J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr


# BGP 데이터셋 분석 및 CyCOP 가시화 방안 연구☆

# Research on BGP dataset analysis and CyCOP visualization methods

정 재 영[1,2] 김 국 진[1,2] 박 한 솔[1,2] 장 지 수[1,2] 신 동 일[1] 신 동 규[1,2,*]
Jae-yeong Jeong Kook-jin Kim Han-sol Park Ji-soo Jang Dong-il Shin Dong-kyoo Shin

**요 약**

기술이 발전함에 따라 인터넷 사용량은 더욱 증가하고 있으며, 이를 통해 네트워크 트래픽 및 통신량이 기하학적으로 증가하고 있다. 인터넷의 핵심 요소 중 하나인 네트워크 경로 선택 프로세스는 이에 따라 더욱 복잡하고 고도화되고 있으며, 이를 효과적으로 관리하고 분석하는 것이 중요하다. 또한 이를 직관적으로 이해할 수 있는 표현 및 가시화 방안이 필요하다. 이를 위해 본 연구에서는 네트워크 경로 선택 방법인 BGP를 사용하여 네트워크 데이터를 분석하고, 상황인식을 위한 사이버 공통작전상황도에 이를 적용하는 프레임워크를 설계한다. 그 후 정보들을 가시화하기 위해 필요한 작전상황도 가시화 요소들을 분석하고 간단한 가시화를 구현하는 실험을 진행한다. 실험에서 수집되고 전처리된 데이터를 기반으로 구현된 가시화 화면들을 통해 지휘관 또는 보안 담당자들이 네트워크 상황을 효과적으로 파악하고 명령 및 제어를 할 수 있도록 도움을 준다.

☞ 주제어 : BGP, CyCOP 프레임워크, 머신러닝, 사이버 3계층, 이상징후 탐지

**ABSTRACT**

As technology evolves, Internet usage continues to grow, resulting in a geometric increase in network traffic and communication volumes. The network path selection process, which is one of the core elements of the Internet, is becoming more complex and advanced as a result, and it is important to effectively manage and analyze it, and there is a need for a representation and visualization method that can be intuitively understood. To this end, this study designs a framework that analyzes network data using BGP, a network path selection method, and applies it to the cyber common operating picture for situational awareness. After that, we analyze the visualization elements required to visualize the information and conduct an experiment to implement a simple visualization. Based on the data collected and preprocessed in the experiment, the visualization screens implemented help commanders or security personnel to effectively understand the network situation and take command and control.

☞ keyword : BGP, CyCOP Framework, Machine Learning, Cyber 3 Layer, Anomaly Detection

## 1. 서 론

인터넷은 현대 사회에서 필수적인 통신 수단이 되었으며, 이로 인해 인터넷 기반의 서비스 및 응용프로그램의 중요성은 나날이 증가하고 있다. 이와 함께, 인터넷의 핵심 요소 중 하나인 네트워크 경로 선택 프로세스는 더욱 복잡해지고 있으며, 이를 효과적으로 관리하고 분석하는 것이 중요하다. BGP(Border Gatway Protocol)는 전세계 인터넷의 경로 선택에 사용되며, 신뢰성과 보안 측면에서 중요한 역할을 한다. BGP 데이터셋을 통해 경로 선택 프로세스를 이해하고 개선하기 위한 풍부한 정보를 얻을 수 있다. 하지만 대용량의 BGP 데이터를 효과적으로 분석하고 이를 토대로 네트워크 성능 최적화를 수행하는 것은 매우 어렵다. BGP 데이터셋을 활용하여 네트워크 상태를 모니터링하고 관리하는데 도움이 되는 새로운 접근 방식을 탐구하며, 이를 위해 CyCOP 프레임워크를 설계한다. CyCOP(Cyber Common Operational Picture)은 고급 분석 및 보안 조치를 지원하는 효율적인 도구로, 본 연구에서는 CyCOP을 BGP 데이터셋 분석에 적용하여 네트워크 안정성을 향상시키고 효율성을 향상하는 방법을 제시한다.

1 Dept. of Computer Science, Sejong University, Seoul, 05006, Korea
2 Dept. of Convergence Engineering for Intelligent Drones, Sejong University, Seoul, 05006, Korea
* Corresponding author (shindk@sejong.ac.kr)
[Received 15 October 2023, Reviewed 24 October 2023(R2 14 November 2023), Accepted 15 November 2023]
☆ 본 연구는 2023년 국방과학연구소에서 주관하는 미래도전국방기술 연구개발사업(915092101)의 지원을 받아 수행되었습니다.

## 2. 관련 연구

### 2.1 사이버 공통작전 상황도(CyCOP)

COP(Common Operational Picture)는 특정 작전 환경에 대한 정보들이 통합된 시각적 표현을 말한다. 군사분야에서 지휘관이 다양한 정보를 통해 상황을 포괄적이고 적절하게 파악하는 SA(Situational Awareness) 가 필수적인데, COP은 보다 적절하고 신속한 의사 결정을 하는데 도움을 주는 역할을 한다 [1-8].

CyCOP(Cyber Common Operational Picture)은 사이버공간에서 실행되는 작전에 COP를 적용한 것으로, 사이버 환경을 시각화하거나 표현하여 지휘관의 네트워크 및 시스템 및 사이버 위협의 상태를 실시간으로 이해할 수 있도록 한다. 이러한 CyCOP의 시각화를 통해 상황인식을 쉽게 할 수 있다면 이는 명령 및 제어를 위한 효과적인 시스템이라고 할 수 있다 [9-12].

#### 2.1.1 사이버 3계층 정의

사이버 작전은 사이버공간에서 수행되며, 공간을 구성하는 요소에 따라 사이버공간은 3개의 계층으로 분류되며, 각 계층은 다음과 같이 기술할 수 있다 [13-14].

1) 물리적 네트워크 계층: 현실에 부피와 질량을 가지고 실재하는 요소들을 포함하는 계층이다. 라우터, 컴퓨터, 휴대폰 등이 이에 해당된다.
2) 논리적 네트워크 계층: 물리적 네트워크 계층이 사용하는 네트워크 구성요소와 네트워크 노드 사이에 존재하는 논리적 연결 등을 포함하는 계층이다. 응용 어플리케이션 및 OS 등이 이에 해당된다.
3) 페르소나 네트워크 계층: 사이버공간 내에서 업무를 수행하거나 수립하는 사용자의 정보 등을 포함하는 계층이다. 사람 이름이나 이메일, 또는 비밀번호나 인증서 같은 민감 정보등이 이에 해당된다.

#### 2.1.2 작전환경 분석

美, ATP(Army Techniques Publication) 2-01.3 IPB (Intelligence Preparation of the Battlefield) [15] 에는 작전에 미치는 영향을 결정하기 위해 관심 지역의 적, 지형, 날씨 및 시민 고려사항의 임무 변수를 분석하는 체계적인 절차가 기술되어 있다. 절차의 목적은 적의 의도를 파악하는 것이다. 각 절차마다 가시화해야 할 투명도 및 템플릿 예제를 제공하고 있으며, 사이버공간에서 전장정보분석 절차는 아래와 같다.

1) 작전환경 (OE: Operation Environment): 작전지역 내에서 사이버공간의 3계층(JP 3-12 [13]) 구성요소 및 위협의 현재 물리적 위치를 그래픽으로 가시화한다. 사이버공간 3계층 식별사항은 표 1과 같다.

(표 1) 사이버공간의 계층별 식별 사항
(Table 1) Identifying layers of cyberspace

| 사이버 계층 | 식별 사항 |
|---|---|
| 물리 계층 | 사이버 네트워크 노드 및 네트워크 장치 (PC, 노트북, 서버, 라우터, 지능형 허브 등), IDS/IPS 등 |
| 논리 계층 | Software, 취약점 정보, 파일 시스템, 서버자원, 운영체제, 웹사이트 등 |
| 페르소나 계층 | 집단이나 조직, 사용자 정보, 개인키, 해시 값, 계정 비밀번호 등 |

2) 작전에 미치는 환경적 영향 설명: 수정된 종합 장애물 투명도(MCOO: Modified Combined Obstacle Overlay)는 사이버공간 3계층을 반영하여 가시화한다. 이는 인터넷을 사용할 수 있는 외부망과 인터넷을 사용하지 않는 내부망(폐쇄망)으로 구분된다. 그리고 외부망과 내부망을 잇는 접점(방화벽) 등을 고려하여 가시화한다.
3) 위협평가: 위협 특성 최신화, 위협 모델 생성, 광범위한 위협 대응 방책 개발, 고가치표적의 식별이 되어야한다. 또한 사이버공격 구조 및 공격자의 과거 패턴 분석 시, 위협에 대한 상황을 이해하는데 도움이 되어야한다. 그리고 위협이 선호하는 내부 이동 공격 기법, 위협 요소에 사용되는 모든 멀웨어(Malware)를 볼 수 있어야 하며, 위협이 작전 또는 지점을 수행할 수 있는 능력에 중요 자산(고가치표적)을 식별해야 한다.
4) 위협 대응 방책 결정: 위협에 대한 대책을 선택할 때는 예상되는 행동(경로)을 시각적으로 표현해야 한다. 환경 영향을 고려하고 특정 방어 조치를 활성화하는 위협을 묘사하기 위해 MCOO를 계층화해야 한다.

#### 2.1.3 CyCOP 가시화

사이버공간에서 이루어지는 행동들은 빠르고 신속하게 이루어지므로 이를 대응하기 위해서 빠른 상황인식이

필요하다. 또한 직관적으로 상황을 인식하기 위한 쉬운 가시화가 필요하다.

K.J, Kim et. al[8] 은 시스템에서 일어날 수 있는 항목과 실시간 응답 시스템의 실행항목을 분석하여 단순, 복잡 응답시간의 정의를 내리고 실험을 통해 CyCOP 응답시간을 측정하였다.

CHO, Sungyoung, et [16]은 사이버 작전을 분석하고 사이버 킬체인 모델을 CyCOP에 적용하여 간단한 가시화를 구현하였다.

ESTEVE, Manuel, et al [17] 은 지휘관의 상황인식에 기반한 지휘 통제용 정보 시스템을 위해 CyHSA(Cyber Hybrid Situational Awareness) 생성을 목표로 CyCOP을 개발하고 도구에 아키텍처에 대한 접근 방식을 제안하였다.

T. Pahi, et al. [18] 은 CyCOP을 지원하는 효율적인 정보 관리 프로세스에 필요한 정보 유형과 출처에 대해 설명하고, 관리 프로세스의 적용 뿐만 아니라 국가 의사 결정에 이르는 전체 프로세스의 예시를 들어 CyCOP이 지원하는 잠재적 의사 결정에 대해 설명한다.

L. Jiang, et al. [10] 은 사이버 상황 인식(CSA) 시각화에 관련된 논문을 분석하고 데이터를 추출하여 정보 유형, 데이터 소스 및 시각화 기법을 조사하였다. 대부분의 논문이 위협 시각화에 초점을 두고 있으며, 공유된 정보를 시각화 하는 것이 부족하다고 말하고 있다.

## 2.2 네트워크 데이터 수집 및 분석

### 2.2.1 BGP(Border Gateway Protocol)

컴퓨터 네트워크의 AS(Autonomous System)는 인터넷을 공통 라우팅 정책을 통해 관리하는 단일 조직 또는 관리자의 제어 하에 있는 네트워크 단말 노드 및 라우터의 집합이다. AS들은 라우팅 및 도달 가능성 정보 교환을 용이하게 하기 위한 핵심 라우팅 프로토콜 BGP(Border Gateway Protocol)를 사용한다. 단일 네트워크 내에서 작동하는 IGP(Interior Gateway Protocol)와 달리 자율 시스템이 인터넷에서 데이터 트래픽을 라우팅하는 방법에 대해 정보에 입각한 결정을 내릴 수 있도록 지원한다. 네트워크 관리자는 경로 길이, 네트워크 접두사, AS 경로와 같은 다양한 속성을 기반으로 라우팅 정책을 정의하여 라우팅 결정을 세밀하게 제어할 수 있다.

BGP 데이터는 RFC 6396에 기술되어 있는 MRT (Multi-threaded Routing Toolkit) 형식으로 저장되며, 그 중 MRT Type 필드의 값에 따라 OSPF, TABLEDUMP, BGP4MP등으로 구분한다. BGP 데이터 중 네트워크 연결정보를 파악하기 위해서 RIB(Routing Information Base) 데이터와 Update 데이터를 사용하여 분석을 진행한다. RIB 데이터는 라우팅 프로토콜을 토대로 목적지 네트워크에 대한 경로 정보가 존재하는 데이터다. Update 데이터는 BGP 피어 간의 라우팅 정보를 전달하는데 사용하는 데이터로, 경로 광고 또는 철회 업데이트 메시지를 주로 전송한다.

### 2.2.2 BGP 수집처

BGP 데이터는 공개된 저장소에서 수집하며, 잘 알려진 2개의 저장소인 Route Views[19], RIPE NCC RIS[20]를 사용한다.

Route Views Projcect는 Oregon 대학에서 호스팅하는 프로젝트로, 전 세계에 위치한 수집기에서 배포된 BGP 라우터 세트를 제공한다. UTC (Universal Time Coordinated) 0시를 기준으로 2시간을 주기로 RIB 파일을 업로드 하며, 15분 주기로 라우팅 정보 update 내역을 업로드한다.

RIS(Routing Information Service)는 RIPE NCC (Réseaux IP Européens Network Coordination Centre)에서 관리하는 프로젝트이다. RIPE NCC는 유럽 및 주변 지역에 서비스를 제공하는 RIR(Regional Internet Registry)이다. RIS는 Route Views와 같이 전 세계 여러 지역에서 BGP 라우팅 정보를 수집하며 rib 및 update 파일을 UTC 0시를 기준으로 각각 8시간, 5분 주기로 업로드한다.

## 2.3 인공지능을 사용한 이상징후 탐지

### 2.3.1 BGP 분석 기법 분류

B. Al-Musawi et. al [21]은 BGP 이상징후 분석에 사용되는 20여 가지의 기법에 관해서 서술하였다. 이상징후를 Direct intended anomaly, Direct unintended anomaly, Indirect anomaly, Link failure의 4가지로 분류하였으며, 분석기법은 탐지에 사용되는 기법, 다른 형태의 이상징후를 알아내는 능력, 사용된 데이터 소스, 발견된 BGP 속성, 이상징후의 원천을 알아내는 능력의 5가지 측면을 기반으로 모두 5종류의 기법들로 분류하였다.

C. Yang et al. [22]은 경로 기반 BGP 라우팅 알고리즘을 제시하면서, 해당 라우팅 알고리즘을 사용하여 하이재킹 공격을 자동으로 검출하는 방식에 대한 연구 결과를 제시하였다.

#### 2.3.2 기계학습 기반의 분석

M. Cheng et al. [23]는 시계열 데이터 분석에 특화된 결과를 보여주는 Long Short-Term Memory (LSTM)을 개량하여 multi-scale LSTM 알고리즘을 제안하고 해당 실험 결과를 서술하였다. 해당 논문에서는 SVM, Naive Bayes (NB)와 AdaBoost와의 비교 결과에서 각각 76.9, 51.9, 83.8의 정확도를 보여주었고, 제시한 알고리즘은 86.8%로서 가장 우수한 결과를 보여주어, 시계열 데이터의 특성을 고려한 알고리즘이 좋은 결과를 보여준다는 연구 결과를 제시하였다.

A. H. Muosa et al. [24]은 LSTM와 오토인코더를 결합한 알고리즘을 이용하여 BGP 이상징후를 분석하였다. LSTM을 각각 오토인코더의 인코더/디코더로 사용하여 오토인코더 알고리즘을 개량하여 실험한 결과를 보여주고 있다. 해당 논문에서는 모두 11개의 이상징후 데이터셋을 이용한 실험을 진행하였다.

Q. P. Nguyen et al. [25]은 비교사 학습 방식의 딥러닝 알고리즘에 해당하는 변분 오토인코더 (VAE: Variational autoencoder)을 이용해서 이상징후를 알아 내고, 그래디언트 기반 지문 기법 (gradient-based fingerprinting technique)를 이용하여 해당 이상징후를 설명하는 시스템에 대한 연구결과를 제시하고 있다. 해당 논문에서는 Gaussian Based Thresholding (GBT) 기법과 일반 오토인코더에 대한 실험 결과를 함께 제시하고 있으며, 각각 90.6% 및 90.2%의 AUC-ROC 결과를 보여주었다. 최종적으로 VAE는 96.2%의 수치를 보여주어 가장 우수한 결과를 보여주었다.

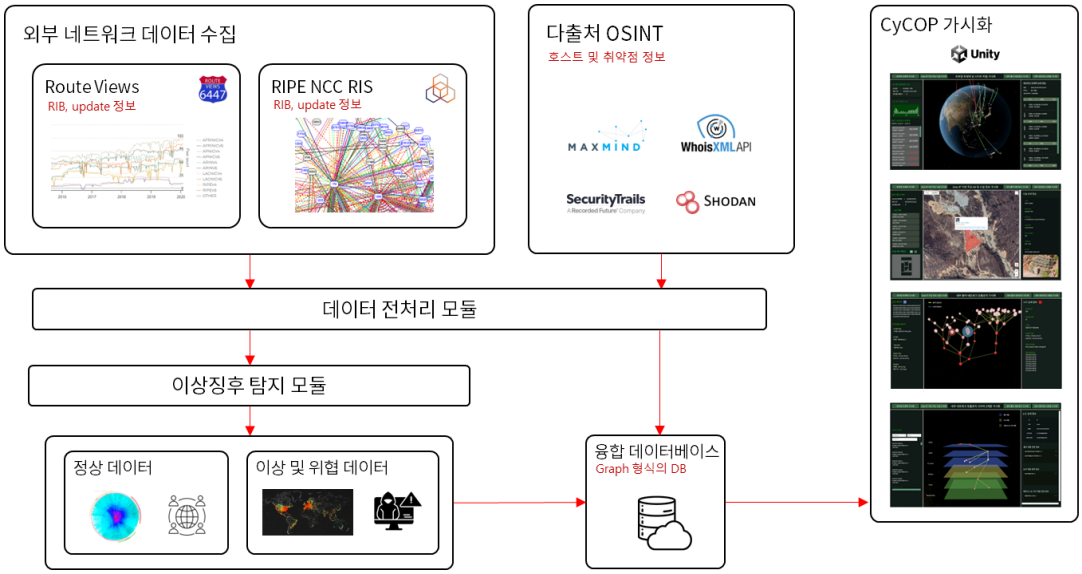
K. Hoarau et. al. [26]은 Graph Neural Network(GNN)을 적용한 알고리즘을 BGP 이상징후를 분석하였으며, 실험 결과에 따르면 GNN 레이어가 4개일 경우에 약 96%의 정확도를 보여주며 가장 우수한 결과를 나타냈다.

## 3. CyCOP 프레임워크 설계 및 구현

### 3.1 네트워크 데이터 수집 및 전처리

#### 3.1.1 BGP 데이터 수집

Route Views와 RIPE NCC RIS 두 개의 공개된 BGP 데이터 저장소에서 네트워크 데이터를 수집한다. Route Views의 경우 데이터를 bz2 확장자의 압축 파일로 저장하며, 평균 데이터 용량은 rib 파일은 101MB, update 파일의 경우 2.2MB에 해당한다. RIS의 경우 gz 확장자의 압축 파일로 데이터를 저장하며, rib 파일은 414MB, update 파일은 5.6MB의 용량을 가지고 있다.




(그림 1) CyCOP 프레임워크 구조 설계
(Figure 1) Designing the CyCOP Framework Structure

본 연구에서 RIB 데이터 중 BGP 라우팅 테이블 정보를 담는 TABLEDUMP 타입의 데이터를 사용하며, 데이터 필드는 표 2와 같다. 수집된 데이터는 0과 1로 이루어진 이진 형태의 파일로 저장되어 있으므로, 사람이 읽을 수 있는 형태로 변환한다.

(표 2) TABLEDUMP의 데이터 필드
(Table 2) Data Field of TABLEDUMP

| 필드명 | 크기 | 필드 설명 |
|---|---|---|
| Time | 32 | 데이터 수신 시간 |
| Peer IP | 32 or 128 | BGP 라우팅 정보를 교환하는 IP 주소 |
| Peer AS | 16 | BGP 라우팅 정보를 교환하는 AS 번호 |
| Prefix | 8 | BGP 테이블에 저장된 대상 네트워크 주소 |
| AS Path | variable | 데이터가 교환되며 지나온 AS 번호의 집합 |
| Next Hop | 32 or 128 | 해당 데이터의 다음 목적지 정보 |
| Origin | 8 | 라우팅 정보의 전송 범위 (EGP, IGP, Incomplete) |

### 3.1.2 다출처 OSINT 수집

네트워크 데이터 수집기를 통해 수집된 BGP 데이터는 물리 계층의 정보 중 네트워크 정보만을 포함하고 있어 논리 및 페르소나 계층의 데이터를 추가적으로 수집하는 방안이 필요하다. CyCOP을 통해 더 정확하고 상세한 정보를 제공하기 위해 OSINT(Open Source Intelligence)를 사용한다. OSINT란 공개적으로 사용 가능한 데이터를 가공, 분석 및 획득한 지능형 정보이다.

특정 지식을 전달하여 행동과 의사결정에 사용할 수 있도록 하며, 손쉬운 접근성과 속도, 다양한 주제의 거대한 정보, 낮은 수집 비용 등의 장점이 있다[27-28].

수집한 BGP 데이터의 IP 주소 및 AS 번호 등의 필드를 사용하여 다양한 출처의 API를 사용하여 데이터를 추출한다. 물리 계층의 지리적 위치 정보를 수집하기 위해 MaxMind사에서 제공하는 GeoIP2 Database[29]를 사용한다. 해당 정보와 BGP Attribute 필드의 Path 값을 연동하여 GeoJSON LineString 데이터를 생성하여 네트워크의 연결도를 생성한다. GeoJSON은 위치 정보를 갖는 점을 기반으로 지형을 표현하는 공개 표준 형식으로, IETK(Internet Engineering Task Force)에서 관리한다는 점에서 다른 표준인 GIS(geographic information system)와의 차이점을 가진다. JSON(JavaScript Object Notation)의 형식을 사용하며 데이터 레이어 통합에 사용한다. ASN(Autonomous System Number)의 정보의 정확도를 높이기 위해 CAIDA(the Center for Applied Internet Data Analysis) ASRank DB[30]에 GraphQL을 사용하여 AS의 정보를 추가한다.

논리 및 페르소나 계층의 데이터를 수집하기 위해 SecurityTrails, WhoisXMLAPI, Maxmind, Shodan, Censys 의 A

(표 3) 다출처 OSINT 수집 데이터
(Table 3) Multi-Source OSINT Ingestion Data

| 수집처 | API | 수집정보 |
|---|---|---|
| SecurityTrails | DNS Record | IP주소, 레코드종류, 호스트이름 |
| | Reverse IP lookup | 도메인이름, 호스팅공급자, 메일공급자, 사이트개수, 연결된IP 주소 |
| WhoisXMLAPI | WHOIS | 도메인생성/수정/만료날짜, 등록기관정보(이름, 주/도시, 국가), 관리자정보(전화번호, 이메일, 이름, 소속기관명), 호스트이름, 네임서버, 도메인가용여부, 최상위도메인 |
| | IP Geolocation | 위치(국가, 위/경도, 표준시간, geonameID, postal), 관리AS(ASN, 라우팅정보, 도메인), ISP |
| | Email verification | 도메인, 이메일주소, 이메일형식, smtp 사용여부, mx 레코드, 메일감사(활성화여부, 생성/수정시간) |
| | SSL Certificates | 접근포트번호, 인증서정보(발행기관, 시리얼번호, 인증알고리즘, pem, 공개키) |
| MaxMind | City | IP 주소의위치정보(위/경도), 국가정보(대륙, 이름, geonameID), 도시정보(주/시정보, postal, zip), 위치범위(Accuracy Radius) |
| | AS, ISP | ASN, AS 이름, 관리기관정보(이름, 기관번호, 주소등), ISP(이름, 모바일네트워크번호) |
| Shodan | Networks | IP주소, 운영체제, 도메인이름, 위치정보, 서비스세부정보(개방포트, 사용프로토콜, 버전), 보안관련정보(SSL/TLS 인증서, 서비스배너, 취약프로토콜및S/W) |
| Censys | Hosts | IP 주소, 도메인이름, 위치정보, 서비스세부정보(웹/메일서버, FTP 서버, 프로토콜및버전), 서비스배너, 응답헤더, 연관메타데이터 |
| | Certificates | SSL/TLS 인증서, SSH 키 |

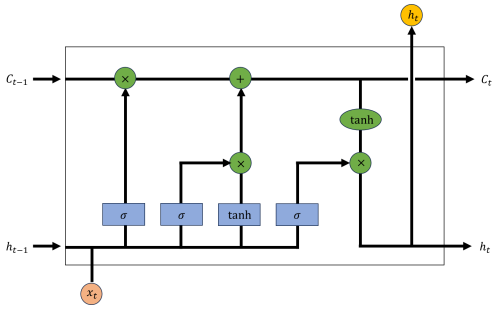
PI를 사용하며, 각 계층의 데이터에 포함되는 내용만을 수집하고 중복되는 내용은 비교를 통해 데이터를 최신화하고 보다 정확하게 보강한다. 수집처별 API 및 수집 가능한 정보는 표 3과 같다.

## 3.2 수집 데이터 이상탐지 및 분석

이상탐지 및 분석을 위해 특징 추출(Feature Extracition)을 진행한다. BGP 데이터는 Volume(number of BGP Announcements)과 As-Path(maximum edit distance)로 분류한다 [21]. Volume의 NLRI(Network Layer Reachability Information)와 OSINT 수집을 통해 얻은 데이터를 연결하여 최종적으로 데이터 분석에서 사용하는 feature는 표 4와 같다. 데이터 분석을 위해 RNN(Recurrent Neural Network)의 한 종류인 LSTM(Long Short-Term Memory)을 사용한다.

LSTM은 시계열 데이터나 연속적인 데이터 패턴을 학습하는데 유용하기 때문에 BGP 데이터를 분석하기에 적합하다고 할 수 있다. LSTM의 구조는 그림 2와 같으며, RNN처럼 신경망 모듈이 반복되는 체인의 형태를 가지지만, 이 신경망 모듈이 다른 구조를 가진다는 특징이 있다. 은닉층의 계산식이 복잡하고 셀 상태(cell state)라는 값을 추가하여 사용한다. LSTM을 사용하여 BGP 데이터를 분석할 경우 얻을 수 있는 정보는 다음과 같다.

1) 이상 탐지: BGP 경로 변화나 BGP 업데이트 패턴에서 이상한 동작을 탐할 수 있다. 시계열 패턴의 일반적인 특성을 학습할 수 있으므로 예상치 못한 변화나 비정상적인 행동을 탐지하는 데 유용하다.
2) 경로 예측: BGP 업데이트의 미래 패턴이나 AS 경로의 변화를 예측할 수 있다.
3) 정책 분석: BGP 업데이트의 변화나 적용이 BGP 경로 변화에 어떻게 영향을 미치는지 분석할 수 있다.




(그림2) LSTM 모델의구조
(Figure 2) Structure of LSTM Model

(표 4) 특징 추출을 거친 Features
(Table 4) Extracted Features

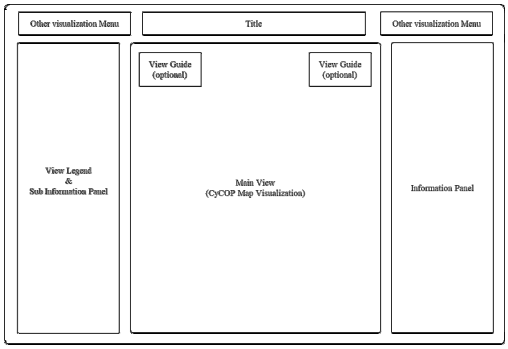
| no | Definition | Category |
|---|---|---|
| 1 | Number of announcements | volume |
| 2 | Number of withdrawals | volume |
| 3 | Number of announced NLRI prefixes | volume |
| 4 | Number of withdrawn NLRI prefixes | volume |
| 5 | Average AS-PATH length | path |
| 6 | Maximum AS-PATH length | path |
| 7 | Average unique AS-PATH length | AS-path |
| 8 | Number of duplicate announcements | volume |
| 9 | Number of duplicate withdrawals | volume |
| 10 | Number of implicit withdrawals | volume |
| 11 | Average edit distance | volume |
| 12 | Maximum edit distance | AS-path |
| 13 | Inter-arrival time | volume |
| 14-24 | Maximum edit distance = n, where n = (7, ..., 17) | AS-path |
| 25-33 | Maximum AS-path length = n, where n = (7, ..., 16) | AS-path |

## 3.3 CyCOP 가시화

관련 연구를 통해 도출된 반응속도 및 가시화 객체 관련 내용들을 바탕으로 CyCOP 인터페이스를 설계한다. 다만 CyCOP의 가시화 화면들은 표현하고자 하는 주제나 의도, 사용하는 데이터의 특성에 따라 조금씩 형태가 변화한다. 화면 내에 나타나는 객체의 경우 美, MIL-STD-2525D [31]의 기준을 준수한다. 객체 크기는 美, MIL-STD-1472H [32]를 바탕으로 13 pixel 크기를 기준으로 하여 설계 및 구현한다. 이를 통해 설계된 UI는 그림 3과 같이 나타낼 수 있다. 가시화 화면 간의 연동을 위해 추가적으로 화면 전환을 위한 버튼을 구성한다. 위의 내용을 기반으로 CyCOP을 구현하며, 구현에 사용한 하드웨어 및 소프트웨어 정보는 표 5와 같다.

(표 5) CyCOP 구현 환경
(Table 5) CyCOP Implementation Environment

| 구 분 | 사용 하드웨어 및 소프트웨어 |
|---|---|
| OS | Windows 11 Pro 64 bit |
| CPU | AMD Ryzen 7 3700X |
| Memory | 65536MB |
| VGA | NVIDIA GeForce RTX 2070 SUPER 8GB |
| Development Languages | Python 3.11.4, C# 7.3 |
| S/W | Unity 2022.3.9, Google Maps, 3D WebView for Windows, Oracle 21c 1.0.0 |

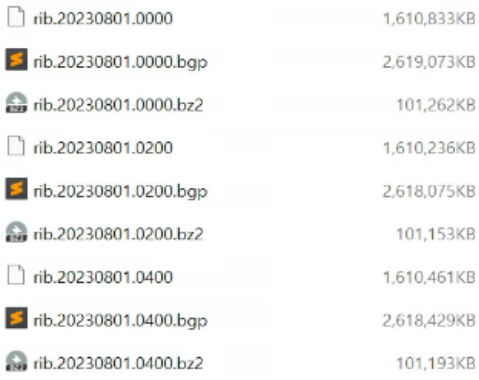



(그림 3) CyCOP 가시화 UI 설계
(Figure 3) Designing the CyCOP Visualization UI

# 4. 실 험

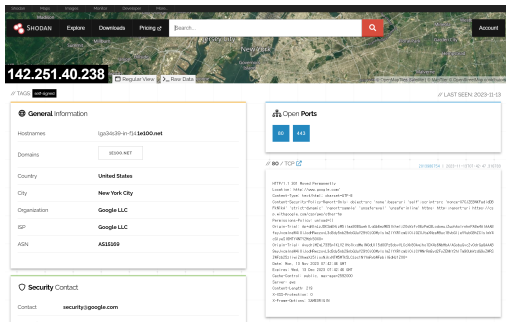
## 4.1 BGP 데이터 수집

본 연구에서는 2023.08.01 ~ 2023.08.30. 기간의 30일 BGP 데이터를 사용하였다. Route Views와 RIPE NCC RIS 데이터를 하루 단위로 결합하여 실험을 진행하였다. 저장소에서 가져온 데이터는 처음 bz2 형식의 압축파일로 되어 있으므로, python mrtparse 라이브러리를 사용하여 전처리하며, 압축해제 파일, 전처리를 거친 파일은 그림 4와 같이 저장된다. 30일 간의 데이터 크기는 Route Views와 RIS 각각 920GB, 513GB에 달한다.




(그림 4) 이진 BGP 및 변환 데이터
(Figure 4) Binary BGP and converted data

변환된 BGP 데이터의 IP 주소, AS 번호를 기준으로 다출처 OSINT에서 추가적인 정보를 수집한다. 그림 5와 6은 추가 수집한 일부 물리 계층의 정보와, 논리 및 페르소나 계층의 데이터를 보여준다.




(그림 5) Shodan을 사용한 OSINT 데이터
(Figure 3) OSINT data with Shodan

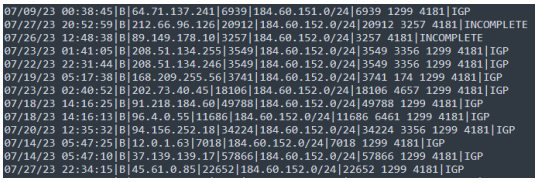
## 4.2 BGP 데이터 수집

특징 추출과 전처리를 거친 데이터는 그림 7과 같이 나타나며, 데이터는 약 144억 row에 달한다. 이 많은 데이터를 분석하는 것은 불가능에 가까우므로 특정 ASN을 포함하는 일부 데이터만을 이상탐지에 사용하여 정확도와 속도를 높였다. 그 결과로 추출된 데이터는 86,413개에 달하며, 해당 데이터를 가지고 실험을 진행하였다. 그 결과 이상 데이터로 판별된 데이터는 239개이며, 그림 8과 같다.

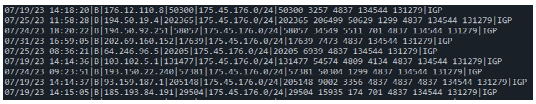

```
{
  "ip": "223.195.38.205",
  "services": [],
  "location": {
    "continent": "Asia",
    "country": "South Korea",
    "country_code": "KR",
    "city": "Seoul",
    "postal_code": "03141",
    "timezone": "Asia/Seoul",
    "province": "Seoul",
    "coordinates": {
      "latitude": 37.566,
      "longitude": 126.9784
    }
  }

  "autonomous_system": {
    "asn": 9769,
    "description": "SEJONG-AS Sejong University",
    "bgp_prefix": "223.195.0.0/18",
    "name": "SEJONG-AS Sejong University",
    "country_code": "KR"
  },
```


(그림 6) Censys를 사용한 OSINT 데이터
(Figure 6) OSINT data with Censys



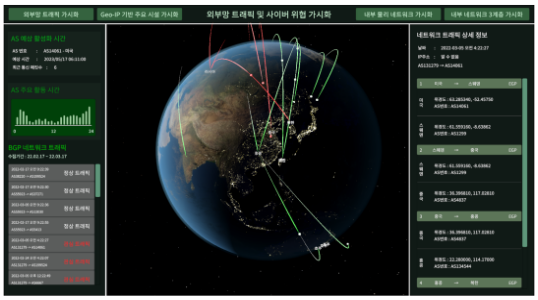

(그림 7) 전처리를 거친 BGP 데이터
(Figure 7) Preprocessed BGP data




(그림 8) LSTM 모델을 통해 감지된 이상 데이터
(Figure 8) Anomaly data detected by the LSTM model

## 4.3 네트워크 가시화

그림 9는 수집한 BGP 데이터의 경로 정보를 기반으로 지도 상에 네트워크 트래픽을 가시화 하고, 좌측에 분석된 정보를, 우측에 추가 수집을 통해 얻은 정보를 함께 표시하는 화면을 보여준다. 네트워크 트래픽이 어떤 경로를 통해 전달되었는지 확인이 가능하며, 분석을 통해 얻은 정보를 바탕으로 공격의 주체나 중간 경로인 교두보를 확인 하는 것이 가능하다. 또한 AS의 활동 시간과 다음 활동 예측시간을 함께 표기하여 작전을 설립하거나 수행하는데 도움을 준다.

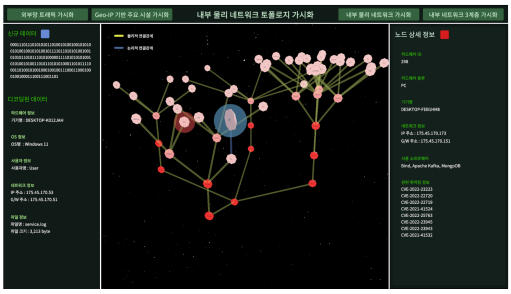


(그림 9) 네트워크 트래픽 기반 BGP 가시화
(Figure 9) Network traffic-based BGP Visualization

그림 10은 수집한 데이터를 사이버 3계층으로 분류하고, 연결관계를 중심으로 각 계층의 요소를 가시화한 것을 보여준다. 해당 정보와 어떤 정보들이 연관되어 있으며, 연결된 객체들을 강조하여 관련 정보를 빠르게 파악할 수 있다.

(그림 10) 사이버 3계층 기반 주요 노드 가시화
(Figure 10) Cyber 3 layer-based critical nodes Visualization

그림 11은 네트워크 peer를 기반으로 각 노드의 연결과 관련 정보를 요약하여 빠르게 노드를 파악할 수 있도록 도와주는 가시화이다. 특정 노드로 접근하기 위한 경로를 파악할 수 있으며, 관련 취약점이나 페르소나 정보를 함께 보여주어 이후 작전에 필요한 요소들을 정확하고 빠르게 식별할 수 있다.



(그림 11) 네트워크 피어 기반 가시화
(Figure 11) Network peer-based Visualization

## 5. 결 론

본 연구의 목적은 BGP 데이터셋을 기반으로 CyCOP을 설계하고 구현하여 보다 신속하고 정확하게 정보를 얻을 수 있게 하는데 있다. 이를 위해 CyCOP을 분석하여 가시화에 필요한 요소들을 식별하고 관련 연구들을 바탕으로 CyCOP 프레임워크를 설계하였으며, 각 구성요소를 간단하게 구현하는 방안을 제시하였다. 마지막으로 실험을 통해 실제 네트워크 데이터를 수집하고 가공하여 이를 가시화에 적용하였다. 본 연구에서 제시한 프레임워크를 통해 CyCOP을 개발한다면, 지휘관이나 보안 담당자

들이 보다 신속하고 정확하게 정보를 얻고, 효과적으로 네트워크 상황을 파악할 수 있을 것이다. 향후 연구로는 다양한 이상탐지 기법이나 분석하여 얻은 정보를 사용하는 것 뿐만 아니라 해당 정보를 공유하고 빠르게 전파하여 활용할 수 있는 방안이 필요할 것으로 보인다.

## 참고문헌(Reference)


[1] Wolbers, Jeroen, and Kees Boersma. “The common operational picture as collective sensemaking,” Journal of Contingencies and Crisis management, vol. 21, no.4, pp.186-199, 2013. http://dx.doi.org/10.1111/1468-5973.12027

[2] Mittu, Ranjeev, and Frank Segaria. “Common operational picture (COP) and common tactical picture (CTP) management via a consistent networked information stream, Proc. Command Control Res. Technol. Symp,” pp. 3-7, 2000. https://apps.dtic.mil/sti/citations/ADA461803

[3] Keuhlen, Daniel T. et al., “The common operational picture in joint vision 2020: a less layered cake,” NATIONAL DEFENSE UNIV NORFOLK VA JOINT AND COMBINED WARFIGHTING SCHOOL, 2002. https://apps.dtic.mil/sti/citations/ADA421610

[4] Baar, Dr David, and Garth Shoemake, “Pliable Display Technology for the Common Operational Picture,” IDELIX Software Inc, 2004. https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/RTO-MP-IST-043/MP-IST-043-13.pdf

[5] Copeland, Jeffrey L. “Emergency response: Unity of effort through a common operational picture,” U.S. Army War College, Carlisle, PA, Strategy Research Project, Mar 2008. https://apps.dtic.mil/sti/citations/ADA479583

[6] Wreski, Erin E., and Erik A. Lavoie, “A concept of operations for an unclassified common operational picture in support of maritime domain awareness,” Naval Postgraduate School Monterey United States, 2017. https://apps.dtic.mil/sti/citations/trecms/AD1046180

[7] Mittu, Ranjeev, and Frank Segaria, “Common operational picture (cop) and common tactical picture (ctp) management via a consistent networked information stream (cnis),” NAVAL RESEARCH LAB, Washington, DC, USA, 2000. https://apps.dtic.mil/sti/citations/ADA461803

[8] Kookjin Kim, et al., “A study on the cyber common operation picture for situational awareness in cyberspace,” Journal of Korean Society for Internet Information, vol.23, no.5, 2022. https://doi.org/10.7472/jksii.2022.23.5.87

[9] Llopis, Salvador, et al., “A comparative analysis of visualisation techniques to achieve cyber situational awareness in the military,” International Conference on Military Communications and Information Systems (ICMCIS), pp. 1-7, 2018. http://dx.doi.org/10.1109/ICMCIS.2018.8398693

[10] Jiang, Liuyue, et al., “Systematic Literature Review on Cyber Situational Awareness Visualizations,” IEEE Access, vol. 10, pp. 57525-57554, 2022. http://dx.doi.org/10.1109/ACCESS.2022.3178195

[11] Doucette, H., “Identifying Requirements for a Cyber Common Operating Picture (CyCOP): Information Collection,” Defence Research and Development Canada, Ottawa, Canada, March 2020. https://cradpdf.drdc-rddc.gc.ca/PDFS/unc343/p811884_A1b.pdf

[12] Dillabaugh, C., and D. Bennett, “CyberCOP: Cyber Situational wareness Demonstration Tool,” Defence Research and Development Canada, Ottawa, Canada, Feb. 2020. https://cradpdf.drdc-rddc.gc.ca/PDFS/unc343/p811728_A1b.pdf

[13] The Joint Staff, “Joint Publication (JP) 3-12, Cyberspace Operation,” Washington, DC, USA, June 2018. https://irp.fas.org/doddir/dod/jp3_12.pdf

[14] Ducheine, Paul, and Jelle Van Haaster, “Fighting power, targeting and cyber operations,” 2014 6th International Conference On Cyber Conflict (CyCon 2014), IEEE, 2014. https://doi.org/10.1109/CYCON.2014.6916410

[15] Headquarters, Department of the Army, “Army Techniques Publication (ATP) 2-01.3, Intelligence Preparation of the Battlefield,” Washington, DC,

USA, Jan. 2021.
https://home.army.mil/wood/application/files/8915/5751/8365/ATP_2-01.3_Intelligence_Preparation_of_the_Battlefield.pdf

[16] CHO, Sungyoung, et al., “Cyber kill chain based threat taxonomy and its application on cyber common operational picture,” 2018 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA). IEEE, p. 1-8, 2018.
https://doi.org/10.1109/CyberSA.2018.8551383

[17] ESTEVE, Manuel, et al., “Cyber Common Operational Picture: A Tool for Cyber Hybrid Situational Awareness Improvement,” North Atlantic Treaty Organization (NATO) Science and Technology Organization (STO), Technical Report STO-MP-IST-148, 2016.
https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-IST-148/MP-IST-148-11.pdf

[18] PAHI, Timea et al., “Preparation, modelling, and visualisation of cyber common operating pictures for national cyber security centres,” Journal of Information Warfare, vol.16, no.4, pp.26-40, 2017.
https://www.jstor.org/stable/26504116

[19] https://www.routeviews.org/routeviews/

[20] https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris

[21] Al-Musawi, Bahaa et al., “BGP anomaly detection techniques: A survey,” IEEE Communications Surveys & Tutorials, vol.19. no.1, pp.377-396, 2016.
https://doi.org/10.1109/COMST.2016.2622240

[22] YANG, Chang; JIA, Wenli, “BGP anomaly detection-a path-based approach,” 2023 3rd Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), IEEE, pp. 408-414, 2023.
https://doi.org/10.1109/ACCTCS58815.2023.00100

[23] CHENG, Min, et al., “MS-LSTM: A multi-scale LSTM model for BGP anomaly detection,” 2016 IEEE 24th International Conference on Network Protocols (ICNP). IEEE, pp.1-6, 2016.
https://doi.org/10.1109/ICNP.2016.7785326

[24] Muosa, Ali Hassan, and A. H. Ali, “Internet Routing Anomaly Detection Using LSTM Based Autoencoder,” 2022 International Conference on Computer Science and Software Engineering (CSASE) IEEE, pp.319-324, 2022.
https://doi.org/10.1109/CSASE51777.2022.9759613

[25] NGUYEN, Quoc Phong, et al., “GEE: A gradient-based explainable variational autoencoder for network anomaly detection,” 2019 IEEE Conference on Communications and Network Security (CNS). IEEE, pp.91-99, 2019.
https://doi.org/10.1109/CNS.2019.8802833

[26] Hoarau, Kevin, et al., “Bgnn: Detection of bgp anomalies using graph neural networks,” 2022 IEEE Symposium on Computers and Communications (ISCC). IEEE, pp. 1-6, 2022.
https://doi.org/10.1109/ISCC55528.2022.9912989

[27] Glassman, Michael, and Min Ju Kang, “Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT),” Computers in Human Behavior, vol.28, no.2, pp.673-682, 2012.
https://doi.org/10.1016/j.chb.2011.11.014

[28] Ivanjko, Tomislav, and Tomislav Dokman, “Open Source Intelligence (OSINT): issues and trends,” INFuture 2019: knowledge in the digital age, pp. 191-196, 2019.
https://repozitorij.ffzg.unizg.hr/islandora/object/ffzg:7025

[29] https://www.maxmind.com/en/geoip2-databases

[30] https://asrank.caida.org/

[31] Department of Defense, United States of America, “Military-Standard (MIL-STD)-2525D, INTERFACE STANDARD, JOINT MILITARY SYMBOLOGY,” Washington, DC, USA, Nov. 2008.
http://everyspec.com/MIL-STD/MIL-STD-2000-2999/MIL-STD-2525D_50933/

[32] Department of Defense, United States of America, “Military-Standard (MIL-STD)-1472H, DESIGN CRITERIA STANDARD, HUMAN ENGINEERING,” Washington, DC, USA, Jan. 2019.
http://everyspec.com/MIL-STD/MIL-STD-1400-1499/MIL-STD-1472H_57041/

## ◐ 저 자 소 개 ◑

**정 재 영(Jae-yeong Jeong)**
2021년 숭실대학교 정보보안학과(학사)
2021년~현재 세종대학교 대학원 컴퓨터공학과(석사과정)
관심분야 : 네트워크, 사이버보안, 인공지능, etc
E-mail : jaeyeong@sju.ac.kr

**김 국 진(Kook-jin Kim)**
2017년 서울호서전문학교 정보보호학과(학사)
2019년 ㈜엠투소프트 전자문서사업부 주임
2023년 세종대학교 대학원 컴퓨터공학과(박사)
관심분야 : 사이버전, 사이버 지휘통제, 정보보호, 인공지능, etc.
E-mail : kjkim@sju.ac.kr

**박 한 솔(Han-sol Park)**
2021년 숭실대학교 정보보안학과(학사)
2023년 세종대학교 대학원 컴퓨터공학과(석사)
2023년~현재 세종대학교 대학원 컴퓨터공확과(박사과정)
관심분야 : 이상 탐지, 딥러닝, 사이버전, etc.
E-mail : miro9303@sju.ac.kr

**장 지 수(Ji-soo Jang)**
2017년 서울호서전문학교 정보보호학과(학사)
2023년 세종대학교 대학원 컴퓨터공학과(석사)
2023년~현재 세종대학교 대학원 컴퓨터공학과(박사과정)
관심분야 : 국방정보시스템, 사이버보안, etc.
E-mail : wekki96@sju.ac.kr

## ◐ 저 자 소 개 ◑

**신 동 일(Dong-il Shin)**
1988년 연세대학교 전산학과(학사)
1993년 Washinton State University 대학원 컴퓨터공학과(석사)
1997년 University of North Texas 대학원 컴퓨터공학과(박사)
1998년 시스템공학연구소 선임연구원
1998년~현재 세종대학교 컴퓨터공학과 교수
관심분야 : 멀티미디어, 머신러닝, 보안 기술, etc.
E-mail : dshin@sejong.ac.kr

**신 동 규(Dong-kyoo Shin)**
1986년 서울대학교 계산통계학과(학사)
1992년 Illinois Institute of Technology 대학원 컴퓨터과학과(석사)
1997년 Texas A&M University 대학원 컴퓨터과학과(박사)
1998년~현재 세종대학교 컴퓨터공학과 교수
관심분야 : 사이버전, 사이버보안, 사이버 지휘통제, 인공지능, 정보보호, etc.
E-mail : shindk@sejong.ac.kr