을 기술적으로 보완하는 메커니즘을 제안하고 해당 메커니즘이 보안 측면에서 위험하지 않음을 검증한다. 이를 통해 망 분리 환경에서 다양한 인터넷을 활용할 수 있는 기술적 방법적인 방향성을 제시한다.

## 1.3. 연구의 구성

본 연구는 국가 정보보안 정책 안에서 업무효율을 끌어올리는 방법을 제안하기 위해 크게 6개의 장으로 구성된다.

첫 장, 서론에서는 연구의 배경이 되는 망 분리를 둘러싸고 있는 환경과, 코로나19로 인하여 변화하는 업무처리 방식, 그리고 급격한 클라우드 전환에 따른 변화를 알아보고 문제와 연구의 목적 및 범위를 설정한다.

2장 관련 연구에서는 본 연구를 수행하는 데 필요한 제도나 규제를 살펴보고, 망 분리의 다양한 방식을 구분하고 장단점을 분석한다. 지금까지 발전한 망 분리 기술에 필요한 기술의 연구 방향과 대안으로 제시할 망 연계를 활용한 웹 스크래핑 브라우징 기법에 요구되는 기술들을 정리한다.

3장에서는 본격적으로 망 분리 환경에서 웹 스크래핑 방식에 대한 프로그램 설계를 위해, 요구사항을 분석하고 구현에 필요한 부분들을 종합하여, 모델을 만들고 악성코드를 분석하기 위한 가설을 수립한다.

4장에서는 프로그램 구현 및 분석을 통해 실제 망 분리 환경에서 웹 스크래핑 브라우저가 동작함을 확인한다. 위험성이 존재하는 웹페이지를 비 악성화 처리하여 악성코드에 대한 부분을 분석한다.

5장에서는 본 연구를 요약하고 결과를 분석하고 고찰하여, 연구 결과가 가지는 가치를 평가하며, 개선사항을 분석한다.

6장에서는 본 연구의 시사점과 한계점을 통해 결론을 도출 하고 향후 연구 방향과 과제에 대해 제시한다.


[그림 Ⅰ-3] 연구의 구성

## 1.4. 연구의 범위와 방법

본 연구의 목표를 달성하기 위해, 기존 연구 자료들을 활용하였으며, 인터넷을 통한 자료 검색과 정부 정책 가이드라인을 참조하였다.

또한 실제 망 분리가 적용된 기관의 정보보안 구성을 분석하여, 망 분리의 개선점을 찾아 방향성을 찾아보았다.

1) 본 연구는 망 분리 규제에 관한 내용을 분석하고, 업무 친화적인 망 분리 변화를 이끌 방안을 찾기 위해서 웹 스크래핑 장치를 설계, 구현, 분석의 과정을 수행한다.

2) 해당 장치가 망 분리 체계를 위배하지 않고 웹의 정보를 비 악성화 한다는 것을 증명하기 위해서 연계 데이터 정보를 상세 분석하여, 실제 악성코드가 비 악성화 됨을 확인한다.

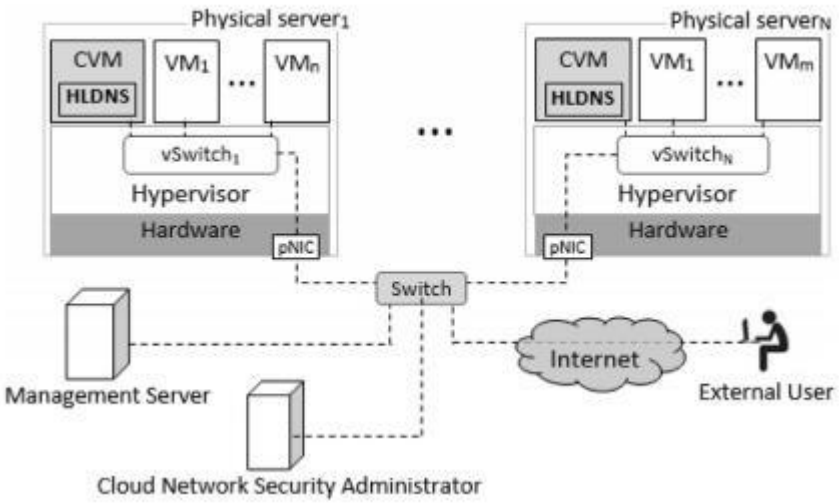
3) 웹을 이용한 취약점과 이런 취약점을 활용한 공격기법은 다양하고, 계속 생성되기 때문에, MITRE, OWASP 지표 중 망 연계 이후 데이터 교환에서 가장 잘 나타날 수 있는 경우를 일부 악성 웹으로 한정하고 본 연구를 진행한다.

# II. 이론적 배경

## 2.1. 관련 연구

Rajendra 등은 현재의 보안이 실현 가능한 HLDNS(Hypervisor Level Distributed Network Security) 모델을 제안하여 네트워크 트래픽 특징을 식별하고 공격을 정확하게 탐지하는 하이퍼바이저(Hypervisor) 수준의 분산 네트워크 모델의 정의 하였다.

네트워크를 탐색하고 식별하면서 명확하게 표적화된 공격은 방화벽, 침입 탐지 및 보호 시스템과 같은 솔루션이 확장되어야 한다. [그림 Ⅱ-1]은 네트워크 격리의 안정성을 보여주며, 이는 물리적으로 독립적인 서버가 다음의 가상화를 통해 논리적으로 격리된다는 것을 보여준다[7].




[그림 Ⅱ-1] 제안된 클라우드 기반 HLDNS 프레임워크[7]

망 분리 구조와 다른 점은 외부와 내부를 연동하는 장치에서 강화된 보안 검역을 통해 안정성을 확보해야 하는 점에서 중앙 스위치(Switch)에 역할이 중요한 차이점은 [표 Ⅱ-1]과 같다.

[표 Ⅱ-1] HLDNS와 본연구의 장단점 비교

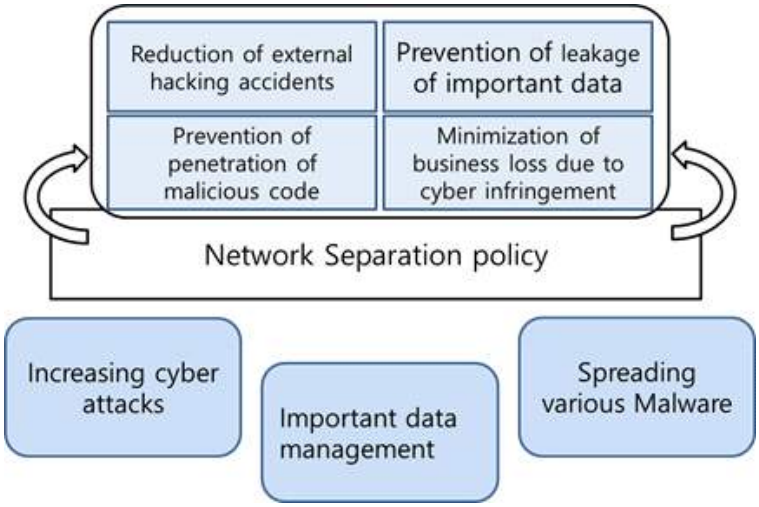
| 구분 | HLDNS 방식 | 스크래핑 브라우징 방식 |
|---|---|---|
| 방식 요약 | 관리 서버가 중앙 스위치를 통제하여 안전하지 않은 접근을 차단 | 외부에 자료를 안전한 형태로 변환하여 내부로 전송하는 방식 |
| 장 점 | 관리 서버를 통해 선제 대응이나, 기관이나 회사 업무 특성에 맞는 제어가 가능하다. 인프라 이식성이 뛰어나서 클라우드나 다양한 환경에 적용할 수 있다. | 내부에서 필요한 외부 정보를 빠르게 적용할 수 있으며, 기존 보안 정책의 변경 없이 적용할 수 있다. |
| 단 점 | 관리자가 통제하지 못한 제어는 보안 위협이 된다. 국내에서는 보안 정책을 변경하거나 보안 수준을 낮춰야 도입할 수 있다. | 특정 업무 환경에서 불편함이 발생할 수 있다. 양방향 정보를 원하는 경우에 불편함이 발생한다. |

## 2.2. 망 분리의 정의와 세부 내용

망 분리는 외부의 공격으로부터 내부의 자원을 보호하기 위한 정보보안 방어 전략의 하나로, 기관의 네트워크를 내부망과 외부망으로 분리하여 공격의 접근을 원천 차단하는 것을 말한다.

망 분리는 「국가·공공기관 망 분리 및 자료전송 보안 가이드라인」을 기준으로 각 기관에 맞는 망 분리 방법을 적용하도록 안내하고 있다. 가이드라인은 국가·공공기관에 필요한 망 분리 구성기준과 보안 고려사항을 제공하며, 가이드라인의 근거는 「국가 정보보안 기본지침」에 있다.

정보통신 발달로 ICT(Information and Communications Technology) 기술은 급속하게 발전하여, 개인의 일상으로부터 다양한 산업 전반까지 사회의 대부분에 IT가 적용되고 있다. ICT의 확산에 비례하여 사이버 공격도 그 방식과 규모가 진화하고 있으며, 이런 사이버 공격이나, 악성코드로 인한 정보 유출, 랜섬웨어(Ransomware), DDoS(Distributed Denial of Service) 공격 등 다양한 공격으로 인한 피해가 증가하고 있다. 정리해보자면 ICT 기술이 인간과 사회의 편리한 변화를 가져왔지만, 편리해 진만큼 해킹의 피해도 비례하여 증가하였다고 할 수 있다. [그림 Ⅱ-2]과 같이, 고도화되는 취약점, APT(Advanced Persistent Threat) 공격 기법을 활용한 랜섬웨어 유포 등의 모든 보안 위협에 대응하는 원천적인 망 분리 대책이 필요했다. 특히, 우리나라에서는 대규모 사이버 공격과 테러가 발생함에 따라, 금융권과 공공기관에 우선으로 내부망과 외부망을 원천적으로 분리하는 개념이 정립되고 적용되기 시작하였다.




[그림 Ⅱ-2] 망 분리 정책의 방어체계

개인정보보호법에서는, 고객정보를 다루거나, 정보통신 기반시설, 정보통신망, 그리고 금융정보를 다루는 기업들은 망 분리 조처를 할 것을 규정하고 있다.

「(개인정보보호위원회) 개인정보의 기술적·관리적 보호조치 기준」(제2020-5호)은 「개인정보보호법」 제29조 및 같은 법 시행령 제48조의2제3항에따라 개인정보의 안정성 확보를 위한 최소한의 기준을 정하는 것을 목적으로 한 기준이다. 이 기준에 따르면 접속 권한에 대한 제한, 외부에서 개인정보처리시스템 접속에 필요한 통신의 안전성, 마지막으로 물리적·논리적으로 망 분리를 하여야 한다고 명시되어있다.

“ 제4조(접근통제) ④ 정보통신서비스 제공자등은 개인정보취급자가 정보통신망을 통해 외부에서 개인정보처리시스템에 접속이 필요한 경우에는 안전한 인증 수단을 적용하여야 한다.”

“ 제4조(접근통제) ⑤ 정보통신서비스 제공자등은 정보통신망을 통한 불법적인 접근 및 침해사고 방지를 위해 다음 각 호의 기능을 포함한 시스템을 설치·운영하여야 한다.

1. 개인정보처리시스템에 대한 접속 권한을 IP주소 등으로 제한하여 인가받지 않은 접근을 제한

2. 개인정보처리시스템에 접속한 IP주소 등을 재분석하여 불법적인 개인정보 유출 시도를 탐지”

“ 제4조(접근통제) ⑥ 전년도 말 기준 직전 3개월간 그 개인정보가 저장·관리되고 있는 이용자 수가 일일평균 100만명 이상이거나 정보통신서비스 부문 전년도(법인인 경우에는 전 사업연도를 말한다) 매출액이 100억원 이상인 정보통신서비스 제공자등은 개인정보처리시스템에서 개인정보를 다운로드 또는 파기할 수 있거나 개인정보처리시스템에 대한 접근권한을 설정할 수 있는 개인정보취급자의 컴퓨터 등을 물리적 또는 논리적으로 망 분리 하여야 한다.”

이 기준에서는 개인정보가 저장·관리 될 때 이용자 숫자와 전년도 매출액을 기준으로 개인정보의 의무를 정의하고 있다. 이용자 수가 일일 평균 100만 명 이상이거나 매출액인 100억 원 이상이면 개인정보 취급자의 단말기에 접근통제를 적용하여야 한다. 기준에서는 모든 개인정보 취급자가 대상이 아니라, 개인정보를 다운로드할 수 있거나, 개인정보를 파기 또는 접근 권한 설정을 변경할 수 있는 경우에 의무를 부과하고 있다.

개인정보의 기술적·관리적 보호 기준에서 정의하는 접근통제는 망 분리 적용으로 안전한 인증 수단, 유해 접근차단 그리고 망 분리 조건을 모두 충족할 수 있다.

전자금융감독규정(금융위원회고시 제2018-36호)에 따르면 내부통신망과 연결된 내부 업무용 시스템은 인터넷 등 외부 네트워크와 차단이 되어야 한다. 그리고 정보처리시스템에 접근하는 단말기 중 운영, 개발, 보안 목적일 때 외부 통신망으로부터 물리적으로 분리되어야 한다.

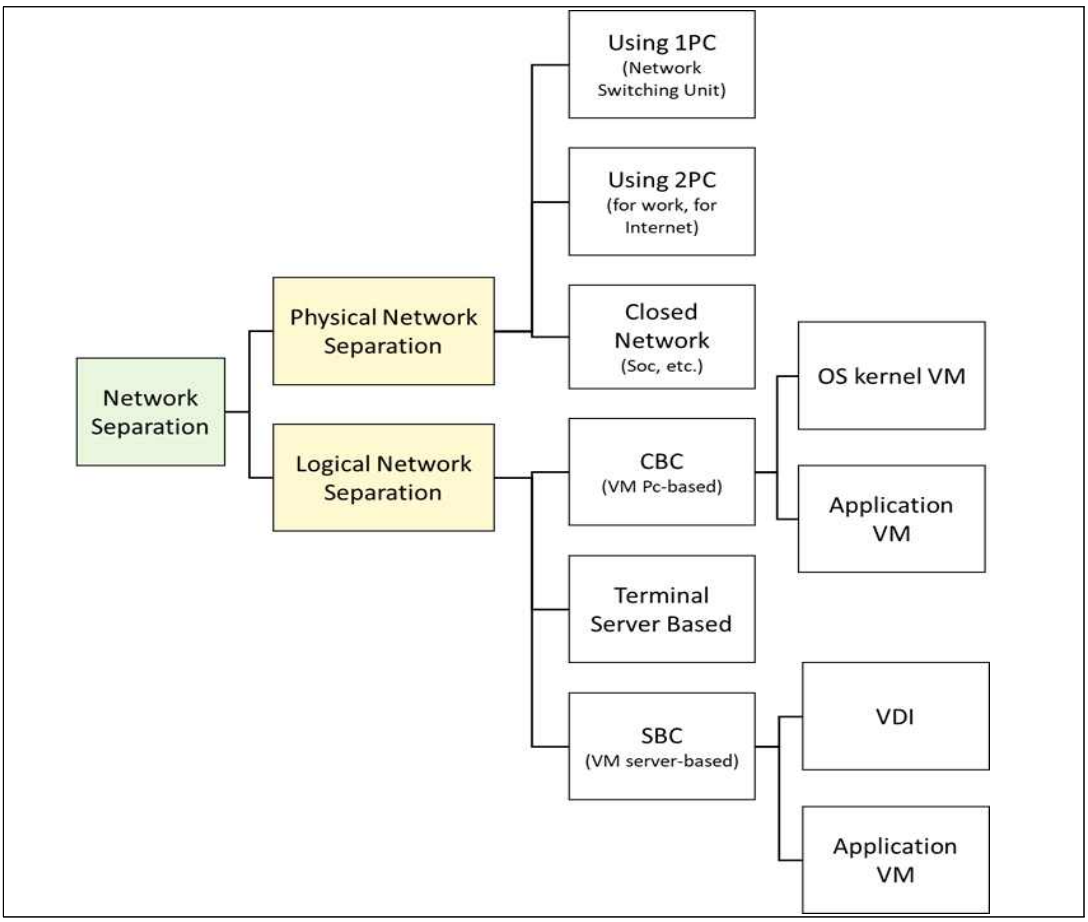
“ 제15조(해킹 등 방지대책) 3. 내부통신망과 연결된 내부 업무용시스템은 인터넷(무선통신망 포함) 등 외부통신망과 분리·차단 및 접속 금지(단, 업무상 불가피하여 금융감독원장의 확인을 받은 경우에는 그러하지 아니하다)

5. 전산실 내에 위치한 정보처리시스템과 해당 정보처리시스템의 운영, 개발, 보안 목적으로 직접 접속하는 단말기에 대해서는 인터넷 등 외부통신망으로

부터 물리적으로 분리할 것(단, 업무 특성상 분리하기 어렵다고 금융감독원장이 인정하는 경우에는 분리하지 아니하여도 된다.) “

대부분 공공기관에서는 매년 국정원으로부터 정보보안 실태평가를 받게 된다. 국정원 정보보안 실태평가에서 주요 사항으로 망 분리를 집중적으로 점검하기 때문에 실태평가 고득점을 위해서는 망 분리를 할 수밖에 없게 강제하고 있다.

위에서 살펴본 것과 같이 망 분리 환경을 구축하기 위해서는 기관이 보유한 정보시스템을 식별하고, 망 분리의 방법과 영역을 결정하여야 한다. 그래야 적합하고 안전한 네트워크 설계가 가능하다. 이런 기준이 필요한 이유는 명확한 정책과 기준이 설정된 이후에 정보시스템이 배치되어야 새로운 취약요소가 발생하지 않기 때문이다. [그림 Ⅱ-3]은 망 분리 방식에 대한 구분을 보여준다.




[그림 Ⅱ-3] 망 분리 방식과 유형 구분
(개인정보 기술적 관리적 보호조치 해설서, 방통위, KISA)

a) 물리적 망 분리 방식의 특징

물리적 망 분리는 크게 3가지 방식으로 분류된다. 「2 PC 사용」방식은 내부망용과 외부망용 PC를 각각 하나씩 사용하는 방식이다. 높은 보안성을 제공하지만, 네트워크 구축 비용, PC 구매 비용 및 유지보수 비용 등 많은 비용을 요구한다. 또한 공간이나 전기를 소모하는 양도 증가한다. 일반적으로 책상에 PC를 두 대를 놓아야 하므로 업무 장소가 협소해질 뿐만 아니라, 보안 체계를 유지하기 위해서는 별도의 추가 보안 장비가 요구된다.

「1 PC 사용」 방식은 한 대의 PC를 사용하지만, 네트워크 전환 장치를 이용하여, 내부망과 외부망을 전환하여 사용하는 방식이다. 이 방식은 높은 보안성을 제공하고, 직원들(특히, 사무직 지원)에게 책상 등 넓은 사무공간을 제공한다는 장점이 존재하지만, 높은 비용이 필요하다는(망 전환 장치 설치, 네트워크 구축 비용 등) 단점과 망 전환 시 많은 시간이 소요되거나, 업무의 흐름이 단절되어 업무효율 저하로 이어진다.

「폐쇄망 구성」 방식은 SOC(Security Operation Center) 같은 산업 시설에 사용되는 방식이다. 외부망을 제공하지 않기 때문에, 높은 보안성을 보장하지만, 인터넷이나 네트워크를 활용하는 효율적인 업무에는 부적절하여, 보안이 무엇보다 중요한 경우에 제한적으로 사용되는 방식이다[8].

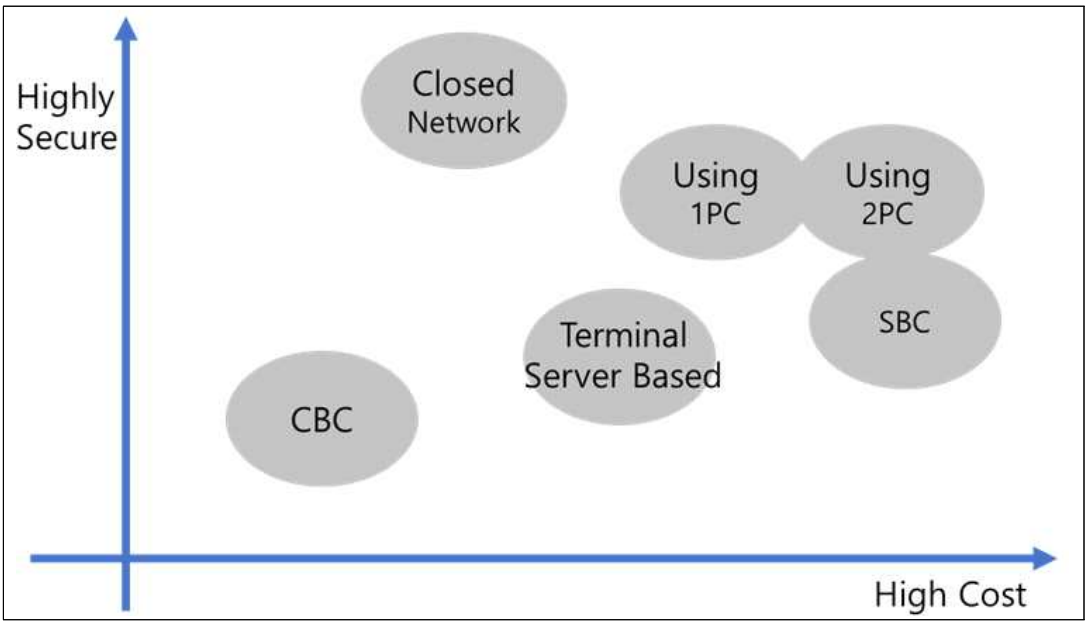
b) 논리적 망 분리 방식의 특징

논리적 망 분리는 OS(Operating System)나 커널(Kernel)의 사용 방식으로 OS 커널 가상화, Application 가상화, VDI 등 다양한 방식으로 분류되는 예도 있으나, 본 논문에서는 SBC(Sever Based Computing), CBC(Client Based Computing), 터미널 서버 기반 방식의 3가지 방식으로 분류하여 살펴보도록 한다.

「SBC 방식」은 서버에 가상화 데스크톱을 올려 서비스하는 경우를 의미한다. 장점으로는 각 사용자에게 같은 보안 정책을 일관되게 배포하고 관리할 수 있다는 이점이 있으며, 악성코드 감염을 최소화할 수 있다. 또한, 업무 데이터를 중앙 집중형으로 관리하여, 정보 유출을 최소화할 수 있다. 하지만 단점으로는 높은 비용이 필요하고 네트워크 트래픽이 증가한다. 일부 보안 프로그램이 가상화된 PC에서 동작하지 않는 경우가 있어 호환성 확보가 어려운 부분이 있다.

「CBC 방식」은 PC를 가상화로 구분하여 사용하는 방식이다. 장점으로는 낮은 비용과 사용자에게 정책을 관리하고 적용하기에 쉽다는 점이다. 반대로 단점은 보안 프로그램이 동작하기에 성능이 충분하지 않으며(OS, App 등의 호환성 문제), 다른 네트워크 장치가 있어야 구성을 할 수 있다는 점이다.

「터미널 서버 기반」 방식은 각 터미널의 보안 관리를 통합하여 관리할 수 있어 터미널 서버의 보안 수준만큼 유지하는 것이 가능하지만, SBC 방식에서의 단점인 높은 비용과 네트워크 트래픽 증가하는 단점을 같이 가지고 있다. 이 밖에도 악성코드나 취약점을 직접 대응해야 한다[8].



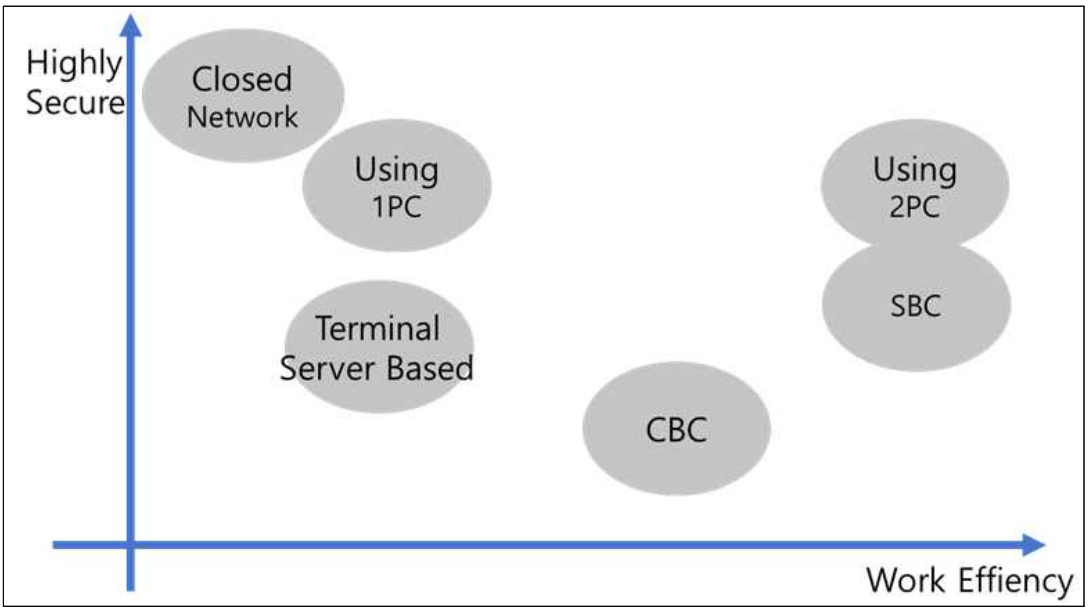

[그림 Ⅱ-4] 망 분리 방식 & 유형별 비용과 보안 수준

망 분리를 도입하고 운영하기 위해서는 다양한 기술과 비용, 효율성, 그리고 호환성 등 고려사항이 다양하다. [그림 Ⅱ-4]는 기대되는 보안 수준과 필요한 비용을 표현한 그래프이다. 대부분의 망 분리는 큰 비용이 요구된다.

그중에서도 「2 PC 사용」과 「SBC」는 특히 큰 비용을 요구한다. 하지만 물리적인 망 분리 형태인 「2 PC 사용」이 더 높은 보안성 확보가 가능하다.

비용을 줄이는 방법에는 「터미널 서버 기반」, 「CBC」 방식이 있지만, 높은 수준의 보안을 보장하기는 어렵다. 그중에서 CBC 방식은 가장 적은 비용으로 구축 가능한 망 분리 방식이다.

「폐쇄망 구성」은 내부망의 망 분리 구성과 적용되는 분야가 다르므로 본 논문에서는 논외로 하기로 한다.



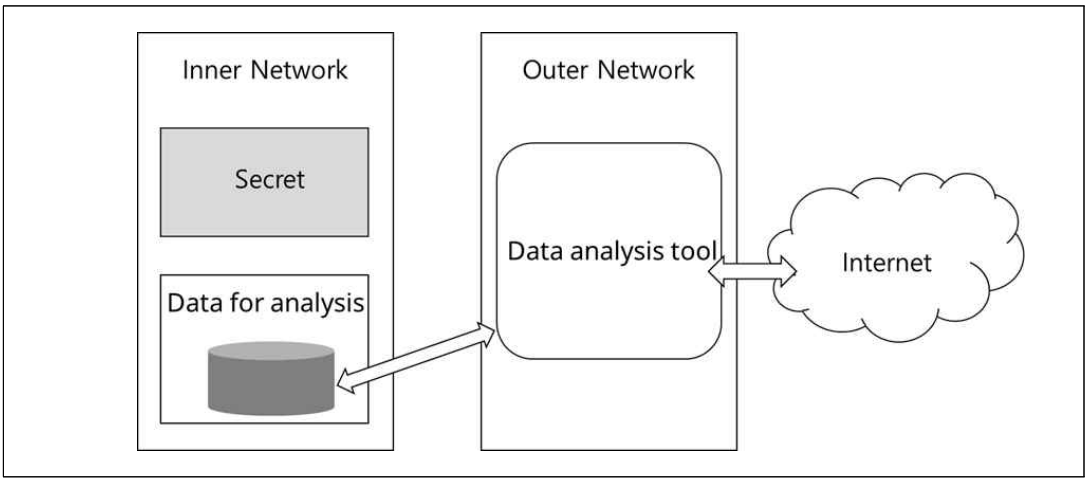

[그림 Ⅱ-5] 망 분리 방식 & 유형별 업무효율성과 보안 수준

[그림 Ⅱ-5]는 망 분리 방식 & 유형별 업무효율성과 보안 수준을 정리한 내용을 보여주는 그래프이다.

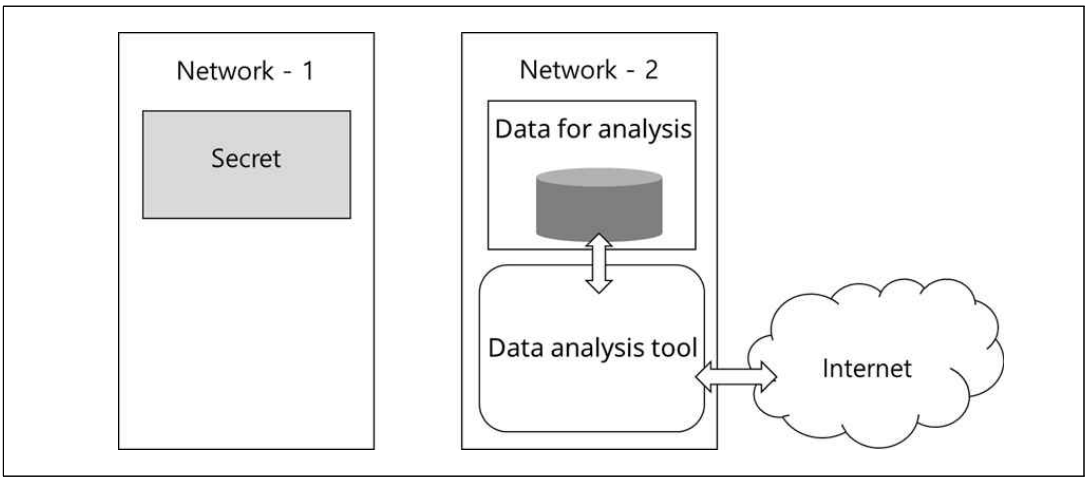
「폐쇄망 구성」은 가장 높은 보안성을 제공하지만, 업무와 관련된 자료의 전송을 허용하지 않기 때문에, 업무효율을 평가할 수 없다. 「1 PC 사용」 방법과 「터미널 서버 기반」 방식은 비슷한 업무 효율성을 제공하지만, 보안성에 대한 수준은 서로 차이가 난다.

「CBC」방식은 중간 정도의 업무 효율성을 가지며, 낮은 수준의 보안성을 가진다. 「2 PC 사용」과 「SBC」는 높은 비용이 요구되나 높은 보안 수준을 가진다. 또한, 업무 효율적인 측면에서도 높은 수준의 효과성을 보여준다. 그 이유는 「2 PC 사용」 방식의 경우 두 개의 PC를 내부망과 외부망을 각각 띄워 업무를 할 수 있기 때문이며, 서버 기반의 「SBC」역시 VM(Virtual Machine)을 윈도우 창의 형태로 구동시키기 때문에 VM을 사용하여 논리적으로 분리 PC와 분리함으로써 내부망(VM)-외부망(PC)의 구성으로 「2 PC 사용」과 같은 형태로 업무수행이 가능하다.

[그림 Ⅱ-6]는 도메인 중심으로 망을 분리할 경우를 [그림 Ⅱ-7]는 데이터 중심의 망 분리의 구성을 예시로 보여준다. 현행의 망 분리 제도에서 망 분리가 도메인 중심의 망 분리라고 할 수 있으며, 이런 구성은 데이터와 분석 도구가 분리되어 데이터 활용에 비효율적인 부분이 존재한다. 그리고 기밀 데이터와 일반 데이터가 같은 망에 있어, 데이터를 이용하기 위해 망에 접근한 사용자가 기밀 데이터에도 접근할 가능성이 있어 보안 위협 요소가 발생할 수 있다. 도메인 중심의 망 분리 기법에서는 클라우드, 스마트워크, 오픈소스 활용 등 통신기술 활용이 불가능하다. 하지만 데이터 중심의 망 분리는 데이터와 데이터 분석 도구가 같은 공간에 있어 효율적인 데이터 활용 및 분석을 할 수 있어, 혁신적인 데이터 분석이나 아이디어를 형상화하여 추진하는 것이 가능한 구조이다. 기밀 데이터를 일반 데이터와 분리하기 때문에 기밀 데이터의 방어전략을 추진하기에도 쉬우며, 보안 위협으로부터 상대적으로 안전하지만, 분석데이터는 도메인 중심의 망 분리 구조보다 취약해진다. 분석데이터에 개인정보나, 가공하기 전 상태의 기밀 데이터 소스가 존재하는 경우, 한정된 예산을 이용하여 방어전략을 추진하는 데 많은 어려움을 가진다. 하지만, 클라우드, 스마트워크, 오픈소스 등 신기술의 활용이 가능하므로 도메인 중심의 망 분리와 비교해 효율적이다.




[그림 Ⅱ-6] 도메인 중심의 망 분리

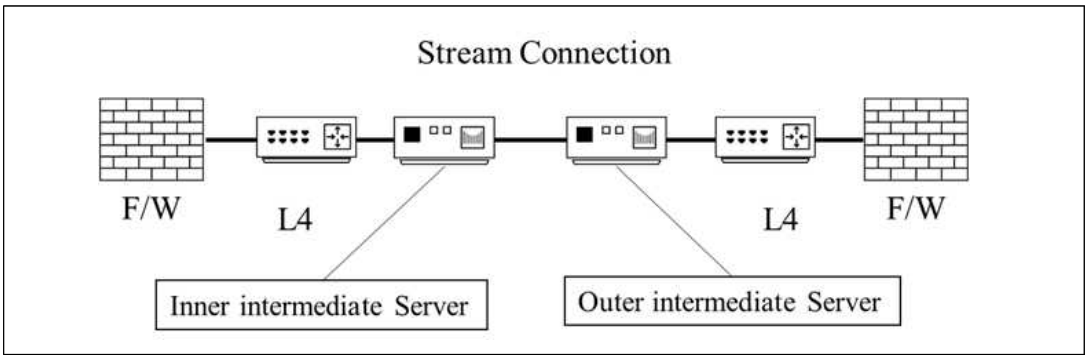



[그림 Ⅱ-7] 데이터 중심의 망 분리

망 분리는 구현하는 기술 방식에 따라 ① 2대의 PC를 이용한 PC 이중화, ② 망 전환 장치를 활용한 망 분리, ③ 서버 기반 논리적 망 분리 그리고, ④ 컴퓨터 기반의 논리적 망 분리 방식이 존재한다.

망 분리 구성은 산업시설제어 망에 필요한 완전히 차단된 네트워크를 사용하는 것이 아니다. 필요한 메일이나 업무에 필요한 정보를 내·외부에 전송하여야 한다.

[그림 Ⅱ-8]처럼 망 분리 환경에서 안전한 방법으로 내·외부 자료를 전송하기 위한 기본적인 구조를 가진다.




[그림 Ⅱ-8] 기본적인 망분리 환경의 망 연계 장치의 구조

내부망에 존재하는 PC에서 외부망에 연결되어있는 PC로 자료를 옮길 수 있는 초기 방법은 [그림 Ⅱ-9]과 같이 USB(Universal Serial Bus)를 이용한 이동이다.

이 방식에는 다양한 보안 위협이 존재하고 있다. USB를 허용하면, 업무용 자

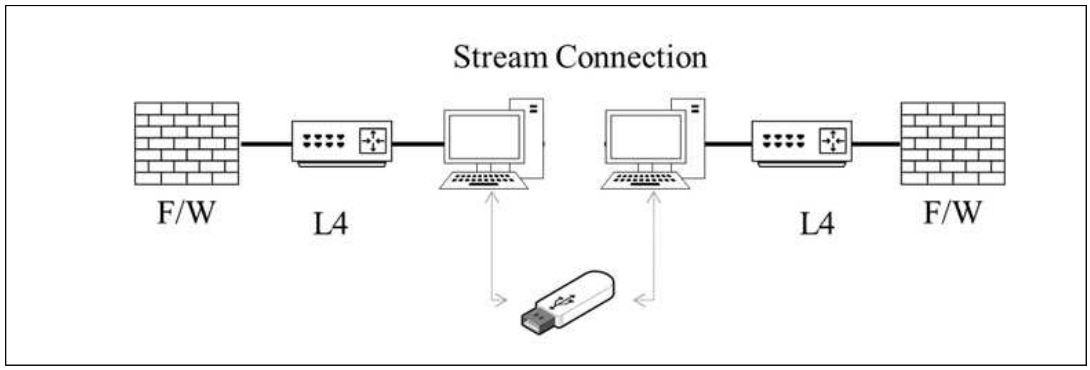
료의 유출을 제어하지 못하고, 인터넷의 악성코드가 여과 없이 내부망으로 들어올 수도 있다. 이 밖에도 물리적인 USB 분실의 가능성도 존재한다. 이를 보완하기 위한 보호 대책으로 다음과 같은 방식은 적용할 수 있다.

① 매체제어 장치를 통해 허용되지 않은 USB는 접근을 차단한다.

② 보안 USB 저장장치만 허용한다. 보안 USB 저장장치란 저장되는 데이터를 암호화 저장하여, 분실이 되어도 접근 ID/PW를 모르는 경우 내부의 내용 유출을 방지할 수 있는 USB 저장장치를 말한다.

③ 매체제어 장치에 악성코드를 탐지하는 스캐닝 기능을 추가하여, USB 담긴 파일의 악성코드 여부를 확인한다.

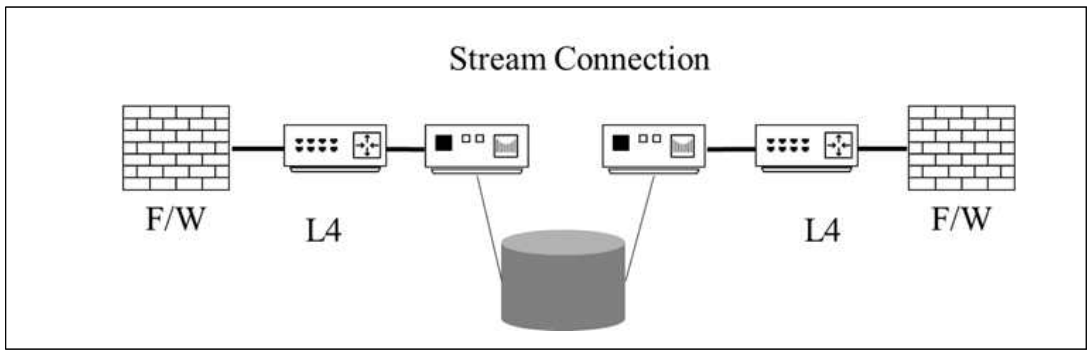
이 밖에도 안전한 USB 저장장치를 이용한 망간 자료전송을 위해서는 USB 사용 이력을 기록하고 주기적으로 점검하는 정보보안의 기본 활동이 필요하다.




[그림 Ⅱ-9] 보안 USB를 이용한 망간 자료전송 방법

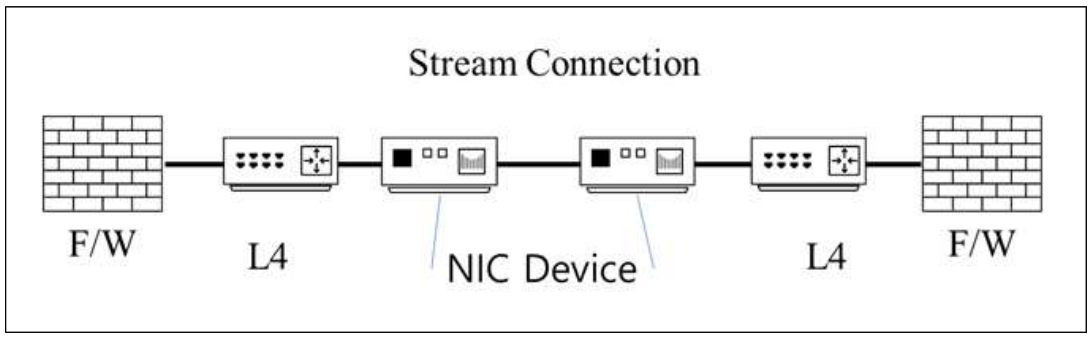
초기 망 분리는 대부분 물리적 망 분리를 적용하였기 때문에 내·외부망 자료전송에 보안 USB를 사용하였지만, 이후 실시간 자료를 전송하기 위한 스토리지형 망 연계 방식으로 발전되었다. [그림 Ⅱ-10]은 스토리지형 망 연계 시스템의 방식이다. 이 방식은 망 사이에 공유 스토리지를 구성하여, 스토리지를 이용하여 자료나 데이터를 전송하는 방식으로, 전통적인 망 연계 방식으로 물리적인 보안성이 우수하여, 군사시설 등 높은 보안이 요구되는 기관에서 사용되지만, 초기 구축 시 높은 비용이 필요하다는 단점이 존재한다. 스토리지는 내부영역과 외부영역을 구분하여 읽기/쓰기의 권한을 분류하여 두 영역의 전송사용자가 사용할

수 있게 구성되었다. 하지만 외부망을 통해 접근한 공격자에 의해서 공유 스토리지가 해킹당했다고 가정하였을 때, 접점이 발생하여, 내부 자료 유출의 보안 위협이 존재한다.

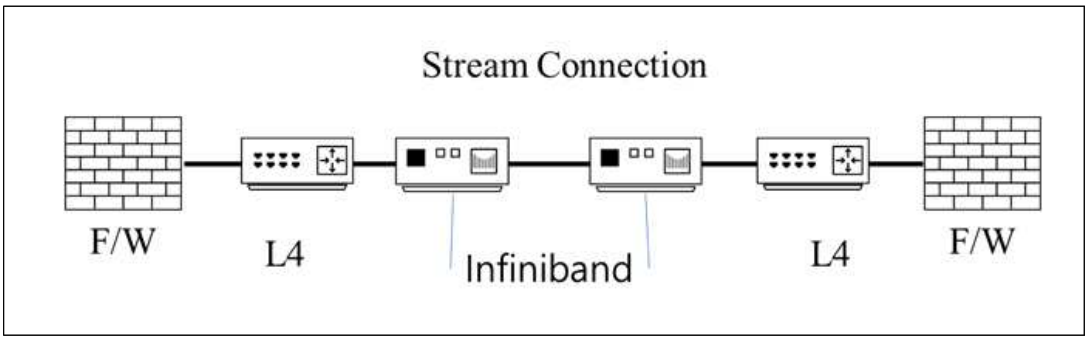



[그림 Ⅱ-10] 스토리지 형 망 연계 시스템 구조

스트리지형 망 연계 시스템 이후에 실시간 연동이 가능한 비 스토리지 형 망 연계 장치가 개발되었다. 두 개의 PC 혹은 서버 간에 통신을 허용한다. 통신은 non-TCP/IP로 정의되지 않은 방식의 통신을 사용한다. [그림 Ⅱ-11]와 [그림 Ⅱ-12]은 동일한 구성이지만, 망 연계 방식이 하나는 소켓 통신을 지원하고 하나는 고성능 통신인 인피니 밴드(InfiniBand)를 지원하는 부분에서 차이가 난다. NIC(Network Interface Card) 디바이스와 소켓 통신을 수행하는 구조에서는 물리적 보안성은 낮지만, 구축 비용이 저렴하여 소규모 기관에서 사용하기 적합하고, 인피니 밴드는 전송 응답속도가 높은 만큼 대규모 기관이나. 대국민 서비스를 시행하는 기관에서 구축할 수 있다.




[그림 Ⅱ-11] NIC 디바이스 망 연계 장치의 구조




[그림 Ⅱ-12] 인피니 밴드 망 연계 장치의 구조

인피니 밴드는 사실상 표준으로 사용되고 있는 고속통신 기술 중 하나라고 하며, 채널 기반 데이터 통신을 호스트당 1,600만 개의 통신을 생성하여, 10~100Gb/s의 전송 성능을 제공한다[9]. 인피니 밴드는 대규모 데이터를 처리하는 빅데이터, 딥러닝, 인공지능의 다양한 분야에서 사용되고 있는데, 두 개의 노드를 하나의 노드처럼 붙여 ACL 설정을 통해, 망은 연계하는 방식으로 사용된다.

## 2.3. 웹 스크래핑(Web Scraping) 기술

웹 스크래핑은 웹사이트에서 원하는 정보를 컴퓨터나 소프트웨어 기술로 추출하는 것을 의미한다. 우리는 인터넷을 통해 원하는 정보를 추출하여, 별도의 데이터로 만들어, 새로운 기획에 활용하거나, 현재 있는 데이터를 분석하기 위해 활용할 수 있다. 웹에서 원하는 데이터를 추출, 복사 또는 수집하는 과정을 컴퓨터나 소프트웨어가 자동으로 수행하게 만드는 기술이 바로 웹 스크래핑이라고 할 수 있다.

최근에는 웹 스크래핑 기술을 활용하여, 데이터 수집 및 분석, 정보 및 의사결정 및 연구 관련 활동으로 과학, 기술, 경영 분야에 광범위하게 사용히고 있다.

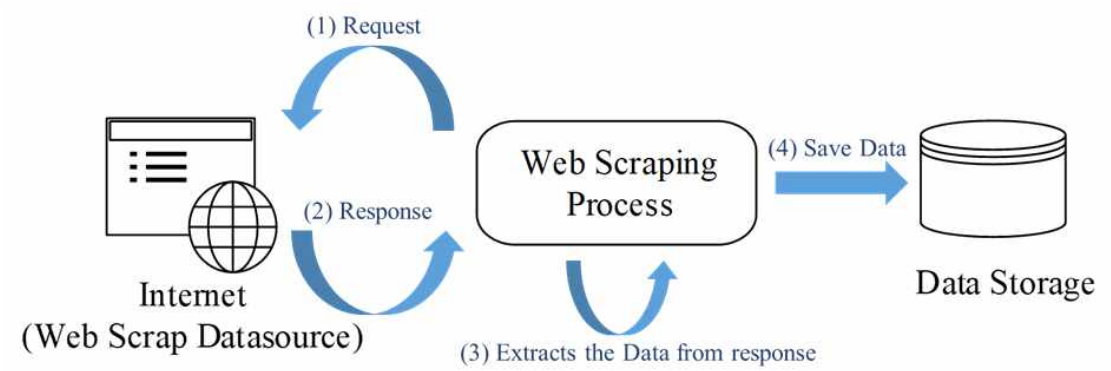
일반적으로, 웹 스크래핑과 웹 크롤링은 의미를 교환하여 사용하기도 하지만, 본 논문에서는 웹 크롤링은 웹 사이트에 있는 링크를 탐색하여 탐색의 범위를 분류하는 과정으로 정의하고, 웹 스크래핑은 웹 사이트에서 콘텐츠를 추출하는 과정으로 정의한다.

웹 스크래핑의 과정을 이해하기 위해서 사용자가 웹 사이트에 접근하여 정보

를 수집 하는 과정을 살펴볼 필요가 있다. 먼저, 웹 브라우저 실행하고, 원하는 사이트의 URL을 입력한다. 웹 브라우저는 해당 URL을 해석하기 위해서 DNS(Domain Name System)에 쿼리를 날리게 되는데, 쿼리를 받은 DNS는 URL에 해당하는 IP를 응답하게 된다. 브라우저는 응답 받은 IP로 웹페이지를 요청하고, 웹서버로부터 응답받은 내용을 브라우저를 통해 웹페이지로 구성하는 과정이다. 사용자는 필요한 경우에 해당 웹페이지를 콘텐츠 형태로 변환하여 저장하거나, 데이터로 만들 수 있다.

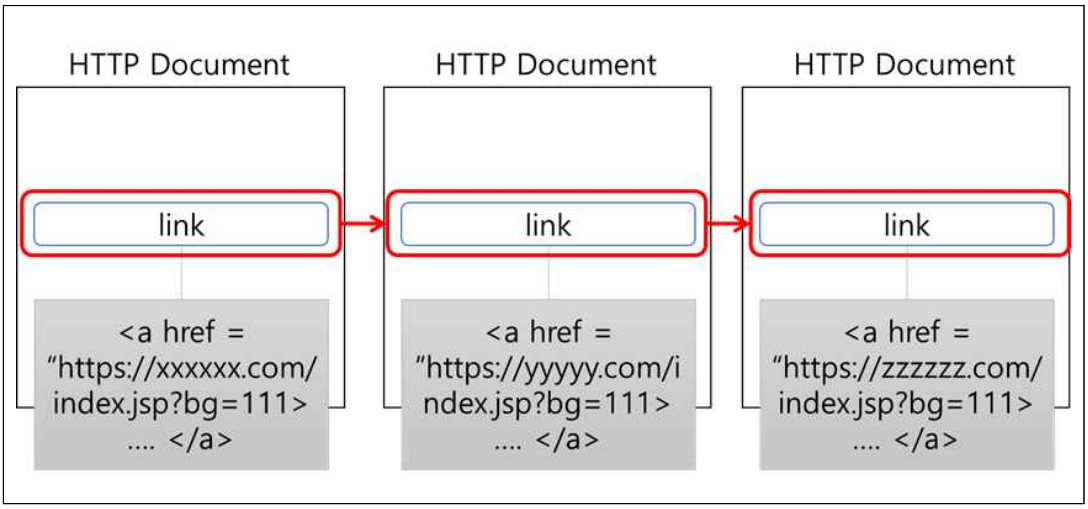
웹페이지의 내용을 저장하는 방법은 화면을 캡처하는 방법이 가장 쉽게 사용할 수 있는 방법이지만, 텍스트나 링크가 이미지로 고정 되어 저장된 이미지에서 사용할 수 없는 단점이 있다. 이미지가 아닌 HTML, PDF, JSON 등의 형태로 변환 저장하면, 웹의 형태를 유지하게 시켜 텍스트나 링크를 활용할 수 있는 형태로 변환 할 수 있다. 지금 까지 설명한 웹에서 콘텐츠를 저장하는 방식을 [그림 Ⅱ-13]에서 보는 것과 같이 자동화 처리하는 방법을 웹 스크래핑이라고 하며, 이런 스크래핑을 하는 프로그램을 스크래퍼(Scraper)라고 한다.

스크래퍼의 동작 방식은 특정 웹 사이트에 HTTP GET 요청을 보낸다. 웹 사이트가 응답하면 스크래퍼는 HTML 문서를 분석해서 특정 패턴의 데이터를 찾는다. 추출된 데이터는 스크래퍼가 원하는 특정 형식으로 변환하여 저장된다.

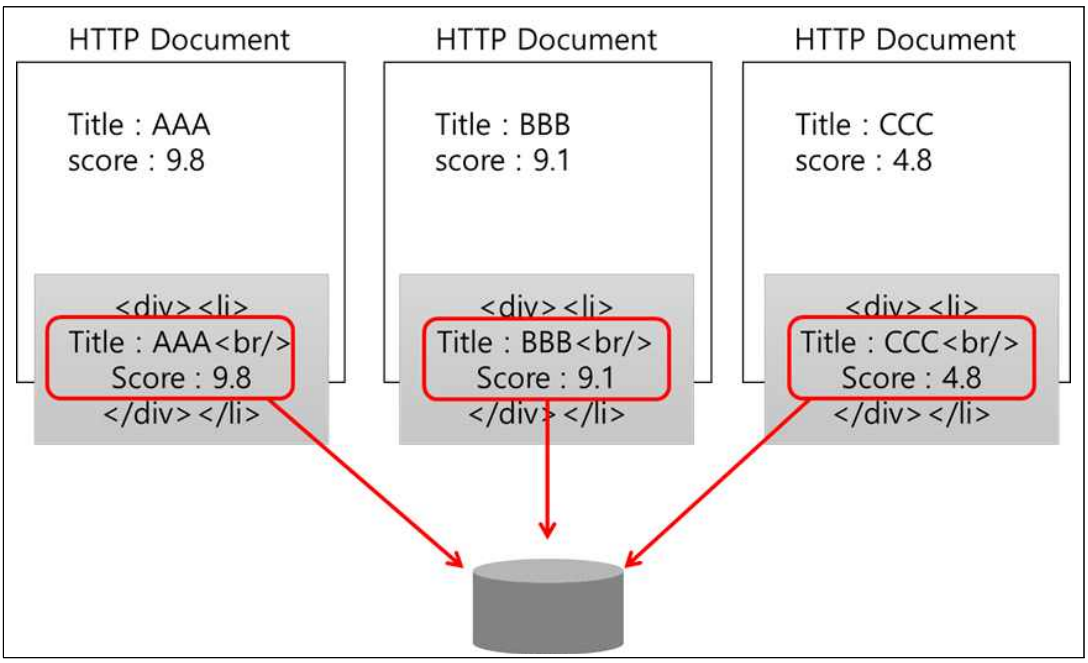



[그림 Ⅱ-13] 웹 크롤링과 스크래핑 동작 방식 다이어그램

[그림 Ⅱ-14]와 같이 웹 크롤러는 웹페이지에 있는 링크 파일을 대기 큐(Waiting Queue)에 넣고 원하는 동작이 종료되면, 대기 큐에 남아 있는 링크로 가서 다시 링크를 수집하는 동작을 수행한다. 웹 스크래퍼는 필요한 정보나 내용을 웹페이지에서 [그림 Ⅱ-15]처럼 수집하는 역할을 하게 된다.




[그림 Ⅱ-14] 웹 크롤러의 동작 방식




[그림 Ⅱ-15] 웹 스크래퍼의 동작 방식

## 2.4. 헤들리스 브라우저(Headless-browser)

헤드리스 브라우저(Headless Browser)의 정의는 「GUI(Graphical User Interface) 환경 없는 브라우저」이다. 일반적으로 사용되는 브라우저(IE, 크롬, 파이어폭스 등)의 구동 방식은 크게 브라우저 엔진과 렌더링 엔진으로 구분된다.

브라우저 엔진은 웹 브라우저의 핵심 구성 요소이다. 브라우저 엔진은 사용자 인터페이스와 렌더링 엔진 사이에 인터페이스를 통한 중개자 임무를 수행하며, 사용자 인터페이스에서 받은 입력을 렌더링 엔진을 전달하고 처리한다. 익숙하게 사용되는 브라우저 엔진은 [표 Ⅱ-2]와 같다[10].

[표 Ⅱ-2] 자주 사용되는 브라우저 엔진[10]

| 이름 | 설명 |
|---|---|
| 게코(Gecko) | 모질라 재단에서 만든 레이아웃 엔진으로 파이어폭스, 모질라 선더버드, 시몽키 등이 이를 탑재 |
| 블링크(Blink) | 웹키트에서 파생된 레이아웃 엔진으로 크롬, 오페라 등이 이를 탑재 |
| 트라이던트 (Trident) | 마이크로소프트의 레이아웃 엔진으로 인터넷 익스플로러, 아웃룩 익스프레스, 마이크로소프트 아웃룩, 그리고 윈앰프, 리얼플레이어의 미니 브라우저 등이 이를 탑재 |
| 웹키트(Webkit) | KHTML에서 파생된 레이아웃 엔진으로 사파리 등이 탑재 |

이름에서 알 수 있듯이 이 구성 요소는 사용자가 요청한 특정 웹페이지를 화면에 렌더링하는 역할을 한다. CSS(Cascading Style Sheets)를 사용하여 스타일이 지정되거나 형식이 지정된 이미지와 함께 HTML 및 XML 문서를 해석하고 최종 레이아웃이 생성되어 사용자 인터페이스에 표시된다.

요청된 HTML 페이지는 렌더링 엔진에 의해 외부 CSS 파일이나 스타일 요소를 포함한 청크(Chunk)로 구문 분석됩니다. 그런 다음 HTML 요소는 DOM(Document Object Model) 노드로 변환되어 "콘텐츠 트리" 또는 "DOM 트리" 를 형성한다.

동시에 브라우저는 렌더 트리도 생성하는데, 이 트리에는 스타일 정보와 요소가 표시되는 순서를 정의하는 시각적 지침이 모두 포함된다. 렌더 트리는 콘텐츠가 원하는 순서로 표시된다.

또한 렌더 트리는 레이아웃 프로세스를 거치는데, 렌더 트리가 생성될 때 위치 또는 크기 값이 할당되지 않는다. 원하는 위치를 평가하기 위한 값을 계산하는 전 과정을 레이아웃 과정이라고 한다. 이 과정에서 모든 노드에 정확한 좌표가 할당되며, 이렇게 하면 모든 노드가 화면의 정확한 위치에 표시된다.

마지막 단계는 화면을 그리는 단계이다. 여기서 렌더 트리가 탐색 되고 렌더러의 paint() 메서드가 호출되어 UI 백엔드 레이어를 사용하여 화면의 각 노드를 그리는 과정을 수행한다.

헤드리스 브라우저는 실제 브라우저 애플리케이션 없이 사용자 인터페이스로 실행되는 브라우저이다. 그러므로 더 빠르고, 더 적은 메모리를 소비하면서 동작하게 되며, 자동화된 동작을 통해서 부하 없이 안정적인 연결이 가능하다. 결론적으로 말하면 실제 브라우저보다 속도가 빠르다고 할 수 있다.

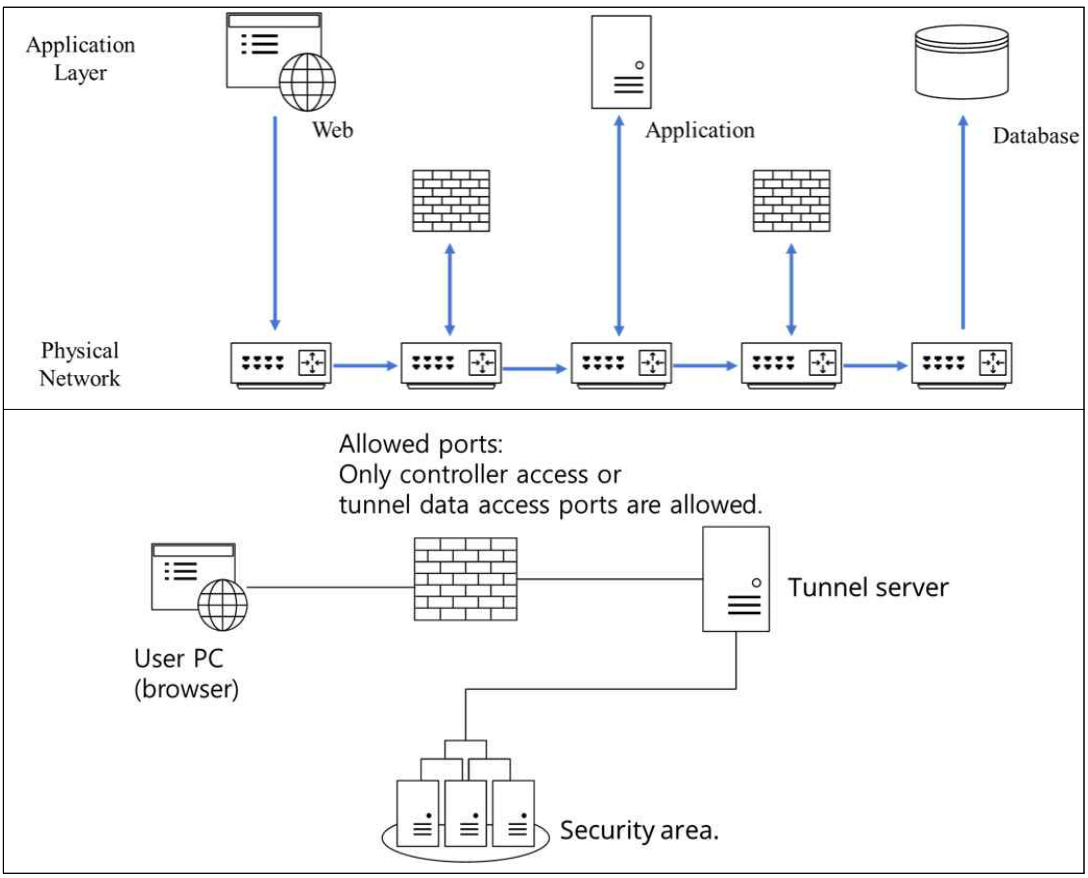
헤드리스 브라우저는 인터넷의 자료를 수집, 활용하는 방법으로 이용되며, 개발자가 필요한 방향으로 설계 활용이 가능하다. 가장 일반적으로 사용되는 스크래핑이나 크롤링 관련 라이브러리는 selenium과 python의 requests 이며, 이 두 개의 라이브러리와 비교자료는 [표 Ⅱ-3]와 같다.

[표 Ⅱ-3] requests 과 selenium 비교

| 플랫폼 | Reauests | Selenium |
|---|---|---|
| 주요 동작 | 웹페이지(html) 읽어오기 | 웹페이지 자동화 |
| 속도 | 빠르다 | 느리다 |
| 동적 웹페이지 | O | X |
| 기타 | 주어진 URL을 통해 받아온 html에 원하는 정보가 있는 경우 사용 (Restful) | 크롬 버전에 맞는 Chromedriver.exe가 필요 |
| 사용 방법 | res.raise_for_status() | find_elements(s)_by_id<br>find_elements(s)_by_x path 등 |

## 2.5. 네트워크 세분화

ISO 27001에서 정의하는 네트워크 관리방안은 네트워크 세분화(Network Segmentation)이다. 보안과 성능 그리고, 유용성을 위해 네트워크를 작은 단위로 나누어 구성한다. 일반 데이터와 기밀 데이터를 분리하여, 공격자가 모든 네트워크에 연결된 자원에 접근할 수 없도록 한다. ISO 가이드에서 설명하는 네트워크 세분화 구성은 [그림 Ⅱ-16] 에 보는 것처럼 같이 세 가지 영역으로 구분되어 운영된다[11].

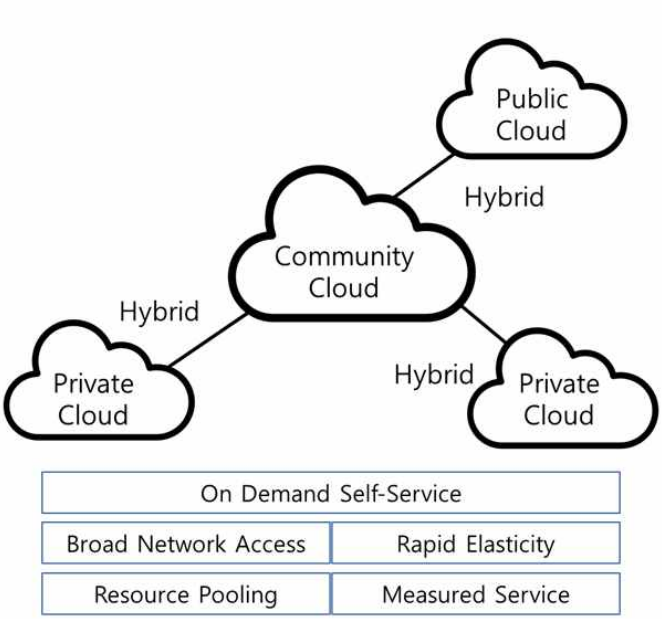



[그림 Ⅱ-16] (상)ISO 27001 세분화 네트워크 구조, (하)세부 구성방안[12]

위 [그림 Ⅱ-16]과 같이 네트워크 방화벽을 사용하여 DMZ(Demilitarized Zone)에 포함된 하위 보안 도메인과의 트래픽을 제한하는 방법에 대한 예를 제공한다. 네트워크 분리는 스위치에 구성된 VLAN(Virtual Local Area Network)을 사용하여 수행할 수 있다. 하지만, ACL(Access Control List)을 설정하여야 허용되지 않은 접근을 통제해야 한다. ISO 27001에서 네트워크 세분화는 조직 단위, PC, 서버 등의 액세스를 기반으로 보안 도메인을 설정하는 것을 의미하고, ISO 27001은 이것이 논리적이거나 물리적으로 완전히 분리된 도메인일 것을 권고하고 있다[12].

## 2.6. 클라우드 서비스

클라우드 운영 모델은 특별한 보안 요구 등 특별한 공동의 목적을 위해 구성한 커뮤니티 클라우드(Community Cloud)로 정의한다. 커뮤니티 클라우드에 대한 전개 모델은 Private Cloud와 Hybrid Cloud, Public Cloud가 있으며, 커뮤니티 클라우드는 중간 인프라 역할을 하여 각 클라우드의 중계를 통해 자원을 공유할 수 있는 구조를 조직하게 한다.




[그림 Ⅱ-17] 커뮤니티 클라우드 구성방안

[그림 Ⅱ-17]처럼 커뮤니티 클라우드는 인프라를 공유하는 구조이며, 이것은 다양한 정보시스템이나 다른 클라우드와 연결 관계를 제어하고 사용한다. 이를 통해 애플리케이션의 보안, 정책 및 효율성 요구사항을 관리하는 역할이 제공되며, 이를 통해 클라우드컴퓨팅의 비용 절감을 실현하게 된다.

커뮤니티 클라우드의 목표는 클라우드에서 비즈니스 프로세스를 통합하는 것이다. 이런 기능은 동시에 하이브리드 배포 모델을 통해 높은 수준의 보안을 유지한다. 커뮤니티 클라우드를 통해 시스템 및 서비스에 접근하기 때문에 보안성과 효율성을 보장하게 된다.

# Ⅲ. 제안하는 프레임워크

## 3.1. 요구사항 분석

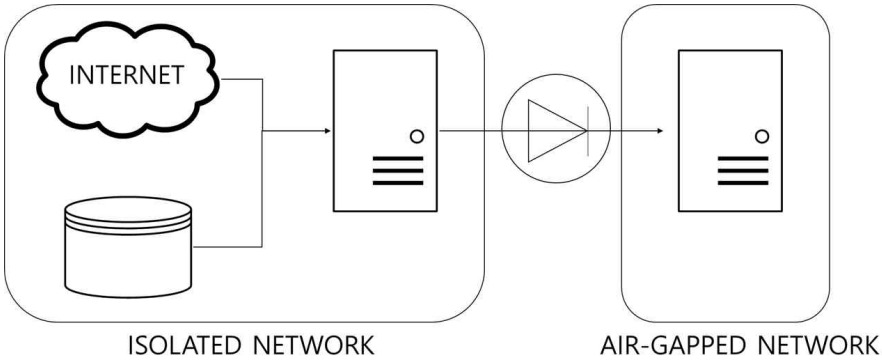
본 장에서는 네트워크 망 분리 구성과 망 연계 장치를 분석하고, 웹 스크래핑 기술을 통해 어떻게 망 분리 환경에서 브라우징이 가능한지에 대한 부분을 구성하고, 스크래핑을 통한 웹의 안전성을 분석하기 위한 취약점에 대해서 정의한다.

a) 망 연계 장치의 분석

SCADA(Supervisory Control And Data Acquisition) 혹은 ICS(Industrial Control Systems)이라고 불리는 제어 컴퓨터 시스템은 폐쇄망으로 운영되거나 일방형 데이터 전송 장치가 있는 환경으로 구축된다. 이런 네트워크 구조에서는 완전 차단을 기본으로 하기 때문에, 외부 위협으로부터 안전한 부분이 있지만 반면에 취약한 부분도 존재한다.

첫 번째로 시스템의 OS를 업그레이드 할 수 없다. OS 패치 파일을 오프라인 상태에서 패치를 진행하는 과정도 쉽지 않은 과정이지만, 시스템 간의 연결이나 시스템 위에서 작동할 프로그램과의 호환성 문제가 더 크기 때문에 OS 패치가 어렵다. OS 업그레이드를 통해 얻는 이득이, 제어시스템 장애로 이어져 안전에 문제가 발생할 경우 생기는 문제 보다 작기 때문에 OS를 업그레이드 하지 않는다.

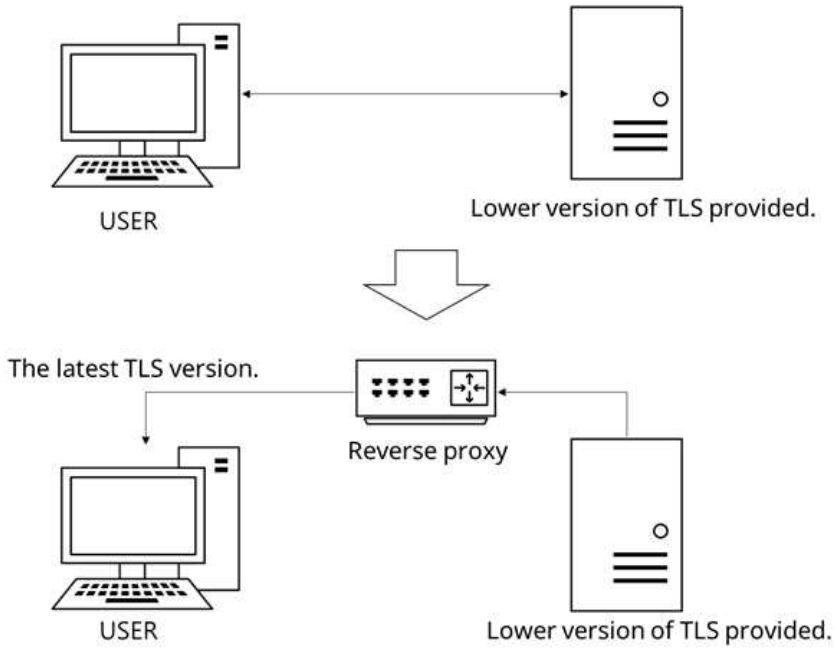
두 번째는 솔루션 제품들에 대한 취약점을 패치 하거나 보완하기가 어렵다. SCADA 구조에서도 상용 솔루션들을 사용하는데, 폐쇄적인 호환성이 맞춰진 시스템 구조에서 일부 시스템을 패치를 하거나 수정하는 작업은 또 다른 위험이 발생시킬 수 있다.




[그림 Ⅲ-1] SCADA 구조에서 사용되는 AIR GAP 구조

[그림 Ⅲ-1]처럼 SCADA 구조에서는 네트워크 간에 에어 갭을 만들고 에어 갭을 연결하기 위해서는 데이터 다이오드(Data Diode)라고 불리는 일방향 장치를 통해서만 통신한다. 즉 왼쪽에서 오른쪽으로 데이터 전달은 가능하지만, 그 반대는 불가능한 구조이다. OS나 보안 시스템 패치가 원활하진 않지만, 물리적으로 차단된 구조는 훌륭한 방어책이 될 수 있다[13].

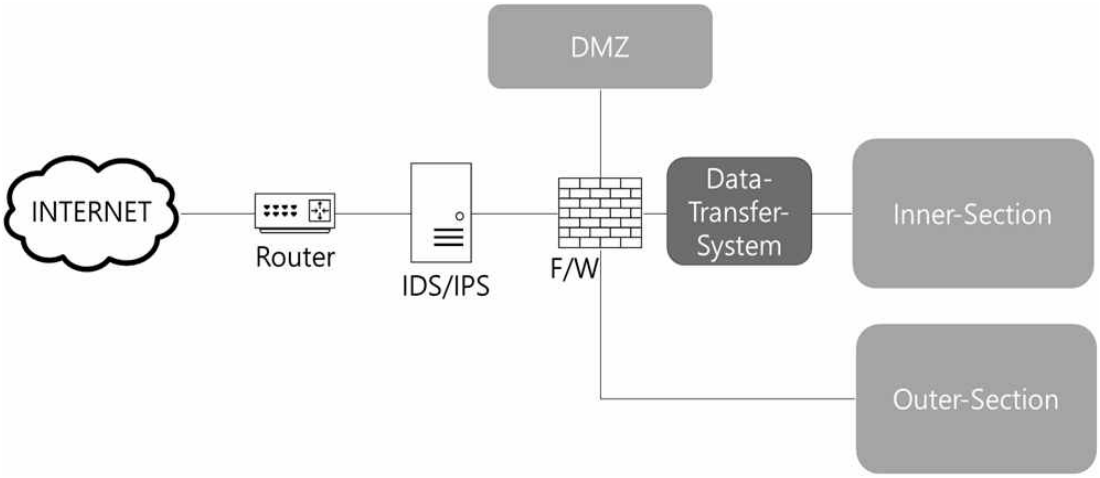
완전 차단 구조에서 일부 통신 취약점을 보완하기 위해 [그림 Ⅲ-2]처럼, BITW(Bump in the Wire)를 통해 해결하고자 하는 시도가 있었다. 이는 기존 중단 간 네트워크 구조를 변형하지 않은 상태로 통신의 무결성, 기밀성, 신뢰성을 향상하게 시키기 위한 시스템 변형 방법이다. 암호화 보안 프로토콜인 TLS(Transport Layer Security)는 주기적으로 최신 버전을 발표하는데, 이전 TLS에서 나타난 취약점을 보완하고 시대에 맞는 암호 통신 인증을 제공하기 위함이다. 하지만, SCADA 구조에서는 최신 TLS를 적용하기 위해서는 OS 버전을 상위 버전으로 올리고, 솔루션이나, 통신 규칙의 최신화를 진행해야 하는데, 위에서 살펴본 것처럼 호환성의 문제가 발생하여 가용성에 문제가 발생할 수 있다. 이런 문제를 해결하기 위해서 프록시(Proxy) 서버를 삽입하는 것이다. 낮은 TLS 버전으로 통신하는 것은 문제가 되지만, 해당 통신 노드 앞쪽에 최신 버전을 활용할 수 있는 서버를 놓아 리버스 프록시(Reverse Proxy) 방식으로 구성을 하면 사용자와의 통신에서는 안전한 방법의 통신이 가능해진다.




[그림 Ⅲ-2] BITW(Bump in the Wire) 예시

위에서 SCADA와 같은 제어시스템을 위한 폐쇄 구조에서 에어 갭을 이용한 망 분리를 살펴보았다. 본 논문에서 말하는 망 분리는 앞서 살펴본 망 분리 기법과는 구조가 다르다. 업무 중심으로 판단하여 업무영역을 결정한다. 업무영역을 결정하기 위해서는 우선 정보시스템 자산의 현황을 파악하는 것부터 시작되어야 한다. 망 분리 환경의 도입을 위해서는 시스템과 소프트웨어 그리고 데이터를 기준으로 중요도 현황을 파악하여야 한다.

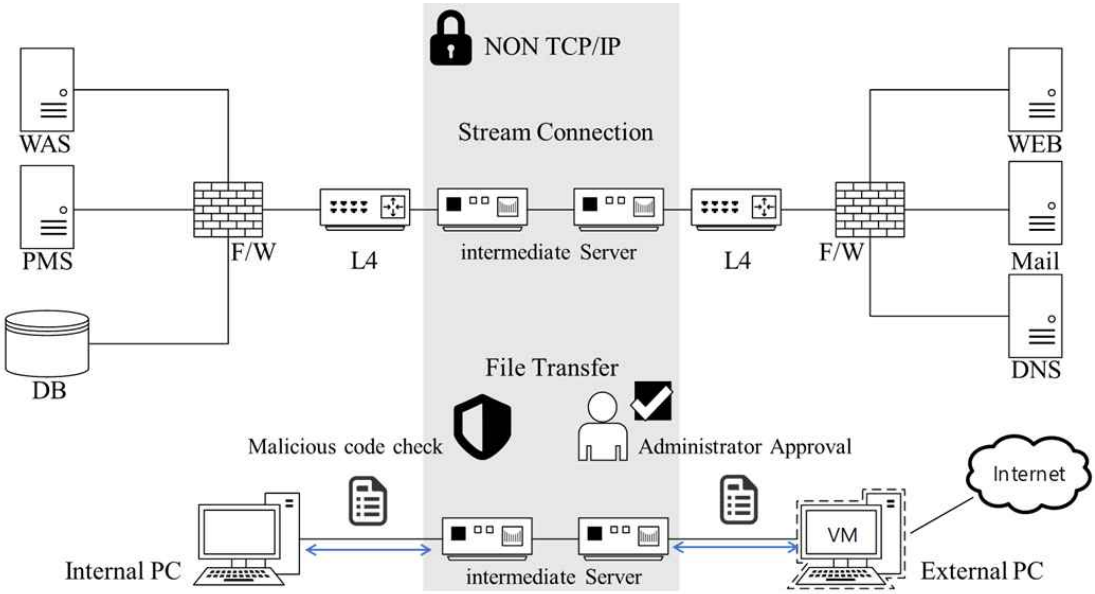
정보자산의 중요도 평가를 정보보안의 3요소인 기밀성, 무결성, 가용성을 기준으로 수행하여, 중요도 등급을 결정한 이후에 가용성 평가 프로세스를 통해 내부망 또는 외부망에 위치할 서비스와 정보자산을 결정하여야 한다. 이때, 「업무 효율성」과 「업무 적절성」을 기관의 기준에 맞게 설정하여, 결정할 필요가 있다. 영역이 결정되면 [그림 Ⅲ-3]과 같이 일반적인 기관 네트워크 구성도를 갖게 된다.




[그림 Ⅲ-3] 물리적 망 분리 네트워크 구성도

[그림 Ⅲ-3]에서 보는 것처럼 인터넷 영역에서 방화벽까지는 기관의 보안 정책에 맞는 보안 장비를 설치하게 된다. 필수로 IPS/IDS가 설치되며, DDoS 장비나, APT, 유해사이트 차단 장비는 기관의 필요에 따라 설치할 수 있다. 방화벽을 기준으로 세 개의 망이 분리되는데 첫 번째는 DMZ 영역이다. DMZ 영역은 웹서버나 메일서버가 위치하며, 그 밖에 인터넷과 연결이 필요한 시스템이 위치되어 공개자료가 포함된 서버를 구성한다. 대부분의 서비스가 웹 기반 서비스를 하기 때문에, 기관에 따라 웹 방화벽을 하나 더 놓은 구조로 운용할 수 있다. 두 번째는 내부망 영역이다. 내부망 영역은 업무를 처리하는 자료에 접근할 수 있는 시스템을 구성한다. 내부망에 접근하기 위해서는 스트리밍으로 불리는 네트워크 일방향 차단 시스템 이나, 특정 포트만 허용해 주는 제한적인 양방향 시스템을 위치하고 내부망에 대한 통제를 수행한다. 세 번째는 외부망 구성이다. 내부망 외에 외부망에서 사용되는 솔루션, 시스템을 위치하며, 인터넷 PC의 에이전트들을 컨트롤 하는 역할을 수행한다. DMZ 영역과 다른 점은 DMZ는 대외 서비스를 위해 구성하는 영역인 반면, 외부영역은 기관 내 외부망을 관리하기 위한 관리 서버가 위치한 장소라는 점이다. 외부망에 주로 위치하는 시스템은 PC를 관리하기 위한 정보시스템으로, 백신 서버, PMS(Patch Management System), 매체 제어 등이 존재한다.

망 분리가 적용된 기관에서, 비 보안 영역과 보안 영역 사이에 데이터나 파일을 전송할 수 있는 서비스는 필요하다. 망 분리는 SCADA처럼 산업제어를 위해 차단된 네트워크를 사용하는 것을 의미하지 않는다. 보안은 유지를 위해 내부망과 외부망을 차단하고 있지만, 업무 필요에 따라 절차에 맞게 외부 자료를 활용할 수 있다. 이를 안전하게 수행하기 위해서는 안전한 망 전송 시스템이 필요하다.




[그림 Ⅲ-4] 망 분리 장치를 통한 정보시스템 구성도

[그림 Ⅲ-4]처럼 자료연계 장치는 일반적으로 네트워크에서 안전한 자료전송을 위하여 전 구간에서 암호화 채널을 이용하고, 내부망 반입 시 백신 프로그램 검사를 통한 악성코드 유입 방지와 외부망 반출 시 관리자 승인을 통한 내부 자료 유출을 방지하는 기능을 지원한다. 스트리밍 연계(Streaming Service)는 분리된 내부망과 외부망 구간을 Non TCP/IP 기반의 자체 프로토콜 암, 복호화 과정을 통해 전송 기능을 제공한다. 연계가 필요한 데이터를 실시간으로 내부망 중계서버에 등록된 보안 정책에 의거 연계 서비스를 제공하며, 고속의 데이터 처리 기술을 탑재하여 최고의 전송속도를 보장한다. 환경에 따라, 강력한 보안 구성이 요구되는 국방, 공공기관에서 실시간 서비스 연계 방식을 서버 간 파일 연계로 간접 연동하는 방식을 제공한다. 외부망과 내부망의 연계를 허용하는 대신에 전달되는 파일의 안전을 확인하기 위해 아래와 같은 절차를 통한 파일만 전달한다.

○ 관리자가 승인하는 파일만 전송한다. (사후 승인을 포함)

○ 바이러스 백신을 통해 알려진 악성코드가 있는지 확인한다.

○ SandBox 기반의 툴을 이용하여 행위기반 이상을 탐지한다.

위의 절차를 통하기 때문에 망 분리를 통한 보안 수준을 확보할 수 있으나, 인터넷의 정보를 활용하여, 업무에 반영하는 경우 제약이 생길 수밖에 없어 업무효율의 저하를 가져올 수밖에 없다.

b) 웹 주요 취약점과 방어체계 정의

본 논문에서는 웹 취약점 관리 체계에 사용되는 OWASP(The Open Web Application Security Project)과 MITRE의 취약 기준을 알아보고, 종단 장치 보안에 영향을 끼치는 항목을 분석하여, 안전성을 확보할 수 있도록 기준을 마련해야 한다.

OWASP은 오픈소스 웹 애플리케이션 보안 프로젝트로, 주로 웹에 관한 정보 노출, 악성 파일 및 스크립트, 보안 취약점 등을 연구한다. OWASP TOP 10은 웹 애플리케이션 취약점 중에서 빈도가 많이 발생하고, 보안상 영향을 크게 줄 수 있는 것들 10가지를 선정하여 보통 2년에 한번 발표하고 있다.

MITRE releases 2020 CWE Top 25(이하, MITRE TOP 25) 문제를 일으킬 가능성이 가장 큰 소프트웨어 문제 (오류, 버그 및 잠재적 공격 벡터). 시스템 하이재킹, 데이터 유출 (및 민감한 데이터 도난), 서비스 거부 공격, 시스템 충돌, 임의 코드 실행 등 MITRE가 선정한 상위 25개 목록을 공개한다.

OWASP TOP 10과 MITRE TOP 25중 위험도와 빈도가 높은 11개 중 공격의 대상이 종단 장치인 4개를 [표 Ⅲ-1]과 같이 선정한다.

대부분의 종단 공격은 페이로드에 의해 악성코드를 다운로드 받고 이후 그 파일을 실행시켜서, 랜섬웨어 등을 감염시키거나 구성을 변형하여 피싱 등의 2차 공격으로 이어지는 것을 알 수 있다. 종단 장치를 방어하기 위해서는 허가받지 않은 파일 다운로드 금지와  스크립트를 통해 실행 명령을 차단 해야 한다.

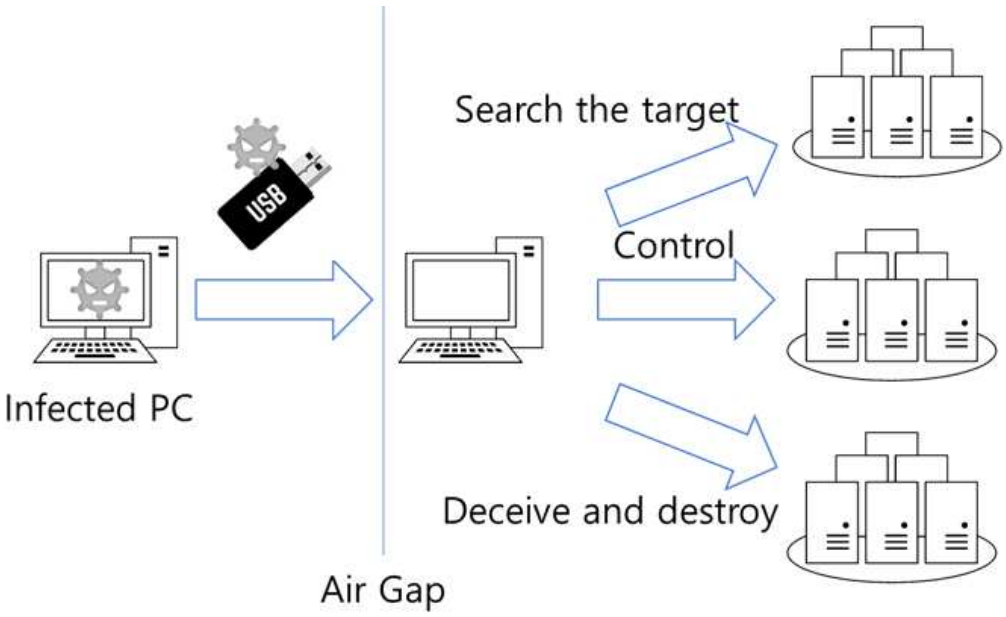
[표 Ⅲ-1] 빈도가 높은 취약점 분석[14][15][16][17][18][19][20]

| 순위 | 취약점 | 세부 내용 | 취약사항 |
|---|---|---|---|
| 1 | CVE-2014-6271 | Execution of arbitrary code through crafted environment | (){:;};ping-c1-p cb18cb3f7bca4441a595fcc1e 240deb0 attacker.com |
| 2 | CVE-2014-6278 | Execution of arbitrary code through crafted environment | () { :;}; /bin/sleep 20\|/sbin/sleep 20\|/usr/bin/sleep 20 |
| 3 | CVE-2014-6277 | Execution of arbitrary code through crafted environment | 'f() { x() { _;}; x() { _;} < |
| 4 | CVE-2017-5638 | Remote execution of arbitrary commands through crafted content-type HTTP header | payload += "(#cmd='%s')." % cmd try:headers={'User-Agent' : 'Mozilla/5.0','Content-Type': payload} |

## 3.2 네트워크 차단 환경에서의 보안과 한계점

공공기관에 적용된 구조는 SCADA와 같이 완벽하게 통신을 차단하는 모델이 아니다. 핵발전소 같은 주요 기반시설에서는 내부 시스템은 외부망과 완벽하게 차단되어야 한다. 만약 외부망이 연결된 경우를 가정하에, 시스템의 취약점이 존재할 경우, 외부 공격 및 악성코드 등으로부터 보호 대책을 반영하기 어렵다.

실제 이란 부셰르 원전을 감염시킨 스턱스넷(Stuxnet)을 살펴보면, 악성코드는 C&C(Command & Control) 서버를 통해서 SCADA의 내부로 침투하여 PLC(Programmable Logic Controller)의 제어 명령을 변조하게 된다. 이를 통해 이란의 핵시설에 원심분리기에 오작동을 유발하게 된다. 스턱스넷은 특정 제품의 통합관리 도구를 찾기 위한 악성코드이며, 찾은 이후에는 필요한 사용자 계정을 조사하여, 결국 산업망에 침투하여 최종 공격에 성공하게 된다[21][22].




[그림 Ⅲ-5] 스턱스넷(Stuxnet) 동작 요약

[그림 Ⅲ-5]와 같이 스턱스넷 공격 사례를 통해 교훈으로 삼아야 하는 부분은 네 가지 관점에서 요약할 수 있다.

첫 번째는 산업기반제어 시스템의 네트워크 통제는 완벽하지만 그래도 예외는 존재한다. 스턱스넷에 감염된 PC에서 USB 저장장치를 통해 메인 PC까지 악성코드가 전달되어 PLC 제어를 하는 권한을 획득하는 과정을 살펴보면 네트워크상에서는 폐쇄망을 유지했지만, USB 저장장치를 통해 악성코드가 침투되었다. 아무리 주요한 산업제어시스템이라고 할지라도, 시스템의 동작에는 주기적인 변경 관리가 필요하므로 어떤 형태의 파일이 지속해서 교환 될 수밖에 없는 구조였다.

두 번째는 폐쇄망을 믿기 때문에 방심하게 되는 구조이다. 폐쇄망은 정보보안의 방어전략으로 아주 효과적인 방법이지만, 모든 위협으로부터 안전함을 보장하지는 않는다. 폐쇄망이지만 신규 악성코드에 대한 주기적인 예방 활동을 통해서 공격을 차단할 수도 있다. 메인 PC에 OS로 제공되는 개인 방화벽을 활성화한다면 내부 제어시스템과 제어 동작의 위험을 낮출 수 있다.

세 번째는 보안 의식 부족이다. 산업제어시설과 같이 원천적인 네트워크가 차단된 환경에서는 작은 보안 수칙들을 간과할 수 있지만, 외부에서 내부 PC로 USB 저장장치가 사용되고 해당 USB 저장장치에 악성코드 검사를 수행하지 않은 점으로 미루어 짐작해보면, 높은 보안 인프라를 신뢰하여 작은 보안 수칙이 만들어지지 않거나, 지켜지지 않았다고 볼 수 있다.

마지막으로 공격자가 내부에 침투하여 데이터를 유출 시키는 것과 내부를 파괴하는 것은 그 동작 방법과 난이도가 상이하다. 보통 인터넷과 웹을 통해 제공되는 서비스는 웹의 취약점을 이용하여, 시스템 내부의 자료를 유출한다. 크게 보면 DB에 접근하거나 이미 주어진 DB에 SQL(Structured Query Language)을 실행할 수 있는 권한을 변경하여 원하는 데이터를 추출하거나, 로그인 프로세스에 있는 취약점을 이용하여 관리자로 로그인하는 방식이다. 하지만, 인터넷으로 제공되는 서비스가 아닌 경우에는 데이터 유출을 위해서는 우선 타겟이 되는 시스템이나, PC 등의 제어권을 획득하고, 관리자 권한을 취득하여 데이터에 접근한 뒤에 데이터를 추출하여 유출하는 방법을 사용한다. SK컴즈 개인정보 유출 사건(2011)과 인터파크 해킹 사건(2016)이 이런 방식을 통해 개인정보가 유출되었다. 대부분 망 분리가 되지 않는 기업에서 발생한 사례이고, 실제 폐쇄망으로 운영되

는 시스템에서는 유출 사고가 현저하게 적다. 그 이유는 개인정보 유출 사건에서 보는 것처럼 내부의 네트워크로 연결되지 않아도 특정 시스템에서 동작할 수 있는 악성코드는 침투할 수 있다. APT 공격을 통해 특정 목표를 표적으로 삼아 오랜 기간 다양한 리소스를 투입하여 공격한다면, 네트워크가 단절된 내부 시스템에도 침투할 수 있지만, 침투 이후에 데이터를 다시 가지고 나오는 것은 침투만큼 더 어려운 과정이다. 그래서 통제된 네트워크에서는 파괴를 위한 공격을 수행한다. 농협 전산망 마비 사태(2011)를 통해 볼 수 있다.

공격자는 농협에 출입하는 외부 직원의 노트북을 APT을 이용하여 감염시킨 이후 노트북을 통해 농협 시스템에 삭제 명령을 수행하여 30분 만에 시스템의 절반이 파괴되었다.

이렇듯, 망 분리가 효과적인 방어 전략이지만, 모든 위협과 위험으로부터 안전을 보장할 수는 없다[21].

## 3.3 네트워크 흐름 설계

본 논문에서 제안하는 모델은 PC 기반과 프록시 기반 모델에서 적용할 수 있다. 물리적, 논리적 망 분리 양쪽 모두에서 사용이 가능한 방식으로 설계되었다.

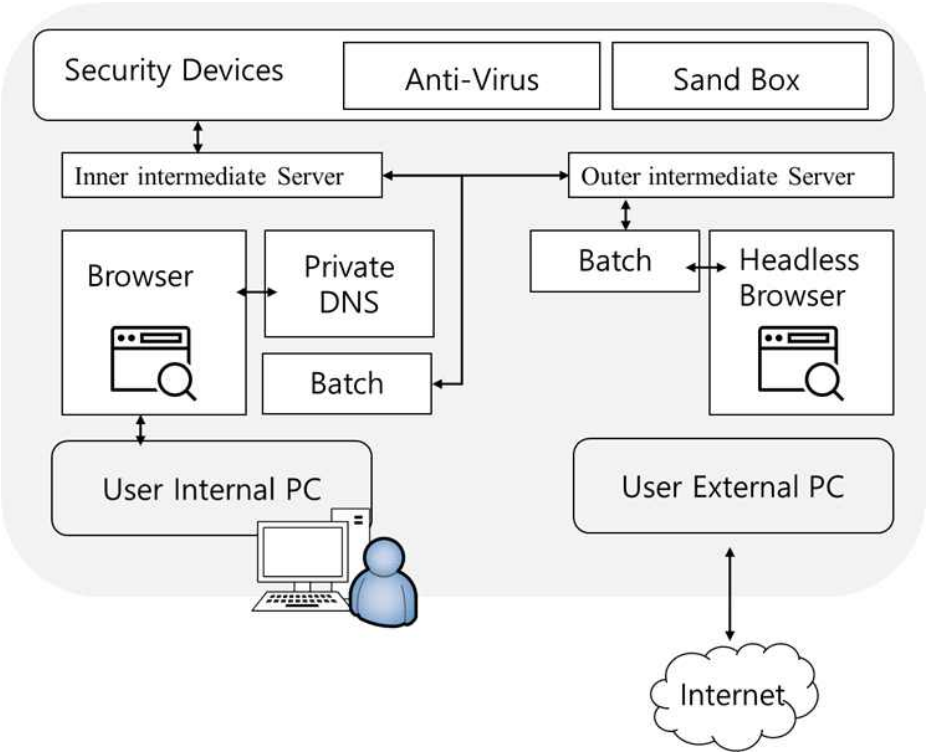
망 분리된 환경에서, 인터넷 정보를 활용하기 위해서는 망 연계를 위한 장치가 필요하다. 공공기관에 망 연계 장치가 도입되기 위해서는, 보안 적합성을 판단하기 위해서 CC 인증 제도가 존재한다. 본 연구를 위해서는 망 연계 장치가 설치된 환경이 안전하다는 가정에서 설계를 진행했다.

내·외부를 통신하기 위해서는 non-TCP 프로토콜을 구간 암호화 통신으로 구현하여 내부 네트워크와 외부 네트워크 간에 파일을 전송하기 위해 사용된다.

외부에서 내부로 전송된 파일의 경우 파일의 서명을 확인하고 백신프로그램을 통해 악성 여부를 판단한다. 마지막으로 YARA 프레임워크가 연동된 SandBox 기반의 APT 탐지 장치를 통해서 행동을 판단하고 악성프로그램 패턴을 인식하여 전달된 자료나 데이터에 악성코드가 있는지를 검증한다.

a) 네트워크 흐름 설계

[그림 Ⅲ-6]에서 보는 것처럼 망 분리 환경에서 내부망에 있는 사용자의 PC에서 웹페이지를 호출하기 위해서는 아래와 같은 설계가 필요하다. 사용자는 접속하기를 원하는 웹페이지의 URL을 브라우저를 통해 실행시키게 된다. 일반적인 망 분리 환경에서는 해당 URL의 웹페이지에 접근하기 위한 IP 정보를 알 수 없어서 브라우저에는 '찾을 수 없는 페이지'라고 보여줄 것이다. 하지만 본 연구에서 수행하고자 하는 내부망에서 웹스크래핑을 위해서는 DNS쿼리가 실패하면 해당 URL을 외부의 에이전트로 보내게 된다. DNS쿼리가 실패했다는 것은 내부 DNS 쿼리로 찾을 수 없는 URL로 판단하여, 인터넷 망으로 보내는 것이다.

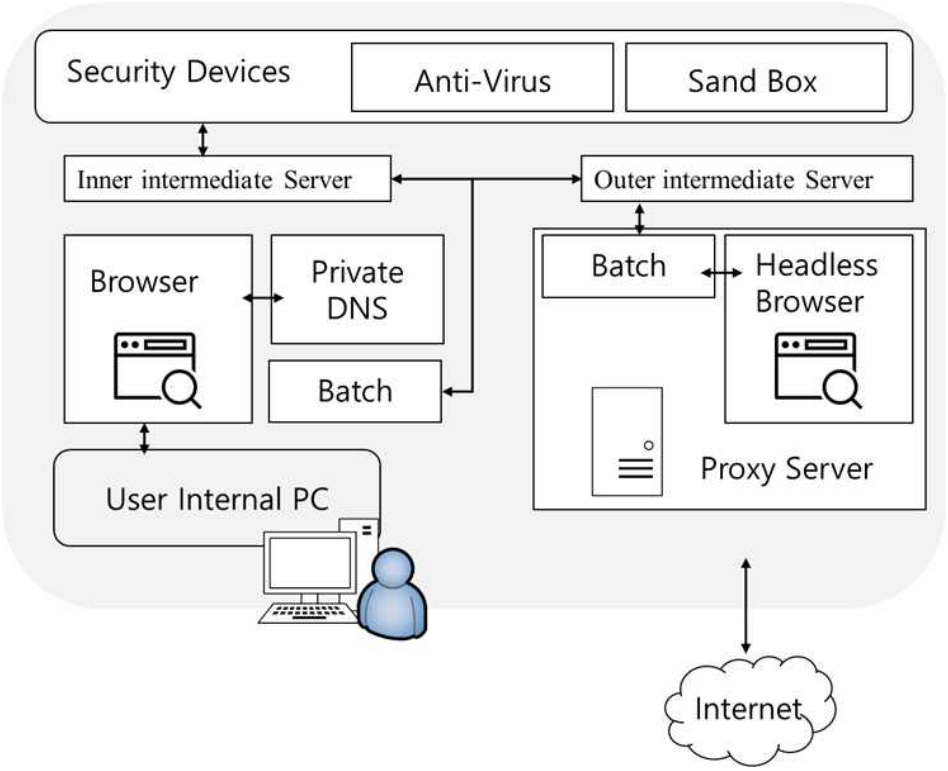



[그림 Ⅲ-6] 제안하는 망 분리 환경에서 웹 스크래핑 모델

외부망 PC에 위치한 에이전트는 보내진 URL에 해당하는 웹페이지를 스크래핑하고 스크래핑 된 파일을 자료교환 시스템을 통해 내부로 보내지게 된다. 전달받은 데이터는 내부망 PC에 위치한 에이전트가 받아서 브라우저로 띄우게 된다.

위에서 살펴본 네트워크 구조는 2PC를 사용하는 망 분리 환경을 기본으로 하였지만, VM을 이용한 구조에서도 사용할 수 있다.

여기서 중요한 점은, 내부망 PC를 사용하는 사용자는 외부망 PC에 별도의 접근 없이 내부망 PC에서만 인터넷 정보를 사용할 수 있었다는 것이고,




[그림 Ⅲ-7] 제안하는 프록시 서버 기반의 웹 스크래핑 모델

외부망 PC의 에이전트 방식을 이용하기 위한 구조의 단점은 외부망 PC가 켜져 있어야 한다는 것이다. 그리고 본인의 내·외부망에 연결된 자료연계 장치가 빠른 속도로 연계 작업을 하여야 한다. 이런 단점을 보완할 수 있는 구조로

[그림 Ⅲ-7]과 같이 프록시 서버 기반의 네트워크 모델로 구성할 수 있다. 프록시 서버에 인터넷 넓은 대역폭과 고성능을 할당하고, 망 연계 등에서 예외적인 핫라인을 제공한다면 좀 더 빠른 서비스가 가능하다, 게다가 사용자의 외부망 PC의 온/오프와 무관하게 동작한다는 장점도 있다.

프록시 구성이 의미가 있는 이유는, 본 논문이 제안하는 방법이 일반화된다면, 망 분리에서 반드시 2 PC를 제공할 필요가 없다. 사용자의 필요에 따라 구분을 나누어서 내부망 1 PC만을 제공하여도 충분히 업무가 가능할 것으로 보인다. 사용자는 내부망 PC에서 업무를 수행하다가 인터넷 정보가 필요한 경우에는 스크래핑 된 웹을 전달받아 사용할 수 있다. 1 PC로 망 분리가 구현된다면 큰 비용과 관리 소요가 없어지게 되어 보안 효율성이 생길 것이다.

b) 망 분리 환경에서 웹스크래핑을 위한 DNS 설계

먼저 내부망에서 인터넷 웹사이트에 접근하기 위해서는 URL을 해석해야 한다. 하지만, 인터넷이 가능한 공개 DNS가 없이는 URL을 IP로 변경할 수 없다. 공개 DNS는 DNS 쿼리를 받으면, IP를 결과로 반환하지만, 내부망에 있는 사설 DNS 서버는 공개된 DNS와 연동이 되어있지 않기 때문에 IP를 반환할 수 없다. 본 논문에서 제안한 방법은 이 부분을 이용하여 자동화된 프로세스를 만들 것이다.

DNS가 일반적인 동작을 간단하게 살펴보면 아래와 같다.

○ Explanation of terms

INPUT : DNS에 질의하여 IP로 변환하기 위한 브라우저에 입력된 URL

e.g.) <https://www.intra.net>

OUTPUT : 요청된 URL을 DNS에 질의 후 받은 값(IP 등) 의 결과 값

Internal network web service : www.intra.net( IP addr : 192.168.10.XXX)

Public web service : www.exter.net( IP addr : 220.111.11.XXX)

본 연구에서는 사설 DNS 쿼리 이후에 답이 오지 않는 경우는 외부로 보내서 웹 캡처를 수행하도록 DNS를 설계하였다. [그림 Ⅲ-8]은 망 분리 환경에서 웹스크래핑을 하기 위한 DNS 구성방안을 알고리즘으로 나타낸 것이다. 먼저 입력된 URL을 받아 내부에 있는 사설 DNS에 쿼리를 수행한다. 응답이 있는 경우는 내부 시스템을 위해 정의된 URL이라고 판단하고 종료한다. 응답이 없는 경우는 외부망으로 보내 스크래핑을 수행하도록 한다.

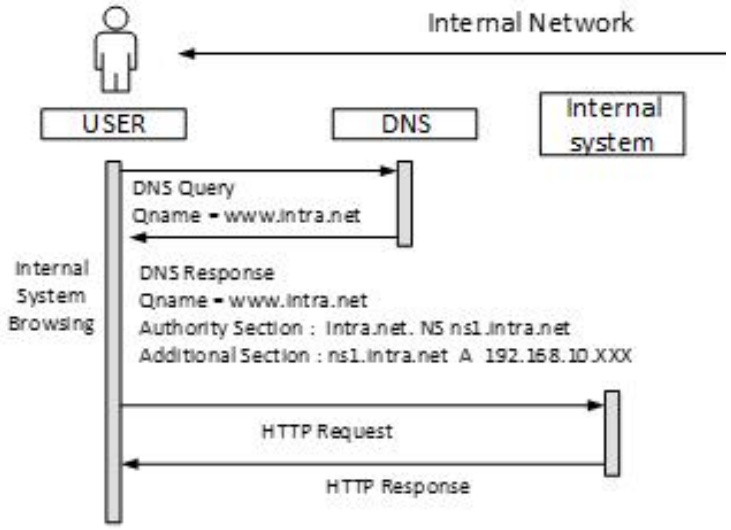
```
Private DNS configuration algorithm

INPUT : User input URL
OUTPUT : IP(or Error message)

1. if  DNS Recursive Name Server[List] ≠ DNS query
2.    Foward to external network agent(via. File Transfer system)
3.    Query a URL in External DNS
4.    if External DNS Recursive Name Server[List] ≠ DNS query
5.       return( Error message : No records, Nonexistent domain)
6.    else
7.       return( Capture Web of URL)
8. else
9.    return( IP of requested URL)
```

[그림 Ⅲ-8] 제안하는 사설 DNS 구성 알고리즘

이 구성의 한계점은 실제 오타나 존재하지 않는 URL을 가려낼 방법이 없다는 한계점이 존재한다.

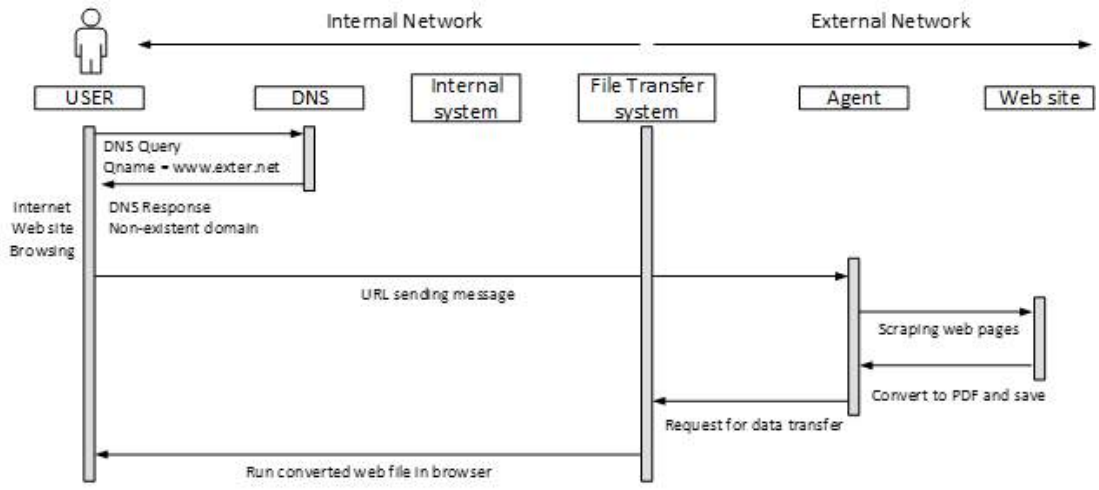
c) 네트워크 시퀀스 다이어그램

내부망에서 내부망 시스템을 이용할 때 [그림 Ⅲ-9]과 같이 기존 망 분리 정책과 같게 웹서비스를 수행한다. 브라우저에 입력된 URL을 IP로 변환하기 위해서 내부망에 있는 사설 DNS에 DNS쿼리를 하게 되고, 쿼리 응답받은 IP 주소로 웹서비스를 수행한다. 내부망에서 외부망의 이용은 기존 망 분리 정책에서 불가능하므로 외부망 PC로 접속을 하여, 인터넷 웹사이트에 접속하여 필요한 정보를 파일로 만들어서, 파일 전송 절차를 수행한 뒤에 내부망으로 파일을 가져와야 한다. 이렇다 보니, 시간과 비효율이 발생한다. [그림 Ⅲ-10] 와 같이 내부망에서 사설 DNS를 조회했을 때, 검색되지 않는 URL은 외부로 전송하여 인터넷 웹사이트에서 PDF 등으로 변환하여 파일을 내부 브라우저에서 보여줄 수 있게 하려고 아래와 같이 네트워크 흐름도를 정의한다.




[그림 Ⅲ-9 내부 시스템 요청 시퀀스 다이어그램

PC 기반의 망 분리 환경에서 웹 스크래핑의 전체적은 시퀀스 다이어그램은 [그림 Ⅲ-12] 과 같이 정의한다.




[그림 Ⅲ-10] 웹 스크래핑 시퀀스 다이어그램

## 3.4. 애플리케이션 설계

a) URL 전송 데이터 포맷 정의

URL을 외부망 에이전트에 전달하는 프로토콜을 정의하여 내부에서 외부로 필요한 정보를 전달하고, 응답으로 스크래핑 된 파일을 전달하는 데 사용한다. URL 메시지 프로토콜은 아래와 같이 json 타입으로 규정한다.

필요한 정보의 속성은 외부망에서 scraping을 할 때 필요한 정보로 해당 URL과 반환 타입을 정의하고 스크랩핑 방식을 지정한다. 마지막으로 authKey를 이용하여 최소한의 발신자를 인증하는 절차를 거친다.

{url:"https://www.intra.net/",outputAsJson:true,renderType:"PDF",authKey:"accd-fgfe-ffeq-dqwe"}

http_URL =
"http:" "//" host [ ":" port ] [ abs_path ] //base64 encoding을 적용한다.
outputAsJson = true | false // 페이지 스크래핑에 관한 정보를 반환한다.
renderType = PDF | jpg | plain Text 페이지를 스크래핑 하는 방식을 지정한다.
authKey = 4ALPHA+DIGIT "-" 4ALPHA+DIGIT "-" 4ALPHA+DIGIT "-" 4ALPHA+DIGIT
사용자를 구분하고, 요청자를 확인하기 위해 미리 부여된 키를 확인한다.

b) Application 동작 설계

Agent는 proxy 방식으로 설계하여 운영할 수도 있고, 외부망에 있는 PC에서 각자 구동시켜도 된다. Agent는 아래와 같은 순서로 작동한다.

① URL 전달 메시지가 있을 때까지 대기한다.

② 메시지가 전달되면, 메시지를 읽는다.

③ phamtomJS 라이브러리를 이용하여 해당 페이지를 렌더링한다.

④ 요청한 renderType에 맞게 웹을 다운로드 한다.

⑤ 해당 파일을 싱크 폴더로 이동시킨다.

⑥ 싱크 폴더는 자료교환을 통해 내부로 이동한다.

⑦ 전달받은 파일을 실행한다. (PDF 파일의 경우 브라우저로 띄운다.)

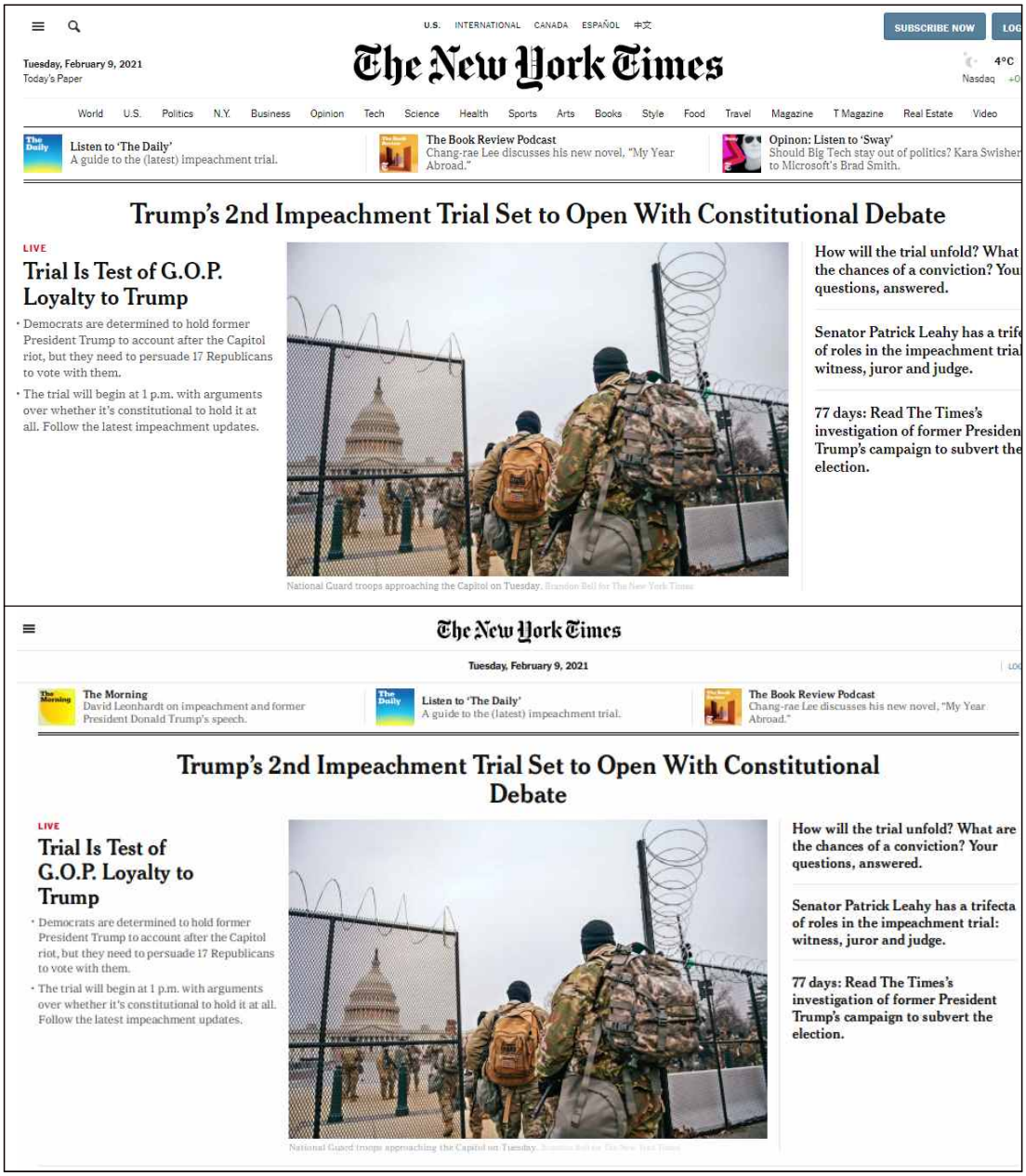
# Ⅳ. 실험 결과 및 분석

## 4.1. 애플리케이션 구현을 통한 실험 수행

New York Times의 웹사이트를 브라우저에서 스크래핑하여 비교를 수행하였다. HTML에 적용된 뷰포트(View Port)에 따라 반응형 웹 디자인(Responsive Web Design)이 적용되어 약간의 랜더링이 다른 것을 볼 수 있지만, 대부분 콘텐츠와 품질이 같은 것을 아래 그림에서 확인할 수 있다. [그림 Ⅳ-1]은 에이전트에서 PhantomJS를 호출하는 부분을 나타낸다.


```
#include <file.au3>

$fileNM = ""
$urlFileNM = ""
Exec()
moveFile()

Func Exec()
    Local $hSearch = FileFindFirstFile("*.adr")
    If $hSearch = -1 Then
        Return False
    EndIf

    Local $urlFileName = ""
    $urlFileName = FileFindNextFile($hSearch)
    $urlFileNM = $urlFileName
    Local $urlFileopen = FileOpen($urlFileName, 0)
    Local $sFileRead = FileRead($urlFileopen)
    FileClose($urlFileopen)
    ;Run application
    Run("cmd.exe")
    ;Wait for CMD to be opened
    WinWaitActive("Administrator: C:\Windows\system32\cmd.exe", "", 1)
    Send('cd O:\사후결재받은함' & "{ENTER}")
    $fileNo = Random(0, 100, 1)
    $fileNM = 'screen' & $fileNo & '.pdf'
    Send('phantomjs rasterize.js ' & $sFileRead & ' screen' & $fileNo & '.pdf' & "{ENTER}")
    sleep(9000)
EndFunc

Func moveFile()
    FileMove($fileNM, 'O:\보낼파일함\' & $fileNM & '.jpg', $FC_OVERWRITE) ; $FC_OVERWRITE (1) = overwrite existing files.
    FileDelete($urlFileNM)
EndFunc
```

[그림 Ⅳ-1] 웹스크래핑 에이전트 장치 소스 코드




[그림 Ⅳ-2] (위) 브라우저 웹 (아래) 스크래핑 웹

[그림 Ⅳ-2]는 실제 브라우징 된 웹과 스크래핑 된 웹을 비교하여 보여준다. 최신 웹 들은 반응형 웹으로 보여주는 크기에 따라 다른 화면을 나타내기 때문에 실제 브라우저에서 실행한 웹페이지의 모습과 스크래핑 된 웹이 약간의 차이가 존재하는 것을 확인할 수가 있다.

파일을 생성하고 전송하고 삭제를 해야 하므로, OS 파일을 제어할 수 있는 프로그램 언어가 필요하다. 본 논문에서는 실험 프로그램 작성을 위해서 요구사항에 맞는 프로그램 언어를 이용하여 개발하였다.

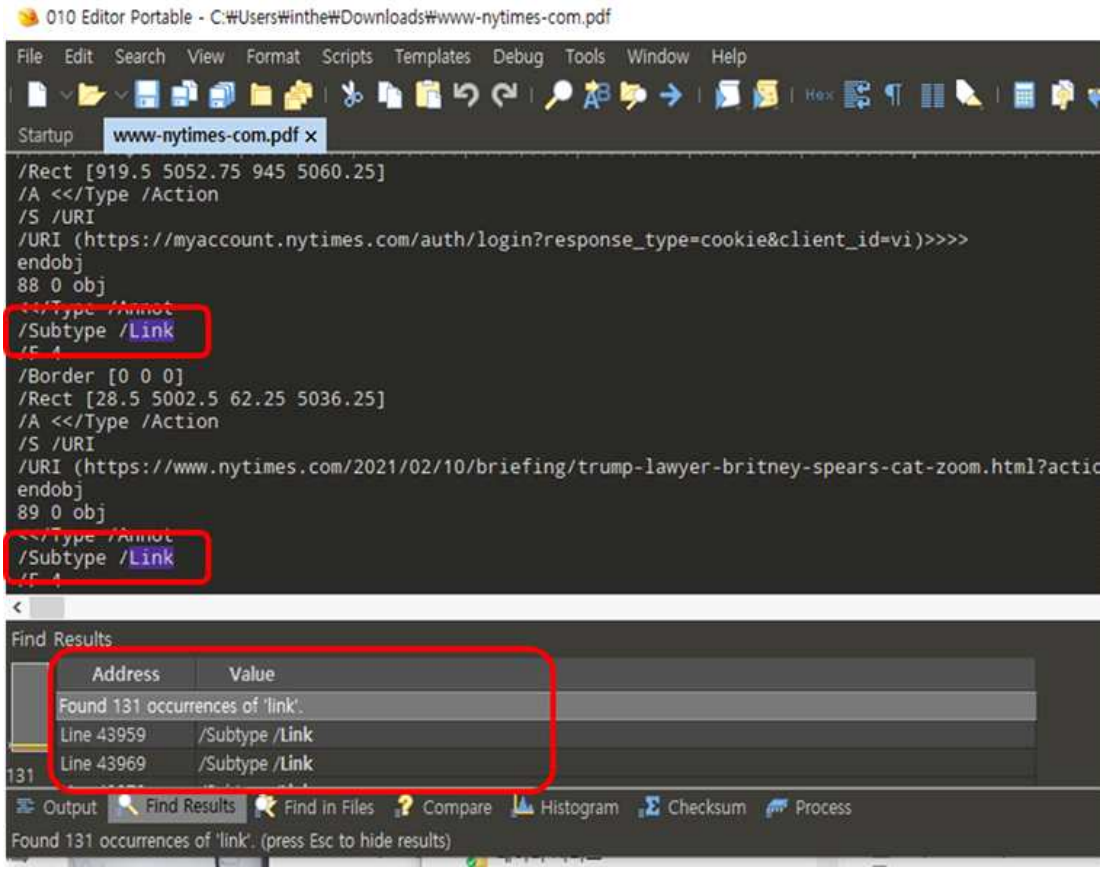
## 4.2. 악성코드 안전성 분석

PhantomJS를 이용하여, 변환된 PDF를 헥사코드로 변환하여 상세 내용을 분석을 수행한 결과 웹페이지에 구현된 하이퍼 링크는 사용이 가능한 상태로 스크래핑했음을 알 수 있다. 이후 추가로 정의된 취약 요소가 남아있는지 분석을 수행하였다.

```
85 0 obj
<</Type /Annot
/Subtype /Link
/F 4
/Border [0 0 0]
/Rect [28.5 4926 931.5 4982.25]
/A <</Type /Action
/S /URI
/URI (https://www.nytimes.com/live/2021/02/10/us/impeachment-
trial/?action=click&module=Spotlight&pgtype=Homepage)>>>>
endobj
```

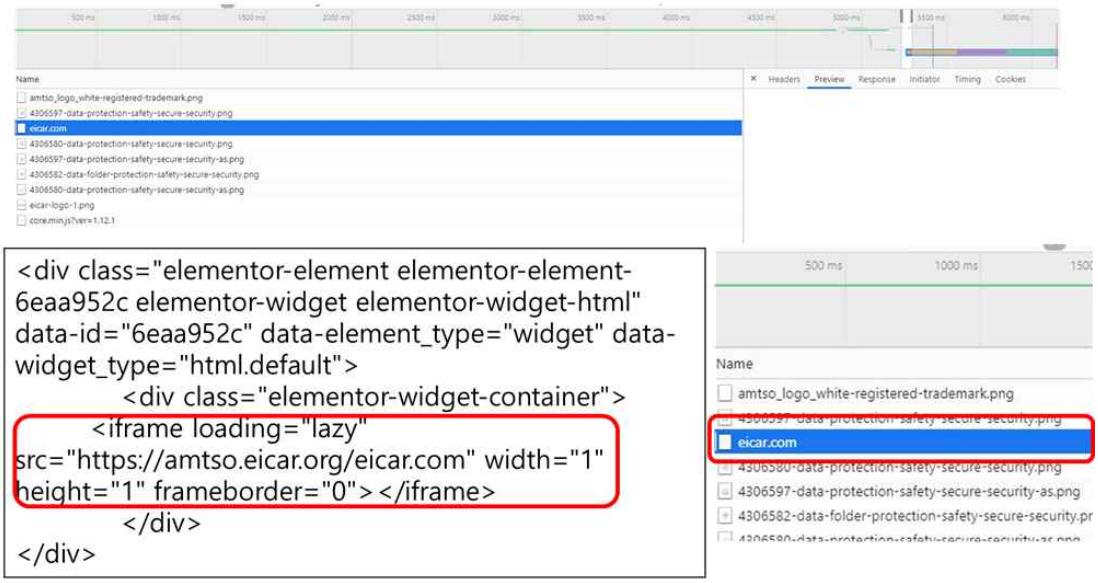
[그림 Ⅳ-3] 스크래핑 파일을 헥사코드로 변환한 결과

위에 [그림 Ⅳ-3]처럼 헥사코드로 변환하여 내부에 있는 코드들을 분석해 본 결과 하이퍼링크(Hyperlink)를 그대로 가져왔음을 분석할 수 있다.

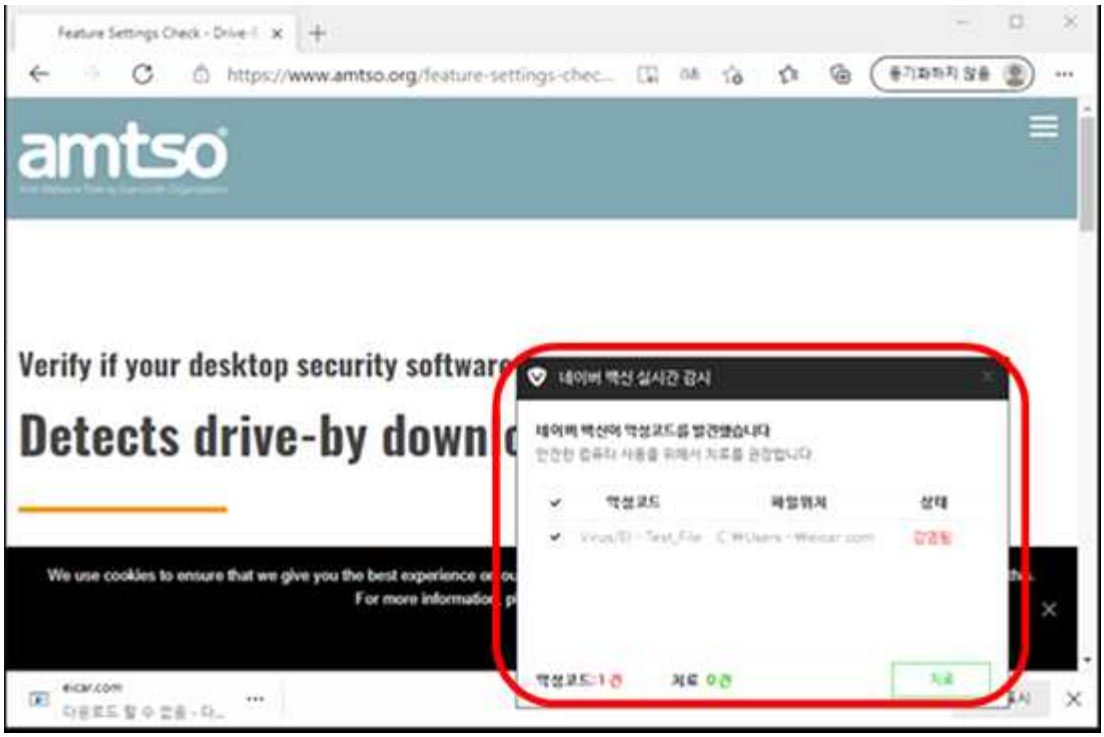



[그림 Ⅳ-4] 스크래핑 파일을 헥사코드로 변환 후 분석

frame 태그를 이용한 Drive-by Download를 구현한 사이트에 접속하면, 브라우저는 frame 태그 안에 URL을 호출한다. 이 URL은 onload 함수에 다운로드 파일을 실행시킨다. 이런 태그를 이용한 공격 방식은 동의 없이 다운로드를 시킬 수 있다. 악성코드가 있었다면 페이로드 된 파일을 실행시켜 PC의 통제권을 갖거나 랜섬웨어를 실행하여 원하는 금전적인 이득을 노릴 수 있는 상황이다. 테스트 웹페이지에서 어떤 다운로드나 저장 버튼을 누르지 않았지만, eicar.com 파일을 다운로드한다. 백신이 해당 파일을 탐지하여 감염 표시를 해주었다. [그림 Ⅳ-5]는 실제 브라우저 네트워크 개발자 도구를 통해 페이로드 되는 과정을 추적한 내용이며, 하단은 frame 태그를 이용하여 Drive-by Download가 수행되는 코드를 나타낸다.



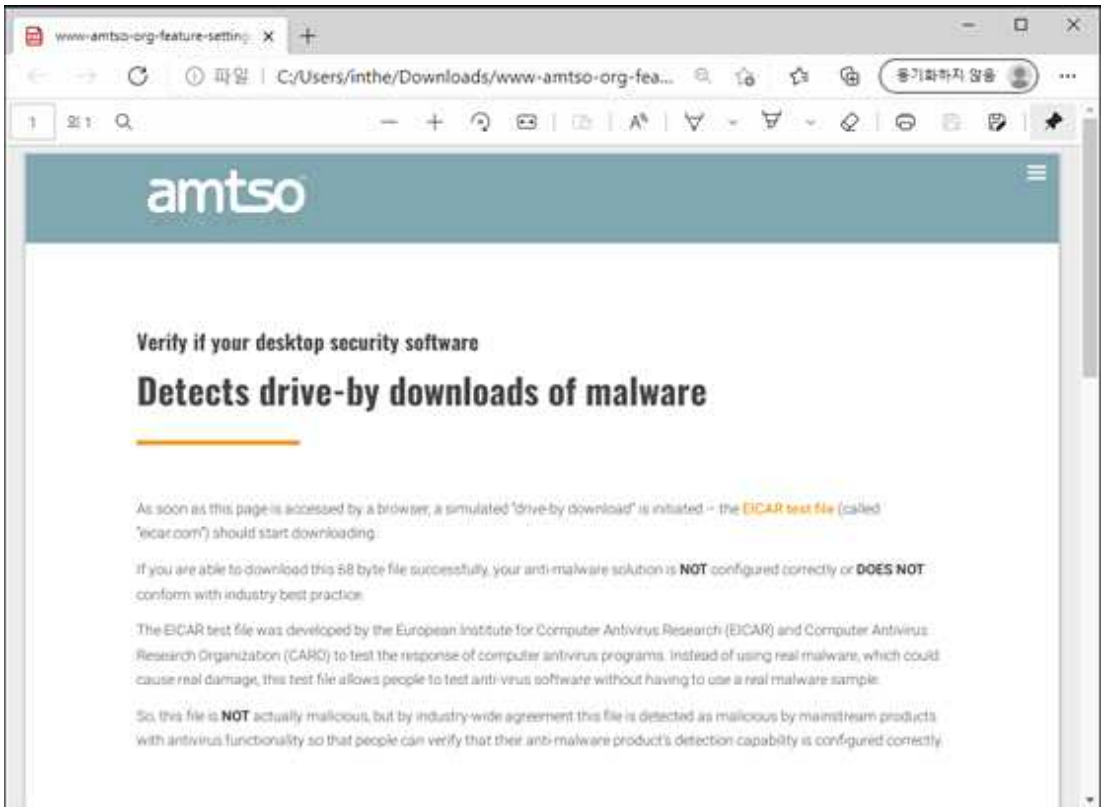

[그림 Ⅳ-5] Drive-by Download 분석




[그림 Ⅳ-6] Drive-by Download 테스트 웹 접속

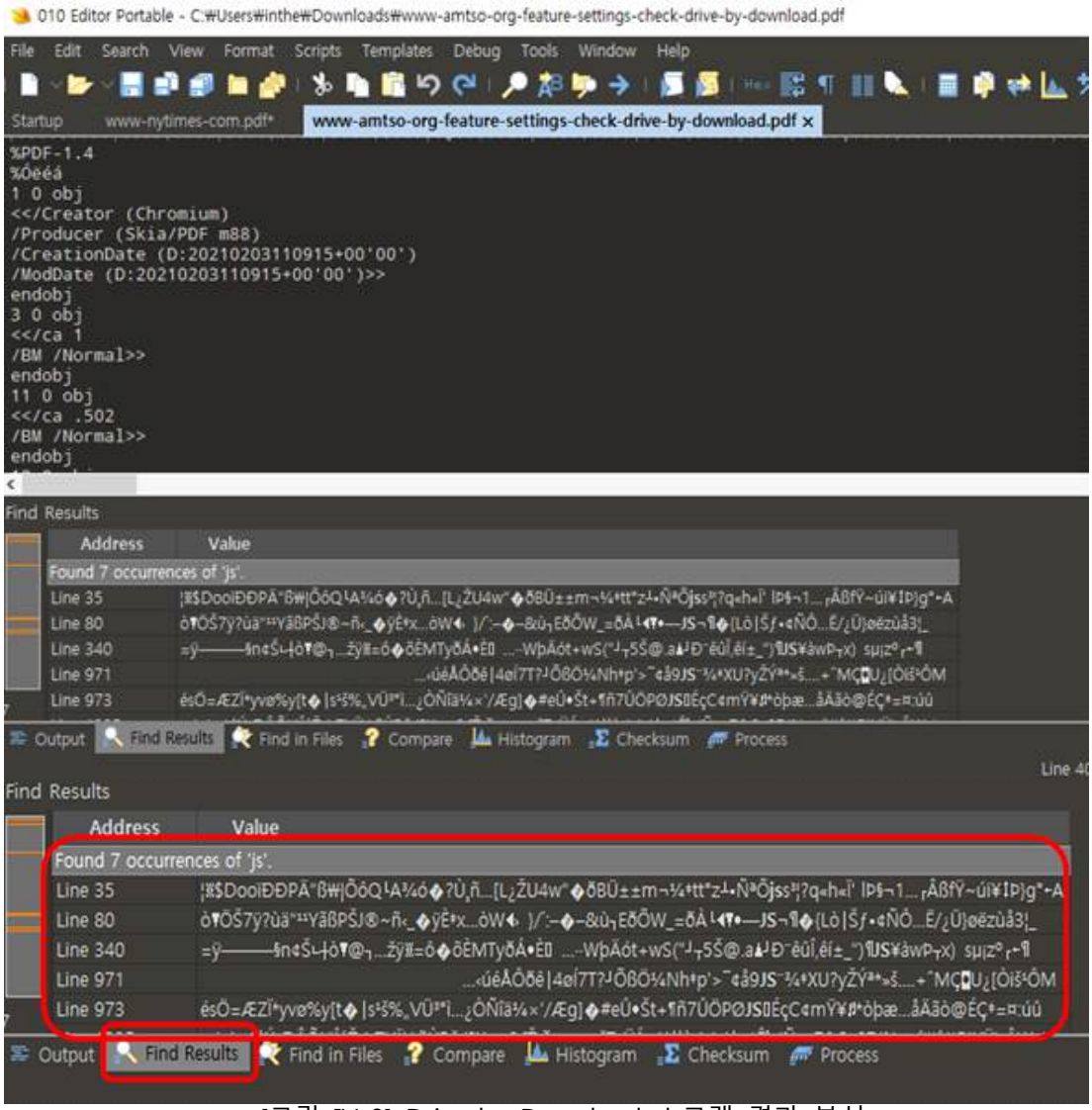
[그림 Ⅳ-6] 보면 Drive-by Download 사이트에 접속하면, 백신 프로그램이 악성 행위를 판단하여 바로 실행 차단이 수행되는 것을 알 수 있다. 백신 프로그

램은 이미 알고 있는 악성코드의 시그니처(Signature)를 보고 판단하기 때문에, 알려지지 않은 악성코드였다면, 페이로드가 성공되어 해커(공격자)가 실행할 수 있는 위험 상황에 있다고 할 수 있다.




[그림 Ⅳ-7] Drive-by Download 테스트 웹 스크래핑 결과

[그림 Ⅳ-7] 과 같이 스크래핑을 이용한 후 PDF로 저장된 파일을 브라우저로 열었을 경우 Drive-by Download가 작동하지 않았으며, frame 태그나 PDF에서도 동작할 수 있는 JS와 Javascript 코드가 제거된 것을 [그림 Ⅳ-8]을 통해 확인할 수 있다. PDF로 변환된 파일은 frame 태그가 캡처되기 때문에 frame 태그 안에 URL이 호출되지 않는다. 호출이 될지라도 브라우저에서 일어나는 동작이 아니기 때문에 파일의 다운로드를 수행할 수 없게 된다.




[그림 IV-8] Drive-by Download 스크랩 결과 분석

[그림 Ⅳ-8]과 같이 악성코드를 강제로 다운로드 시키는 사이트를 PDF로 변환 결과를 다시 헥사코드로 변환하여 분석을 해보았다. 실제 브라우저에서 동작하는 많은 태그들이 정적인 이미지나 텍스트 형태로 변형된 것을 확인할 수 있으며, 작동하는 코드는 링크 태그 정도만 남아있는 것을 확인할 수 있다. 로직을 실행할 수 있는 frame과 Javascript가 제거된 것을 볼 수 있다. 악성코드가 있던 웹페이지가 스크래핑 과정을 통해 비 악성화가 되는 것을 분석을 통해 확인할 수가 있다.

# Ⅴ. 실험 결과 고찰

## 5.1. 실험 결과 고찰

웹 스크래핑을 이용한 브라우징 방법이 망 분리 환경에서 구현을 통해 실현이 가능하다는 것을 확인했으며, 망 분리 정책에서 허용하는 망 연계를 이용하였기 때문에 정책의 위배사항 없이 활용할 수 있다. 특히, 악성코드가 동작하지 않아서, payload나 다른 동작을 수행할 수 없는 상태로 만드는 비 악성화 상태가 되었음을 확인하였다. [표 Ⅴ-1]처럼 제안한 망 분리 방법을 비교하였다.

[표 Ⅴ-1] 제안한 망 분리 방법 비교

| 구분 | 네트워크 세분화 | 망 분리 | 제한된 망 분리 |
|---|---|---|---|
| 차단 방식 | 작은 네트워크 단위로 분리 | 내부망과 외부망을 물리적 또는 논리적으로 분리 | 좌동 |
| 외부 서버와 연동 | ACL 을 관리자가 허용하여 연결 | 망 연계의 스트리밍 장치를 이용하여 연결 | 좌동 |
| 외부 자료 이용 방법 | 허용된 ACL을 통해 데이터를 전송하여 활용 | ① 외부망 PC에서 자료를 수집<br>② 망 연계 시스템을 통해 승인 후 내부로 전송 | 내부망 PC 브라우저를 이용하여 자료 이용 |
| 보안 위협 | 허용된 통신 포트를 통해 해커가 침투하거나, 악성코드 유입 가능 | 외부망 PC가 악성코드에 감염된 경우, 악성코드 유입 가능 | 비 악성화 조치를 통해 악성코드 유입 불가 |

네트워크 세분화의 특장점은 중요한 정보를 지키기 쉬운 구조이다. 작게 나누어진 네트워크 단위별로 접근제어가 설정되어있기 때문에, 접근 가능한 한 두 개의 서버를 장악하여도 그 피해가 확산이 되지 않는다. 하지만 허용된 통신 포트를 통한 공격에는 대응 방안이 없다. 이를 보완 하기 위해서는 관리자가 접근제어 설정을 철저히 관리하고 주기적으로 모니터링 하는 등 일반적인 정보보안 활동의 강화가 필요하고, 전체적인 네트워크 세분화 구조가 업무의 중요도나 비밀유지 여부에 따라 최소화 단위로 구분될 필요가 있다.

반면에, 망 분리는 각 네트워크 단위로 접근제어를 설정하지 않기 때문에 내부망과 외부망이 접점이 되는 부분의 방어가 중요하다. 그래서 외부 자료를 내부자료를 가져오기 위해서는 망 연계를 이용하여 승인받고, 보안 장비를 통해 안전성을 검증받는 과정이 필요하다. 정보보안 담당자는 외부망과 내부망 사이에 있는 망 연계 장치의 보안 정책을 통해 망 분리 전체적인 안전성을 확보해야 한다. 그래서 정보보안 전략적으로 망 연계에 정보보안 장비와 절차들이 집중하게 된다.

제안하는 방법은 두 가지 측면에서 망 분리 업무 효율성을 높일 것으로 생각한다. 외부 자료의 비 악성화를 통해 정보보안 검증 시간의 절약이다. 망 연계 장치를 이용 시, 웹 스크래핑 결과는 이미 비 악성화 조치가 확인되었기 때문에 즉시, 전송처리를 하여 시간과 절차를 절약할 수 있다. 이 밖에도 외부 PC를 거치지 않기 때문에 내부망 단일 구성도 가능하다.

## 5.2. 한계점과 개선사항

망 분리 정책의 배경은 서론에서 살펴본 것처럼 전 국가적인 사이버 위협상황에서 위협요인을 최소화하기 위한 정책이었다. 기술적으로는 랜더링이 종료된 정적인 페이지를 스크래핑하는 기술은 악성 동작을 할 수 없음을 실험을 통해 확인하였지만, 해킹은 정상적인 경우를 비정상인 경우로 만들어 침투 및 정보 유출을 하는 방법이기 때문에 완전하게 안전하다고 할 수 없다.

망 분리 환경에서 웹 브라우징 방법이 지속해서 안전하게 관리되기 위해서는 몇 가지 제도적인 규제와 검증이 필요하며, 기술적인 추가 연구도 필요해 보인다.

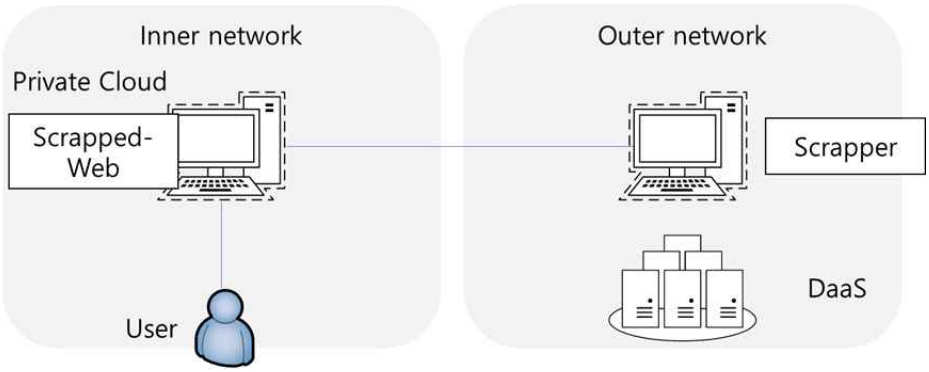
첫 번째, 내부망과 외부망 사이에 프로토콜의 안전한 암호화 적용을 통해 정보 노출의 최소화 하는 것이다. 본 논문에 제안된 프로토콜은 내부망과 외부망의 연결과정에서 통신 되는 정보는 타겟 URL, 포트와 웹 스크래핑을 위한 정보들로 노출되어도 되는 정보지만, 공격을 위해서는 사전 탐지 과정에서 노출되는 정보를 최소화할 필요가 존재한다. 국정원 또는 한국인터넷진흥원이 검증한 암호방식을 통해 요청 프로토콜과 결과물을 암호화할 필요가 있다.

두 번째, 망 연계 시스템 보안 적합성 검토사항에 웹 스크래핑을 통한 브라우징 기능이 있는 경우 보안 적합성 검토 항목에 포함하여, 웹 스크래핑 된 웹 브라우징을 검증할 필요가 있다. 정적인 웹을 스크래핑하여, 재구성하는 방식은 다양할 수 있으므로, 해당 방식을 검증하고 검증된 방식을 승인하는 절차가 필요하다.

세 번째, 동적인 웹을 정적인 웹으로 변환하는 기술의 연구가 필요하다. 마크업(Mark Up) 언어는 Javascript와 CSS를 통해 동적으로 작동하는 웹을 위해 발전하였지만, 이를 통해 다양한 해킹 기법이 같이 발전되어 왔다. 본 논문에서 제안한 방향이 적용되고 발전한다면, 보안상황이나 필요에 따라 동적인 웹을 정적인 웹으로 변환하는 기술, 예를 들면 마크업으로 구성된 웹페이지를 마크다운(Mark Down)으로 스크래핑하여 비 악성화 처리하는 연구가 추가로 필요하다.

## 5.3 클라우드 환경에서 망 분리

공공기관은 행정안전부의 공공클라우드 전환 정책에 따라, 일반 기업은 비용 절감과 관리의 편리함으로 클라우드를 선택한다. 하지만 클라우드를 인프라 구성에서 망 분리 정책으로 수립하기 어렵다. 그 이유는 클라우드는 정보시스템 자산을 소유하지 않고 빌려 쓰기 때문에 인터넷 회선을 통해서 접근해야 한다. 업무에 따라 내부망과 외부망을 나눈다고 해도 사용자와 관리자는 두 개의 망 모두 접근할 때, 전용망이나 인터넷 회선을 가상 사설망(VPN) 변환하여 접속하게 될 것이다. 본 논문에서 살펴본 망 분리의 다양한 구조에서는 SBC 방식의 경우가 서버를 클라우드로 전환하여 사용할 수 있으며, 그 밖에 DaaS의 형태로 사용할 수 있다. 클라우드 전환된 구성에서는 그림[그림 Ⅴ-1]과 같이 웹 스크래퍼를 위치하여, 구성할 수 있다.




[그림 Ⅴ-1] Cloud 환경에서 망 분리와 웹 스크래퍼 구성 예시

일반적으로 인터넷 환경을 고려하면, 외부망에서 내부망을 접속하는 것이 편리한 네트워크 설계이겠지만, 최초 접속을 내부망으로 구성한다면, 재택근무나 스마트워크에 적합한 망 분리 네트워크 구성을 할 수 있다. 사용자는 집이나 사무실 어디서나 동일한 환경에서 업무를 수행할 수 있다.

# Ⅵ. 결론

망 분리는 보안을 위해서, 그리고 관련 규제를 준수하기 위해서 필수로 해야 하지만 큰 비용이 필요하고, 매우 불편한 것으로 인식되어있다. 한국학술지인용색인(KCI)에서 최근 5년간 “망 분리” 키워드로 검색을 하면 53편의 논문이 검색되는데 논문 대부분은 차단에 초점이 맞춰져 있고, 업무 효율성에 관한 연구는 수행되지 않았다.

하지만, 4차 산업 혁명과 코로나19를 거쳐 인터넷의 데이터를 활용하는 것은 이제 선택이 아닌 필수가 되어가고 있으며, 근무 형태도 재택근무를 허용할 수 있도록 유연한 변화가 필요한 상황이다.

본 연구에서 제안하는 방법을 통해 망 분리에서 인터넷상의 웹페이지를 정적으로 활용하는 것이 가능하다는 것과 정적으로 변환되는 과정과 결과물이 비 악성화된다는 것을 실험을 통해 확인하였다. 제안하는 방법이 의미가 있는 이유는, 현재 망 분리 정책과 제도를 준수하고 보안 수준을 유지하는 상황에서 효율성 측면에서 새로운 방법을 제시했다는 것이다. 또한, 비 악성화가 확인된 방법이기 때문에 인터넷에서 데이터를 내부망으로 전송할 때마다. [표 Ⅵ-1]과 같이 시간을 단축할 수 있으며, 1일에 여러 번씩 수백 명이 망 분리 환경에서 업무를 수행한다면, 효율성을 극대화될 것으로 예상된다.

[표 Ⅵ-1] 내부망에서 외부 정보 이용에 필요한 시간

| 외부 정보 이용 구분<br>(망 분리 환경) | 웹페이지<br>렌더링 | 악성코드 검사 | 전송 시간 |
|---|---|---|---|
| 수동 정보 전달 방식 | 3초 이내 | 최소 60초 | 10초 이내 |
| 자동화된 스크래핑 방식 | 상동 | 해당 없음 | 상동 |

향후에는 다양한 환경에도 동일하게 실행 가능한 안전한 웹, 즉, 스크래핑 과정이 생략된 비 악성화 브라우징에 관한 후속 연구를 진행할 예정이다.

# 참 고 문 헌


[1] 김근혜, 박규동, 심미나. (2019). 정보보안 종사자의 조직갈등과 직무이탈 의도에 관한 연구. 정보보호학회논문지, 29(2), 451-463.

[2] 디지털데일리, 또 도마오른 물리적 망분리...망분리와 망연결을 함께하는 이상한 나라 비판 : http://m.ddaily.co.kr/m/m_article/?no=196998(접속일 : 21년 11월 8일)

[3] 2021 스타트업얼라이언스아젠다세미나. "규제가 핀테크산업 개발환경에 미친 영향" 토론회 (접속일 : 21년 11월 8일)

[4] 행정안전부(디지털정부기반과) 보도자료(2021), "2025년까지 모든 행정·공공기관 정보시스템 인터넷 기반 자원 공유(클라우드)로 전환"

[5] 클라우드 보안 인증 제도 : https://isms.kisa.or.kr/main/csap/intro/index.jsp (접속일 : 21년 11월 8일)

[6] 김소희, 이유림, 이일구. (2021). 공공 클라우드 기술과 정책의 개선방안에 대한 연구. 디지털융복합연구, 19(8), 11-20.

[7] Rajendra Patil, Harsha Dudeja, Chirag Modi(2019), Designing an efficient security framework for detecting intrusions in virtual network of cloud computing, Computers & Security,Volume 85, 2019, Pages 402-422, https://doi.org/10.1016/j.cose.2019.05.016.

[8] 정다빈, 나현대, 김선우. (2020). 디지털 금융 산업 활성화를 위한 망 분리 규제개혁 정책연구(데이터 중심의 핵심가치 보호와 개발 효율성을 중심으로). KPC4IR REPORT No. 11

[9] 김봉재, 정진만, 민홍, 허준영, 정혜동. (2017). 인피니밴드 기반 고성능 클러스터를 위한 효율적인 데이터 선반입 기법. 정보과학회 컴퓨팅의 실제 논문지, 23(5), 293-298.

[10] Headless Browser란? :
https://devahea.github.io/2019/04/13/Headless-Browser%EB%9E%80/
(접속일 : 21년 11월 8일)
[11] 네트워크 세그멘테이션 소개(cisco) :
https://www.cisco.com/c/en/us/products/security/what-is-network-segmentation.html
(접속일 : 21년 11월 8일)
[12] ISO 27001 Network Segmentation Overview. :
https://iso27001guide.com/iso?27001?network?segmentation?overview?iso27001?guide?iso27001?guide.html
(접속일 : 21년 11월 8일)
[13] 송동훈, 임현종, 박수진, 신익현. (2018). 원자력시설 사이버보안 강화를 위한 관리적 보안조치 검증 방법론 연구. 한국통신학회 학술대회논문집, 1048-1049.
[14] CVE-2014-6271. Available online:
https://cve.mitre.org/cgi-bin/cvename.cgi?name=cve-2014-6271
(접속일 : 21년 11월 8일)
[15] CVE-2014-6277. Available online:
https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-6277
(접속일 : 21년 11월 8일)
[16] CVE-2014-6278. Available online:
https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-6278
(접속일 : 21년 11월 8일)
[17] CVE-2017-5638. Available online:
https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2017-5638
(접속일 : 21년 11월 8일)
[18] CWE-20: Improper Input Validation. Available online:
https://cwe.mitre.org/data/definitions/20.html (접속일 : 21년 11월 8일)
[19] CWE-190: Integer Overflow or Wraparound. Available online:
https://cwe.mitre.org/data/definitions/190.html (접속일 : 21년 11월 8일)
[20] Won-Chi Jung, Jinsu Kim, Namje Park (2021). Web-Browsing Application Using Web Scraping Technology in Korean Network

Separation Application. Symmetry 2021, 13, 1550. https://doi.org/10.3390/sym13081550

[21] 악성코드의 새로운 패러다임, Stuxnet. 안랩 Special Report, http://www.ahnlab.com_kr_site_securityinfo_secunews_secuNewsView.do_menu_dist=2&seq=16852 (접속일 : 21년 11월 8일)

[22] 이선호, 한민수. (2015). 산업망에서 APT(지능형 지속위협) 침투경로 분석 및 대응 방안 고찰 - 스턱스넷 사례를 중심으로-

[23] 조병주, 윤장호, 이경호. (2015). 금융회사 정책의 효과성 연구. 정보보호학회논문지, 25(1), 181-195.

[24] 김동훈, 손인수. (2019). 네트워크 침입탐지 기술 연구 동향. 전자공학회논문지 56(8), 2019.8, 3-12(10 pages)

[25] 길아라. (2013). 스니핑 공격에 대응하는 애드-혹 무선 센서 네트워크를 위한 보안 라우팅 프로토콜. 정보과학회논문지 : 정보통신 40(1), 2013.2, 26-35(10 pages)

[26] 김현우. (2013). 베이지안 정리를 이용한 싱크홀 공격 탐지 기법,의사결정학 연구 21(2), 2013.12, 115-124(10 pages)

[27] 박준호, 성동욱, 유재수(2011). 무선 센서 네트워크에서의 에너지 효율적인 선택적 전송 공격 탐지 기법, 한국정보과학회 학술발표논문집 38(1D), 2011.6, 248-251(4 pages)

[28] 최재영, 백현철, 김상복, 심종채, 박재흥. (2014). 세션 하이재킹 공격에 대한 TCP Sequence Number 암호화. 한국지식정보기술학회 논문지, 2014, vol.9, no.6, pp. 707-714 (8 pages)

[29] 김성기, 장종수, 민병준. (2010). 제로데이 공격 대응력 향상을 위한 시그니처 자동 공유 방안. 정보과학회논문지 : 정보통신 37(4), 2010.8, 255-262(8 pages)

[30] 오일석. (2019). 미국 정보기관 제로데이 취약성 대응 활동의 법정책적 시사점. 미국헌법연구, 30(2), 143-185.

[31] 김영선, 서춘원. (2018). 클라우드 컴퓨팅의 웹 스크래핑을 이용한 텍스트 데이터 분석에 대한 연구. 대한전자공학회 학술대회, 1445-1447.

[32] 최동근, 송미선, 임종인, 이경호. (2015). 정보보호담당자의 역할이 조직의 정보보호수준에 미치는 영향. 정보보호학회논문지, 25(1), 197-209.

[33] 박준경, 김범수, 조성우. (2011). 기업정보보호 활동을 위한 조직 구성원들의

태도와 주요 영향 요인. 경영학연구, 40(4), 955-985.

[34] 정원치. (2020) "망분리 적용 및 정착을 위한 기술적 방안에 관한 연구." 국내석사학위논문 제주대학교

[35] 정원치, 박정훈, 박남제. (2019). Safe Web Using Scrapable Headless Browser in Network Separation Environment. 한국컴퓨터정보학회논문지, 24(8), 77-85.

[36] MinSu Kim, Byoungcheon Lee. (2020). A Study on the Assessment Measures for Availability of Information Assets. Jouranl of Information and Security, 20(2), 53-58.

[37] Anish Chapagain. (2019). Hands-On Web Scraping with Python : Perform Advanced Scraping Operations Using Various Python Libraries and Tools Such As Selenium, Regex, and Others

[38] 최재운, 김세헌. (2010). 무선 센서 네트워크에서 통계적 기법을 활용한 웜홀 공격 탐지 한국경영과학회 학술대회논문집 , 2010.6, 1584-1587(4 pages)

[39] Won-Chi Jung, & Namje Park (2020). A Safe Web in Network Separation Environment. Journal of Computational and Theoretical Nanoscience, 17, 3243-3249.

[40] 국가정보원, 국가·공공기관 망 분리 및 자료전송 가이드라인

[41] 디지털데일리, 또 도마오른 물리적 망분리...망분리와 망연결을 함께하는 이상한 나라 비판 : http://m.ddaily.co.kr/m/m_article/?no=196998 (접속일 : 21년 11월 8일)

# ABSTRACT

## A Web-browsing Technique using Web Scraping in Network Separation Environments

Won-chi Jung

Convergence Information Security
Graduate School, Jeju National University
Jeju, Korea

(Supervised by professor Namje Park)

As the threat of hacking increases, strong methods are needed. In Korea, due to a series of information security incidents involving government public institutions. Network separation is a powerful method of separating the internal network from the external network and fundamentally blocking intrusion from external attackers, This method(network separation) had a great effect as a defense security strategy.

However, the type of work has changed due to COVID-19, and the collection and sharing of information has become more important through the development of IT technology..In terms of information security, network separation is one of the ways to ensure security. On the other hand, it is a factor that lowers work efficiency.

Although a technology for detecting attacks by analyzing request data has been studied, it is difficult to detect and defend all intrusion. so it is difficult to use it as a security function.

Therefore, The proposed technique is non-malicious external data connected to the Internet and transmits it to the inside. The technology proposed in this paper utilizes web scraping technology. This technology securely communicates data between the external network and the internal network. This technology will improve the user's work efficiency. It becomes possible to browse the data of the external network from the internal network. This method improves the security and work efficiency of network separation.