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

BGP 네트워크 데이터 내의 이상징후 감지를 위한 인터랙티브 시각화 분석 기법☆

Interactive Visual Analytic Approach for Anomaly Detection in BGP Network Data

최 소 미[3] 김 선 영[1*] 이 재 연[1] 고 장 혁[2] 권 구 형[2] 주 재 걸[3]
So-mi Choi Son-yong Kim Jae-yeon Lee Jang-hyuk Kauh Koo-hyung Kwon Jae-gul Choo

요 약

지난 2020년부터 세계는 COVID-19 확산으로 인해 사회적 거리두기와 재택근무를 시행함에 따라 인터넷을 활용한 비디오 및 음성 관련 콘텐츠 서비스와 클라우드 컴퓨팅 활성화로 인터넷에 대한 의존도가 늘어나면서 라우팅 프로토콜 기반 실시간 스트리밍 세션이 증가하고 있다. BGP는 가장 많이 사용되는 라우팅 프로토콜로써 보안성을 향상시키기 위해 많은 연구들이 지속되고 있으나 분석의 실시간성과 알고리즘의 오탐을 판단하기 위한 시각적 분석이 부족하다. 본 논문은 정상 및 이상으로 분류된 BGP 데이터를 수집 및 전처리 후 통계적 기법과 Rule-based 기법을 융합한 이상징후 감지 알고리즘을 활용하여 실 데이터 기반으로 분석한다. 더불어 지도 및 Sankey Chart 기반 시각화 기법으로 알고리즘의 분석 결과와 직관적인 시각화 방안으로 인터랙티브한 시공간 분석 방안을 제시한다.

☞ 주제어 : Border Gateway Protocol, 이상징후 감지, 인터랙티브 시각화 분석, 시공간 분석

ABSTRACT

As the world has implemented social distancing and telecommuting due to the spread of COVID-19, real-time streaming sessions based on routing protocols have increased dependence on the Internet due to the activation of video and voice-related content services and cloud computing. BGP is the most widely used routing protocol, and although many studies continue to improve security, there is a lack of visual analysis to determine the real-time nature of analysis and the mis-detection of algorithms. In this paper, we analyze BGP data, which are powdered as normal and abnormal, on a real-world basis, using an anomaly detection algorithm that combines statistical and post-processing statistical techniques with Rule-based techniques. In addition, we present an interactive spatio-temporal analysis plan as an intuitive visualization plan and analysis result of the algorithm with a map and Sankey Chart-based visualization technique.

☞ keyword : Border Gateway Protocol, Anomaly Detection, Interactive Visual Analytic, Spatio-temporal Analysis

1. 서 론

지난 2020년부터 세계는 COVID-19 확산으로 인해 사회적 거리두기와 재택근무를 시행함에 따라 인터넷을 활용한 비디오 및 음성 관련 콘텐츠 서비스와 클라우드 컴퓨팅 활성화로 라우팅 프로토콜 기반 실시간 스트리밍 세션이 증가하고 있다 [1]. Border Gateway Protocol(BGP)는 가장 많이 사용되는 라우팅 프로토콜로써 Internet Service Providers에서 제공한 라우팅 정책을 기반으로 Autonomous Systems(AS)간 Network Reachability Information을 전달한다 [2]. 하지만 악의적으로 트래픽을 유도해 라우팅 경로에 불필요한 지연을 추가하는 공격이 빈번히 발생한다. [3]에 의하면 발생한 BGP 공격들 중 20%는 2분 이내 인터넷의 90%를 오염시킬 만큼 인터넷의 성능과 안정성을 크게 위협한다.

BGP의 보안성을 향상시키기 위해 많은 연구들이 지속되고 있다. [4]에서는 취약성에 대한 대응방안으로 Public

1 Ground Control · Cyber Team, Hanwha Systems, 13524, 188 Pangyoyeok-Ro Bundang-Gu Seongnam-Si Gyeonggi-Do Korea.
2 Cyber & Network Technology Center, Agency for Defense Development, 05744, San25 Songpa-Gu Seoul Korea.
3 Korea Advanced Institute of Science and Technology, 34141, 291 Daehak-Ro Yuseong-Gu Daejeon Korea.
* Corresponding author (sonyong.kim@hanwha.com)
[Received 30 August 2022, Reviewed 17 September 2022(R2 20 October 2022), Accepted 25 October 2022]
☆ 이 논문은 2019년 정부(방위사업청)의 재원으로 국방과학연구소의 지원을 받아 수행된 연구임(UC190039ED)

Key Infrastructure 암호화, 의심스러운 라우팅 업데이트 또는 이상징후 경로 차단, 불특정 경로 차단, BGP 트래픽 hijacking 이상탐지 총 4가지 범주로 정리한다. 이중 BGP 트래픽 기반 이상탐지 기법은 높은 정확도를 보이지만 분석의 실시간성과 알고리즘의 오탐을 판단할 수 있는 시각적 분석이 부족하다 [4].

본 논문에서는 BGP 네트워크 데이터 내의 이상징후 감지를 위한 인터랙티브 시각화 분석을 연구하기 위해 Réeseaux IP Européens(RIPE) Network Cooordinate Centre(NCC) 및 BGPstream에서 제공하는 실제 BGP 정상 및 이상(hijacking, bgp leaks, 그리고 outage) 데이터를 각각 수집해 제시하는 이상탐지 알고리즘에 적용한다. 더불어 분석 결과를 직관적인 시공간적 시각화 방안으로 사용자에게 제공하여 이상징후를 인터랙티브하게 감지할 수 있는 인터페이스를 제공한다. 2장에서는 BGP 개념을 적용한 이상징후 감지 연구와 BGP 데이터의 주요 특징을 분석한다. 3장에서는 본 논문에서 제시하는 이상징후 감지를 위한 인터랙티브 시각화 분석에 대한 설계를 설명하고, 4장에서는 분석 결과를 기술한다. 끝으로 결론에서는 본 논문에서 제안하는 시각화 분석을 통해 기대할 수 있는 성과와 향후 연구 방향을 기술한다.

2. 관련 연구

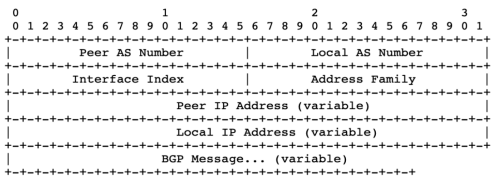
2.1 Border Gateway Protocol(BGP) 개요

2.1.1 BGP 개요

AS는 하나의 행정 기관으로부터 관리되는 IP 네트워크들의 집합으로 일관된 라우팅 계획과 Autonomous System Number(ASN)이라는 고유 식별자를 통해 다른 AS와 라우팅 정보를 송수신 한다 [4]. BGP는 AS간 원활한 통신을 위해 최적의 경로를 찾아 네트워크 속도를 개선하며 라우팅 테이블을 주기적으로 업데이트 후 메시지를 송수신 한다. [4]에 의하면 BGP 메시지는 그림 1과 같이 이웃 ASN, 로컬 ASN, 이웃 IP주소, 로컬 IP주소, 그리고 BGP 메시지로 구성된다.

2.1.2 BGP 데이터 유형 및 특징

BGP raw 데이터 종류는 control plane과 data plane이 있다 [5]. Control plane 데이터는 BGP speakers간에 교환되는 Routing Information Base(RIB) 또는 BGP 업데이트 메시지를 의미하며 RouteViews 프로젝트 [4], RIPE NCC [4], BGPmon [6] 리포지토리에서 수집할 수 있다. 반면 data plane 데이터는 BGP speaker와 peer 간에 패킷이 사용하는 경로를 의미하며 control plane 데이터와 달리 에이전트를 통해서만 수집이 가능하다.



(그림 1) Border Gateway Protocol 메시지 구성
(Figure 1) Border Gateway Protocol Message Format

더불어 [7]에 의하면 일련의 BGP 메시지 속에서 비정상 데이터 감지는 쉽지 않다. 이를 해결하기 위해 [8]에서는 단일 BGP 데이터보다 다수의 BGP volume을 기반으로 표 1에 제시된 AS-PATH 기반 BGP 특징들을 활용해 탐지한다. 본 논문은 3장에서 제시하는 이상징후 알고리즘 분석 요소들로 [8]에서 제시한 방안들을 활용한다.

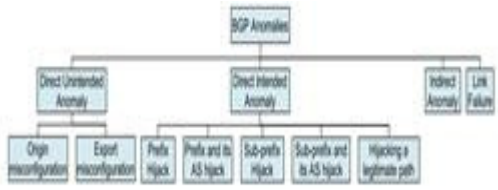
(표 1) AS-PATH 속성에 의한 BGP 특징
(Table 1) Possible BGP Features from AS-PATH Attribute

구분	특징 종류
1	AS-PATH 길이
2	최대/평균 AS-PATH 길이
3	최대/평균 유일 AS-PATH 길이
4	긴/짧은 경로에 대한 알림
5	경로에서 회귀 AS 관찰
6	경로에서 최대/평균 회귀 ASes
7	지리적 위치에 따른 AS-PATH 변경
8	Prefix 원점 변경
9	이전 경로 철회 후 생성된 경로의 수
10	새로운 경로 공지의 수

2.2 BGP 이상징후 감지 기법

BGP 이상징후 감지를 위한 방안은 활발히 연구되고 있다. [4]에서는 BGP 이상징후 유형을 그림 2처럼 정리하며, 여러 분야에서 다양한 BGP 이상징후 유형을 탐지하는 기법들이 상용화되었다. 해당 기법들은 크게 시계열, 통계적, 역사적, 그리고 머신러닝 기법으로 분류할 수 있다. 그러나 대부분의 연구들은 제한된 이상징후 유형 탐

지 및 오직 정확도 중점으로 연구가 되어 사용자가 탐지 결과에 대한 직관적인 시각적 분석 방안이 부족하다.



(그림 2) BGP 이상징후 유형
(Figure 2) Taxonomy of BGP Anomalies

2.2.1 시계열 접근 BGP 이상징후 감지 기법

시계열 접근으로 [9]에서는 라우팅 업데이트 주기 기반 Fast Fourier Transform 기법을 활용하였다. 해당 연구는 수집된 BGP 데이터 내 신규, 재선언, 중복, 그리고 제한 경로에 대한 특징들을 활용하여 탐지하였으나 발생 시점과 원인을 분석하기에는 제한적이다. [10]은 BGP volume을 Wavelet Transform 기법에 활용하였으나 장시간의 분석이 요구되는 단점이 있다. [11]에서는 BGP volume에 AS-PATH의 평균 길이의 특징을 Recurrence Quantification Analysis 기법에 활용해 Direct Intended Anomaly(DIA)를 탐지하였다. 본 논문은 [11]에서 제시하는 기법을 이상징후 알고리즘 설계시 참고하였다.

2.2.2 통계적 접근 BGP 이상징후 감지 기법

통계적 접근으로 [12]에서는 PCA 기법을 활용하였으며, [8]에서는 Generalized Likelihood Ratio Test 기법을 활용해 정확도를 높이고 computational cost를 낮추었으나, 두 기법 모두 소요시간 때문에 제한적이다. [13]에서는 AS-PATH 빈도를 지리적 공간과 매핑한 통계적 기법을 통해 Indirect Anomaly 탐지하였다. [14]에서는 통신사 Link Telecom에서 발생한 BGP 데이터를 z-score를 활용해 AS-PATH 길이의 분포도가 평균 대비 특정 표준편차 이상일 때 DIA로 분류하였다. 본 논문은 [13]에서 제시하는 지리적 공간 매핑 기법과 [14]에서 활용한 z-score 기법을 이상징후 알고리즘 설계시 참고하였다.

2.2.3 역사적 접근 BGP 이상징후 감지 기법

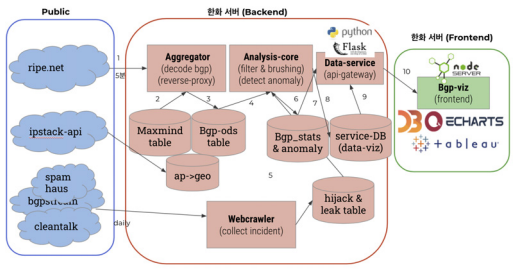
역사적 접근으로 [3]에서는 지난 2개월 간 발생한 RIB 테이블 및 BGP 데이터 업데이트 내역을 최신 정보와 비교하여 신규로 식별되는 데이터를 이상으로 분류하였다. 역사적 접근은 설정된 주기 내 수집 데이터를 기반으로 분석하기에 단순 역사적 접근은 기록 주기에 따라 오탐이 발생하거나 다르게 해석이 될 수 있어 제한적이다.

2.2.4 머신러닝 접근 BGP 이상징후 감지 기법

머신러닝 접근으로 [15]에서는 Decision tree, Naive Bayes, SVM 알고리즘 기반 학습 및 이상징후 감지를 수행하였으며, [16]에서는 Winnowing 알고리즘을 활용해 BGP 속성들을 학습시켜 분석하였다. 그러나 DIA 식별 및 발생 원인과 시점을 구분하기에는 제한적이다.

3. 설계 및 구현

본 논문은 BGP 데이터를 수집 및 전처리 후 제시하는 이상징후 감지 알고리즘을 통해 실 BGP 데이터로 분석해 지도 및 Sankey Chart 시각화 기법으로 직관적인 시각화 방안을 제시한다. 그림 3은 본 논문에서 제시하는 시공간 시각적 분석을 위한 소프트웨어 설계도이다.

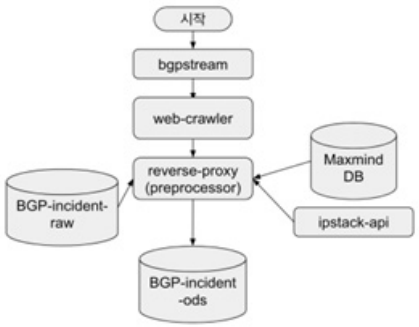


(그림 3) 시공간 시각적 분석을 위한 소프트웨어 설계도
(Figure 3) Spatio-temporal Visualization Analysis Software Architecture

시공간 시각적 분석 소프트웨어는 크게 Public, 백엔드, 그리고 프론트엔드로 구분한다. Public 단계에서는 외부에서 제공하는 정상 BGP raw 데이터, 이상으로 식별된 BGP 데이터, 그리고 IP와 지리적 연관관계 정보를 수집 후 저장한다. 백엔드에서는 수집한 BGP raw 데이터를 전처리 후 이상탐지 알고리즘을 적용해 이상 유무를 판단해 저장한다. 프론트엔드에서는 분석 결과를 다양한 시각화 툴을 기반으로 사용자가 직관적으로 해석할 수 있게 지도 및 Sankey Chart를 도시 후 종료한다.

3.1 BGP 데이터 수집 및 전처리

BGP 데이터 수집 및 전처리에 대한 소프트웨어 설계도는 그림 4와 같다.



(그림 4) BGP 데이터 수집 및 전처리 소프트웨어 설계도
(Figure 4) BGP Data Collection and Processing Software Architecture

3.1.1 BGP 데이터 수집

데이터 수집 및 전처리 서비스는 reverse-proxy가 처리한다. Reverse-proxy는 IP주소를 위도와 경도로 매핑하는 기능이다. File Transfer Protocol로 글로벌 지역으로 구성된 RIPE NCC BGP 데이터를 BeautifulSoup4 web-crawler 서비스로 수집한다. Prefix Hijacking이 일어났는지의 여부는 실제로 AS 소유자만 알 수 있기에 BGPstream, Spamhaus, Cleantalk에서 제공하는 BGP outage, leak, hijack, spam IP 및 AS 데이터를 web-crawler 서비스로 수집한다.

3.1.2 BGP 데이터 전처리

RIPE NCC 및 BGPstream, Spamhaus, 그리고 Cleantalk에서 수집된 비정형 BGP 데이터를 각각 brt2exabgp 및 mrt2bgpdump 함수를 사용해 암호화를 디코드하여 정형화된 BGP 데이터를 MariaDB에 적재한다. 더불어 수집된 RIPE 및 RouteView BGP 데이터에 대한 지리적 정보를 찾기 위해 IP주소를 기준으로 Maxmind DB 내 위도와 경도가 맞는 지리적 정보를 찾아 매핑 한다. Maxmind DB에서 조회가 안 되는 IP주소는 추가로 ipstack-api를 활용해 지리적 정보를 변환한다. 이때 ipstack-api의 제한된 request으로 API limit을 방지하기 위해 이전에 변환된 지리적 정보는 DB에 적재해 중복으로 조회되는 BGP 데이터는 별도로 API 사용 없이 변환한다. 전처리가 완료된 BGP 데이터는 표 2와 3의 테이블 형태로 DB에 저장된다.

(표 2) BGP Raw 데이터 수집 테이블 설계
(Table 2) BGP Raw Data Collection Database Table Architecture

칼럼 명	타입	설명
claimed_at	TIMESTAMP	발생 시간
origin_ip	VARCHAR	원천 IP주소
origin_as	INT	AS 번호
as_path	VARCHAR	AS 경로
next_hop_ip	VARCHAR	AS가 선언한 IP
status	VARCHAR	상태
region	VARCHAR	나라
latitude	DECIMAL	위도
longitude	DECIMAL	경도
country_iso_ code	VARCHAR	나라 코드
country_name	VARCHAR	나라 명
city_name	VARCHAR	도시 명
pk_as_dt(claimed_at, origin_as, as_path)	PRIMARY KEY	Composite Key로 구성한 PK

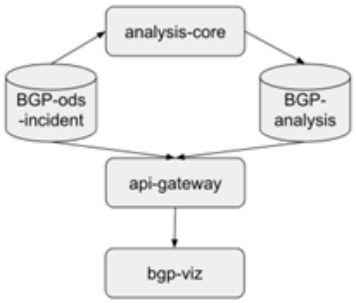
(표 3) 이상 BGP 데이터 테이블 설계
(Table 3) Anomalous BGP Data Collection Database Table Architecture

칼럼 명	타입	설명
event_type	TIMESTAMP	발생 시간
country	VARCHAR	원천 IP주소
asn	INT	AS 번호
as_path	VARCHAR	AS 경로
starttime	TIMESTAMP	시작 시간
endtime	TIMESTAMP	종료 시간
leaked_prefix	VARCHAR	누설된 IP
leaked_by_asn	INT	누설한 AS 번호
leaked_to_asn	INT	우회된 AS 번호
example_path	VARCHAR	AS 경로
expect_prefix	VARCHAR	예상된 IP
expect_asn	INT	예상된 AS 번호
detect_prefix	VARCHAR	감지된 IP주소
detected_asn	VARCHAR	감지된 AS 번호
detected_path	VARCHAR	감지된 AS 경로
pk_as_dt(claimed_at, origin_as, as_path)	PRIMARY KEY	Composite Key로 구성한 PK

4. BGP 이상징후 감지 인터랙티브 시각화

4.1 BGP 이상징후 감지 및 시각화

BGP 이상징후 감지 및 시각화에 대한 소프트웨어 설계도는 그림 5와 같다.



(그림 5) BGP 이상징후 감지 및 시각화 소프트웨어 설계도
(Figure 5) BGP Anomaly Detection and Visualization Software Architecture

Data-service는 api-gateway를 활용한 백엔드와 프론트엔트의 미들웨어 웹 서비스이다. analysis-core에서 최종 service DB 모델을 구성한 후 api-endpoint를 정의한다. Python Flask 프레임워크 기반으로 analysis-core에서는 MariaDB에 적재된 정보 중 필요한 칼럼을 추려낸다. BGP 이상징후 감지를 위해 사전에 적재된 정보를 통계적 및 Rule-based 접근을 적용해 지도 및 Sankey Chart 기반 이상징후 인터랙티브 시각화에 필요한 정보를 취합한다. 해당 정보는 bgp-viz 서비스를 통해 데이터 브러싱 후 Tableau와 오픈소스인 D3.js와 E-Charts로 시각화 한다.

4.2 지도 기반 이상징후 감지 인터랙티브 시각화

본 논문은 이상징후 점수 합산 기법으로 기존 통계적 기법과 Rule-based 로직을 융합한다. 3장에서 수집한 정보에서 표 1에 제시한 AS-PATH 속성들을 추출해 표 4에 제시된 Rule-based를 통해 이상징후 점수 부여한다. BGP outage 또는 leak를 한 이력이 있을 경우 50점을 부여한다. [14]에서 제시한 통계적 기법으로 AS-PATH 길이가 z-score 2.33 표준편차를 넘길 경우 30점을 부여한다. 더불어 AS volume이 이전 경로 대비 흔치 않을 경우와 잠재적인 hijacking 이력이 있을 경우 각각 20점을 부여한다.

(표 4) 이상징후 점수 부여 방식
(Table 4) Anomaly Scoring Method

이상징후 점수	이상징후 원인
50점	BGP outage 또는 leak를 한 이력이 있을 경우
30점	AS-PATH 길이가 z-score 2.33 sigma를 넘길 경우
20점	AS volume이 이전 일주일과 다르게 흔치 않은 경로가 발생할 경우
20점	BGPstream에서 수집한 잠재적인 hijacking 이력이 있을 경우

(표 5) 이상징후 등급 부여 방식
(Table 5) Anomaly Stage Classification Method

이상징후 등급	이상징후 점수	색 구분
상	60점 이상	빨간색
중	30점 이상 60점 이하	노란색
하	30점 이하	초록색

표 4를 통해 합산된 이상징후 점수를 지도에 도시하기 위해 표 5로 이상징후의 등급과 표현할 색을 구분한다. 합산된 이상징후 점수가 60점 이상일 경우 등급은 상으로 분류되며 빨간색으로 도시한다. 더불어 이상징후 점수가 30점 이상 60점 이하일 경우 등급을 중으로 분류해 노란색으로 표현한다. 끝으로 이상징후 점수가 30점 이하일 경우 등급을 하로 구분 후 초록색으로 표현한다.

(그림 6) 지도 기반 BGP 이상 데이터 시각화
(Figure 6) Anomaly BGP World Map Visual Analysis

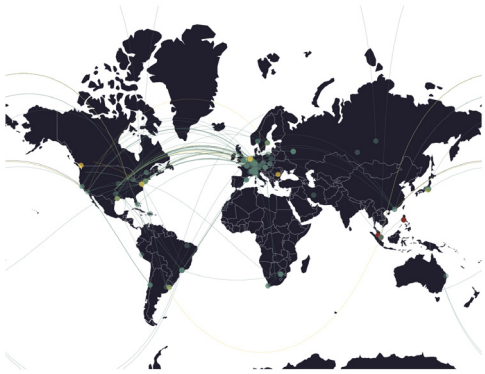
그림 6은 표 4와 5의 점수화 기법을 기반으로 BGPStream, Cleantalk, 그리고 Spamhaus에서 수집한 BGP 이상 데이터를 시각화한 결과다. BGPStream에서 BGP leak로 구분된

데이터는 빨간색 노드, Possible Hijack로 구분된 데이터는 노란색 노드로 도시된다. 더불어 부모 AS 중 BGP leak 이력이 존재하는 데이터의 엣지 또한 빨간색으로 구분된다.

(그림 7) 기간 구분 없이 중복된 ASN 정상-이상 BGP 데이터
(Figure 7) Coincided Normal-Anomalous BGP ASN Regardless of Period

그림 7은 기간 구분 없이 수집된 모든 이상 BGP 데이터의 ASN을 정상으로 판단되는 BGP 데이터의 부모 ASN과 겹치는 데이터만 시각화한 결과다. 하지만 수집된 BGP 데이터 중 오직 14%만의 ASN이 겹쳤을 뿐더러 부모 ASN 단일 조건으로 분석한 결과 BGP 이상탐지에 대한 false positive가 다수 발생한다. 이를 방지하기 위해 1차로 중복된 ASN 데이터 중 부모 AS 및 자식 AS 관계가 동일한 데이터 조건을 추가해 2차 필터링 과정을 걸친다. 더불어 기간을 일주일 단위로 분석하는 것이 오탐을 최소화할 수 있어 3차 필터링 과정을 걸친다. 그림 8은 위 3가지 필터링 절차를 걸친 시각화 화면이며 도시된 결과는 그림 7 대비 오탐률이 적은 동시에 실제 BGP prefix hijacking이 방생한 이상 데이터를 관련 등급과 색에 맞게 도시된 것을 볼 수 있다.

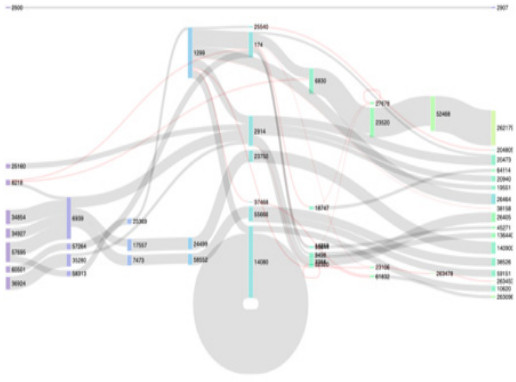
본 논문은 지도 기반 이상징후 감지 인터랙티브 시각화 기법으로 수집된 정상 및 이상 BGP 데이터를 단순 알고리즘의 분석 결과에만 의존하는 것이 아닌, 융합된 분석 기법과 BGP의 지리적 경로를 지도를 통해 사용자가 직관으로 이상 유무, 이상징후 점수, 그리고 등급에 관한 정보를 인지하고 위치한 상황에 따른 의사결정을 할 수 있게 지원한다.



(그림 8) 3가지 필터링 과정을 걸친 지도 기반 이상징후 감지 인터랙티브 시각화 결과
(Figure 8) 3-Ways Filtered Interactive World Map Visual Analysis

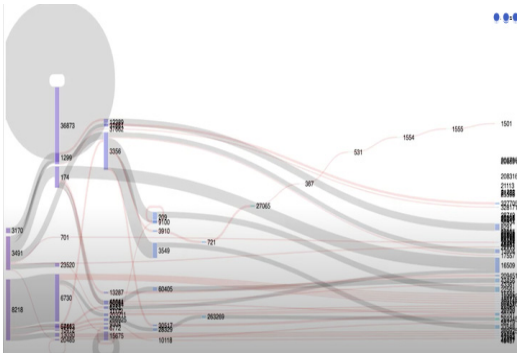
4.3 Sankey Chart 기반 이상징후 감지 인터랙티브 시각화

Sankey Chart를 활용한 시각화 기법은 해당 차트의 특성을 살려 BGP 정보와 융합해 AS-PATH간 발생 빈도가 낮은 경로를 이상징후로 분류한다. 본 기법도 지도 기반 이상징후 감지처럼 AS-PATH 데이터를 활용하지만 변화율과 degree를 활용한다. 그림 9는 BGP volume이 낮은 데이터를 이상징후로 분류하는 특징을 Sankey Chart에 적용한 화면이다. 사용자가 직관적으로 흔치 않은 BGP 경로가 빨간색으로 식별이 되었음을 알 수 있다.



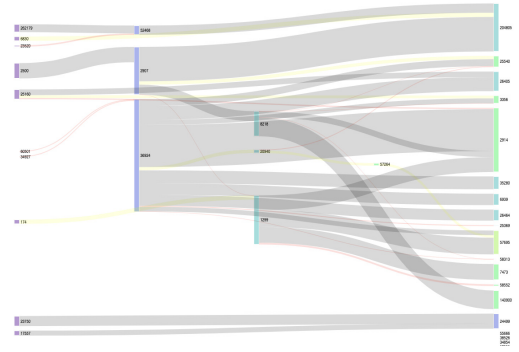
(그림 9) Sankey Chart 기반 BGP 이상징후 시각화 결과
(Figure 9) Anomalous BGP Data Using Sankey Chart Visual Analysis

그림 9에 분석한 BGP 데이터는 degree가 낮은 경로가 자주 발생하지 않아 앳지간의 false positive가 높기 않았으나, 과련 데이터 보유시 그림 10과 같이 다수의 false positive가 발생한다. 단순 흔치않은 AS-PATH를 기준으로 Sankey Chart를 활용한 BGP 이상 데이터를 표현하기는 제한적이다.



(그림 10) False Positive가 높은 Sankey Chart 시각화
(Figure 10) Highly Rated False Positive Sankey Chart Visual Analysis

이를 방지하기 위해 동일 기간에 발생한 BGP 데이터에 대한 분석 요소로 흔치않은 AS-PATH에 이어 AS 경로의 degree를 z-score를 적용하여 통계적으로 이상 유무를 분류 과정을 추가한다. 즉 AS degree 경로가 흔치 않으면서도 AS-PATH가 설정된 평균 길이 대비 유독 짧거나 너무 길면 이상징후로 분류한다. 그림 11은 기존 rule-based Sankey Chart에 통계적 기법을 적용한 시각화 화면이다.



(그림 11) 통계적 기법을 적용한 Sankey Chart 이상징후 인터랙티브 시각화 결과
(Figure 11) Statistical Interactive Sankey Chart Visual Analysis

본 논문은 Sankey Chart 기반 이상징후 감지 인터랙티브 시각화 기법으로 수집된 정상 및 이상 BGP 데이터를 단순 AS-PATH degree에만 의존하는 것이 아닌, 오탐률을 낮추기 위해 AS-PATH degree에 대한 z-score 통계적 기법을 추가한다. 이를 기반으로 흔하지 않은 경로가 발생할 경우 이상징후로 분류해 해당 경로를 빨간색으로 식별해 그림 11과 같이 사용자에게 가시적으로 유용한 시공간적 인터랙티브한 분석 결과를 제시한다.

5. 결론 및 향후 연구

본 논문에서는 BGP 네트워크 데이터 내의 이상징후 감지를 위한 인터랙티브 시각화 분석을 연구하기 위해 RIPE NCC 및 BGPstream에서 제공하는 실제 BGP 데이터와 hijack 데이터를 각각 수집해 통계적 및 Rule-based 기법 이상탐지 알고리즘에 적용 후 이상징후에 대한 결과를 지도 기반 및 Sankey Chart를 활용해 인터랙티브한 시각화 방안을 제시했다. 현재 RIPE에서 공개한 BGP 데이터 기반으로 각 이상징후 감지 알고리즘을 비교한 논문이 없기 때문에 앞으로 사이버보안 솔루션을 위한 이상징후 알고리즘을 고도화할 때 본 연구 결과를 활용할 수 있다. 향후 더 많은 BGP Incident 데이터베이스를 구성하여 labeled 데이터셋에 대한 구성 및 검증을 통해 이상징후 감지율을 개선하고 네트워크를 보호하는데 기여할 것이다.

참고문헌(Reference)

[1] da Silva, Carlos Alexandre Gourvea, et al., “The Behavior of Internet Traffic for Internet Services during COVID-19 Pandemic Scenario,” arXiv preprint arXiv_2105.04083, 2021. https://doi.org/10.48550/arXiv.2105.04083

[2] Rekhter, Yakov, Tony Li, and Susan Hares, “A border gateway protocol 4 (BGP-4),” No. rfc4271, 2006. https://tools.ietf.org/html/rfc4271

[3] Shi, Xingang, et al., “Detecting prefix hijackings in the internet with argus,” Proceedings of the 2012 Internet Measurement Conference, pp.15-28, 2012. https://doi.org/10.1145/2389776.2398779

[4] Al-Musawi, Bahaa, Philip Brach, and Grenville Armitage, “BGP anomaly detection techniques: A

survey," IEEE Communications Surveys & Tutorials, Vol.18, No.1, pp.377-396, 2016. https://doi.org/10.1109/COMST.2016.2622240

[5] Biersack, Ernst, et al., "Visual analytics for BGP monitoring and prefix hijacking identification," IEEE Network, Vol.26, No.6, pp.33-39, 2012. https://10.1109/MNET.2012.6375891

[6] Yan, He, et al., "BGPmon: A real-time, scalable, extensible monitoring system," 2009 Cybersecurity Applications & Technology Conference for Homeland Security, 2009. https://doi.org/10.1109/CATCH.2009.28

[7] Feldmann, Anja, et al., "Locating internet routing instabilities," ACM SIGCOMM Computer Communication Review, Vol.34, No.4, pp.205-218, 2004. https://doi.org/10.1145/1030194.1015491

[8] Deshpande, Shivani, et al., "An online mechanism for BGP instability detection and analysis," IEEE transactions on Computers, Vol.58, No.11, pp.1470-1484, 2009. https://doi.org/10.1109/TC.2009.91

[9] Labovitz, Craig, et al., "Internet routing instability," IEEE/ACM transactions on Networking, Vol.6, No.5, pp.515-528, 1998. https://doi.org/10.1109/90.731185

[10] Mai, Jianning, et al., "Detecting BGP anomalies with wavelet," NOMS 2008-2008 IEEE Network Operations and Management Symposium. IEEE, pp.465-472, 2008. https://doi.org/10.1109/NOMS.2008.4575169

[11] Al-Musawi, Bahaa, et al., "Detecting BGP instability using recurrence quantification analysis (RQA)," 2015 IEEE 34th International Performance Computing and Communications Conference (IPCCC). IEEE, pp.1-8, 2015. https://doi.org/10.1109/PCCC.2015.7410340

[12] Huang, Yiyi, et al., "Diagnosing network disruptions with network-wide analysis," ACM SIGMETRICS Performance Evaluation Review, Vol.35, No.1, pp.61-72, 2007. https://doi.org/10.1145/1269899.1254890

[13] Theodoridis, Georgios, et al., "A novel unsupervised method for securing BGP against routing hijacks," Computer and Information Sciences III, pp.21-29, 2013.

[14] Witten, Ian H., and Eibe Frank, "Data mining: practical machine learning tools and techniques with Java implementations," Acm Sigmod Record, Vol.31, No.1, pp.76-77, 2002.

[15] de Urbina Cazenave, et al., "An anomaly detection framework for BGP," 2011 International Symposium on Innovations in Intelligent Systems and Applications. IEEE, pp.107-111, 2011. https://doi.org/10.1109/INISTA.2011.5946083

[16] Lutu, Andra, et al. "Separating wheat from chaff: Winnowing unintended prefixes using machine learning," IEEE INFOCOM 2014-IEEE Conference on Computer Communications. IEEE, pp.943-951, 2014. https://doi.org/10.1109/INFOCOM.2014.6848023

◐ 저 자 소 개 ◑

최 소 미(So-mi Choi)
2016년 퍼듀대학교 컴퓨터공학과 (공학사)
2017년 조지아 공과대학교 컴퓨터공학과 (공학석사)
2020년 10월 ~ 2022년 5월 신한AI
관심분야 : 데이터시각화, 데이터분석
E-mail : smchoi257@gmail.com

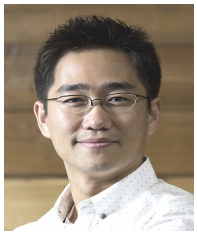
◐ 저 자 소 개 ◑

김 선 영(Son-yong Kim)
2019년 2월 고려대학교 전기전자전파공학(공학사)
2019년 1월~현재 한화시스템(주) 재직
관심분야 : 인공지능보안, 디지털포렌직, 사이버보안
E-mail : sonyong.kim@hanwha.com

이 재 연(Jae-yeon Lee)
2002년 2월 가톨릭대학교 정보통신(공학사)
2004년 2월 광주과학기술원 정보통신(공학석사)
2004년 2월~현재 한화시스템(주) 재직
관심분야 : 사이버보안
E-mail : jaeyeon46.lee@hanwha.com

고 장 혁(Jang-hyuk Kauh)
1996년 2월 광운대학교 컴퓨터과학과(공학사)
1998년 2월 광운대학교 컴퓨터과학과(공학석사)
2018년 8월 광운대학교 컴퓨터과학과(공학박사)
1998년 3월~현재 국방과학연구소 재직
관심분야 : 사이버보안
E-mail : jhkauh@add.re.kr

권 구 형(Koo-hyung Kwon)
2001년 2월 고려대학교 전기전자전파공학(공학사)
2003년 2월 고려대학교 전파공학학과(공학석사)
2006년 7월~현재 국방과학연구소 재직
관심분야 : 사이버보안
E-mail : koohyung@add.re.kr



주 재 걸(Jaegul Choo)
2001년 서울대학교 전기공학부 (공학사)
2009년 Georgia Tech, Electrical and Computer Engineering (공학석사)
2013년 Georgia Tech, Computational Science and Engineering (공학박사)
관심분야 : 인공지능, 컴퓨터비전, 자연어처리
E-mail : jchoo@kaist.ac.kr