석사학위논문

인터넷 라우터 로그를 활용한 SSH 무차별 대입 공격 탐지 및 접근통제

Detection and Access Control of SSH Brute-Force Attack Using Internet Router Logs

제주대학교 대학원

융합정보보안학협동과정

박 정 훈

2020년 8월

인터넷 라우터 로그를 활용한 SSH 무차별 대입 공격 탐지 및 접근통제

지도교수 박 남 제

박 정 훈

이 논문을 융합정보보안학협동과정 석사학위
논문으로 제출함

2020년 6월

박정훈의 융합정보보안학협동과정 석사학위
논문을 인준함

심사위원장 변 영 철 ㊞

위 원 조 정 원 ㊞

위 원 박 남 제 ㊞

제주대학교 대학원

2020년 6월

목 차

표 목 차

그 림 목 차

요 약

최근 정보통신의 발전은 상상을 초월할 정도로 빠르게 발전하고 있다. 이를 통해 우리는 많은 양질의 정보를 빠르게 찾고 공유하며 수집할 수 있게 되었지만, 그 역기능으로 사이버공격 또한 급속도로 증가하고 있다. 증가하는 사이버 공격은 기업 및 기관의 중요한 정보를 탈취하고 가용성을 파괴하며, 신뢰도를 하락시켜 막대한 금전적 피해를 입히고 있다.

이런 사이버 공격을 100% 막을 수 있는 방법은 없으며 모든 보안정책은 외부 공격이 내부로 유입된다는 가능성을 반드시 고려하여야 하고 예방을 통해 이를 수용 가능한 수준으로 낮추는 것이 보안의 가장 큰 역할이다.

접근통제 정책은 사이버 위협을 예방할 수 있는 좋은 방법이며 본 연구에서는 인터넷 라우터에 생성되는 로그를 활용하여 접근통제 정책을 수립하고자 한다.

우선 IT 인프라의 구조와 취약점을 파악하여 인터넷 라우터에 대용량 로그가 생성되는 원인을 파악한다. IT 인프라를 구성하는 네트워크 장비는 역할과 오가는 트래픽을 분석하여 적정 규모로 구축한다. 또한 보안장비는 방화벽을 기반으로 기본적인 통신의 통제를 구축하고, 웹을 보호하기위해 웹 방화벽을 구축하며, 네트워크 대역별 접점 부분이나 해당 네트워크에 속한 모든 장비를 관찰하고 정책을 적용하기 용이한 위치에 침입방지시스템 등 적절한 보안장비를 구축한다.

하지만 IT 인프라 구조상 방화벽 상단에 존재하는 인터넷 라우터는 이런 보안장비의 보호를 받기 힘들며 많은 사이버 위협에 노출되어 있다. 때문에 인터넷 라우터는 불필요한 서비스를 차단하고 필요한 서비스의 경우 적절한 보안 조치를 취해 취약점이 발생하지 않도록 조치한다. 이런 조치를 통해 인터넷 라우터에 접근하는 비인가 접속은 차단되며 많은 로그를 남기게 된다.

비인가 접속으로 차단되는 로그의 대부분은 SSH 무차별 대입 공격이며 이를 통해 공격 근원지에 대한 IP 정보를 수집가능하다. 또한 생성된 로그는 월, 일, 시간, 년도, 장비명, 메시지의 필드로 구분되며 이를 가공하면 빈도수, IP, 접속국가 등 많은 정보를 만들 수 있다.

수집된 인터넷 라우터 로그를 통해 접근통제 정책을 생성하기 위해서는 로그를 단편화하고 분석하여야 한다. 단편화는 블랙리스트를 탐지하기 위한 특징인 월, 일, 시간, 년도, IP로 필드를 구성되며 중복되는 문장을 삭제하여 생성한다. 분석은 ElasticSearch를 통해 진행하고 Kibana로 시각화 하여 악의적인 공격 근원지에 대한 보다 세밀한 정보를 확인한다. 또한 공격 빈도수, 날짜별 접근, 공격 국가 등 블랙리스트 IP를 선택하기 위한 특징들을 정하여 임계치 이상의 IP를 블랙리스트로 정하고 방화벽을 통한 차단 정책의 객체로 삼는다.

IT 인프라를 안정적으로 관리하기 위해서는 적절한 규모의 네트워크와 용도에 맞는 보안장비가 필요하지만 이를 관리하기 위한 정책도 필요하다. 인터넷 라우터의 로그를 활용한 접근통제 정책은 IT 인프라의 보안을 더욱 강화할 수 방법이며 명확한 표준의 부재로 인하여 직관이나 경험에 의존보안 로그의 분석에 작은 도움이 될 것이다.

주제어 : SSH 무차별 대입 공격, ELK Stack, IT 인프라, 로그, 접근통제

Ⅰ. 서 론

1. 연구의 필요성

최근 정보통신의 발전은 상상을 초월할 정도로 빠르게 발전하고 있다. 이를 통해 우리는 많은 양질의 정보를 빠르게 찾고 공유하며 수집할 수 있게 되었지만, 그 역기능으로 사이버공격 또한 급속도로 증가하고 있다. 증가하는 사이버 공격은 중요한 정보를 탈취하고 가용성을 파괴하며, 기업의 신뢰도를 하락시켜 막대한 금전적 피해를 입히고 있다.

이런 사이버공격은 각 단계별 위협행위들로 나눌 수 있다. 일반적인 공격절차는 가장 먼저 공격대상에 대한 정보수집 단계이며, 그 다음 수집한 정보를 바탕으로 시스템 침입단계를 거치게 된다. 그리고 지속적인 침입 및 다른 시스템의 공격을 위한 공격전이 단계를 거치게 된다. 사이버 공격을 100% 막을 수 있는 방법은 없으며, 모든 보안정책은 외부 공격이 내부로 유입된다는 가능성을 반드시 고려하여야 하고, 예방을 통해 이를 수용 가능한 수준으로 낮추는 것이 보안의 가장 큰 역할이다.

사이버 공격을 수용 가능한 수준으로 낮추기 위한 많은 보안장비가 IT 인프라에 구축되고 있으며 심층적 보안을 통해 중요한 정보 자산을 방어하고 있다. 이런 보안장비는 가장 기본적인 역할을 하는 방화벽을 시작으로 허용된 트래픽을 좀 더 세밀하게 분석하는 IPS 등 많은 장비로 중첩되어 구축되며 사용자 PC까지 보안 소프트웨어를 설치하여 생길 수 있는 보안 사고에 대비한다.

하지만 IT 인프라의 중첩된 보안 장비의 구축에도 인터넷과 접점에 위치한 인터넷 라우터는 보호하는 장비를 따로 두는 경우가 드물다. 라우터는 서로 다른 네트워크를 연결해주는 장비로 패킷의 최적 경로를 지정하고 전달한다. 이런 기능은 외부에서 ICMP나 IP Scanning을 통해 쉽게 라우터나 네트워크의 정보를 확인할 수 있으며 파악된 정보를 통해 많은 악의적 접근들이 발생한다.

기존의 사이버공격은 웜, 바이러스, 분산서비스거부공격(DDoS) 등 네트워크상의 시스템의 가용성을 파괴하는 공격이 주를 이뤘지만, 최근에는 정상적인 패킷을 가장하여 특정시스템을 지속적으로 침입을 시도하는 공격으로 바뀌고 있다. 이러한 접근 중에는 시스템에 로그인하여 원격으로 명령을 실행하고 파일을 복사할 수 있는 프로토콜인 SSH(Secure Shell)를 이용한 정상적인 접근도 포함되어 있으며, 특정 시스템에 지속적으로 접근을 시도하여 권한을 탈취하려고 한다. 이러한 행위는 SSH 무차별 대입 공격(SSH Brute-Force Attack)에 해당하며 정상 접근을 가정하고 있어 보안장비에서 탐지가 쉽지 않다. 또한 그 공격위치가 보안장비로 보호받지 못하는 네트워크 장비라면 더욱 공격을 받을 수밖에 없는 취약한 구조일 것이다.

본 연구에서는 인터넷 라우터에 가해지는 SSH 무작위 공격을 로그를 통해 탐지하고 공격 근원지의 정보를 수집하려고 한다. 이 근원지 정보는 직접적인 공격자의 IP 정보이거나 정보 노출을 숨기기 위한 봇넷(botnet)의 가능성이 높으며 해당 IP를 어떻게 접근 통제 할 것인지에 대하여 연구하고자 한다.

2. 연구의 목적 및 문제 설정

본 연구는 IT 인프라를 관리하는 관리자가 인터넷 라우터에서 생성된 SSH 무차별 대입 공격 로그를 통해 해당 공격의 근원지에 대한 접근 통제를 설정함으로써 발생할 수 있는 사이버 위협을 차단하고자 한다. 본 연구의 목적을 달성하기 위해서 다음과 같은 문제를 설정하였다.

첫째, IT 인프라의 구조적 취약점이 있는가?

적절한 규모의 네트워크를 구성하고 중첩된 형태의 보안장비를 구성하여 기업과 조직의 중요 정보자산을 보호하고 있는 IT 인프라지만 어떤 구조적 취약점이 있는지 확인하고자 한다.

둘째, 취약한 IT 인프라에 대한 방어

취약한 구조의 IT 인프라를 보안장비로 보호하지 않는 이유와 이 장비를 적절

하게 보호하는 방법에 대해 알아보고자 한다.

셋째, 어떠한 공격이 발생하며 어떤 징후가 있는지 확인

취약한 구조에 위치한 장비로의 접근은 해당 장비의 적절한 보안 설정에 따라 로그를 남기며, 로그는 공격자의 정보 및 어떤 형태의 공격인지 확인할 수 있다.

넷째, 공격 근원지에 대한 정보와 어떻게 막을 것인가?

로그를 통해 공격 근원지에 대한 IP 정보를 수집가능하다. 이를 통해 접근의 빈도수와 국가별 접근 등 여러 정보를 확인할 수 있으며 이 정보를 토대로 외부에서의 접근과 내부에서 공격 근원지에 대한 접근을 동시에 차단하고자 한다.

3. 연구의 범위와 방법

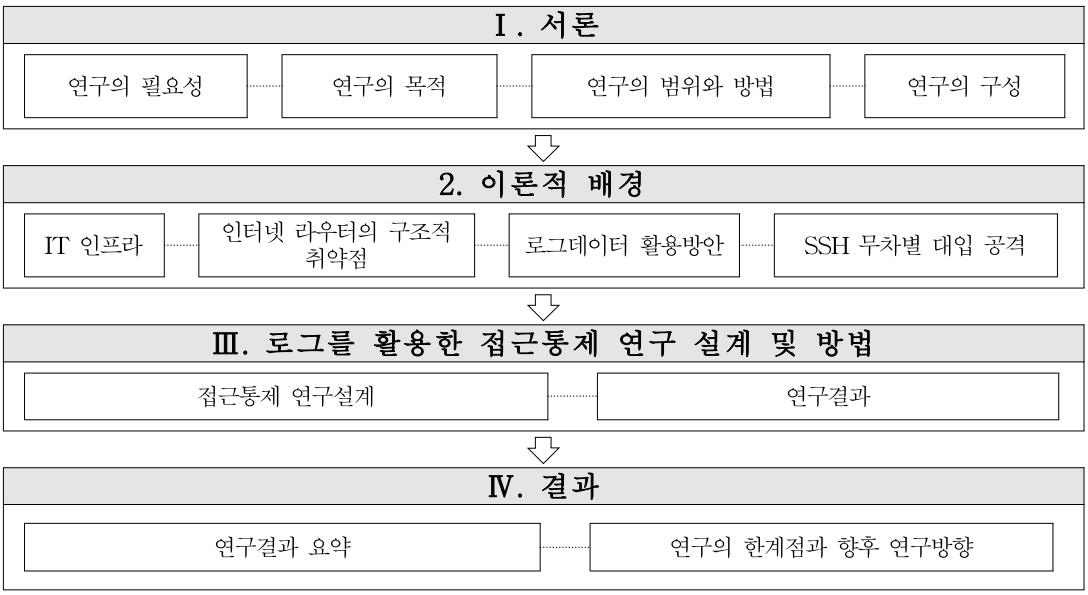
본 연구의 목적을 달성하기 위해서는 기존 논문을 참고하여 이론적 배경을 작성하였다. 이론적 배경은 본 연구와 관련된 논문, 동향보고서, 가이드라인 및 인터넷을 통한 자료 검색 등 IT 인프라와 라우터 및 로그에 대한 자료를 찾아보았다. 이를 통해 연구 목적을 위한 기초 자료인 IT 인프라의 구성과 보안 장비의 구축, 인터넷 라우터의 취약점 및 가설 검증을 위한 발생 로그를 수집하였다.

실증 분석을 위한 자료 수집은 IT 인프라의 네트워크 장비인 인터넷 라우터의 로그를 수집하였으며 공격 로그만 선택하여 단편화 시켰다. 선택 방법은 공격 패턴의 문자열을 가진 로그를 특정하여 분리하고 로그 메시지 중 중복되는 문구를 제거함으로써 로그를 단편화 시켰다. 공격 로그만으로 단편화된 자료는 월별로 가공하여 1년 분량을 수집하고 월, 일, 시간, 년도, IP의 필드로 구성된 CSV 파일로 저장하였다.

수집 저장된 공격 로그는 빅 데이터 분석 도구인 ElasticSearch를 이용하여 분석하고 Kibana를 통해 시각화하여 공격 IP에 대한 가시성을 확보하였다. 분석된 공격 IP와 적절한 임계치를 통해 블랙리스트 IP를 정하고 방화벽을 통한 접근통제 정책에 적용까지 연구의 범위로 한다.

4. 연구의 구성

본 연구의 구성은 총 5장으로 구성된다. 제Ⅰ장 서론에서는 연구의 필요성, 목적, 범위와 방법, 구성을 밝힌다. 제Ⅱ장 이론적 배경에서는 IT 인프라에 가해지는 사이버 위협과 이를 통해 구축되는 보안장비 및 그 구성에 대해 알아보고 안전한 네트워크 구조와 유형별 네트워크 구조를 통해 네트워크가 가지는 구조적 취약점을 확인한다. 또한 로그의 수집과 분석 및 시각화를 통해 취약한 장비의 로그 데이터를 분석하는 방법을 정리하고자 한다. 로그의 공격 패턴인 SSH 무차별 공격에 대한 이론적 검토와 블랙리스트 IP를 어떻게 관리하는지 확인한다. 제Ⅲ장은 인터넷 라우터 로그를 활용한 접근통제 연구 모형을 만들고 가설을 설정한다. 연구 설계에서는 원시 로그의 분류하고 단편화하여 분석도구에 사용하기 적절한 형태로 만들어 분석한다. 이를 통해 블랙리스트 IP를 탐지하고 접근통제 정책을 만들어 적용한다. 제Ⅳ장 결론에서는 접근통제를 통한 연구의 시사점과 이로 인한 한계점을 밝히고 향후 어떤 방향으로 연구가 진행되어야 하는지 방향을 제시한다.



[그림 Ⅰ-1] 연구의 구성

Ⅱ. 이론적 배경

1. IT 인프라(IT Infrastructure)

우리는 IT 인프라를 통해 안전하고 편리하게 정보서비스를 이용하고 있다. 이를 위해 기업과 기관에서는 안전한 구조의 네트워크와 적절한 수준의 보안장비를 구축한다. 하지만 모든 시스템에는 구조적인 취약점이 존재할 수 있으며 이를 보안하거나, 그 징후를 사전에 예방한다면 더욱 안전하게 정보서비스를 이용할 수 있을 것이다. 기본적인 IT 인프라는 어떻게 이루어지며 장비별 역할을 통해 발생 가능한 취약점을 알아볼 수 있다.

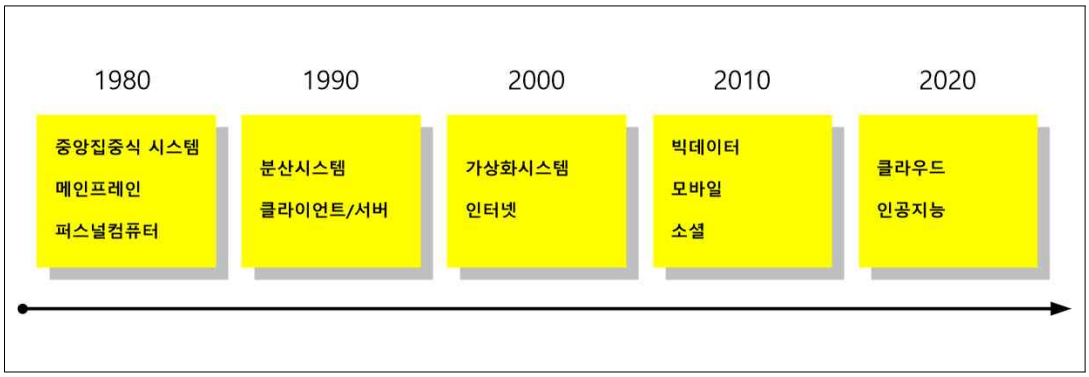
1) IT 인프라 정의

기업 및 기관에서 제공하는 어플리케이션을 가동하기 위한 물리적인 하드웨어, 운영체제, 네트워크 등 기반이 되는 시스템을 말한다. 이 중에서 네트워크는 각 계층을 담당하는 물리적인 네트워크 장비를 통해 사용자와 사용자를 연결하거나 시스템을 사용할 수 있도록 통신서비스를 제공하는 통신망을 의미한다. 이를 위해 서로 다른 네트워크를 연결하는 라우터(Router)와 물리적인 광케이블을 연결하여 회선의 가용성을 보장해주는 ROADM(Re-configurable Optical Add-Drop Multiplexer)과 같은 전송장비로 통신 기반을 구성한다. 또한 네트워크는 Data, Voice 등 용도별로 구분할 수 있으며 그에 맞는 객관적인 장비 규모를 산정하여 구성하여야 한다[11].

IT 인프라를 구성하는 보안장비는 네트워크의 구조 및 규모에 따라 적절한 형태로 구축된다. 생성되는 각종 주요 정보의 기밀성, 가용성, 무결성을 유지하고 기술적, 물리적, 관리적 조치를 통해 인프라의 안전성을 확보한다.

초기의 IT 인프라는 가용성이 강조된 아주 단순한 구조의 형태를 가졌지만 정

보 통신의 발전으로 많은 사이버공격들이 발생하여 그 구조도 변하게 되었다. IT 인프라도 이에 맞추어 변화하고 있으며 서비스별 네트워크를 분리하고 각 접점에 보안장비를 구축하여 발생하는 사이버공격을 방어하고 있다.



[그림 Ⅱ-1] IT 인프라 변천사

[그림Ⅱ-1]과 같이 IT 인프라는 1980년대 메인프레임을 중심으로 한 중앙집중식 계층화 구조의 인프라부터 1990년대의 클라이언트/서버, 2000년대의 인터넷과 가상화 시스템, 2010년의 빅데이터, 모바일, 소셜 네트워크, 그리고 현재의 인공지능과 클라우드를 중심으로 한 디지털 혁신 플랫폼까지 비즈니스 모델의 변화에 따라 기업 및 조직이 사용하는 핵심 애플리케이션의 종류와 형태, 그리고 이를 뒷받침하는 IT 인프라의 구조도 변화해 왔다. 변화하는 IT 인프라에 따라 많은 취약점이 발생하였으며 이를 공격하는 사이버 위협 또한 크게 증가하게 된다[27].

IT 인프라에 가해지는 대표적인 공격은 랜섬웨어, APT(Advanced Persistent Threat), 웹 공격, DDoS(Distributed Denial of Service) 등 여러 가지 공격이 있다.

랜섬웨어는 사용자 PC의 정보를 암호화 하고 이를 해제하는 방법으로 금품을 요구하거나 암호화 상태를 유지시켜 정보의 사용을 제한하는 목적으로 피해를 주는 공격이다. 주된 감염지는 인터넷 사용과 이메일 등 일반적인 IT 인프라의 이용에 의한 것으로 예방을 위해서는 사용자 PC와 시스템에 최신 보안 업데이트 및 데이터 백업을 진행하고 접속 사이트의 안전 상태 확인이 있을 것이다[13].

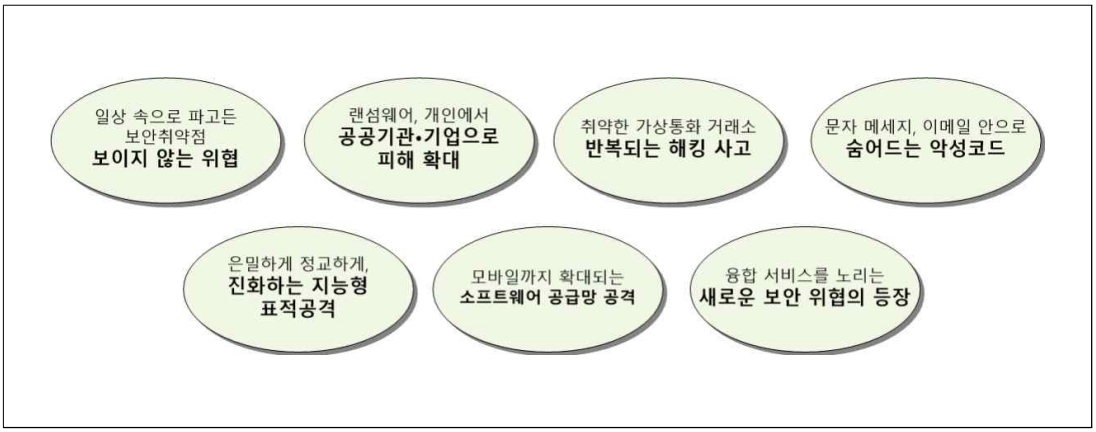
APT(Advanced Persistent Threat)는 지능형 지속 위협이라 불리며 하나의 목적을 두고 장기간 지속적으로 가하는 공격을 말한다. 장기간 준비로 공격은 은밀하고 정교하며 다양한 공격 방법을 동원하여 집요한 공격이 특징이다. 해당 공격의 피해는 정보 취득 및 파괴를 유발하는 규모가 큰 공격이 많아 피해 또한 대규모로 발생한다[13].

웹 공격은 제공 웹서비스의 취약성 때문에 항상 존재하고 위험성이 높은 공격이다. 간단한 Scan 공격부터 Injection, XSS 등 많은 시도를 허용된 서비스 내에서 진행되고 있어 웹 어플리케이션에 대한 보안이 중요하다. OWASP(The Web Application Security Project)는 웹 어플리케이션에 대한 10 취약점을 발표하고 해당 공격 항목을 조치할 수 있는 예방 방법을 도출하는데 도움을 준다[13].

DDoS(Distributed Denial of Service)는 IT 인프라의 가용성을 파괴하는 공격이다. 공격의 방법이 비교적 단순하며 노력과 비용이 적게 들어 높은 잠재력을 가진 공격이라고 할 수 있다. 또한 IT 인프라가 허용한 서비스에 대한 정상적인 접근으로, 이를 감당할 수 있는 규모를 고려한 인프라 설계 및 방어가 필요하다[13].

위와 같이 정형화된 공격 이외에 암호화된 통신 트래픽의 증가로 인한 IT 인프라의 위협이 증가하고 있다. 암호화된 통신 트래픽은 내부를 확인할 수 없어 트래픽 속 악성 프로그램이나 의심스러운 행위를 찾아내기 불가능하다. 이런 통신 트래픽의 비율은 계속 증가하고 있으며, 이에 따라 암호화된 트래픽의 가시성 확보 또는 악의적인 근원지와의 통신 단절이 필요하다. 하지만 암호화된 트래픽을 복호화하여 그 내용을 확인하는 행위는 개인의 민감한 정보를 침해할 수 있고 복화화 과정에서 발생하는 시스템의 지연은 전체 네트워크 속도에 영향을 미친다. 때문에 유해한 공격 근원지를 미리 차단하여 사이버 위협을 예방하는 정책이 가장 적절한 보안 방법일 것이다[1][6][13].

[그림 II-2]는 인터넷진흥원에서 전망한 2020년 사이버 공격전망으로 앞에서 언급한 공격 외에 공급망과 융합서비스를 노리는 공격이 증가하고 있음을 보여준다. 이런 공격들은 집중화되며 점차 자동화 되어 인력으로 방어하기에는 힘들며 취약점에 대한 철저한 보안과 근원지에 대한 차단이 필요하다고 생각된다.

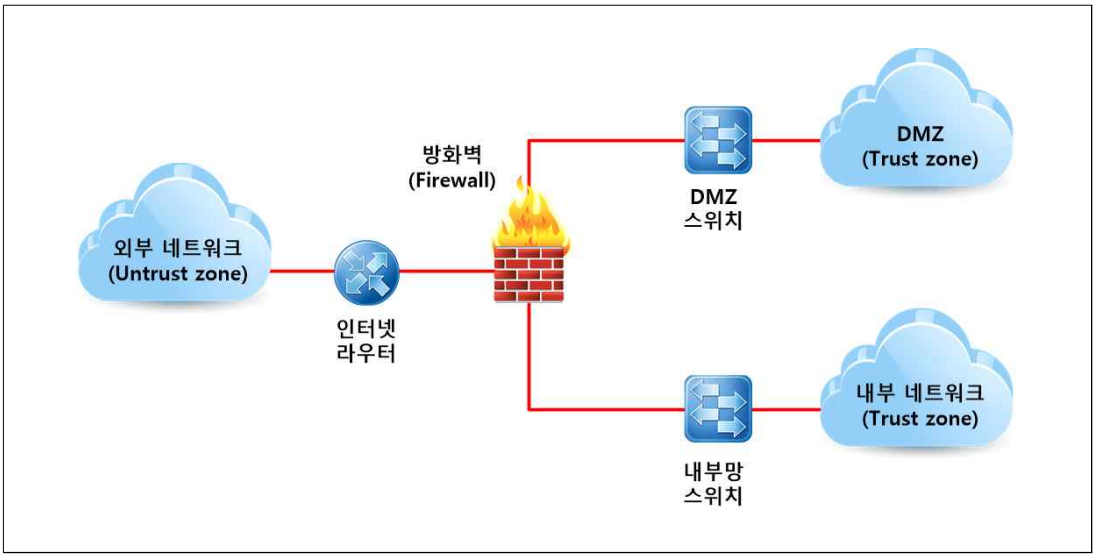


[그림 Ⅱ-2] 2020년 7대 사이버 공격전망 (인터넷진흥원)

2) 네트워크 용량 산정

기업과 기관에서는 정보서비스의 가용성, 무결성, 기밀성을 위해 용도에 맞는 네트워크 하드웨어와 소프트웨어를 갖춘다. 이를 위해 다른 네트워크와 통신을 위한 라우터를 최상단에 위치하고, 다음에 방화벽을 위치시켜 내부망과 외부망으로 분리하여 사용자와 시스템을 보호한다. 그리고 외부에서 접근되어야 할 필요가 있는 서비스를 위해 DMZ를 운영한다. 외부에 서비스되는 대표적인 서비스는 웹, 메일, DNS 등이 있으며 이를 보호하기 위해서 웹 방화벽, DDoS 방어 장비, 침입방지시스템(IPS) 등을 추가 설치하는 경우도 있다. [그림 II-3]은 방화벽을 기준으로 네트워크를 나누는 일반적인 네트워크의 구조이며, 이를 바탕으로 제공하는 서비스 또는 사용량에 따른 적절한 네트워크 규모를 산정한다.

IT 인프라를 구성하기 위한 세부적인 네트워크장비로는 라우터, 백본스위치, 서버팜 스위치, 사용자 대역 워크그룹스위치 등 주어진 역할에 따라 분류된다. 한국정보통신기술협회(TTA)에서는 네트워크 구축 시 참고할 수 있는 “정보시스템 하드웨어 규모산정 지침”을 제공하고 있으며 이를 바탕으로 객관적인 장비의 규모산정을 통해 적절한 구축을 할 수 있다. 이는 네트워크 장비를 설계할 때 아주 중요한 지표가 되며 과도하게 설계되어 불필요한 예산이 낭비되거나 혹은 부족한 용량으로 설계되어 정보서비스의 지연 및 병목현상(Bottleneck)이 발생되는 것을 방지할 수 있다[28].



[그림 Ⅱ-3] 방화벽 Zone 구성

네트워크 구축 시 장비는 위치와 형태 및 역할에 따라 그 규모를 다르게 산정하여야 하며, 오가는 트래픽을 분석하여 적정 규모의 IT 인프라를 구축해야 한다. 기존 네트워크를 사용 중이면 해당 트래픽을 분석하여 규모 산정이 쉽게 가능하지만 신규의 경우는 트래픽을 추정하여야 한다. 따라서 표준으로 제시하는 서비스별 보정치와 장비의 역할별 보정치를 활용하여 네트워크 규모를 산정 할 수 있다.

[표 Ⅱ-1] 주요서비스 소요 대역폭 및 보정계수

주요서비스 구분	단위	크기 [대표크기]	희망 응답 시간	요구 대역폭	보정계수	
					FE(F)	GE(G)
일반업무(WEB기반)인터넷검색포함	건	수백KB~수MB[10MB]	3초	27Mbps	0.3	0.003
문서전송을 포함한 범용서비스메일/문서시스템 등	건	수백 KB~수MB[100MB]	10초	80Mbps	0.8	0.08
전화(인터넷전화)	통화	64Kbps[100Kbps]	실시간	100Kbps	0.001	0.0001

주요서비스 구분	단위	크기 [대표크기]	희망 응답 시간	요구 대역폭	보정계수	
					FE(F)	GE(G)
보안업데이트	건	수백 KB~수 MB[10MB]	3초	27Mbps	0.3	0.03
Full HD 영상 영상회의/CCTV	채널	4.5~9Mbps [10Mbps]	실시간	10Mbps	0.1	0.01
3D 지도검색 지역/각도에 대한 크기 차이	건	10~60MB [60MB]	10초	48Mbps	0.5	0.05
무선 LAN(Wi-Fi) AP 장치접속용	대	802.11n[15~150Mbps] 802.11ac[88~867Mbps]	실시간	150Mbps 867Mbps	- -	0.15 0.87

[표 Ⅱ-2] 접속형 L2/L3 스위치 규모산정 항목 및 보정치

항목	내용	입력값 범위	일반 값
다운링크 포트 용량	필요 포트의 이론적 최고 용량	1 ~ xxx	1Gbps
소요 다운링크 포트 수	다운링크 활용되는 포트수 총합	1 ~ xxx	
다운링크 확장계수	다운링크의 향후 사용이 예상되는 포트수	100%~200%	
다운링크 안정성계수	예기치 못한 네트워크 증가 및 시스템의 안정적 운영을 위한 여유율	100%~150%	120%
다운링크별 보정계수	주요 서비스 보정계수	0 ~ 1	120%
업링크 포트 이중화	이중화 구현 기술에 따른 소요 포트 비율	100%~200%	0.1
이론적 패킷 처리상수	1GE에서 이론적 최대 패킷 처리상수		1,488,095
양방향상수	Tx/Rx를 고려한 양방향상수		2

항목	내용	입력값 범위	일반 값
산정식	· 다운링크 포트 전체 수량 = 다운링크 소요 포트수 × 다운링크 확장계수 × 안정성 계수 · 업링크 포트 용량 = 다운링크 포트 용량 × 다운링크 포트 전체 수량 × 다운링크별 보정계수 · {업링크 단위 용량, 업링크 포트수량} = F{업링크 용량} : 이중화 고려 · 스위칭 용량 = {∑(i의 이론적 최대 용량 × I의 포트 수량)} × 양방향상수 단, I∈다운링크 & 업링크 I/O 인터페이스 종류 = {10/100Base-Tx, 100/1000Base-Tx, 1000Based-Fx 등} · Throughput = {∑다운링크 & 업링크 인터페이스별 용량(i)} × 이론적 패킷 처리상수		

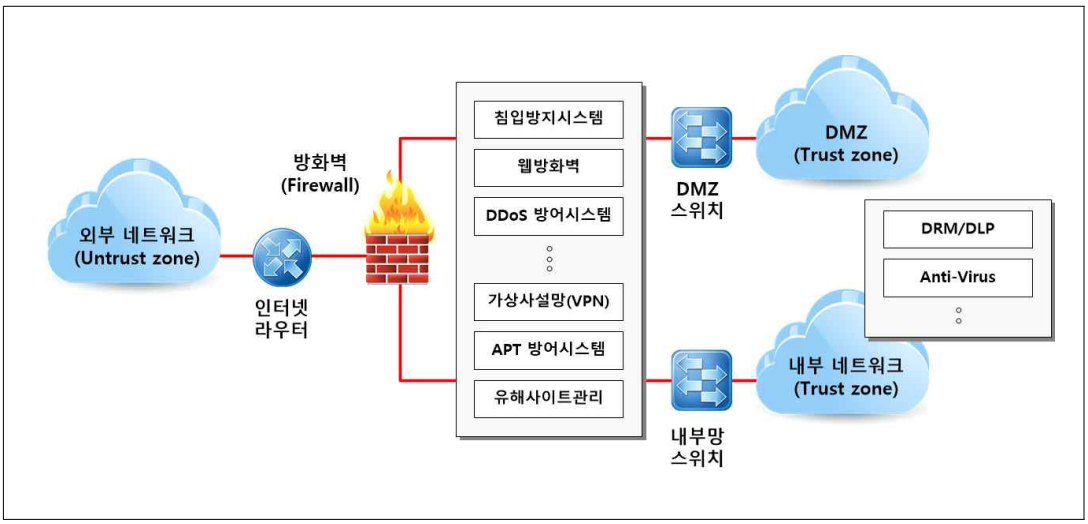
[표 II-1]과 [표 II-2]를 통해 트래픽과 네트워크의 규모를 산정할 수 있다. 네트워크 장비별 규모 산정과 이중화 여부를 통해 적절한 장비가 선택되면 용도에 맞는 네트워크 설계를 진행하여야 한다. 서비스 및 업무별 용도에 맞는 대역을 나누는 것은 전체 IT 인프라 성능 및 보안에 아주 중요한 역할을 한다. 예를 들어 사용자 대역과 업무 서버군의 네트워크가 구분이 없다면 사용자가 업무서버를 이용하는 행위에는 변화가 없지만 업무서비스 외 접근이 언제든지 가능하고 잘못된 조작이나 실수로 업무 시스템을 파괴할 수도 있을 것이다. 업무별 분리된 네트워크를 통해 용도에 맞는 서비스를 제공하고 그에 적절한 보안장비를 구축하여 필요에 따른 접근만 허용한다면 제공하는 정보서비스에 대한 가용성과 기밀성을 보장할 수 있을 것이다[28].

3) 보안 동향 및 장비 구성

적절한 용량산정으로 가용성을 확보한 IT 인프라는 보안장비를 통해 발생할 수 있는 사이버 공격들을 방어해야 한다. 많은 공격들을 방어하기 위해서는 용도에

맞는 보안장비가 필요하며 적절한 위치에서 정보 자산의 심도 있는 부분까지 보호하여야 한다. 방화벽을 기반으로 기본적인 통신의 통제를 구축하고, 웹을 보호하기위해 웹 방화벽을 구축하며, 네트워크 대역별 접점 부분이나 해당 네트워크에 속한 모든 장비를 관찰하고 정책을 적용하기 용이한 위치에 침입방지시스템 등 적절한 보안장비를 구축한다. 핵심적인 정보 자산을 보호하기 위해서는 IT 인프라 내 보안 장비를 심층적으로 배치하여 허용된 사람에게(기밀성) 적절한 형태로 적시에(가용성) 정확한 정보를(무결성) 제공하여 정보의 가치를 유지하여야 한다.

[그림 II-4]는 IT 인프라 구조상 보안장비가 어떤 위치에 구축되는지 개념적으로 설명한 그림이다. 기업 및 기관에서 제공하는 서비스에 따라 보안 장비가 달라질 수 있으며, 내부망에서 주로 사용하는 업무 및 인터넷 서비스에 따라서도 보안장비가 달라질 수 있다.



[그림 II-4] IT 인프라 보안 시스템

웹서비스를 제공하고 있다면 DMZ 구간에 웹을 보호하기 위한 웹 방화벽과 DDoS 방어를 위한 시스템 그리고 방화벽을 통해 허용된 포트를 통해 들어오는 알려진 공격의 패턴을 차단하기 위한 침입방지시스템(IPS)이 필요할 것이다.

내부망도 사용하는 업무의 기밀성이 강조된다면 VPN(Virtual Private Network)

을 이용하여 오가는 트래픽을 암호화하고, 단말에는 문서유출방지를 위한 소프트웨어와 백신을 통한 단말 보호를 한다. 또한 NAC(Network Access Control)를 통해 모든 IP를 통제하고 사용자 PC의 필수 프로그램 설치를 보장한다.

IT 인프라의 발전과 더불어 사이버 위협의 변화에 따라 보안장비도 변화하는 트렌드, 이슈사항에 따라 바뀌고 있다. 따라서 변화하는 보안전망을 통해 어떤 보안장비가 필요하며 이에 따른 IT 인프라의 구성은 어떻게 이루어지는지 확인해볼 필요가 있다.

보안장비의 경우 매해 새로운 트렌드, 이슈사항이 발생하고 있으며 여러 전망들을 제시되고 있다. 이런 전망은 한국정보보호산업협회의 정보보호산업동향에서 분석한 '2019 정보보호 분야 주요 이슈사항 분석'을 통해 알 수 있다[22].

미국의 정보기술 연구 및 자문회사인 가트너사(Gatner)는 10대 전략 트렌드로 자율 사물, 증강 분석, 인공지능 등을 제시하고 있으며, 마이크로소프트(MS)에서는 프라이버시 보호, SNS를 둘러싼 가짜뉴스 논란, 국가주도의 사이버공격, AI가 직면한 새로운 과제 등을 제시하였다.

[표 Ⅱ-3] 가트너, MS, 글러벌보안기업, KISA의 2019년 보안 전망

구분	가트너	마이크로소프트	글로벌 보안기업	한국인터넷진흥원
1	Autonomous Things	프라이버시 보호	공급망 공격	크립토재킹 확산
2	Augmented Analytics	SNS를 둘러싼 가짜 뉴스 논란	IoT	SNS를 이용한 악성코드 유포
3	AI-Driven Development	보호무역주의 확산	AI	엔드포인트 보안취약점 겨냥한 공격
4	Digital Twins	사이버 공격 논의	개인정보 보호	지능화된 스피어피싱과 APT 공격

구분	가트너	마이크로소프트	글로벌 보안기업	한국인터넷진흥원
5	Empowered Edge	AI가 직면한 새로운 과제, 윤리	간편해진 공격도구	사물인터넷을 겨냥한 신종사이버위협
6	Immersive Experience	AI와 경제, 그리고 AI와 일자리	표적공격	소프트웨어 공급망 관련 사이버공격
7	Blockchain	사람을 위한 기술	클라우드	악성행위탐지를 우회하는 공격기법 진화
8	Smart Spaces	서서히 좁혀지는 지역 격차		
9		주권, 인권 그리고 클라우드		
10		기술의 발전과 지역사회		

[표 II-3]은 가트너, 마이크로소프트, 글로벌보안기업 및 한국인터넷진원에서 발표한 2019년 보안전망으로 이를 통해 예측 가능한 기술 트렌드, 이슈, 보안위협 등을 알 수 있으며 다음과 같은 몇 가지 특징을 가지고 있다.

첫째, 인프라의 변화다.

IT 인프라는 사물인터넷(IoT), 클라우드(Cloud), 블록체인(Blockchain)등 기술의 발전으로 변화하고 있다. 급격한 IoT 장비의 보급으로 관리 대상이 늘어나고 그로인해 계정관리 등 보안이 취약하게 된다. 이를 악용하여 DDoS 공격의 Agent로 이용하는 사례가 늘어나고 있으며, 지속적인 잠재적 위협으로 자리잡고 있다. 또한 클라우드는 분산된 환경에서 가상화된 컴퓨터 자원을 요구하는 형태의 접근으로 관리의 중요성이 강조된다.

둘째, 인공지능 보안장비의 증가

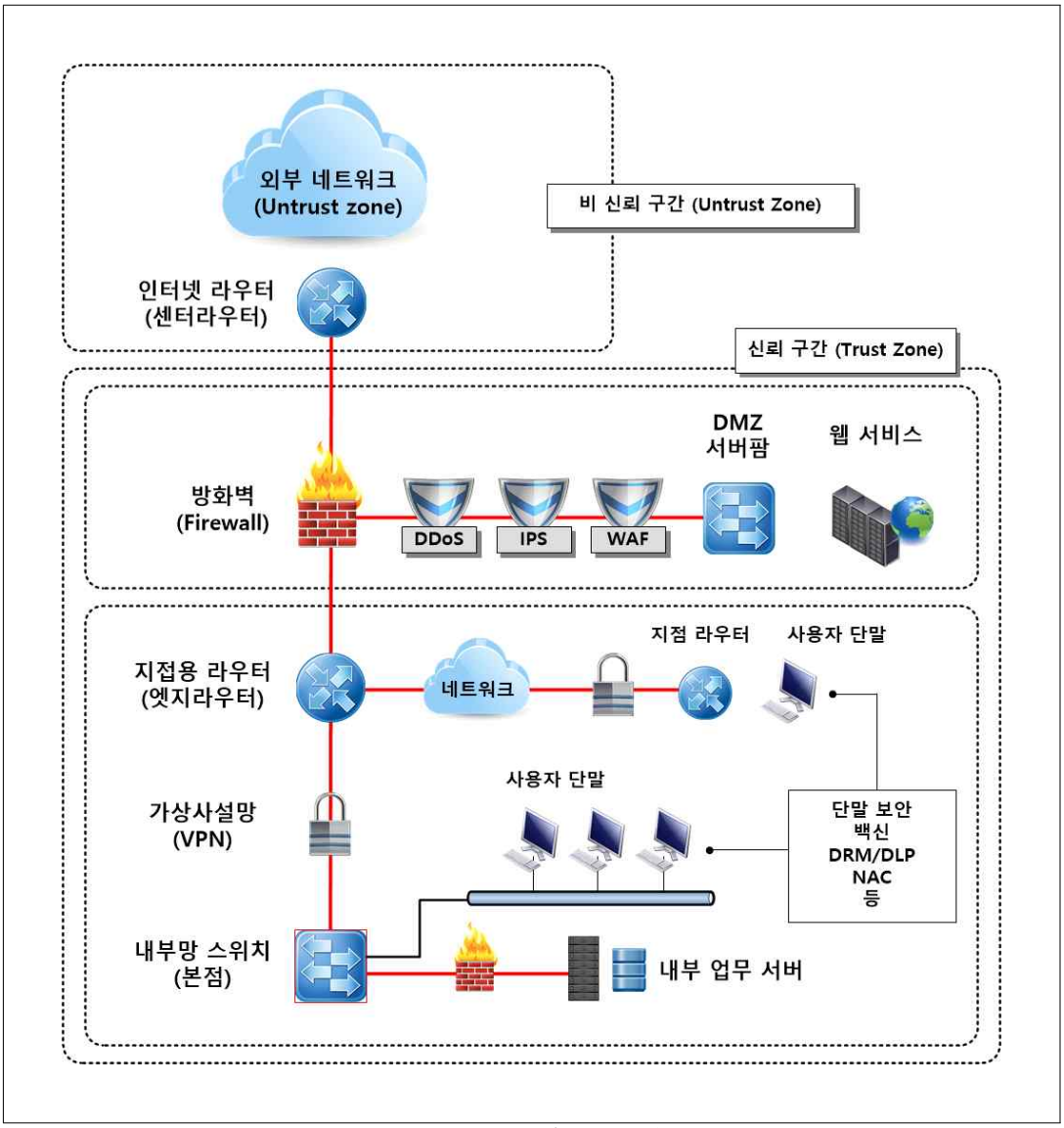
인프라의 변화는 다양한 환경의 기기들의 등장과 이를 이용한 다양한 서비스의 증가로 이어지고 이를 모두 탐지하고 대응 하는데 기존 보안장비로는 불가능하다. 때문에 보안전문가들은 보안에 인텔리전스(Inteligence)를 더하여 기능을 더욱 발전시켰다. 특히, 이 작업은 보안전문가 뿐만 아니라 인공지능, 네트워크, 알고리즘 등 여러 분야 전문가의 협업이 필요한 작업이다.

셋째, 개인정보보호 관련 규제 강화

IT 인프라를 통해 생산되는 많은 DATA가 단순한 자료가 아닌 가공을 통해 새로운 의미의 정보로 바뀌고 있다. 따라서 개인이나 기업 및 기관 등 많은 곳에서 생산되는 자료는 상당히 중요하며 가공하여 개인을 식별할 수 있는 개인정보에 대해서는 법률을 통한 규제가 강화되고 있다.

트렌드 및 이슈를 종합해보면 인프라의 변화에 따른 관리 대상 장비의 증가와 이를 이용한 DDoS 증가와 접근통제의 중요성이 강조되었고, 공격의 다변화와 자동화에 따른 방어대책과 개인정보의 중요성에 따른 자료의 중요성이 올라갔다. 이러한 위협을 토대로 IT 인프라를 구성하는 보안장비를 구축한다면 안전한 IT 인프라가 될 것이다[22].

[그림 III-5]는 IT 인프라를 구성하는 네트워크와 보안장비를 어떻게 구축하는지 보여주는 구성도이다. 변화하는 인프라와 그에 따른 사이버 위협의 증가 그리고 중요성이 높아지는 많은 자료들 보호하기 위해 많은 장비들이 구축된다. 이런 장비들을 계층적으로 나열하여 정보자산을 보호하지만 취약한 부분은 존재하며 그 부분을 보강하거나 아니면 그로 인해 발생하는 징후를 확인한다면 좀 더 높은 수준의 보안을 할 수 있을 것이다.

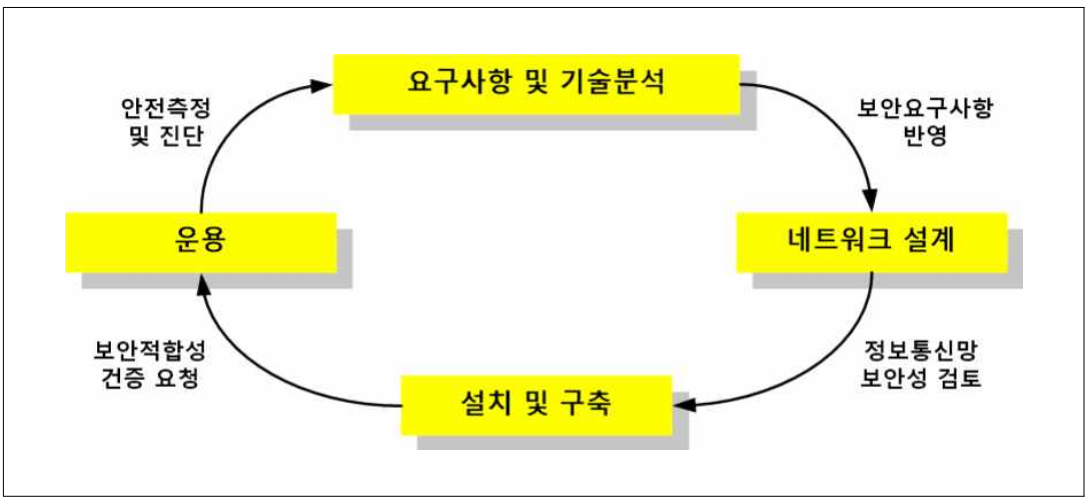


[그림 II-5] IT 인프라 보안시스템 구성안

2. 인터넷 라우터의 구조적 취약점

1) 안전한 네트워크 구축 절차

규모에 맞는 가용성이 확보된 네트워크 장비와 변화하는 트렌드에 맞는 적절한 보안장비를 통해 네트워크를 구축해야 한다. 즉, 안전한 네트워크는 사용목적과 요구사항에 맞도록 구축되어야 한다. 하지만 편리하고 효율적인 네트워크 구축도 안전성이 결여된다면 외부 위협에 노출되어 막대한 피해를 입을 수 있다. 따라서 네트워크 구축 시 외부의 위협으로부터 정보자산을 보호하기 위해서는 일정한 기준을 통해 네트워크 장비와 보안장비를 구축하여야 한다. 국가정보원은 "안전한 정보통신 환경 구현을 위한 네트워크 구축 가이드라인"을 통해 네트워크 구축 시 준수해야 할 구성기준과 보안 고려사항을 제공하여 안전한 네트워크 구축을 지원한다.



[그림 II-6] 안전한 네트워크 구축절차 (국가정보원)그림

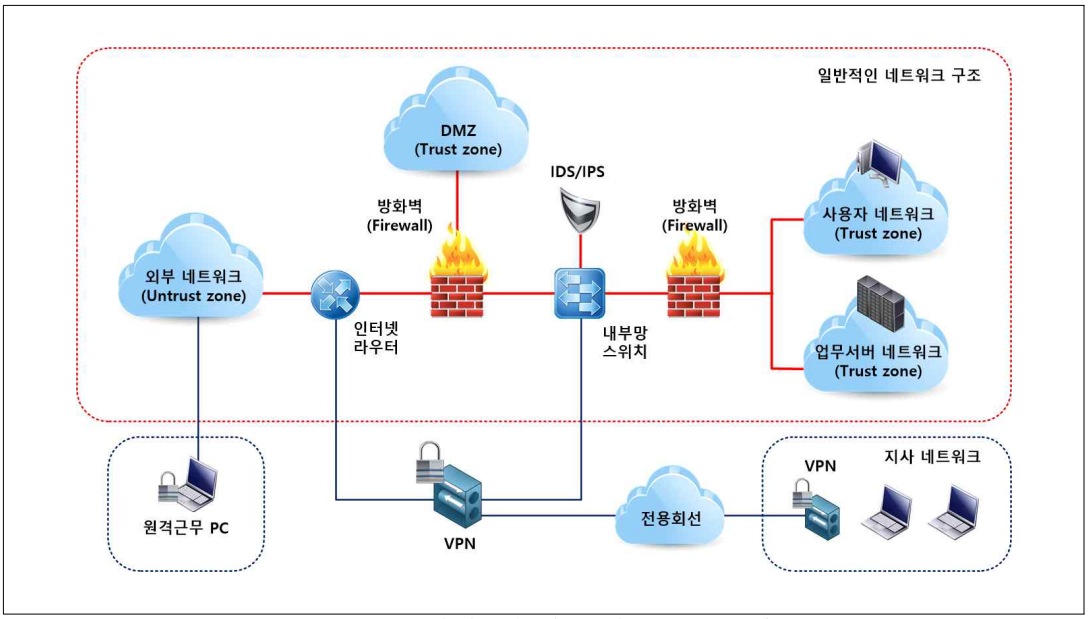
[그림 II-6]은 안전한 네트워크 구축절차를 나타낸다. 구축 절차는 사용목적에 따른 요구사항을 분석하고 기술인 분석을 진행하여 요구사항을 만족하는지 검토

한다. 네트워크 설계는 가이드라인에서 제시하는 네트워크 구성기준과 보안고려사항을 준수하는 안전한 네트워크 설계안 중 용도에 맞는 설계안을 채택하여 설계한다. 설계가 완료된 네트워크는 구축된 장비에 대해 보안성 검토를 요청하여 도입되는 보안장비의 CC인증 여부 및 암호화 제품의 경우 안정성 검증을 받는다. 이후 보안기능과 안전성 검증이 완료된 네트워크는 운용이 가능하며 주기적인 안전 진단을 수행함으로써 발생할 수 있는 취약점을 제거한다[21].

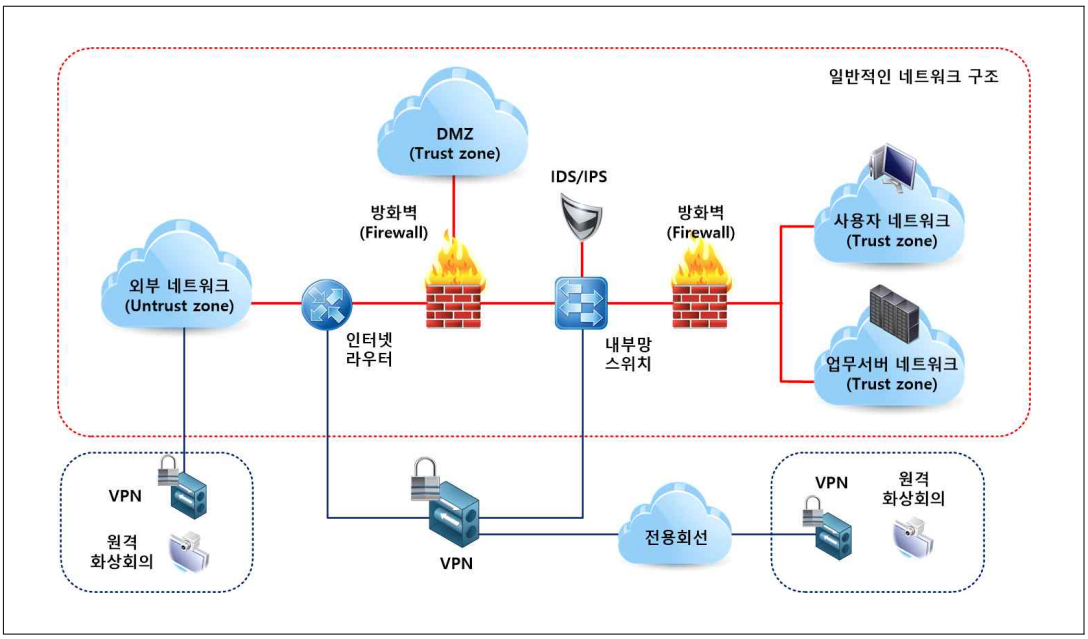
2) 유형별 네트워크 보안

네트워크 및 IT 인프라의 일반적인 형태는 방화벽을 기준으로 인터넷영역과 업무영역을 분리하는 방식으로 구성된다. 때문에 업무영역의 안전성은 방화벽이 책임지게 되며 방화벽의 접근통제 정책은 매우 중요한다.

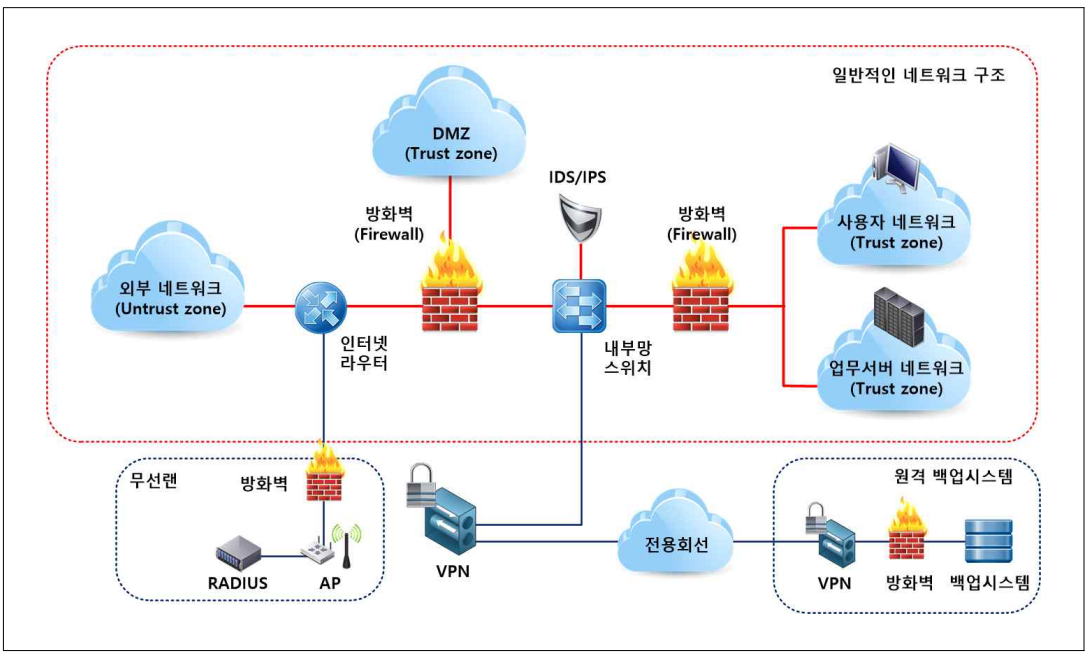
일반적인 네트워크 구성요소는 인가받은 사용자들의 업무 수행을 위한 사용자 대역, 정보 공유 및 저장 등 내무 업무를 위한 서버들의 내부망 서버대역, 외부로부터 접근을 허용한 웹 서버 및 메일서버 등 공개용 서버들이 위치한 DMZ 구간이 있다. 이를 위해 방화벽, 웹 방화벽 IPS/IDS, DDoS, VPN등 많은 보안장비를 구축한다.



[그림 II-7] 지사 및 유관기관 연동 네트워크



[그림 II-8] 원격 화상회의시스템 네트워크



[그림 II-9] 무선랜 및 원격 백업시스템 네트워크

[그림 II-7],[그림 II-],[그림 II-9]는 국가정보원에서 제시한 유형별 네트워크 구조이다. 지사 및 유관기관 연동, 원격 화상회의시스템 운용, 원격 백업시스템

운용, 무선랜 운용 등 여러 유형의 안전한 네트워크 구조를 제시하고 있다.

지사 및 유관기관의 접속의 경우 내부 사용자에 준하는 보안을 위해 VPN을 활용하여 접속을 암호화하고 방화벽 또는 내부망 스위치에 연결하여 내부에 접속을 허가한다. 원격화상회의시스템도 전용회선의 경우 VPN을 통해 내부와 직접연결하거나 공인망의 경우 인터넷 라우터를 경유하여 VPN을 통해 내부망과 연결된다. 원격백업은 공인망이나 전용회선으로 인터넷 라우터를 경유하여 VPN을 통해 장비를 연결하고 무선랜의 경우는 RADIUS를 통해 AP를 인가해준다[21].

제시되는 네트워크 구조는 모두 안전한 구조의 예시로 내부망과 연결되는 외부의 접근을 통제하고 암호화하여 내부로의 비인가 접속을 차단한다

3) 네트워크 구조적 취약점

국가정보원에서 제시하는 여러 유형의 네트워크 연결과 계층적이고 심층적인 보안 구성에도 외부 네트워크와 첫 관문인 인터넷 라우터는 보안장비로 보호하지 않는다. 많은 이유가 있지만 중첩되는 보안장비의 구축에 따른 비용의 부담과 외부 네트워크와 순수한 연결을 위해 따로 보안장비를 두지 않을 것이다.

비용적인 부담은 인프라와 기술의 발전에 따른 많은 보안 위협이 나타나고 있지만, 기업과 조직이 가진 정보자산의 가치보다 높은 투자를 보안장비에 할 수 없는 이유에 있다. 즉, 정보자산의 가치보다 높은 보안을 구축할 수 없으며 적절한 수준의 보안을 유지하여 위험을 수용 가능한 범위로 낮춰야 한다. 또한 외부 네트워크와 접점에 위치한 인터넷 라우터는 보안장비 연결로 인한 지연 및 통신 에러가 발생되면 인프라 전체에 영향을 줄 수 있으므로 순수한 통신을 위해 전송장비 또는 네트워크 장비와 직접 연결되어야 한다.

비용적인 부담을 배제하고 방화벽이 라우터 역할을 하는 경우나 라우터 상단에 방화벽을 놓는 방법이 있지만 실제 구축되는 경우는 드물다. 우선 방화벽이 라우터 역할을 하는 경우에는 라우터와의 성능차이가 있다. 라우터는 OSI 7계층에서 2~3계층(데이터 링크 계층, 네트워크 계층)의 프로토콜을 빠르게 처리하고 많은 라우팅 프로토콜과 스위칭의 성능이 높은 장비이다. 반면 방화벽은 라우터

에 비해 4계층(전송 계층)의 프로토콜 더 처리해야 하고 제공되는 라우팅 프로토콜 및 스위칭 용량도 라우터에 비하여 적다. 라우터 앞단에 방화벽을 놓는 경우도 규모에 비해 과한 투자로 인한 비용 증가의 부담이 있다. 방화벽이 외부 네트워크와 인터넷 라우터 사이에 구축되려면 최소 두 가지를 만족해야 한다. 첫째, Transparent 모드로 동작하며 외부 네트워크를 구성하는 전송장비 및 네트워크 장비와 인터넷 라우터 사이의 통신에 문제가 없음을 보장해야 한다. 둘째, 조직 및 기업의 사용 용량을 만족하는 규모여야 하며, 하단의 방화벽이 존재할 경우 라우터만을 위한 접근통제용 보안장비의 구매에 대한 부담을 감수해야 한다. 따라서 인터넷 라우터는 전송장비 및 네트워크 장비 등 외부 네트워크와 원활한 통신을 위해 인터넷 관문에 설치되어 조직과 기관의 가용성을 확보하며 자체적인 보안설정으로 외부에서의 불법적인 접근을 차단해야 한다.

네트워크 구조상 방화벽은 들어오는 트래픽을 모니터링 하여 허용된 트래픽만 통과시킨다. 허용된 트래픽은 다음 보안장비를 통해 더욱 세밀하게 검사하고 악의적인 공격을 탐지하며 이상이 없으면 통과된다. 그리고 사용자 또는 시스템에 전달되며 사용자의 PC 및 시스템 내에서도 백신 등 보안 소프트웨어를 통해 들어온 트래픽을 감시하고 이상이 있는 경우 격리시킨다. 하지만 방화벽 상단에 존재하는 인터넷 라우터는 많은 이유로 보안장비의 보호를 받기 못한다. 인터넷 라우터는 조직 전체의 인터넷 사용의 가용성을 보장해주는 중요한 장비이지만 많은 기업과 기관은 라우터를 통과한 트래픽에만 관심이 있으며, 방화벽을 비롯한 각종 보안장비와 사용자의 PC까지 보안 소프트웨어를 설치한다. 그래서 방화벽 하단의 보안은 보안 감사를 받는 조직이라면 심지어 완벽에 가깝게 중첩보안을 하고 있는 실정이다.

4) 인터넷 라우터 취약점 보강

인터넷 라우터는 OSI 7 계층 중 네트워크 계층에 속하는 장비로 경로 결정(Path Determination)과 스위칭(Switchin)을 하는 장비이다. 패킷이 원하는 목적지로 찾아갈 수 있도록 최상의 경로를 찾아낸 뒤 패킷을 보내준다. 이를 위해 많은 종류의 라우팅 프로토콜(RIP, OSPF, IGRP, EIGRP, BGP 등)과 고가용성을

위한 여러 이중화 프로토콜로 네트워크의 안전성을 보장한다. 이는 다른 보안장비가 지원하지 못하는 부분으로 라우터를 대체할 수 없는 기능중 하나다.

라우터는 크게 세 종류로 나눌 수 있는데, 인터넷 서비스 제공자(ISP)가 네트워크를 서로 연결하는 코어라우터, WAN 회선을 거쳐 기업과 기관의 본점과 지점을 서로 연결하거나 인터넷 서비스 제공자와 기업의 네트워크를 연결하는 센터 라우터, 기업의 본점과 지점, 영업소를 연결하는 엣지 라우터가 있다. 이중 인터넷 라우터는 센터 라우터로 네트워크 전체의 인터넷을 담당하는 라우터이다.

기업과 기관의 인터넷을 담당하는 대형 라우터의 경우 많은 기능을 가지고 있다. 관리를 위한 웹서비스, 원격 접속을 위한 SSH 서비스 외 Finger 서비스, TCP/UDP Small 서비스, CDP 서비스 등 많은 서비스들이 있으며 라우터의 기본 기능과는 거리가 먼 부가적인 기능이다. 하지만 인터넷 라우터는 IT 인프라의 위치 상 모든 접근에 허용되어 있는 위치에 있어 불필요한 서비스를 통해 라우터의 정보를 외부에 알려주는 취약점이 존재한다. 따라서 불필요한 서비스를 차단하고 필요한 서비스의 경우 적절한 보안 조치를 취해 취약점이 발생하지 않도록 조치해야 한다.

인터넷 라우터 자체적인 보안설정은 한국인터넷진흥원 “주요정보통신기반시설 기술적 취약점 분석 평가 상세 가이드”를 통해 최소한의 보안을 적용할 수 있다.

해당 가이드는 계정관리, 접근관리, 패치관리, 로그관리, 기능관리의 5개 분류로 되어 있으며, [표 II-4]와 같이 38개의 점검항목으로 나누어져 있다. 이를 통해 사용하지 않는 서비스를 차단하고 서비스에 대한 접근을 제한한다[24].

[표 Ⅱ-4] 네트워크 취약점 분석·평가 항목

분류	점검항목	항목 중요도	항목코드
계정관리	패스워드 설정	상	N-01
	패스워드복잡성 설정	상	N-02
	암호화된 패스워드 사용	상	N-03

분류	점검항목	항목 중요도	항목코드
	사용자·명령어별 권한 수준 설정	중	N-15
접근관리	VTY 접근(ACL) 설정	상	N-04
	Session Timeout 설정	상	N-05
	VTY 접속 시 안전한 프로토콜 사용	중	N-16
	불필요한 보조 입·출력 포트 사용 금지	중	N-17
	로그온 시 경고 메시지 설정	중	N-18
패치관리	최신 보안 패치 및 벤더 권고사항 적용	상	N-06
로그관리	원격 로그서버 사용	하	N-19
	로깅 버퍼 크기 설정	중	N-20
	정책에 따른 로깅 설정	중	N-21
	NTP 서버 연동	중	N-22
	timestamp 로그 설정	하	N-23
기능관리	SNMP 서비스 확인	상	N-07
	SNMP community string 복잡성 설정	상	N-08
	SNMP ACL 설정	상	N-09
	SNMP 커뮤니티 권한 설정	상	N-10
	TFTP 서비스 차단	상	N-11
	Spoofing 방지 필터링 적용	상	N-12
	DDoS 공격 방어 설정	상	N-13
	사용하지 않는 인터페이스의 Shutdown 설정	상	N-14
	TCP keepalive 서비스 설정	중	N-24
	Finger 서비스 차단	중	N-25
	웹 서비스 차단	중	N-26
	TCP/UDP Small 서비스 차단		N-27
	Boot 서비스 차단	중	N-28

분류	점검항목	항목 중요도	항목코드
	CDP 서비스 차단	중	N-29
	Directed-broadcast 차단	중	N-30
	Source 라우팅 차단	중	N-31
	Proxy ARP 차단	중	N-32
	ICMP unreachable, Redirect 차단	중	N-33
	identd 서비스 차단	중	N-34
	Domain lookup 차단	중	N-35
	pad 차단	중	N-36
	mask-rely 차단	중	N-37
	스위치, 허브 보안 강화	하	N-38

주요정보통신기반시설 기술적 취약점 분석 평가 상세 가이드의 중요 설정은 다음과 같다.

(1) 계정 관리

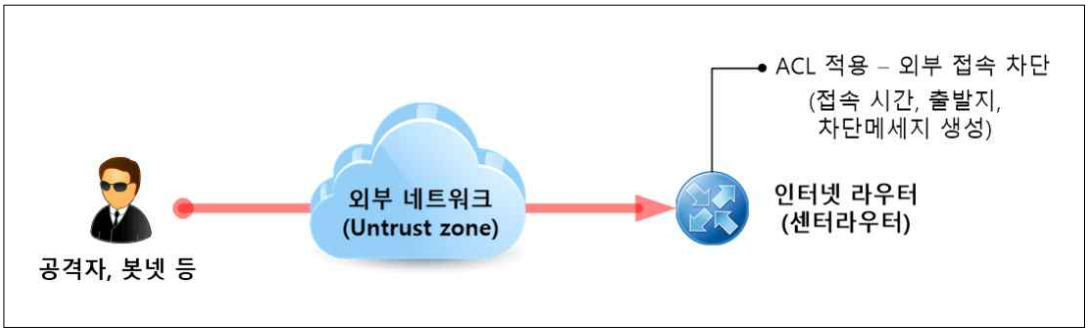
장비 출고 시 설정된 기본 패스워드는 벤더사 별로 인터넷을 통해 쉽게 검색이 가능하여 운용중인 장비에 접근하여 네트워크에 영향을 미칠 수 있다. 또한 패스워드의 복잡성 정책이 적용되어 무작위 대입 공격 및 사전 대입 공격 등에 패스워드가 탈취되지 않도록 한다.

(2) 접근 관리

지정된 IP만 라우터에 접근하도록 설정하여 비인가자의 접근 시도를 차단한다. 라우터의 ACL(Access-List)를 통해 인가자의 IP를 정할 수 있으며 해당 인가자만 VTY(Virtual Type Terminal)을 통해 Telnet 및 SSH 등 원격 접속 프로토콜을 이용할 수 있다.

(3) 로그 관리

라우터는 로그를 저장하기 위한 버퍼 메모리를 가지고 있다. 하지만 정해진 용량의 버퍼를 초과하는 로그는 저장할 수 없으며, 침해사고의 발생의 근거인 로그를 잃게 된다. 따라서 별도의 원격 로그 서버에 보관하도록 설정하여 장비 상태, 서비스 정상 여부 파악 및 보안사고 발생 시 원인 파악 등 각종 침해 사실에 대한 확인이 가능하다. 또한 NTP(Network Time Protocol) 서버를 연동 설정하여 발생된 로그의 신뢰성을 제공한다.



[그림 Ⅱ-10] 라우터 접근통제

이중 가장 중요한 부분은 계정 관리와 접근 관리일 것이다. 불필요한 서비스를 기능관리 부분에서 차단하고 최소한의 서비스는 허용하여 장비를 관리해야 한다. 장비의 상태 확인 및 세팅을 위해 필요한 원격접속은 가장 필요한 부분이지만 동시에 취약한 부분 중 하나일 것이다. 이를 방지하기 위한 최소한의 보안은 계정관리를 통한 복잡한 암호 사용과 [그림II-7]과 같이 접근 관리를 통한 원격 접속에 ACL(Access-List)을 적용하는 것이다. 하단에 방화벽이 존재할 경우 NAT IP를 통해 인터넷 관문라우터로 접속하며 해당 IP만 허용함으로써 모든 접근을 막아 장비의 권한이 넘어가지 않도록 조치한다.

네트워크 장비 취약점 분석·평가 항목을 통해 최소한의 보안조치를 취하지만 해당 장비로의 악의적인 접근이 사라지는 것은 아니며, 많은 접속이 허용된 원격 접속을 통해 지속적으로 발생할 것이다. 이런 접속은 장비의 ACL에 의해 차단되

며 그 이력을 로그로 남기게 된다. 다른 네트워크 장비와 다르게 인터넷 라우터에만 지속적으로 원격접속에 대한 차단 로그가 발생하는 것이다.

공격자 혹은 자동화된 봇넷 IP들의 지속적이고 대용량의 접속은 인터넷 라우터가 구조상 취약하다는 의미이며 우리는 해당 공격의 차단 로그로부터 공격의 진원지 정보를 확인할 수 있다. 생성되는 로그는 악의적인 공격 근원지를 확인할 수 있는 좋은 데이터이며 이를 통해 접근통제 정책을 수립할 수 있다.

3. 로그데이터 활용 방안

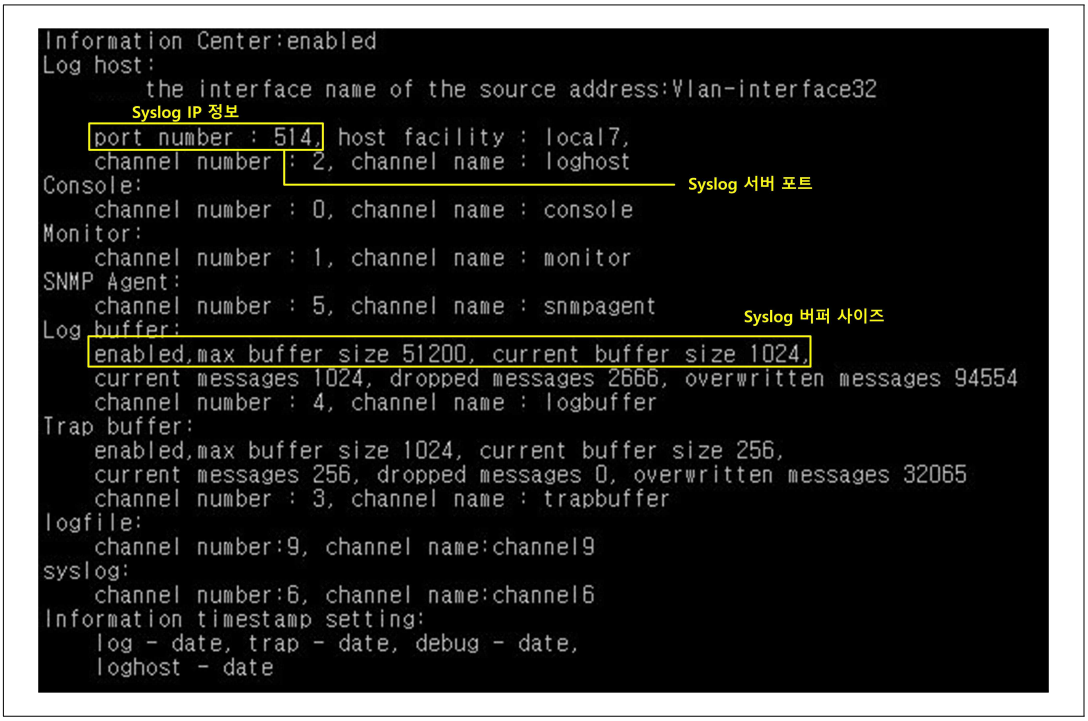
수많은 장비는 각각의 고유 로그를 만든다. 이러한 로그 데이터는 장비의 상태를 나타내는 지표로서 어떻게 사용하느냐에 따라 가치가 달라진다. 로그 자체로의 의미보다 그로 인해 발생할 수 있는 예측을 할 수 있다면 그 가치는 높아질 것이다. IT 인프라를 관리하는 관리자는 매일 쏟아져 나오는 네트워크 장비, 보안장비, 각종 서버들의 로그를 분석하고 중요한 로그를 분류하여 의미 있는 정보를 만들기 위해 노력한다[5].

하지만 많은 장비가 쏟아내는 로그를 모두 검사하는 것은 너무 힘든 일이며 사고가 발생한 후 검증을 위해 확인하는 정도가 다일 것이다. 또한 로그를 수집하고 한곳에 모으는 것이 필수일 것이다.

1) 로그 수집

로그분석의 첫 번째는 각종 장비의 로그를 모으는 일이다. 장비에 접속해서 직접 로그를 확인할 수 있지만 장비마다 로그를 저장하기 위해 정해진 용량의 공간을 할당하는데, 이 용량을 넘어서는 로그는 덮어쓰기 되어 이전의 로그를 확인할 수 없다. 즉 시스템에 쌓이는 로그들을 별도로 저장할 공간을 제공하는 시스템을 구축하여야 한다. 이 시스템은 완벽한 보안이 요구되는데 모든 RPC 데몬들

이나 다른 기타 서비스들도 암호화되지 않고서는 접근을 허락하지 않고 데이터들은 오직 UDP 514 포트를 통해서만 전송을 허락한다. 즉 별도의 로그 서버를 통해 IT 인프라 장비별 로그를 저장하고 사후 검증 및 사전 예방을 위해 사용해야 한다.



[그림 Ⅱ-8] 네트워크 장비 로그 정보

[그림 II-8]은 네트워크 장비의 로그 상태를 보여주는 그림으로 장비가 가지고 있는 로그를 저장 가능한 버퍼(Buffer)와 생성된 로그를 로그서버로 보내기 위한 설정이 있다. 이처럼 로그수집 대상이 되는 IT 인프라 장비는 자체에 로그를 저장할 수 있는 버퍼(Buffer)를 따로 배정한다. 이 버퍼(Buffer) 용량은 장비마다 다르며 해당 용량을 초과하는 로그는 지난 로그부터 삭제되어 기존의 상태를 확인할 수 없다. 따라서 장비별로 로그서버를 지정할 수 있는 설정이 있으며 해당 서버로 자신의 로그를 전송한다. 또한 중요한 부분은 로그가 생성되는 시간인데, 현재까지 사용된 가장 오랜 인터넷 프로토콜 중 하나인 네트워크 타임 프로토콜

(NTP, Network Time Protocol)을 통해 장비의 시간을 표준시간과 동기화 시켜 생성되는 로그가 표준시간에 생성되는 것을 보장한다.

로그 분석의 첫 단계인 로그 수집은 로그를 단순히 한곳에 모으는 일이뿐이다. 로그를 통해 문제점을 찾고 보안 위협으로부터 정보 자산을 보호하기 위해서는 분석과 예방 조치가 필요하다. 로그가 단지 사후 검증을 위한 수단으로 쓰인다면 단순한 자료에 그칠 뿐이다. 로그를 통해 보안 예방의 기능을 더한다면 보안의 수준을 한 단계 올릴 수 있다.

또한 많은 양의 이벤트가 발생한다면 로그서버의 적절한 용량 산정이 필요하며, 압축화를 통해 정해진 기간을 보관하는 것이 안전한 로그의 수집일 것이다.

2) 로그 분석 정책

모아진 로그의 효율적인 사용은 가장 중요한 일이며 이를 위해서는 로그 분석에 관한 정책이 필요하다. IT 인프라 관리자가 관리하는 장비는 매우 많으며 수집된 로그를 제때 분석하기란 쉽지가 않다. 이로 인해 사이버 위협의 초기 단계인 정찰 단계의 이상 징후를 포착하지 못하고 지나치게 된다. 로그는 수집이 목적이 아니라 주기적 분석이 필요하며 이를 통해 발행하는 이슈나 예방 정책을 세울 수 있다.

주기적인 로그 분석을 통해 사고에 대한 원인 파악이 신속해 지며 반복되는 로그를 통해 의미 있는 정보를 추출할 수 있다. 이를 통해 로그의 이해도가 높아지며 중요한 로그를 빠르게 확인 가능할 것이다. 이처럼 로그의 모니터링과 분석은 보안 관리 측면에서 아주 중요하며 정책을 통해 효율적인 관리가 필요하다.

첫째, 주기적인 로그 데이터 분석

IT 인프라의 장비가 장애가 났거나, 보안상 문제가 발생했을 때 가장 먼저 확인하는 부분은 로그일 것이다. 하지만 장애나 보안상 특별한 문제가 아니라면 사이버 공격에 대한 전조현상이 있을 것이며 로그를 통해 표현될 것이다. 주기적 로그 관리는 발생할 수 있는 장애나 보안 문제를 억제할 수 있는 수단이며 예방인 것이다. 장비별 특징적인 로그를 주기적 분석을 통해 신속하고 빠른 장애 탐

지와 대응이 가능할 것이다.

둘째, 로그의 특징(의미) 분석

장비별 다양한 로그가 존재한다. 방화벽은 정책에 따른 허용과 거부에 대한 로그가 존재하며 그 의미는 확실하다. 침입방지시스템도 알려진 공격에 대한 차단 로그가 공격명과 함께 제공된다. 이처럼 의미가 분명한 로그도 있지만 일상적이고 특징적이지 않는 로그도 존재하며, 일상적인 로그도 어떤 반복적 의미가 있는지 확인이 필요하다. 또한 장비에서 특정 로그의 비중이 높거나 지속적이고 반복적인 로그가 있는지 확인하여야 하며, 이를 통해 그 로그가 지니는 의미를 분석하여 장애 예방을 위한 자료로 가공해야 한다.

마지막으로 로그의 단순화

다양한 로그의 표현 형식으로부터 의미 있는 구문을 분리하여 가공한다면 분석이 더욱 쉬워지고 관리 또한 용이해진다. 아무리 긴 문장의 메시지도 의미하는 내용은 간단할 수 있다. 긴 문장의 의미를 단순화 시키거나 필요 없는 중복 문장을 제거하고 필요한 내용만 남기면 로그가 단순해지고 분석의 시간은 줄어들 것이다.

4. 로그 가공 및 시각화

수집된 대용량의 로그는 특정 패턴을 가지고 있으며 분석을 통해 원하는 데이터로 단편화 작업을 할 수 있다. 장비별 로그의 형태는 다르지만 일반적으로 로그의 발생시간, 발생 지점(경로나 인터페이스 등), 장비명(hostname), 메시지(발생 원인) 등 그 의미를 구분할 수 있는 구문으로 구성되어 있다. 메시지 부분에서 반복되는 문구를 단순화 한다면 긴 문장의 구문을 단순한 의미로 가공할 수 있다.

가공된 로그는 분석이 필요하며 많은 로그 분석도구가 있지만 본 연구에서는 오픈소스를 활용한 ELK Stack을 이용하여 로그분석을 진행하고자 한다. ELK

Stack은 오픈소스를 이용함으로써 비용의 부담이 없고 많은 기술공유 자료와 라이선스의 제약 없이 자유롭게 이용할 수 있다는 장점이 있다. ELK Stack은 ElasticSearch, Logstash, Kibana의 오픈소스 프로젝트 세 개의 머리글자를 딴 이름이다. 단순 반복적인 작업을 최소화할 수 있고 최적화된 검색 엔진을 사용하여 시각화가 용이하다. 이를 통해 대량의 로그를 분석하여 내재된 위협을 찾고 사전에 예방할 수 있다.

1) ElasticSearch

ElasticSearch는 검색 및 분석을 진행하는 엔진으로 정형, 비정형, 위치정보, 메트릭 등 여러 형태의 데이터를 원하는 방법으로 검색을 수행하고 결합할 수 있다. JSON 문서 기반의 안정성과 관리의 편리성을 제공하며 RESTful API를 통한 다양한 환경에서 사용이 가능하다. 방대한 데이터를 신속하고 실시간으로 저장, 검색 분석할 수 있도록 오픈소스인 Apache Lucene 기반으로 개발되었다.

ElasticSearch의 구성요소는 물리적인 서버인 노드, 다수 노드의 집합인 클러스터, 데이터를 저장할 때 구분 짓는 하나의 조각인 샤드, 비슷한 특성을 가진 도큐먼트(Document)의 집합인 인덱스, 인덱싱될 수 있는 정보의 기본 단위인 도큐먼트로 구성된다. ElasticSearch를 구성하는 요소들은 [표 Ⅱ-5]와 같은 기능을 갖는다.

[표 Ⅱ-5] ElasticSearch 구성요소

구성요소	기 능
클러스터 (Cluster)	모든 데이터를 함께 가지고 있는 한 개 또는 그 이상의 노드의 집합 연합된 인덱싱과 모든 노드를 검색할 수 있는 기능을 제공 유일한 이름(unique name)으로 판별(identified)
노드 (Node)	클러스터의 일부로 단일서버 데이터를 보관하고 클러스터 인덱싱과 검색 능력에 관여

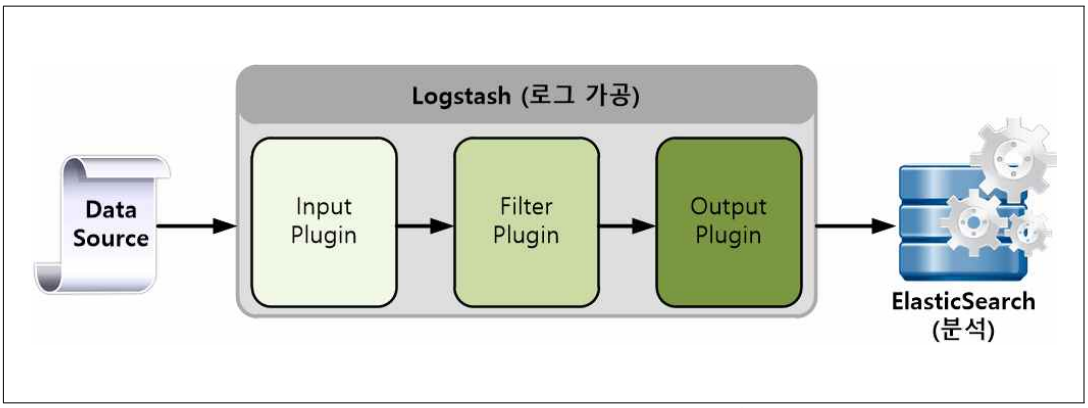
구성요소	기 능
인덱스 (Index)	비슷한 특성을 가진 도큐먼트(Document)의 집합
도큐먼트 (Document)	인덱싱될 수 있는 정보의 기본 단위 유비쿼터스 인터넷 데이터 교환 포맷인 JSON으로 표현
샤드 (Shards)	인덱스를 샤드라 불리는 여러개의 조각으로 다시 나눌 수 있는 기능을 제공 어떤 노드에서도 관리될 수 있는 독립된 "인덱스" 명령을 분배, 병렬적 처리로 성능 향상
레플리카 (Replicas)	인덱스의 샤드에 대한 한개 이상의 복사본 노드가 깨졌을 때를 대비하여 높은 가용성을 제공

ElasticSearch은 실시간 분석, 분산 시스템, 높은 가용성, 멀티 테넌시, 전문검색, JSON 문서, RESTFUL API 사용 등의 특징을 가지고 있다. 분석 기능은 저장된 데이터를 검색하기 위해 별도의 재시작이나 갱신이 필요하지 않고 색인 작업이 완료됨과 동시에 즉시 검색이 가능하다. 또한 데이터를 색인하고 검색을 수행하는 단위 프로세스인 여러 개의 노드로 구성되어 분산 처리가 가능하다. 데이터를 각 노드에 분산 저장하고 복사본을 유지하여 각종 충돌로부터 노드 데이터의 유실을 방지하는 높은 가용성을 가진다. 데이터는 여러 개로 분리된 인덱스에 저장되며 서로 다른 인덱스의 데이터를 하나의 질의로 검색하고 여러 검색 결과를 하나의 출력으로 도출가능하다. HTTP 프로토콜 기반의 RESTFUL API를 이용하여 JSON(Javascript object notation) 문서의 다양한 제어가 가능하다.

이처럼 막강한 기능의 ElasticSearch는 삼성 SDS, 포스코, 네이버, 11번가 등 많은 국내 기업에서 뛰어난 검색 기능과 분석을 위해 사용되고 있으며, 기업 뿐만 아니라 많은 개인 사용자들도 기존 데이터 분석의 어려움을 ElasticSearch를 통해 쉽게 구현하고 있다[2][15].

2) Logstash

Logstash는 다양한 형태의 로그 데이터를 읽어 분석 가능한 형태로 정제하여 ElasticSearch로 전달한다. 실시간 파이프라인 기능을 가진 오픈소스 데이터 수집엔진으로 INPUTS, FILTERS, OUTPUTS 구성된다. INPUT을 통해 DB, kafka, file 등 다양한 데이터를 가져오고, FILTER를 통해 가져온 데이터들의 다양한 전처리가 가능하다. 그리고 OUTPUT을 통해 데이터들을 어떤 곳으로 전달해줄지 설정한다. 대부분 ElasticSearch로 보내서 인덱싱 처리한다. [그림 II-9] 로그 데이터가 Logstash의 전처리 과정을 거쳐 분석도구인 ElasticSearch로 전달되는 과정을 보여준다. 여기서 처리된 분석 데이터는 Kibana를 통해 시각화된다.



[그림 Ⅱ-12] ELK Stack 로그 처리 과정

Logstash는 실시간 파이프라인 기능을 가진 오픈소스 데이터 수집 엔진으로 입력, 필터, 출력이라는 세가지 타입으로 구성된다. 200개 이상의 플러그인을 통해 서로 다른 소스의 데이터를 통합하고 다양한 목적지로 데이터를 정규화하여 전송할 수 있다. Logstash의 필터는 로그를 다양하고 유용한 형태로 가공할 수 있으며 그중 Grok 플러그인은 로그 데이터에서 원하는 정보만 추출할 수 있게 도와준다. 또한 Grok 플러그인은 많이 사용되고 있는 로그(Apache Web log, Cisco Network Log, Syslog 등)들의 미리 정의된 정규 표현식을 제공하고 정의되지 않은 패턴들도 로그가 가지는 특성을 구분하여 정규식 표현을 할 수 있다.

이외에 Geoip 플러그인을 통해 IP 정보를 지리정보와 매핑할 수 있다. Maxmind는 전세계 IP 데이터베이스 서비스를 제공하는 업체이며 IP의 좌표(위도, 경도) 정보를 제공한다. Geoip는 이런 Maxmind 데이터베이스를 이용하여 추출하고 싶은 IP주소 정보를 지도상에서 위경도 정보, 국가코드, 지역코드 등 다양한 지리정보를 표현해 준다. 이처럼 Logstash는 확장할 수 있으며, 실시간 데이터 파이프라인을 구축하는데 유용한 데이터 흐름 엔진이다. Logstash를 활용하면 대용량 로그 데이터와 각종 데이터 형식을 통합하고 정규화하는 프로세스를 구축할 수 있다[2][15].

2) Kibana

Kibana는 데이터를 시각적으로 탐색하고 실시간 분석이 가능하다. 또한 웹서버를 내장하고 있어 별도의 소프트웨어를 설치하지 않고도 웹을 통한 시각화가 가능하다. 검색엔진을 통해 수집된 많은 데이터들은 Kibana가 제공하는 많은 그래프와 차트를 통해 보여주며, 이를 통해 자료간 연관성을 파악할 수 있다.

누적된 많은 양의 로그는 ElasticSearch를 통해 분석되고 Kibana의 시각화를 이용하여 텍스트 상태에서 확인할 수 없었던 특이점을 찾아낼 수 있다. 또한 과거 로그와 새로 생성되는 실시간 로그의 비교가 가능하여 앞으로 발생할 수 있는 미래를 예측할 수 있는 데이터를 제공한다.

Kibana는 ElasticSearch에서 분석된 로그 데이트를 이용하여 각종 그래프와 차트(히스토그램, 라인 그래프, 파이 차트, 선버스트 등), 지도를 활용한 위치정보, 시계열 데이터 분석, 그래프 관계 탐색 등 다양한 시각화와 탐색 기능을 제공한다. 또한 시각화로 제공되는 각종 그래프는 필터를 통해 더욱 세분화하여 분석이 가능하고 분석된 자료는 CSV 파일로 저장 가능하다.

[그림 II-10]은 Kibana에서 제공되는 샘플 데이터를 활용하여 대시보드를 생성한 화면이다. 각각의 데이터를 원하는 차트, 그래프, 지도 등 여러 시각화를 통해 확인하여 발생하는 상황을 빠르게 대처할 수 있다[2][15].

[그림 Ⅱ-10] Kibana 샘플 대시보드

4) IT 인프라 로그 데이터

IT 인프라를 구성하는 가장 기본적인 보안장비는 방화벽이다. 방화벽의 기본적인 기능은 신뢰 수준이 다른 네트워크 구간을 나눠 해로운 트래픽을 차단하는 역할을 한다. 여기서 신뢰 수준이 낮은 구간(Untrust Zone)은 외부망이 되며 높은 신뢰도를 갖는 구간(Trust zone)은 내부망이라고 한다. 이 밖에도 웹서비스 등 외부에 서비스를 제공하는 서버들을 위한 DMZ 구간을 구성할 수 있으며 외부망으로부터 내부망으로 침입을 차단하고 내부망에서는 외부망을 통해 인터넷을 자유롭게 이용할 수 있도록 해준다.

(1) 보안장비 로그

방화벽은 신뢰도가 서로 다른 구간가의 접근 통제가 이루어진다. 관리자는 DMZ에 있는 많은 서비스를 모두 허용하는 것이 아니라 최소한 정해진 서비스만 허용함으로써 생길지도 모르는 위협에 대비한다. 그리고 NAT(Network Address Translation)를 통해 내부망을 사설 아이피로 숨겨서 외부에서 내부로의 침입을 차단한다. 또한 신뢰도가 낮은 구간(Untrust Zone)에서 DMZ를 접근할 때 정해진 정책만 허용하고 나머지는 모두 차단한다. 따라서 방화벽 로그는 정책에 의한

차단과 허용 로그만 발생한다.

IPS나 IDS의 경우는 네트워크 대역별 접점 부분이나 해당 네트워크에 속한 모든 장비를 관찰하고 정책을 적용하기 용이한 위치에 설치된다. 이 위치에서 네트워크를 흐르는 트래픽을 모니터링하고 사이버공격의 징후가 관찰되면 이를 탐지하여 차단하거나 알람을 띄운다. 이런 공격의 징후를 탐지하는 기법은 정해진 시그네이쳐 기반(Misuse/Knowledge 기반) 탐지기술과 비정상 탐지(Anomaly 기반)기술로 나뉜다. 시그네이쳐 기반 탐지 기술은 백신과 유사한 방식으로, 잘 알려진 정형화된 공격 패턴(시그네이쳐)을 데이터베이스(DB)화 하여 공격 탐지에 활용하는 방식이다. 이 방식은 오탐율이 낮고 효율적이지만 공격 패턴에 없는 공격은 탐지하지 못한다는 단점이 있다. 즉 대용량의 트래픽 중 알려지지 않은 패턴이 있다면 미탐으로 인한 피해가 발생할 수 있다.

비정상 탐지 기술(Anomaly Detection)은 호스트의 활동패턴을 살펴고 정상적이며 평균적인 상태를 기준하여 상대적으로 급격한 변화를 일으키거나 정상패턴을 벗어나면 공격이라고 탐지하는 기법이다. 평균적인 상태에 대비한 이상상태를 추정하여 탐지하는 것으로 실제 공격일 수도 있고 아닐 수도 있어 오탐율이 많다. 하지만 정해진 패턴을 필요하는 것이 아니라 알려지지 않은 공격에 대비할 수 있다는 장점은 존재한다. 따라서 정해진 패턴이 존재하는 경우나 이상현상이 발생하는 현상에 대한 로그만 생성한다.

(2) 인터넷 라우터 로그 데이터

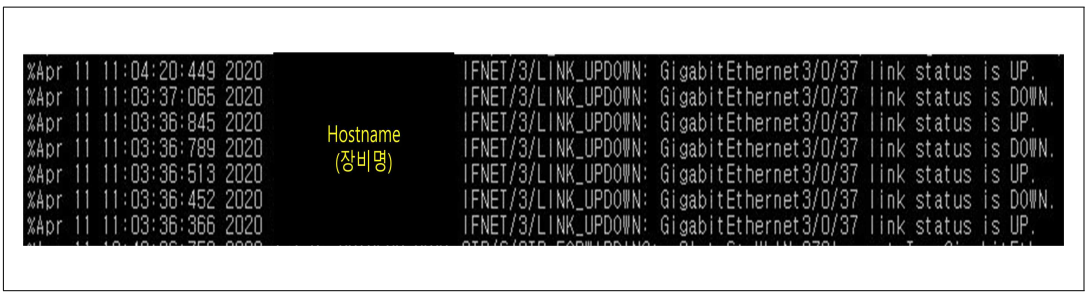
기업 및 기관의 네트워크가 폐쇄망 환경이 아니라면 ISP(Internet Service Provider)로부터 인터넷 회선을 공급받고 인터넷을 사용할 것이다. 이 라우터는 IT 인프라에서 가장 최상단에 존재하며 기업 및 기관의 정보서비스의 가용성을 보장하는 아주 중요한 역할을 한다. 내부망의 모든 정보자산은 방화벽의 NAT(Network Address Translation) 기능으로 보호받으며, DMZ 구간의 서버들은 방화벽의 접근통제를 통해 필요한 서비스만 허용한다. 하지만 라우터의 경우는 다른 네트워크와 접점부분에 존재하며 모든 접근에 대해 응답하고, 경로를 지

정해주며 전달된 패킷을 안전하게 이동하는 역할을 수행한다. 모든 요청에 대해 응답을 하거나 그에 대한 반응을 하는 아주 능동적인 장비인 셈이다.

사이버공격의 초기 단계인 정찰 단계에서는 IT 인프라의 서비스를 확인하기위해 스캐닝(Scanning)을 진행한다. TCP 기반의 프로토콜을 이용하여 기본적인 질의(Request)를 보내서 그에 대한 응답(Response)으로 해당 서비스의 생존 여부를 확인 할 수 있다. 방화벽으로 보호하고 있는 정보서비스는 이런 스캐닝을 차단하겠지만 최상단 인터넷 관문라우터는 이런 스캐닝에도 자신의 존재를 친절히 알려준다. 이를 통해 공격자는 라우터에 공격을 가하게되고 이는 로그로 남게 된다.

최상단 인터넷 관문라우터를 통해 수집되는 로그는 조직의 중요도나 규모에 따라 차이가 난다. 유동 IP를 사용하는 일반 가정집과 고정 IP를 사용하고 사용자를 위한 웹서비스를 제공하는 조직의 로그가 같을 수 없듯이 조직의 규모에 따라 쌓이는 로그 또한 다르다. 로그 확인 방법은 원격 접속 프로토콜을 통해 장비에 접속하는 방법이 보편적이며 이때 방화벽으로 NAT된 IP가 인터넷 라우터에 접속되는 IP가 된다.

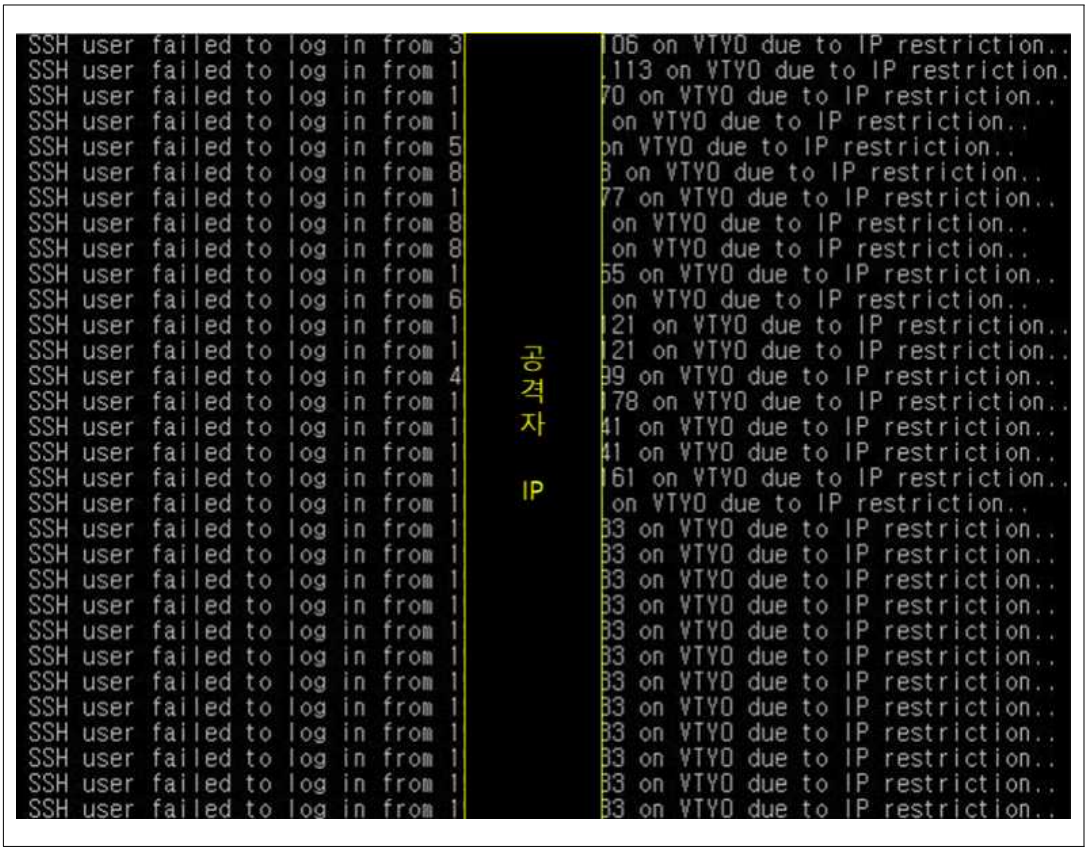
내부망에 존재하는 백본 스위치의 경우는 연결되어 있는 사용자의 PC를 연결하거나 중요한 내부 서버를 연결한다. 이때 발생되는 로그는 [그림 II-11]과 같이 사용자 PC의 ON/OFF로 인한 인터페이스 UP/Down 로그이거나, 직접적인 사용자 PC의 연결이 아닌 워크그룹 스위치를 연결할 경우 스위치와의 연결 로그나 이중화 프로토콜에 대한 로그가 대부분일 것이다.



[그림 Ⅱ-14] 백본 스위치 로그 샘플

하지만 인터넷 라우터는 내부망과 달리 작업 및 설정으로 인한 접속 빈도수가 적고 원격을 통한 접속도 방화벽의 NAT IP를 통해 이루어지기 때문에 사용자의

접속은 확인이 가능하다. 적은 빈도수의 작업과 관리자 접속을 제외한 대부분의 로그는 네트워크 구조적 취약점으로 발생한 외부 비인가자의 원격접속 차단로그이며 그 접속의 패턴은 [그림 Ⅱ-12]와 같이 SSH 무차별 대입 공격이 대부분이다.



[그림 Ⅱ-15] 인터넷 라우터 SSH 무차별 대입 공격 로그

4. SSH 무차별 대입 공격

IT 인프라의 시스템을 관리하기 위해서 직접 연결을 통해 장비에 접속하거나 원격 접속을 통해 장비에 접속하여 설정을 변경하거나 상태를 확인 하는 등 여러 작업을 진행한다. 직접 연결은 물리적인 위치에 제약을 받아 시스템이 설치되어있는 곳에서 직접 연결하는 경우가 대부분이며 이런 불편함으로 인해 원격

접속을 통해 시스템을 관리한다. 원격 접속 자체는 정상적인 접속이지만 원격지에서 시스템의 권한을 획득하여 명령을 내릴 수 있어 특별한 관리를 요한다. 또한 암호화를 이용한 원격 접속으로 내리는 명령이나 계정정보가 노출되지 않도록 한다[8].

1) SSH(Secure Shell)

대표적인 원격 접속 프로토콜인 시큐어 셸(Secure Shell, SSH)은 통신을 통해 다른 컴퓨터에 로그인하거나 원격으로 시스템에 명령을 내려서 실행하고 파일을 복사하는 등 여러 작업을 수행하는 프로토콜이다. 기존 평문 통신을 이용한 텔넷 등 여러 응용 프로그램 및 프로토콜은 암호화가 이뤄지지 않아 계정 정보 및 실행하는 명령이 탈취될 위험이 높았지만, SSH는 강력한 인증 방법 및 암호화를 통한 통신으로 원격의 시스템에 안전한 관리와 통신의 기밀성을 보장한다.

SSH는 TCP 22번 포트를 사용하고 기본적으로 CLI(Command Line Interface)에서 작업을 한다. 또한 공개키 기반의 암호화를 사용하여 통신의 내용이 탈취되더라도 암호화된 내용으로 그 의미를 알 수 없다. 네트워크 장비의 경우 장비 박스를 개봉한 초기에만 콘솔을 통한 직접 연결로 기본적인 설정을 하고 랙에 장착 후에는 대부분 SSH를 이용하여 장비를 설정한다. 따라서 해당 접속은 빈번한 사용이 필요한 서비스이며 정상적인 접근으로 보안장비에서 공격으로 탐지되진 않는다. 하지만 SSH 서비스를 허용하는 장비의 경우 모든 출발지에서의 접속을 허용하는 것은 아니며 관리자로 특정 지을 수 있는 IP에 대해서만 SSH 원격 접속을 허용해야 하며 방화벽에서도 출발지와 목적지를 명확하게 특정지어 정책을 수립해야 한다. 또한 SSH 프로토콜은 국제 인터넷 표준화 기구(IETF)에 의해 신뢰할 수 있는 데이터 스트림에 안전한 파일 전송 기능을 목적으로 지속적으로 버전업이 되고 있다[16].

2) 무차별 공격 (Brute-Force Attack)

무차별 공격은 시스템의 계정 정보를 탈취하기 위해 가능한 모든 값을 대입하

여 접속을 시도하는 공격이다. 공격 방법은 조합 가능한 모든 문자열을 순차적으로 대입하는 무작위 순차 대입 방식과 미리 정의된 문자열을 목록화하여 대입하는 사전 대입 방식으로 나뉜다.

암호를 사용하는 시스템에 모든 가능한 수의 암호를 대입한다면 그 암호는 분명히 풀 수 있다. 효율적인 방법은 아니지만 이론상 충분한 시간만 보장된다면 100% 성공 가능한 공격인 것이다. 따라서 암호를 통한 원격 접속은 무차별 공격에 안전하지 못하며, 충분한 시간이 주어진다면 암호화된 정보를 탈취할 수 있다. 하지만 대부분의 경우 무한한 시간이 제공되지 않으며 그 시도를 위한 많은 컴퓨팅 파워가 소모될 것이다.

무차별 공격은 최근까지 막대한 피해를 입히고 있다. 2018년 6월 23일 우리은행 해킹 사건은 무차별 공격을 받아 고객정보 약 5만 6,000건이 유출된 사건으로 사건 당시 고객들의 민원을 받아 접속 이력을 확인한 후 무차별 공격을 파악하여 해당 IP를 차단하고 사이버수사대에 신고하는 등 즉각적인 조치를 했다. 또한 2017년 12월 18일 보안업체인 워드펜스가 서비스하는 워드프레스 사이트를 대상으로 총 10,000개의 IP 개수와 시간당 1,400만 건에 달하는 큰 규모의 공격으로 19만 개에 해당하는 워드프레스 사이트가 공격 받았다.

이런 무차별 공격을 가하는 공격자가 암호 하나하나를 직접 입력하는 공격 시도를 하기는 힘든 일이며 자동화된 시스템을 통해 공격을 진행할 것이다. 따라서 무차별 공격의 근원지는 자동화된 봇넷의 가능성이 높으며 무차별 공격뿐만 아니라 여러 공격들을 가하는 악성 공격의 근원지일 가능성이 높다.

무차별 대입 공격에 대한 방어 방법은 일정 횟수 이상 암호입력이 틀리면 차단하는 방식이 많이 사용된다. 서버의 경우 계정 장금 정책을 통해 로그인 실패에 대한 임계치를 정하고 임계치를 넘으면 계정을 잠궈서 접속을 제한한다. 또한 암호 정책을 통해 암호의 복잡성, 최소 암호 길이, 최대 암호 사용 기간 등을 설정하여 단순한 암호로 인한 계정 탈취를 막는다. 하지만 네트워크 장비의 경우 특정 임계치 이상의 로그인 실패를 막진 않지만 ACL을 통해 비인가 접속을 막는다. 따라서 인터넷 라우터는 외부의 많은 비인가 접속에 대한 로그를 가지고 있으며 이는 접근통제 정책에 중요한 자료가 된다.

3) 선형회귀 분석기법을 활용한 블랙리스트 탐지

로그를 통해 악의적인 의도를 가진 IP 주소만 구분한다면 먼저 악의적인 IP 주소가 가진 특징들을 파악하여야 한다. 공격 유형에 따른 위험도 파악, 지속적인 공격을 하는 IP 분류(공격 횟수, 공격 기간), IP에 따른 국가별 분류 등은 블랙리스트 탐지를 위한 로그의 여러 특징들이다. 이런 특징을 바탕으로 블랙리스트를 탐지하는 방법에는 선형회귀(Linear Regression) 분석 기법이 있다.

반복적인 선형회귀분석을 이용한 대용량 보안로그의 블랙리스트 IP 분류 연구(전두용, 2018)를 통해 기존의 전문 보안관제 요원의 주관적 분석을 통한 블랙리스팅 기법을 자동으로 IP주소가 블랙리스팅 되어야 하는지 여부를 판단하는 시스템을 설계 및 구현하였다. 이를 위해 IPS, 웹방화벽, 방화벽, 웹서버 로그에서 30개의 특징을 선정하여 목록화 한다.

[표 Ⅱ-6] 선형회귀 기법의 특징 목록

선형회귀 기법의 특징 목록	
국내/해외 IP 여부	웹 방화벽 차단 일수, 횟수
이벤트 발생 일 수	탐지/차단 비율
발생 이벤트 종류 개수	총 이벤트 개수
탐지(차단) 장비 종류 개수	이벤트 발생 시간 수
평일, 업무시간 내 탐지(차단) 비율	이벤트 발생 시간간격 표준편차
평일, 업무시간 외 탐지(차단) 비율	IP Reputation 사이트 제공 Score
휴일 탐지(차단) 비율	Destination IP 개수
방화벽 차단 일수, 횟수	이벤트 발생 네트워크 그룹 개수
웹 정상 접속 일수, 횟수	하루 평균 이벤트 발생 건수
IPS 탐지 일수, 횟수	이벤트 종류 별 평균 발생 건수
IPS 차단 일수, 횟수	IP 소유자 및 ISP
웹 방화벽 탐지 일수, 횟수	발생시킨 이벤트 명

선형회기 기법은 n개의 특징을 선형방정식으로 구성한 후 각 항의 계수를 찾는 방식으로 모델링한다. 즉 n개의 특징들을 xi (0≤i≤n-1)로 정의하고 블랙리스

트 여부를 y(x)로 정의할 경우 아래의 식과 같이 표현된다.

$$y(t) = a0 + a1 \cdot x1(t) + a2 \cdot x2(t) + \ldots + an-1 \cdot xn-1(t)$$

학습데이터를 선형회기 기법에 적용하는 IP별 실수(real number) 하나를 반환한다. 이 반환 값을 해당 IP의 Risk Score로 하였으며, 학습데이터 라벨에는 블랙리스트를 1로 정상사용자를 0으로 정의한다. 따라서 Risk Score가 높을수록 블랙리스트일 확률이 높으며, 낮을수록 정상사용자 일 확률이 높다고 할 수 있다. 선형회기 기법은 직접 분류를 해 주는 기법이 아니므로 블랙리스트와 정상사용자 각 그룹의 Risk Score에 대한 확률밀도함수가 만나는 지점을 이용하여 분류에 활용한다. 두 데이터 그룹 확률밀도함수의 교차점은 결국 블랙리스트와 일반사용자를 구분하기 위해 각각이 수용해야 하는 오류 개수가 가장 적은 지점이 되며 그 지점 이상 Risk Score를 갖는 IP를 블랙리스트, 그 지점 이하를 정상사용자로 정의 한다. 이를 통해 보안관제 요원의 블랙리스트 누락을 64% 감소시켰으며 보안관제 요원이 블랙리스트로 오 등록한 IP 93.8%를 정상으로 판단하였다. 하지만 6.2%의 오탐이 발생하였으며, 이는 일반 사용자가 발생시키는 이벤트의 패턴에 대한 정확한 지식부족으로 발생하였다[9][10].

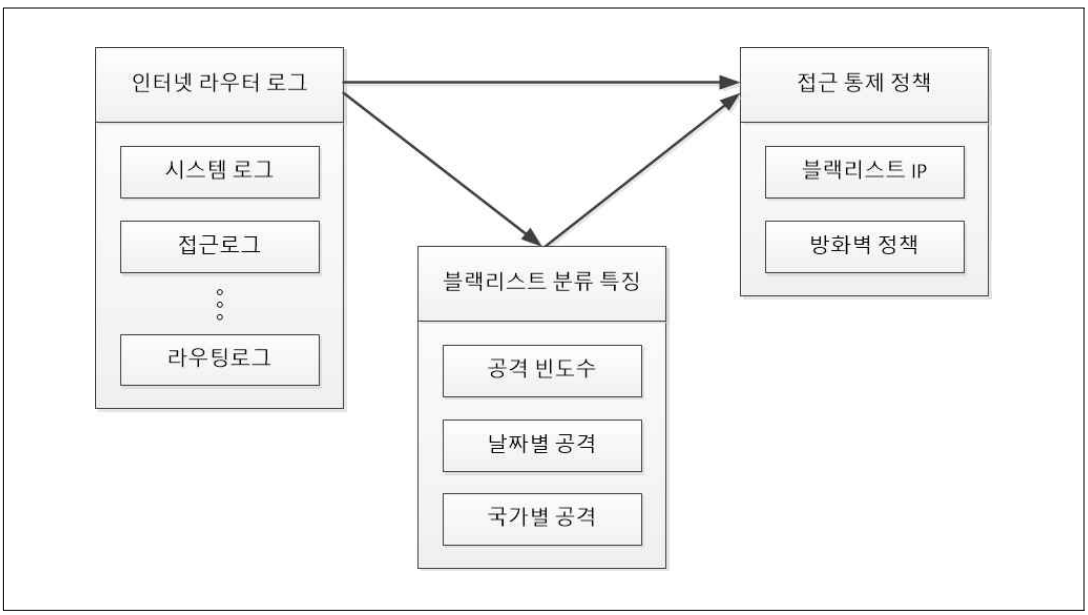
본 연구의 경우는 정상사용자와 블랙리스트 구분을 진행하는 것이 아니라 SSH 무차별 대입 공격만을 대상으로 하기 때문에 오탐이 발생하지 않고 선형회기 분석기법의 특징들에 가중치를 주어 이용한다면 어떤 IP가 더욱 공격적이며 위험한가를 판단할 수 있을 것이다. 블랙리스트 탐지를 위한 기준은 주관적 판단을 배제하기 힘들며, 각 조직이 가진 여건을 고려하여 일정 수의 오류를 불가피하게 수용하여 최적의 임계치를 찾는다. SSH 무차별 대입 공격의 경우도 모두가 공격에 해당하지만 실수로 인한 접근이나 단순 호기심 등 그 빈도수가 적은 경우를 배제하여야 한다.

Ⅲ. 로그를 활용한 접근통제 연구 설계 및 방법

1. 접근통제 연구 설계

1) 연구 모형

본 연구에서는 IT 인프라를 구성하는 장비 중 인터넷 라우터에 발생하는 로그를 통해 악의적인 접근 IP를 탐지하여 접근통제 정책을 만들고자 한다. 이를 위해 라우터의 원시 로그 중에서 SSH 무차별 공격 로그를 탐지하고 단편화 하여 ElasticSearch로 분석한다. 분석을 통해 공격의 빈도수가 높은 IP를 리스트화 하여 인터넷 라우터 및 보안 장비의 차단 정책에 활용한다. 이를 위해 [그림 III-1]과 같이 연구 모형을 설정하였다.



[그림 III-1] 연구모형

악의적인 접근의 근원지 IP는 접근통제 정책에 아주 중요한 지표가 된다. 해당 IP를 지속적으로 공급받을 수 있다면 IT 인프라의 보안 수준은 높아질 것이다.

또한 공격 근원지 자체의 공격을 차단할 뿐만 아니라 해당 근원지를 목적지로 하는 은닉채널을 통한 유출이 있다면 사전에 예방할 수 있다.

2) 연구 가설

인터넷 라우터의 로그를 활용한 접근통제 정책을 수립하기 위해 [그림 III-1]과 같은 연구 모형에 따라 다음과 같은 연구 가설을 설정하고 검증하고자 한다.

(1) 인터넷 라우터에 지속적이고 악의적인 접근 확인

IT 인프라의 구조적 취약점으로 이론적 배경을 통해 인터넷 라우터를 지목하였다. 인터넷 라우터는 외부와의 연결을 보장하기 위해 빈번한 작업이나 접속이 없는 장비이다. 때문에 네트워크의 변화가 없다면 생성되는 로그의 수도 적어야 하며 관리자의 접속만 로그로 남을 것이다.

[표 III-1] 인터넷 라우터 로그 생성 수

구 분	2019년 4월	2019년 5월	2019년 6월	2019년 7월	2019년 8월	2019년 9월	2019년 10월	2019년 11월	2019년 12월	2020년 1월	2020년 2월	2020년 3월
라우터-1	6187	4856	9988	6864	3780	5763	8592	15464	16912	52218	18141	6776
라우터-2	6226	4387	11160	6055	3339	6235	8731	15318	16942	50312	18454	6966

[표 III-1]는 본 연구를 위해 조사한 인터넷 라우터 로그 생성 숫자로 시스템 로그로만 생성될 수 없는 숫자이다. 인터넷 라우터는 기업과 기관의 전체 가용성을 보장하는 장비로 빈번한 작업이 이루어지진 않는다. 라우팅 테이블 변경 및 ISP와 회선 변경 및 전송장비 변경 등 특별한 경우를 제외하고는 특별한 작업이 없다. 하지만 이렇게 많은 로그가 발생되는 원인이 존재할 것이며 대부분을 차지하는 로그는 [표 III-2]와 같이 특정 로그의 반복이다

[표 III-2] 인터넷 라우터 로그 샘플

Apr 1 00:01:07 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 11*.17*.5*.88 on VTY0 due to IP restriction.. Apr 1 00:01:45 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 10*.20*.3*.84 on VTY0 due to IP restriction.. Apr 1 00:03:09 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 15*.23*.24*.63 on VTY0 due to IP restriction.. Apr 1 00:05:25 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 15*.23*.24*.63 on VTY0 due to IP restriction.. Apr 1 00:05:56 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 11*.17*.5*.88 on VTY0 due to IP restriction.. Apr 1 00:10:44 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 11*.17*.5*.88 on VTY0 due to IP restriction.. Apr 1 00:12:00 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 15*.23*.24*.63 on VTY0 due to IP restriction.. Apr 1 00:14:11 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 15*.23*.24*.63 on VTY0 due to IP restriction..

[표 III-2]는 인터넷 라우터 1번의 장비 로그이며 장비의 소유를 유추할 수 있는 hostname 부분을 router-1로 변경하였고 IP 주소 정보를 * 치환하여 IP 정보를 숨겼다. 인터넷 라우터 로그의 메시지 부분을 해석하면 다음과 같은 의미를 나타낸다. “SSH user failed to log in from”는 해석 그대로 SSH 접속 실패를 뜻하며 from 뒤에는 IP 정보가 따라온다. “on VTY0 due to IP restriction..”는 접속 실패가 IP 접속 제한이란 것을 나타낸다. 즉 IP 제한으로 인해 출발지 IP 정보의 SSH 사용자가 VTY0에 로그인 하지 못하였다.

인터넷 라우터는 주요정보기반시설물 취약점 점검을 통해 계정관리, 로그관리, 기능관리 등 취약점을 보강하여 사용한다. 관리자의 VTY 접속을 ACL(Access-List)를 통해 허용하고 그 외 비인가자의 접속은 차단한다. 인터넷 라우터의 대부분을 차지하는 로그는 비인가자의 로그인 것이다. 비인가자의 접속을 전체로그와 비교하면 [표 III-3]과 같이 나타낼 수 있다.

[표 III-3] 인터넷 라우터 공격 로그 분석

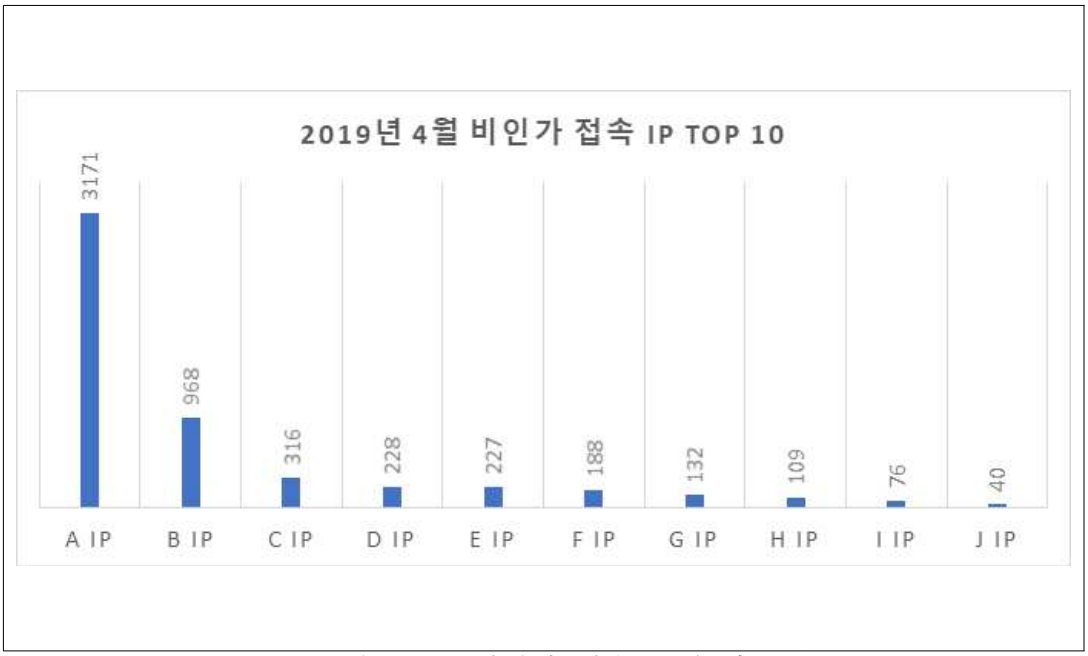
구분	인터넷 라우터-1				인터넷 라우터-2			
	전체 로그수	시스템 로그	비인가 접속	비인가 접속률	전체 로그수	시스템 로그	비인가 접속	비인가 접속률
2019년 4월	6187	37	6150	99.40 %	6226	33	6193	99.47 %
2019년 5월	4856	186	4670	96.17 %	4387	131	4256	97.01 %
2019년 6월	9988	46	9942	99.44 %	11160	30	11130	99.73 %
2019년 7월	6864	164	6700	97.61 %	6055	66	5989	98.91 %
2019년 8월	3780	562	3218	85.13 %	3339	30	3309	99.10 %
2019년 9월	5763	30	5733	99.48 %	6235	29	6206	99.53 %
2019년 10월	8592	32	8560	99.63 %	8731	30	8701	99.66 %
2019년 11월	15464	45	15419	99.71 %	15318	30	15288	99.80 %
2019년 12월	16912	29	16883	99.83 %	16942	29	16913	99.83 %
2020년 1월	52218	30	52118	99.94 %	50312	30	50282	99.94 %
2020년 2월	18141	6	18135	99.97 %	18454	38	18416	99.79 %
2020년 3월	6776	86	6690	98.73 %	6966	35	6931	99.50 %

비인가자의 접속은 전체 로그수와 거의 동일한 수준이며 인터넷 라우터의 로그를 통해 비인가자의 지속적인 접근이 발생함을 알 수 있다.

(2) 비인가자의 접근을 특정하여 탐지

인터넷 라우터로 접속하는 비인가자 접속은 SSH를 이용한 정상적인 접근이

다. 하지만 인터넷 라우터는 관리자 IP만 SSH 접속을 허용하고 나머지 접속은 차단하여 로그를 남긴다. 해당 로그는 [표 III-3]과 같이 엄청난 수이며 비인가자의 SSH 접속은 인터넷 라우터의 계정정보를 확인하기 위한 무수히 많은 시도라고 판단된다. 2019년 4월의 비인가 접속에 대해 확인한 결과 [그림 III-2]와 같이 IP별 접근시도가 집계되었다.



[그림 III-2] 비인가 접속 IP별 차단 수

4월중 인터넷 라우터 1에 접속 시도한 상위 10개의 아이피로 이중 가장 많은 접속을 시도한 A IP(실제 IP를 알파벳 이니셜로 변경함)는 3,171건으로 전제 로그(6,187건)중 51.25%를 차지하고 있다. 정상적인 접속으로 판단하기 어려운 시도 횟수이며 인력으로 하기엔 힘든 수치이다. 또한 이런 접속시도의 다른 특징중 하나는 아주 짧은 간격의 접속 재시도를 진행하고 그 빈도나 간격이 규칙적인 것이다.

[표 III-4] IP 접속 시도 정보 샘플

구분	일	시간	일	시간	일	시간	일	시간	일	시간
1	9	4:40:00	10	12:18:37	11	23:03:41	12	15:39:57	15	11:10:37
2	9	4:40:01	10	12:18:37	11	23:03:41	12	15:39:57	15	11:10:38
3	9	4:40:02	10	12:18:38	11	23:03:42	12	15:39:57	15	11:10:39
4	9	4:40:02	10	12:18:38	11	23:03:42	12	15:39:57	15	11:10:39
5	9	4:40:03	10	12:18:38	11	23:03:42	12	15:39:57	15	11:10:39
6	9	4:40:03	10	12:18:38	11	23:03:42	12	15:39:58	15	11:10:39
7	9	4:40:03	10	12:18:39	11	23:03:42	12	15:39:58	15	11:10:39
8	9	4:40:04	10	12:18:42	11	23:03:42	12	15:39:58	15	11:10:40
9	9	4:40:04	10	12:18:44	11	23:03:43	12	15:39:58	15	11:10:40
10	9	4:40:04	10	12:18:44	11	23:03:43	12	15:39:58	15	11:10:40
11	9	4:40:04	10	12:18:45	11	23:03:43	12	15:39:59	15	11:10:41
12	9	4:40:05	10	12:18:45	11	23:03:43	12	15:39:59	15	11:10:41
13	9	10:33:08	11	2:28:40	12	7:24:22	13	0:18:41	15	11:20:38
14	9	10:33:08	11	2:28:42	12	7:24:22	13	0:18:41	15	11:20:38
15	9	10:33:08	11	2:28:43	12	7:24:23	13	0:18:41	15	11:20:38
16	9	10:33:08	11	2:28:44	12	7:24:24	13	0:18:41	15	11:20:38
17	9	10:33:09	11	2:28:44	12	7:24:24	13	0:18:41	15	11:20:39
18	9	10:33:09	11	2:28:44	12	7:24:24	13	0:18:41	15	11:20:39
19	9	10:33:09	11	2:28:44	12	7:24:24	13	0:18:42	15	11:20:40
20	9	10:33:09	11	2:28:45	12	7:24:25	13	0:18:42	15	11:20:41
21	9	10:33:10	11	2:28:45	12	7:24:25	13	0:18:42	15	11:20:42
22	9	10:33:10	11	2:28:45	12	7:24:25	13	0:18:42	15	11:20:42

구분	일	시간	일	시간	일	시간	일	시간	일	시간
23	9	10:33:10	11	2:28:45	12	7:24:26	13	0:18:42	15	11:20:42
24	9	10:33:12	11	2:28:45	12	7:24:26	13	0:18:42	15	11:20:43

접근 시도가 얼마나 빈번하게 일어나는지는 확인하기 위해 상위 10의 IP중 하나(E IP의 227건중 120건만 샘플링)를 정하여 [표 III-4]과 같이 일자와 시간으로 정렬하였다. 해당 IP는 오름차순 정렬시 가장 먼저 나온 IP로 무작위 선정하였으며, 나머지 상위 10개의 IP도 샘플로 정한 IP와 동일 패턴이 나왔다.

[표 III-4]와 같이 여러 날에 걸쳐 접속 시도가 이루어졌으며 접속 시도 간격도 1초 이하로 아주 짧은 시간이 이루어져 스크립트나 프로그램에 의한 자동 접속이 의심된다. 또한 시도 횟수도 12회를 기준으로 시간과 날짜를 달리해 접속 시도를 진행하여 보안장비의 탐지를 우회하려는 목적이 보인다.

인터넷 라우터로 접속 시도를 진행하는 비인가 IP의 접속은 그 횟수나 접속로그의 메시지 부분에서 특정 지을 수 있으며 로그의 대부분을 차지하고 있어 보안 로그로 충분한 가치가 있다. 이 접속시도는 SSH 무차별공격에 해당하며 이 로그를 활용한다면 악의적인 근원지 IP 정보를 수집할 수 있다.

(3) 비인가자 공격 로그 단편화

인터넷 라우터에 수집되는 SSH 무차별 공격에 대한 로그는 다음과 같은 로그 메시지로 이루어진다.

“Apr 1 00:01:07 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 11*.17*.5*.88 on VTY0 due to IP restriction..”

이중 Apr 1 00:01:07은 월, 일, 시간에 해당하는 부분이고 2019 부분은 년도에 해당한다. router-1은 장비명이고 이하 나머지 부분은 로그의 메시지 부분이다. 구문으로 나눈다면 월, 일, 시간, 년도, 장비명, 메시지로 나눌 수 있다. 메시지를 제외한 부분은 로그가 생성될 때 시스템에서 자동으로 정하여 생성되는 것으로 로그의 메시지 내용을 정확하게 검증해준다. 즉 메시지의 행위가 어떤 장비에서

언제 생겼는지 알려주는 것이다. 중요한 메시지 부분은 IP를 제외한 나머지 부분이 반복되고 있어 이 부분의 중복을 제거하면 IP만 남게 되어 로그의 길이를 줄이고 구문을 확실하게 단편화 할 수 있다.

위 로그를 “Apr 1 00:01:07 2019 router-1 11*.17*.5*.88”과 같이 구분 지을 수 있다. 월, 일, 시간, 년도, 장비명, IP로 나뉘는 구문으로 나뉘며 공란을 기준으로 나누어짐으로 여러 포맷의 파일로 만들 수 있다.

[표 III-5] SSH 무차별공격 로그 단편화

month	day	time	year	hostname	ip
Apr	1	0:01:07	2019	router-1	11*.17*.5*.88
Apr	1	0:01:45	2019	router-1	10*.20*.3*.84
Apr	1	0:03:09	2019	router-1	15*.23*.24*.63
Apr	1	0:05:25	2019	router-1	15*.23*.24*.63
Apr	1	0:05:56	2019	router-1	11*.17*.5*.88
Apr	1	0:10:44	2019	router-1	11*.17*.5*.88
Apr	1	0:12:00	2019	router-1	15*.23*.24*.63
Apr	1	0:14:11	2019	router-1	15*.23*.24*.63
Apr	1	0:15:33	2019	router-1	11*.17*.5*.88

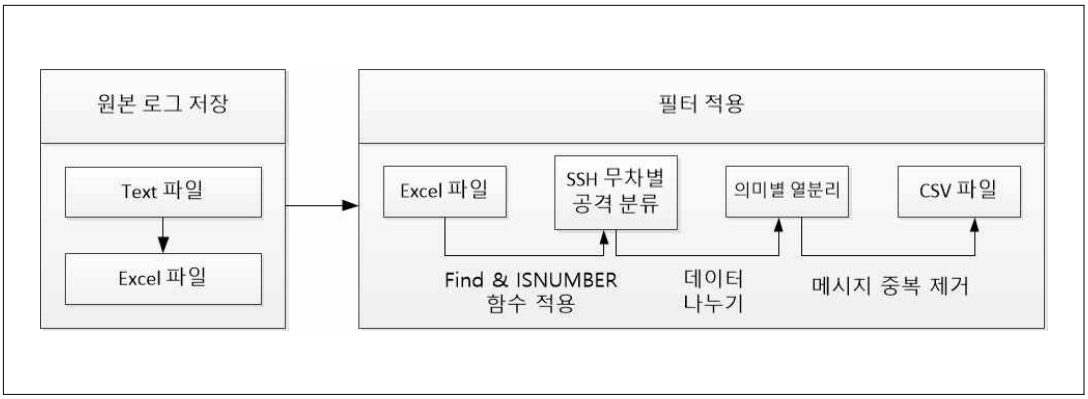
[표 III-5]과 같이 SSH 무차별 공격으로 판단되는 로그를 단편화하여 csv 파일로 만들 수 있고 이를 활용하면 시간별, 날짜별, 년도별 IP에 따른 접속 시도 횟수가 확인이 가능하다. 또한 국가 필드를 추가하여 IP에 따른 국가 코드도 등록하여 접근시도가 어떤 국가에서 발생했는지 확인할 수 있다.

3) 연구 설계

본 연구는 인터넷 라우터의 원본 로그를 가공하여 SSH 무차별 대입 공격 로그만 단순화하여 분석을 통해 접근 통제 정책을 만들어 적용하고자 한다. 이를 위해 원본 로그의 SSH 무차별 공격 로그만 선별 작업이 필요하다. 원 로그는 텍스트 파일형태로 저장 받았으며 해당 파일을 엑셀로 불러와 정리하였다. 이에 엑셀 함수를 적용하여 SSH 무차별 대입 공격 로그와 그 외 로그로 분리하고 중복되는 문자열을 제거함으로써 필요한 정보만 들어있는 CSV 파일로 만든다. 이를 ElasticSearch로 분석하여 Black-List IP를 선별하고 인터넷 라우터의 ACL과 보안장비의 차단 정책에 활용하여 발생할 수 있는 침해사고를 막고자한다.

(1) 원본 로그 가공

인터넷 라우터의 원본 로그는 텍스트 형태로 저장하고 엑셀을 통해 파일을 열 수 있다. 하나의 행과 생성된 로그의 수만큼의 열로 이루어진 로그 문서는 SSH 무차별 대입 공격을 탐지해야 한다. 이를 위해 [그림 III-3]과 같이 필터를 적용하여 단편화 작업을 진행한다.



[그림 III-3] 원본 로그 분류

엑셀의 FIND 함수는 찾는 값이 들어간 셀의 시작점 위치를 반환하는 함수이고, ISNUMBER(식) 함수는 식을 확인하여 참과 거짓을 반환하는 함수이다. 이를 활용하여 엑셀에 쌓인 로그 중 SSH 무차별 대입 공격에 해당하는 메시지 부분

의 문자열을 포함한다면 참(TRUE), 아니면 거짓(FALSE)을 표현할 수 있다. 이를 필터링하면 참인 SSH 무차별공격 로그 혹은 반대의 경우만 추출할 수 있다.

SSH 무차별 공격을 뜻하는 문자열은 “SSH user failed to log in from”과 “ on VTY0 due to IP restriction..”부분으로 해당 문자열을 ISNUMBER의 식으로 삼는다면 [표 III-6]과 같이 로그상 해당 문자열이 있으면 참, 아니면 거짓으로 구분 가능하다.

[표 III-6] FIND & ISNUMBER 함수 적용

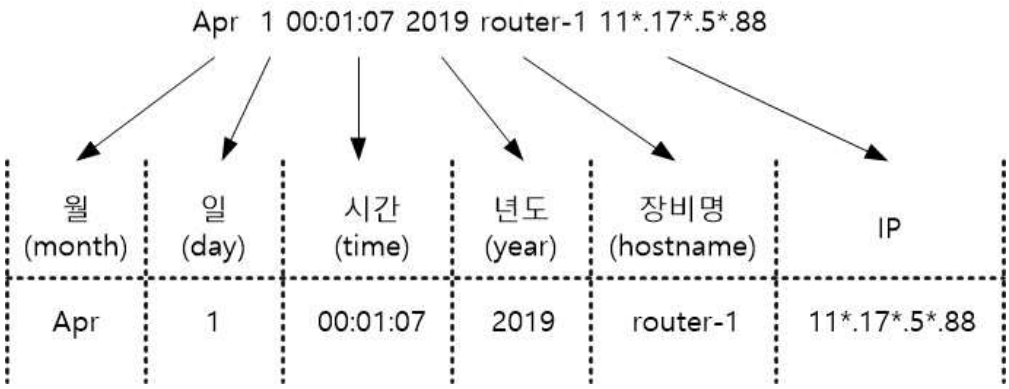
구분	원시 로그 정보
FALSE	Apr 23 12:13:48 2019 router-1 %%10SSH/4/TrapLogoff(t): 1.3.6.1.4.1.25506.2.22.1.3.0.4 SSH user logoff trap information
TRUE	Apr 23 12:38:26 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 5*.21*.12*.66 on VTY0 due to IP restriction..
TRUE	Apr 23 12:50:28 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 6*.24*.20*.20 on VTY0 due to IP restriction..

원시 로그 정보를 참과 거짓으로 구분할 수 있다면 필터를 통해 참의 자료만 또는 거짓의 자료만 추출가능하다. 인터넷 라우터의 로그는 월 적게는 3,000건 많게는 50,000건이 넘는 로그가 쌓였다. 이 로그 중에 시스템로그를 확인하기는 많은 양의 로그를 다 살펴야하는 불편함이 존재한다. 따라서 엑셀 함수를 통해 원본 로그를 분류한다면 원하는 SSH 무차별 대입 공격 로그와 시스템 로그를 구분에서 확인할 수 있다.

(2) 로그 단편화

SSH 무차별 대입 공격의 로그만 탐지된 자료는 쉽게 단편화가 가능하다. 모든 로그의 패턴이 정해져 있어 그 의미별로 행을 나누어 관리할 수 있기 때문이다. 또한 SSH 무차별 공격을 의미하는 문구인 “SSH user failed to log in from”

과 “ on VTY0 due to IP restriction..”부분을 제거하면 IP 부분만 나아서 데이터를 쉽게 나눌 수 있다. 엑셀에서 데이터를 나누는 구분 기호는 쉼표 및 공백 등 여러 패턴이 있어 의미 있는 구문으로 쉽게 나눌 수 있다. 나눌 수 있는 행은 [그림 III-4]와 같이 순서대로 월, 일, 시간, 년도, 장비명, IP로 구분된다.



CSV 파일 샘플

	A	B	C	D	E	F
1	month	day	time	year	hostname	ip
2	Apr	1	0:01:07	2019	router-1	11*.17*.5*.88
3	Apr	1	0:01:45	2019	router-1	10*.20*.3*.84
4	Apr	1	0:03:09	2019	router-1	15*.23*.24*.63
5	Apr	1	0:05:25	2019	router-1	15*.23*.24*.63
6	Apr	1	0:05:56	2019	router-1	11*.17*.5*.88
7	Apr	1	0:10:44	2019	router-1	11*.17*.5*.88
8	Apr	1	0:12:00	2019	router-1	15*.23*.24*.63
9	Apr	1	0:14:11	2019	router-1	15*.23*.24*.63

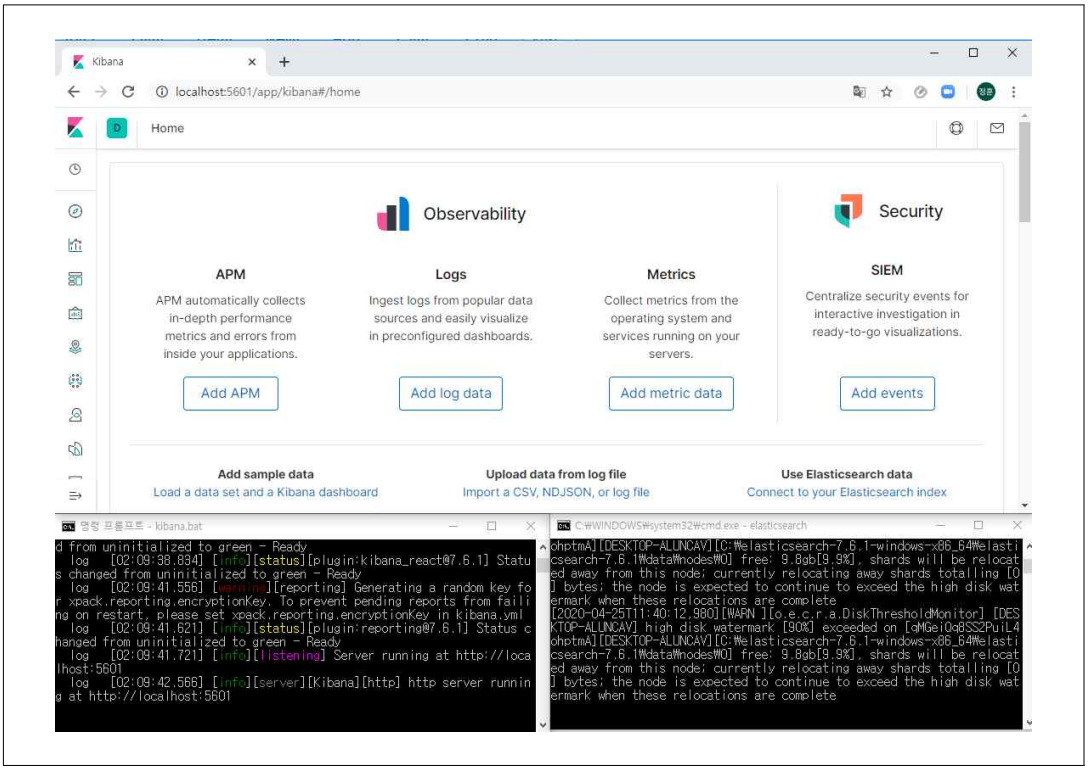
[그림 III-4] 로그 텍스트 나누기

이렇게 나뉜 로그 자료는 장비별, 월별 CSV 파일로 분석이 용이한 형태로 저장하여 다른 분석도구나 엑셀 자체만으로 분석이 가능하다. 이중화된 인터넷 라

우터 로그로부터 1년 분량의 24개의 데이터를 수집하여 가공 후 년-월-장비명.csv로 저장하였다.

(3) 로그 분석

분류되고 단편화된 로그파일은 분석에 적합한 형태로 저장된다. 이 파일은 많은 양의 정보를 가지고 있어 쉽게 분석하기 힘들다. ElasticSearch는 대용량 빅데이터 분석에 사용되는 분석도구로 해당 파일을 분석하기 알맞은 분석도구이다.



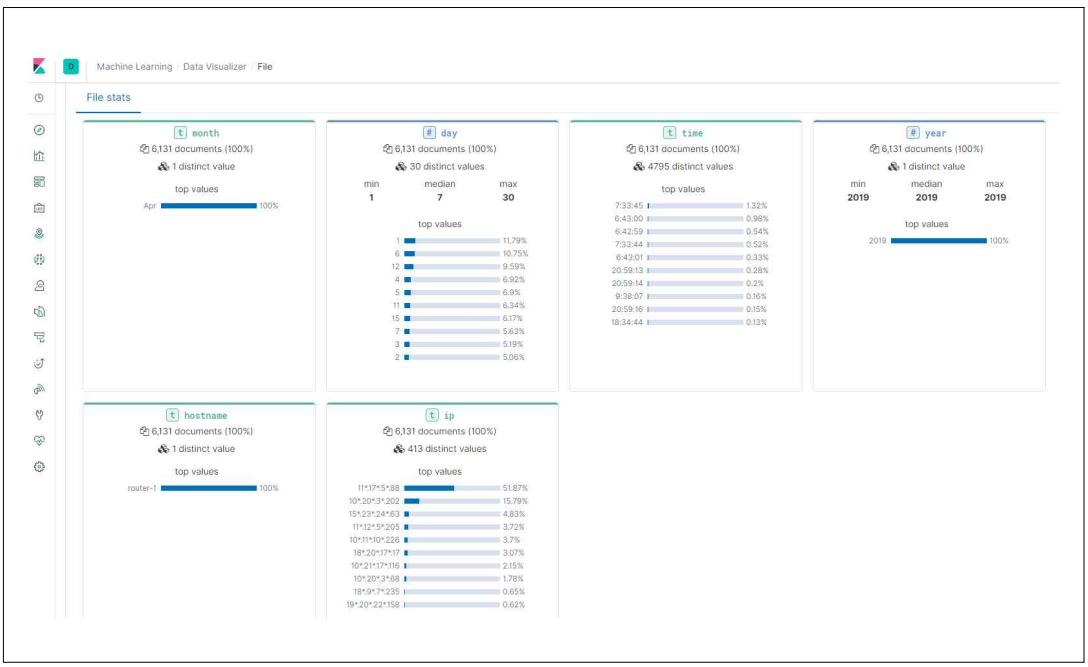
[그림 III-5] ElasticSearch 및 Kibana 구동 화면

ElasticSearch는 원본로그를 실시간 수집 저장 가공하여 분석할 수 있지만 본 연구에서는 인터넷 라우터와 연결하는 부분의 보안 문제 때문에 저장된 로그를 활용하여 분석하고자 한다.

엘라스틱 홈페이지(https://www.elastic.co/kr/elastic-stack)를 통해 JSON 기반의 분산형 검색 및 분석 엔진인 ElasticSearch 및 확장형 사용자 인터페이스로서

데이터를 구체적으로 시각화하는 도구인 Kibans를 다운 받을 수 있다. 윈도우의 경우 압축파일로 제공되며 ElasticSearch와 Kibana 둘 다 압축을 해제한 후 bin 폴더 내 실행 파일인 elasticsearch.bat와 kibana.bat를 실행시킴으로써 분석엔진과 시각화 도구를 실행 할 수 있다. [그림 III-5]는 ElasticSearch와 Kibana를 구동시킨 모습으로 자체 웹 서버를 내장하고 있어 웹 브라우저(localhost:5601)를 통해 분석을 진행할 수 있다.

분석을 위한 자료는 분류와 단편화를 거쳐 SSH 무차별공격으로만 되어있는 CSV 파일을 Kibana에 업로드(Machine Learning/Data Visualizer/File) 함으로써 자료 분석이 진행된다. ElasticSearch는 분석을 위해 제공하는 여러 형태의 로그를 Logstash의 필터를 통해 정형화 시켜 입력을 받지만 본 연구에서 가공된 CSV 파일은 콤마로 분리된 값으로 그 구분과 의미 분석이 용이하다.

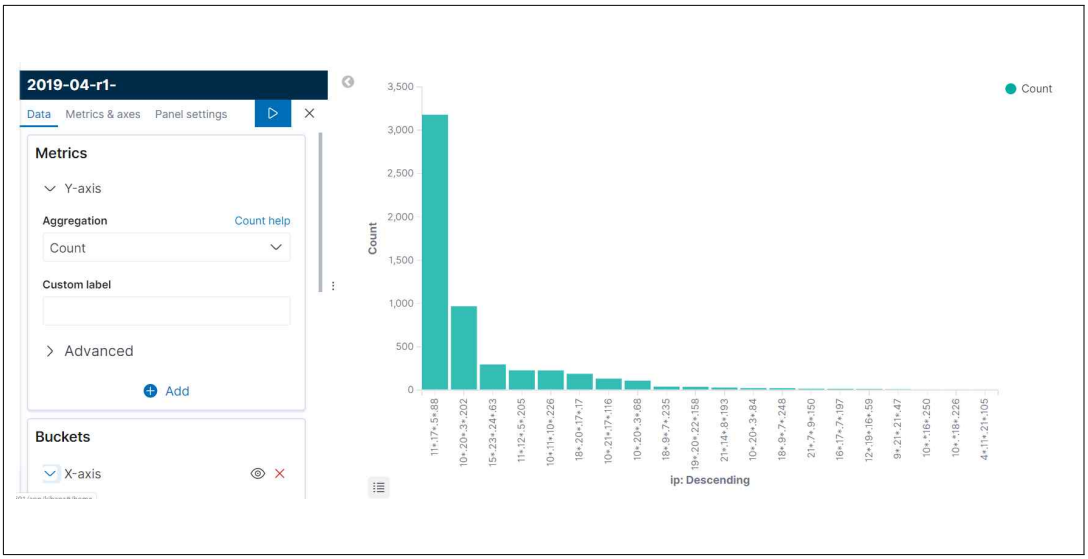


[그림 III-6] Kibana 업로드 파일 상태

[그림 III-6]은 CSV 파일을 Kibana에 업로드한 상태이며, 분류된 자료에 대한 정보가 나온다. 달에 해당하는 month 부분은 Apr이 100%인 6,131 로그 Count가

기록되며 day 부분에서는 1일이 11.79%, 6일이 10.75% 등 로그가 많이 생성된 일자별로 정렬된다. 이중 중요한 ip 필드는 11*.17*.5*.88이 전체 로그 수중 51.87%를 차지하고 수집된 IP 정보는 413개가 수집 되었다고 정보에 표현해준다.

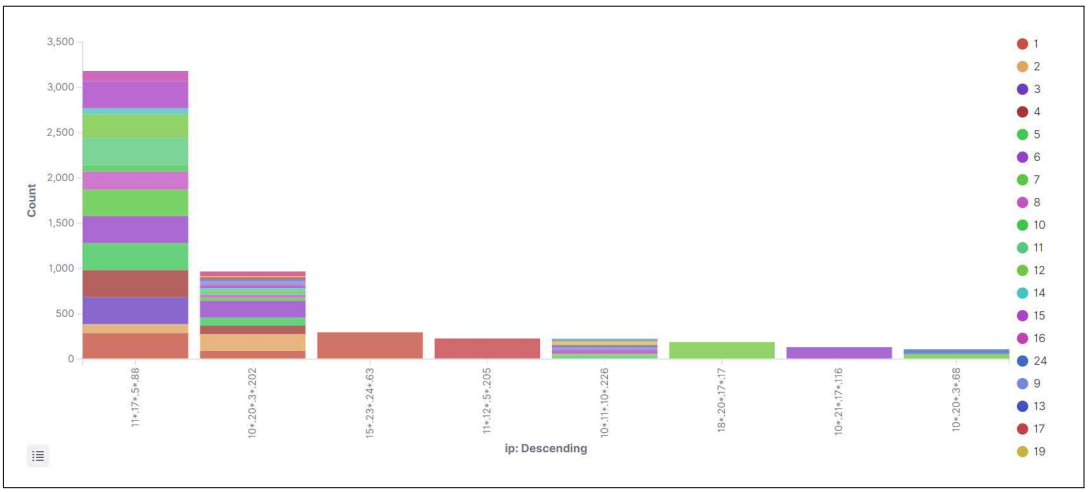
CSV파일을 Kibana에 업로드한 행위 자체로도 많은 정보가 분석되지만 본 연구에서 필요한 부분은 월별 IP의 Count와 이와 연계된 악의성 확인이다. 따라서 업로드된 자료를 Import를 통해 Kibana에 등록하여 좀 더 자세한 분석을 진행한다.



[그림 III-7] Kibana Visualization

Kibana의 Visualization을 통해 분석이 진행되며 X축과 Y축에 측정항목을 정하여 각 집합에 따른 측정값을 시각화한다. [그림 III-7]은 기본적인 시각화 방법으로 Y축의 측정항목을 Count로 설정하고 X축을 ip 필드의 정보를 Count별로 내림차순 정렬한 결과이다. 이를 통해 4월중 공격 IP의 공격횟수를 시각화하여 차단 대상의 임계치를 정할 수 있다.

정해진 임계치의 IP는 해당 IP만 내림차순으로 표현 가능하며 X축에 날짜별 히스토그램을 추가하여 IP의 날짜별 공격 추이도 분석 가능하다.



[그림 III-8] 공격 IP 날짜별 히스토그램

SSH 무차별공격은 많은 빈도수의 공격이지만 날짜별로 분석을 했을 때 같은 날에 행해지는 공격보다는 여러 날에 걸쳐 지속적인 공격의 형태를 띤다. 이런 공격은 특정일에 집중적인 공격보다 그 악의성이 높으며 목표를 지속적으로 공격한다는 행위 차체만으로도 차단의 대상이 되어야 한다.

(4) 블랙리스트(Black-List) IP

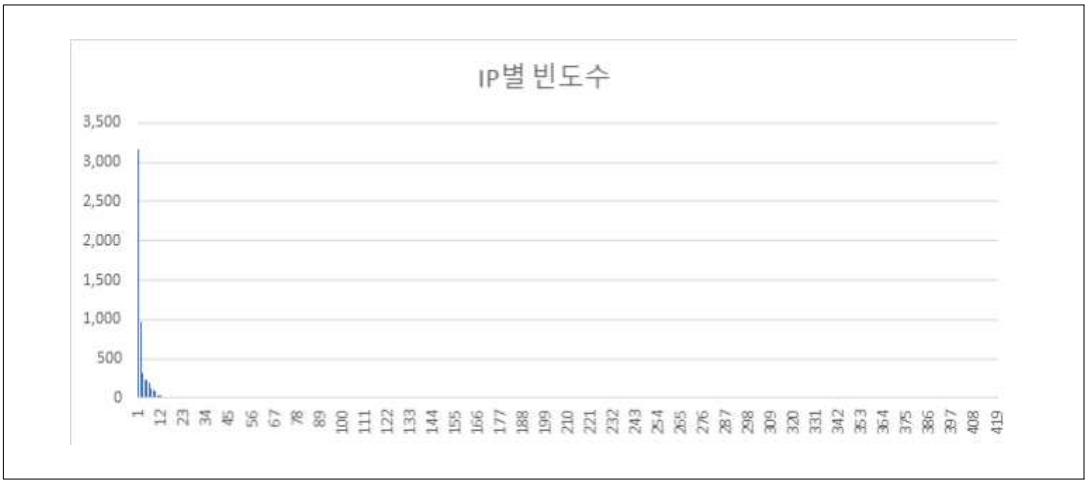
SSH 무차별공격을 가하는 IP는 모두 악의적인 호스트일 것이다. 하지만 공격하는 IP 모두를 블랙리스트로 정하고 차단하는 것은 비효율적인 정책이 된다. 공격 IP중 일부는 단순 일회성 공격이며 더 이상 공격의 의도가 없는 경우도 있다. 따라서 IP의 공격 빈도는 공격의도 입증에 중요한 기준이 된다. 또한 공격의 지속성을 판단하기 위해서는 날짜별 공격을 특징으로 정할 수 있다. 시간과 관련된 특징은 세분화하여 업무시간 구분, 휴일 구분 등 구체화 할 수 있으며 탐지된 IP를 통해 국가별 구분도 가능하다. [4][7]

[표 III-7]는 인터넷 라우터 상에서 악의적인 호스트를 구분하기 위한 특징을 나타내는 것으로 공격 강도인 IP별 차단 빈도수는 블랙리스트 선택에 가장 중요한 지표가 된다[19][17].

[표 III-7] 블랙리스트 IP 선택를 위한 특징

구분	특징
IP 식별	SSH 무차별공격
공격 강도	IP별 차단 빈도수
공격 타이밍	탐지 일 수
공격 위치	국내 / 국외

또한 공격 의도가 없는 일회성 공격은 제외해야 한다. [그림 III-8]은 2019년 4월 router-1의 IP별 차단 빈도수이다. 총 419개의 IP가 인터넷 라우터를 공격했으며 이중 최고 빈도수는 3,171건으로 공격 의도가 분명하다. 하지만 10회 미만의 공격 의도가 불분명하고 일회성 공격은 전체 IP의 96%인 405나 되어서 블랙리스트 여부를 판단하는 자료로는 부적절하다. 따라서 공격 의도가 없는 10회 미만의 공격횟수를 가진 IP를 제외하여 분석을 빠르고 쉽게 진행할 수 있다.



[그림 III-8] IP별 공격 빈도수 (2019년 4월)

이를 기준으로 다른 항목들을 연산하여 수용 가능한 범위의 블랙리스트 판단 기준을 만들고자한다. 3개의 공격 특징들을 x로 정의하고 블랙리스트 판단 여부를 y(x) 정의하여 다음 식과 같이 블랙리스트 여부를 나타낸다.

$$y(x) = x1 \times x2 \times x3$$

[식 III-1] 블랙리스트 판단 여부

각 공격 특징에 대한 정의는 공격 빈도수를 $x1$은 n개의 공격 빈도(0<n)수에 0.01을 곱하여 100개 이상의 공격을 1로 나타낸다. 즉 $x1 = n \times 0.01$이다. 탐지일수인 $x2$는 탐지일수인 $m(0 < m \leq 31)$에 0.1을 곱하여 1일당 0.1의 가중치를 갖게 한다. 즉 $x2 = m \times 0.1$이다. 공격위치인 $x3$은 국내와 국외로 구분하여 국내의 경우 1 나머지는 1.1을 주어 계산한다. 따라서 국내에서 1일 동안 인터넷 라우터에 100건의 SSH 무차별공격을 가한다면 (100×0.01)×(1×0.1)×1 = 0.1의 수치를 갖는다. 블랙리스트로 구분하고자 하는 IP는 최소 100건이상의 공격과 2일 이상의 탐지일수 그리고 해외의 경우로 산정하고자 한다. 이를 수치로 나타내면 0.22의 수치가 나온다. 또한 0.1과 0.22 사이의 수치의 IP를 의심가는 IP로 선택하여 관리한다.

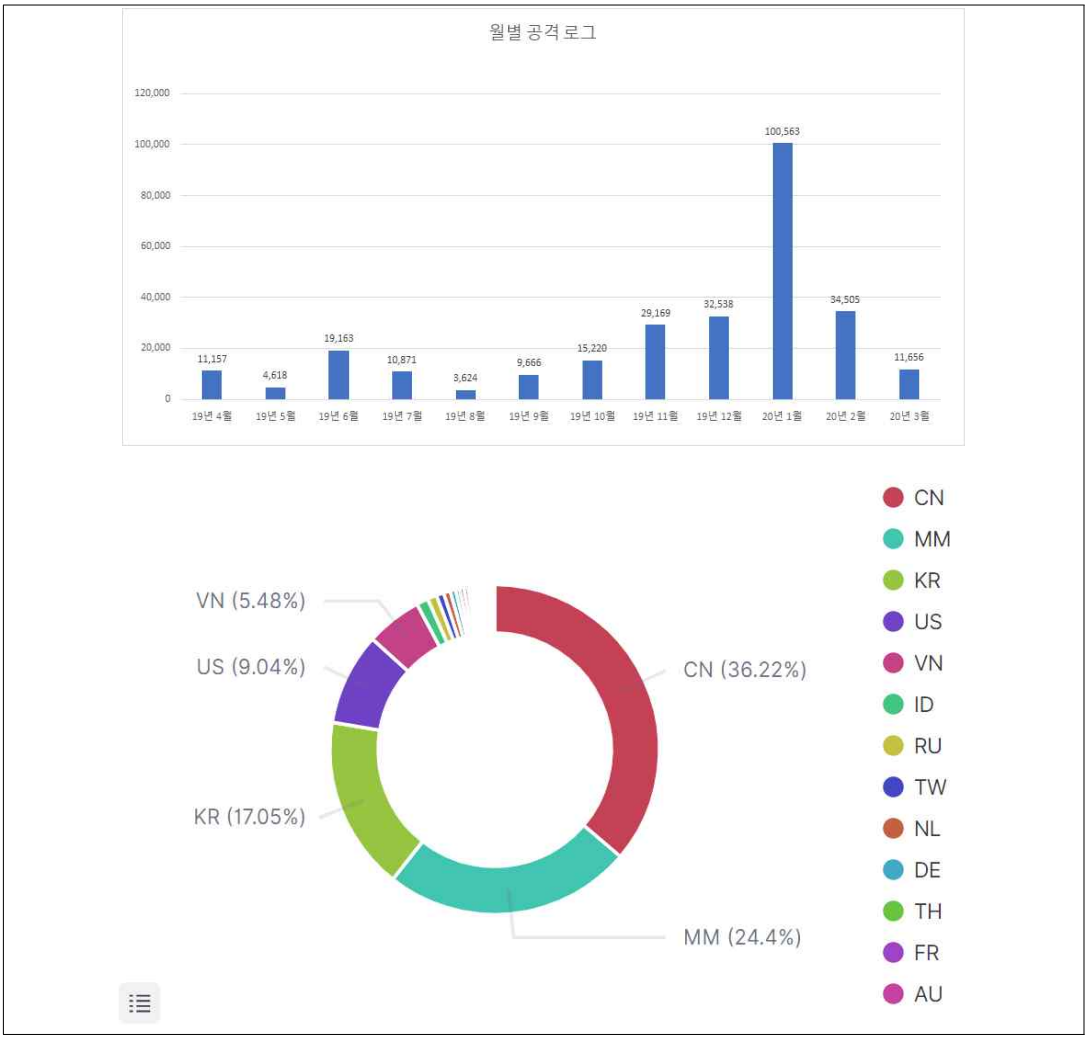
[표 III-8] 블랙리스트 판단 여부 샘플

IP	공격 빈도수 (빈도수×0.01)	탐지 일수 (일×0.1)	위치 (국내1/국외 1.1)	블랙리스트 Rating	블랙리스트 여부
a	1	0.2	1.1	0.22	black-list
b	1.32	0.1	1	0.132	suspect
c	8	0.1	1	0.8	black-list
d	0.9	0.7	1.1	0.69	black-list

[표 III-7]의 샘플과 같이 월별 IP에 대한 판단여부를 표로 작성하여 블랙리스트 여부와 그 수치를 지속 관리하여 접근통제 정책에 활용한다. 표를 통해 작성된 블랙리스트 평가지수는 임계치를 정하기 위한 단순한 실수일 뿐이며 조직의 여건을 고려하여 일정 수의 오류를 수용하는 범위 내에서 임계치를 정하여 블랙리스트 IP로 정하였다.

2. 연구 결과

10건 이상의 공격 빈도를 가진 IP를 분석한 결과 월별 빈도수와 국가별 등 [그림 III-9]와 같이 많은 정보를 확인할 수 있다.

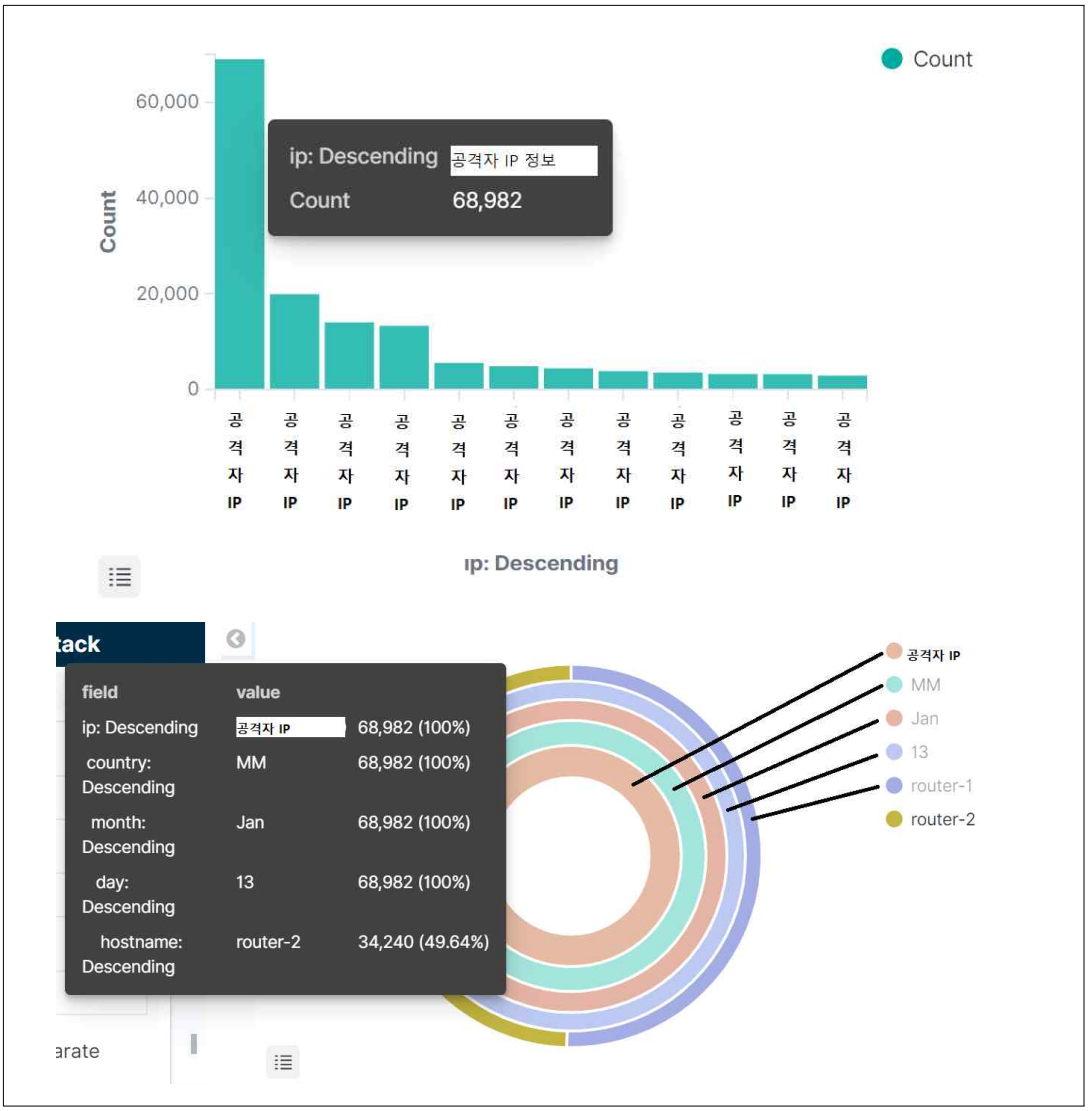


[그림 III-9] 월별, 국가별 공격 빈도수

월별 공격 빈도수는 5월, 8월, 9월을 제외하고는 모두 1만 건 이상의 대량의 로그가 발생하였으며 국가별 공격빈도는 102,408건으로 중국(전체 36.22%)이 가장 높았으며 미얀마(24.4%) 68,982건, 한국(17.05%) 48,199건, 미국(9.04%) 25,548

건, 베트남(5.48%) 15,499건의 순서로 분석되었다.

또한 IP별 빈도수 분석과 이를 통해 특정 IP에 대한 국가, 월, 일, 장비별 분석을 통해 IP의 특이점을 확인할 수 있다. [그림 III-10]은 최고 빈도를 가진 IP에 대한 세분화된 분석으로 68,982건의 공격 빈도와 미얀마를 통해 1일 동안 가한 공격임을 확인 가능하다. 특이점은 미얀마 전체 공격과 일치하는 수치이며 단일 IP에서 하루 동안 지속적인 악의적 공격임을 알 수 있다.



[그림 III-10] IP 상세 분석

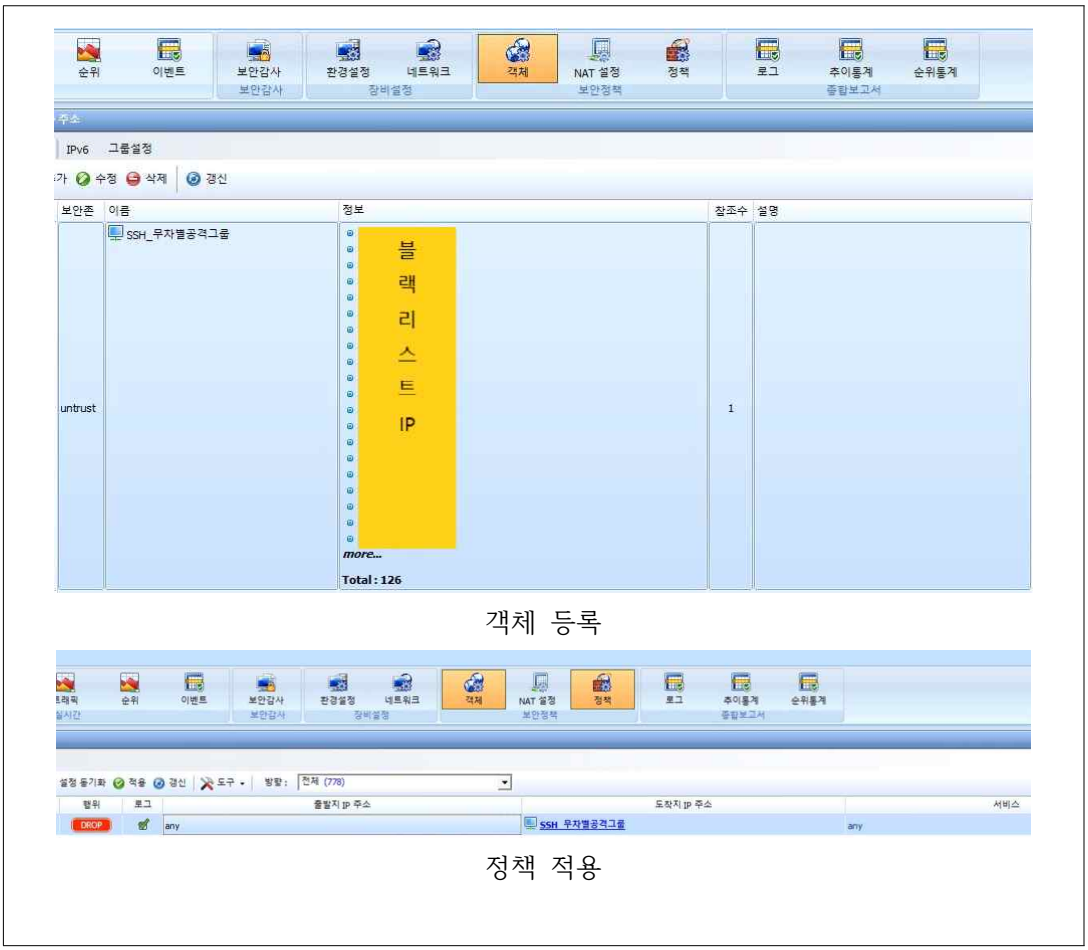
이런 세분화된 분석을 바탕으로 블랙리스트 IP의 빈도수, 날짜별, 국가별 가중치를 통해 접근통제 정책의 객체(object)인 블랙리스트 IP를 선택할 수 있다. [표 III-8]은 블랙리스트 특징을 [식 III-1]에 대입하여 나온 정책의 수를 나타냈으며 이중화된 인터넷 라우터 양쪽 모두를 공격하는 중복 같은 대역의 서로 다른 연속되는 IP는 CIDR(Classless Inter-Domain Routing)로 처리하였다. CIDR로 처리된 블랙리스트 IP대역은 24bit의 subnet을 가지며 공격 빈도수가 높을수록 나타났다.

[표 III-9] 블랙리스트 IP 수

구분	총 블랙리스트 IP (전체수량)	중복 IP	정책 수	비고
19년 4월	11	8	7	
19년 5월	26	4	24	
19년 6월	28	20	18	
19년 7월	17	12	11	
19년 8월	6	4	4	
19년 9월	57	30	14	CIDR 포함
19년 10월	110	110	19	CIDR 포함
19년 11월	95	92	18	CIDR 포함
19년 12월	197	192	22	CIDR 포함
20년 1월	160	154	27	CIDR 포함
20년 2월	150	142	33	CIDR 포함
20년 3월	31	24	19	

블랙리스트 IP에 분류된 공격 대부분이 많은 빈도수로 이중화된 인터넷 라우터 모두를 공격하는 악의적인 공격이며, 연속되는 IP를 통해 자동화된 공격임을

예측할 수 있다. 분석을 통해 나타나는 이런 특징들은 해당 블랙리스트 IP를 접근통제 정책으로 설정하여 차단하기에 충분한 이유가 된다. 또한 자동화된 공격의 근원지를 접근통제 정책으로 차단한다면 내부 PC의 악성코드 감염 등으로 내부에서 외부의 악성 근원지로의 은닉채널이 생기는 것을 차단할 가능성이 있다. 따라서 접근통제 정책은 방화벽을 이용하여 외부에서 신뢰구간으로 접근하는 블랙리스트 IP를 차단할 뿐만 아니라 내부에서 블랙리스트 IP로의 접근 또한 차단하는 정책을 만들어야 한다.



[그림 III-12] 블랙리스트 방화벽 정책 적용

[그림 III-9]는 블랙리스트 IP에 대한 방화벽 정책 적용 화면이다. 방화벽에 적용할 블랙리스트 IP는 총 126개의 IP로 SSH_무차별공격그룹 이라는 객체를 만

들어 관리하고 지속적으로 추가할 예정이다. 이 객체를 목적지로 하는 방화벽 정책은 내부망의 모든 사용자가 출발지이며 목적지는 SSH_무차별공격그룹으로 모든 서비스를 차단하는 정책이다.

[그림 III-13] 방화벽 차단 로그 확인

방화벽의 정책 적용을 통해 발생할 수 있는 사이버 위협의 근원지라 판단되는 목적지를 차단 할 수 있었다. 이를 통해 [그림 III-10]과 같이 내부망에서 블랙리스트 IP를 목적지로 접속하는 통신 테스트가 차단된 것을 확인할 수 있다.

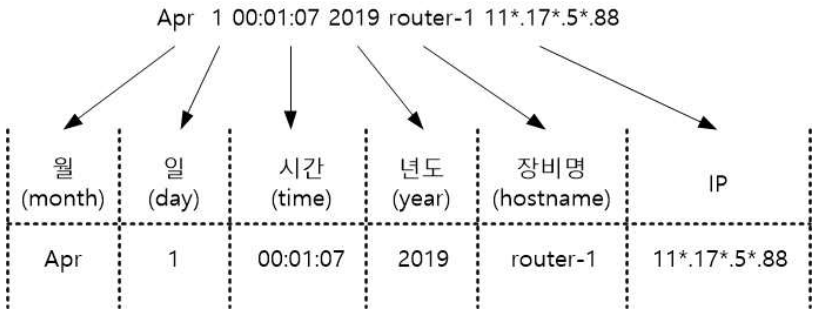
Ⅳ. 결론

1. 연구결과 요약

본 연구에서는 인터넷 라우터에 가해지는 SSH 무차별 공격을 로그를 통해 확인하고, 공격의 특징들을 분석하여 블랙리스트 IP를 선택하였다. 선택된 블랙리스트 IP를 통해 접근통제 정책을 수립하는 것을 목적으로 하였다. 로그를 통해 접근통제 정책을 수립하기 위해 인터넷 라우터에 지속적이고 악의적인 접근, 비인가자의 접근을 특정화, 비인가자 공격 로그 단편화의 가설을 세워 검증하였다. [표 Ⅳ-1]은 블랙리스트 IP 탐지를 위해 선행되어진 가설의 검증 결과를 요약한 결과이다.

[표 Ⅳ-1] 가설 검증 결과

<table>
<tr><td>구분</td><td colspan="6">가설 내용</td><td></td></tr>
<tr><td rowspan="8">가설1</td><td colspan="6">인터넷 라우터에 지속적이고 악의적인 접근 존재</td><td rowspan="8">존재</td></tr>
<tr><td>구 분</td><td>비인가/전체
로그수</td><td>비인가
접속률</td><td>구 분</td><td>비인가/전체
로그수</td><td>비인가
접속률</td></tr>
<tr><td>19년 4월</td><td>12343/12380</td><td>99.44%</td><td>19년 10월</td><td>17261/17323</td><td>99.64%</td></tr>
<tr><td>19년 5월</td><td>8926/9243</td><td>96.57%</td><td>19년 11월</td><td>30707/30782</td><td>99.76%</td></tr>
<tr><td>19년 6월</td><td>21072/21148</td><td>99.64%</td><td>19년 12월</td><td>33796/33854</td><td>99.83%</td></tr>
<tr><td>19년 7월</td><td>12689/12919</td><td>98.22%</td><td>20년 1월</td><td>102400/102530</td><td>99.87%</td></tr>
<tr><td>19년 8월</td><td>6527/7119</td><td>91.68%</td><td>20년 2월</td><td>36551/36595</td><td>99.88%</td></tr>
<tr><td>19년 9월</td><td>11939/11998</td><td>99.51%</td><td>20년 3월</td><td>13621/13742</td><td>99.12%</td></tr>
<tr><td>가설2</td><td colspan="6">비인가자 접근을 특정하여 탐지</td><td>가능</td></tr>
</table>

구분	가설 내용	
	“SSH user failed to log in from [공격자 IP 주소] on VTY0 due to IP restriction..” 의 특정 메시지 반복	
가설3	비인가자 공격 로그 단편화 반복되고 중복되는 메시지 제거 후 필요한 필드만 추출  	가능

우선 인터넷 라우터에 지속적이고 악의적인 접근이 존재한다는 가설은 로그 중 90% 이상이 비인가 접속을 통해 사실임을 알 수 있다. 비인가 공격은 동일 패턴의 메시지로 구성되어 있으며 “SSH user failed to log in from [공격자 IP 주소] on VTY0 due to IP restriction..″ 문구가 반복된다. 이를 통해 두 번째 가설인 비인가자 접근을 특정하여 탐지를 할 수 있다. 비인가자 접근은 SSH 무차별공격을 의미하는 메시지로 해석되며 중복되는 메시지를 제거하여 가설3과 같이 로그를 단편화 시킬 수 있다.

블랙리스트 IP를 탐지하기 위해서 단편화된 공격 로그의 특징들을 수치화 하여 특정 기준치 이상을 블랙리스트 IP로 선별하였다.

[표 IV-2] 블랙리스트 판단

IP	블랙리스트 판단 여부 $y(x)=x1\times x2\times x3$				블랙리스트 여부
	공격빈도수×0.01 ($x1$)	탐지 날짜×0.1 ($x2$)	국내×1,국외×1.1 ($x3$)	블랙리스트 Rating	
a ip	1	0.2	1.1	0.22	black-list
b ip	1.32	0.1	1	0.132	suspect

IP	블랙리스트 판단 여부 $y(x) = x1 \times x2 \times x3$				블랙리스트 여부
	공격빈도수×0.01 ($x1$)	탐지 날짜×0.1 ($x2$)	국내×1,국외×1.1 ($x3$)	블랙리스트 Rating	
c ip	8	0.1	1	0.8	black-list
월	정책 수	월	정책 수	월	정책 수
19년 4월	7	19년 8월	4	19년 12월	22
19년 5월	24	19년 9월	14	20년 1월	27
19년 6월	18	19년 10월	19	20년 2월	33
19년 7월	11	19년 11월	18	20년 3월	19

판단의 기준이 되는 공격 특징들은 공격의 빈도수, 공격 탐지 일수, 공격 국가로 정하고 빈도수×0.01, 탐지 일수×0.1, 국내×1 또는 국외×1.1로 수치화 하여 계산하였다. 이를 통해 국외에서 접속한 공격 중 탐지 일수가 2일 이상 100건의 공격 빈도수를 가진 IP(블랙리스트 Rating 0.22)를 기준으로 블랙리스트를 선별하였다. 월별 나오는 정책 수는 위 [표 IV-2]와 같으며 중복을 제거한 총 정책 126개를 방화벽에 접근통제 정책으로 설정하였다.

본 연구에서는 방화벽 차단정책 설정으로 내부에서 블랙리스트 IP를 목적지로 통신하는 모든 트래픽을 차단하였다. 차단에 활용된 방화벽 객체는 단일 IP이거나 같은 대역의 여러 IP가 있는 경우 그 대역(24bit subnet) 모두를 위험 IP 대역으로 처리한 CIDR이 있다. 5월 방화벽 정책을 수립 후 6월까지 총 46일간 방화벽에서 차단한 횟수는 모두 10,147건이며, 총 29개의 목적지 IP는 모두 중국이었다.

[표 IV-3] 방화벽 차단 로그

내부 IP	차단 수	차단 비중	내부 IP	정책 수	비고
A IP	5,466	53.87%	E IP	45	0.44%
B IP	4,236	41.75%	F IP	30	0.30%
C IP	285	2.81%	G IP	30	0.30%
D IP	54	0.53%	H IP	1	0.01%

[표 IV-3]은 내부 IP별 방화벽 차단 수를 나타내며, 가장 높은 빈도수를 가진 A, B IP의 경우 전체의 90%가 넘는 접속을 보인다. 하지만 내부 A, B IP의 경우 조사를 통해 해외관련 업무를 진행하고 있는 부서였음을 확인했고, 해당 목적지 주소가 블랙리스트 IP와 동일 대역에 있어 차단되어 있었다. 따라서 CIDR 처리된 객체에 차단된 내부 IP는 추가적인 확인 작업이 필요하며 그 의도가 불분명한 경우 차단을 지속해야 한다. 또한 위험성을 고려한 24bit subnet을 기준으로한 CIDR의 경우도 범위를 좁혀 객체를 등록할 필요가 있다.

2. 연구의 한계와 향후 연구과제

IT 인프라의 관리자는 중요한 정보 자산과 많은 수의 시스템을 관리한다. 이 과정에서 생성되는 많은 데이터를 분석해야 하는데 정해진 표준이 없어 경험의 의존하는 경우가 많다. 숙련된 관리자는 어떤 현상을 보고 직감적으로 문제가 있다고 판단하여 그 부분을 면밀히 살피기도 하지만 대부분은 시스템에서 발생하는 방대한 양의 로그와 그 로그가 의미하는 문제의 근원을 찾기 위해 많은 시간과 노력을 소비한다. 특히 중요한 정보자산을 보호하는 시스템에서 탐지되는 보안 위반의 징후들은 빠른 조치가 필요하며 그에 맞는 예방도 필요하다.

본 연구는 인터넷 라우터에 생성되는 1년간의 로그를 분석하였다. 로그의 대부분은 SSH 무차별 대입 공격이었으며, 이를 가공하여 블랙리스트 IP를 정하고 접근통제 정책을 수립하였다. 하지만 이런 접근통제 정책은 실시간으로 가해지는 사이버 위협에 대해서는 대처가 어렵다. 중요한 정보자산 대한 공격은 상시 분석을 통해 실시간으로 접근통제가 필요하기 때문에 로그를 실시간 수집하고 분석을 통해 시각화하여 위협을 방지해야 한다.

실시간 분석을 위해서는 저장된 로그를 분석하는 것보다 실시간 로그를 통해 빠른 판단이 필요하다. 이를 위해 사람의 판단보다는 머신러닝을 통한 자동화로 분석의 성능을 높여야 하며, 머신러닝 학습을 위한 특징들은 더욱 세분화하여 분석의 정확성을 높여야 한다. 본 연구에서는 블랙리스트 분류를 위한 특징들을 공

격 빈도수, 날짜별 접근, 접속 국가로 한정하여 가중치를 주어 일정기준이상을 블랙리스트로 구분하였다. 하지만 실시간 분석에서 이런 특징들로 블랙리스트를 탐지하기에는 정해진 시간이 필요하여 실시간 분석이라 하기 힘들다. 따라서 실시간 분석에 필요한 특징들이 더 요구되며 저장되는 로그와 실시간으로 생성되는 로그의 연관 분석도 필요할 것이다. 또한 IT 인프라의 최상단에 위치한 인터넷 라우터와 직접 연결은 보안상의 문제를 발생시킬 수 있어 실시간 로그를 받기 위해서는 방화벽으로 보호받는 신뢰 구간인 DMZ에서 syslog의 접속만 허용하는 시스템을 구축하고 분석을 진행하여야 한다. 이처럼 실시간 분석을 위한 특징들의 추가 분석과 시스템 구축을 위한 물리적인 위치를 극복한다면 접근통제의 객체로 충분한 가치가 있는 인터넷 라우터의 로그를 활용하여 향후 많은 분석이 가능할 것이다.

IT 인프라의 관리자는 정보보안과 관련된 많은 데이터를 분석함에 있어 정해진 표준이 없어 경험에 의존하는 경우가 많았다. 본 연구에서는 인터넷 라우터 장비에만 발생하는 SSH 무차별 대입 공격이라는 로그를 특정하여 블랙리스트로 만들고 접근통제 정책을 수립하였다. 악의적인 접속 근원지에 대한 차단은 내부에서 발생 할 수 있는 유출의 가능성을 막을 수 있고, 지속적인 접근통제 정책 추가로 인해 생길 수 있는 위협을 낮추어 보안을 강화 할 수 있을 것이다.

참 고 문 헌

[1] 김정욱, “악성도메인 IP 주소 추적을 이용한 효과적인 보안관제 방안 연구”, 2009., 고려대학교 정보경영공학전문대학원

[2] 김태훈, “오픈소스 ELG Stack을 利用한 效率的인 保安管制 方案 硏究”, 2020., 성균관대학교 정보통신대학원

[3] 노정호, “사이버테러에 대한 전조 현상 분석 및 IP 역추적 연구”, 2015., 호서대학교 벤처전문대학원

[4] 문병우, “IoT HoneyPot을 이용한 IoT 악성코드 수집 및 분석”, 2018., 순천향대학교 일반대학원

[5] 문성주, “빅 데이터 기반의 네트워크 로그 분석 및 예측 시스템 설계 및 구현”, 2018., 원광대학교 일반대학원

[6] 유영태, “사이버 범죄에 이용되는 IP주소 수집 및 활용에 관한 연구”, 2015., 고려대학교 정보보호대학원

[7] 이재국, 김성준, 우준, 박찬열, “다중 사용자 컴퓨팅 환경에서 SSH 무작위 공격 분석 및 대응”, 2015., 정보처리학회논문지 컴퓨터 및 통신시스템 4(6): pp.205-212.

[8] 임덕기, “PF(Packet Filter)를 이용한 네트워크 보안성 향상을 위한 연구″, 2009., 동국대학교 국제정보대학원

[9] 전두용, “Black Eye: IP Blacklisting via Threat Feature Extraction and Machine Learning from Security Logs″, 2019., 경북대학교 산업대학원

[10] 전두용, 탁병철, “반복적인 선형회귀분석을 이용한 대용량 보안로그의 블랙리스트 IP 분류″, 한국정보과학회 학술발표논문집 2018(12), 2018., pp.1020-1022.

[11] 주승현, “정보보호관리체계(ISMS)를 이용한 네트워크 인프라 보안 개선 방안 연구″, 2018., 건국대학교 정보통신대학원

[12] 최준용, “실제 대용량 공격 로그의 직접적 분석으로 밝힌 공격형태”, 2015., 서울대학교 대학원

[13] 한성건, “네트워크 인프라보안 관리 개선을 통한 정보보호 강화 연구”, 2016., 동국대학교

[14] 현정훈 (2018), “오픈소스 ELK Stack 활용 정보보호 빅데이터 분석을 통한 보안관제 구현”, 고려대학교 정보보호대학원.

[15] 홍성대, “Sysmon과 ELK Stack를 이용한 윈도우시스템 사이버 위협 탐지 및 가시성 증대에 관한 연구”, 2020., 동국대학교 대학원

[16] Jae-Kook L, Sung-Jun K, Chan Yeol P, Taeyoung H and Huiseung C, “Heavy-Tailed Distribution of the SSH Brute-Force Attack Duration in a Multi-user Environment”, THE JOURNAL OF THE KOREAN PHYSICAL SOCIETY 69(2): 253-258., 2016

[17] Jeon D, “IP Blacklisting via Threat Feature Extraction and Machine Learning from Security Logs”, 2019., The Graduate School of Industry Kyungpook National University.

[18] Lee J-K, Kim S-J, Woo J and Park CY, “Analysis and Response of SSH Brute Force Attacks in Multi-User Computing Environment”, KIPS Transactions on Computer and Communication Systems 4(6): 205-212., 2015

[19] Najafabadi MM, Khoshgoftaar TM, Calvert C and Kemp C, “Detection of SSH Brute Force Attacks Using Aggregated Netflow Data”, In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pp. 283-288., 2015

[20] Najafabadi MM, Khoshgoftaar TM, Kemp C, Seliya N and Zuech R “Machine learning for detecting brute force attacks at the network level”, In 2014 IEEE International Conference on Bioinformatics and Bioengineering, IEEE, pp. 379-385., 2014

[21] 국가정보원, “안전한정보통신환경구현을위한네트워크구축가이드라인”, 2013

[22] 정보보호정책연구소, “정보보호(산업) 동향보고서”, 2019

[23] 한국인터넷진흥원, “2019년 4분기 사이버 위협 동향 보고서”, 2019

[24] 한국인터넷진흥원, “주요정보통신기반시설 기술적 취약점 분석 평가 상세

가이드”, 2017

[25] 한국정보통신산업연구원, “정보통신산업동향 제25권”, 2018

[26] CISCO “2020 글로벌 네트워킹 트렌드 보고서”, 2019

[27] IBM “인공지능을 이용한 애자일 인프라 전략”, 2019

[28] TTA “정보시스템 하드웨어 규모 산정 지침″, 2018