Journal of
Information Systems Engineering
and Business Intelligence

Vol.10, No.2, June 2024
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

The Art of Internet Mapping: A Comprehensive Guide to Regional Internet Topology Mapping at the Autonomous System Level

Timotius Witono [1)]*, Setiadi Yazid [2)], Yudho Giri Sucahyo [3)]

[1) 2) 3)] *Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia*

[1)] timotius.witono@ui.ac.id, [2)] setiadi@cs.ui.ac.id, [3)] yudho@cs.ui.ac.id

[1)] *Faculty of Information Technology, Maranatha Christian University, Bandung, Indonesia*

[1)] timotius.witono@it.maranatha.edu

Abstract

Background: Internet topology is studied more by researchers on global internet coverage compared to limited regional coverage. However, some studies also see the importance of studying internet topology in certain countries or regions. The internet performance of a country or region can be influenced by the structure of its internet topology, and research on internet topology can contribute to improving internet topology in that region.
Objective: This research initially carried out internet topology mapping in a limited region, then used experience from each step in conducting internet topology mapping to create a comprehensive guide on procedures for internet topology mapping at the autonomous system (AS) level in certain regional areas.
Methods: Internet topology mapping is carried out by inferring relationships between ASes through an inference process against border gateway protocol (BGP) table dumps, while the internet topology mapping method chosen is passive mapping.
Results: The entire series of steps involved in the regional internet topology mapping process have been successfully outlined in a detailed guide as a result of this research. Evaluation of the research results was carried out by implementing the application of this comprehensive guide and also through assessments from experts in related fields regarding the results of this research. The results of both evaluations showed that the research results were appropriate.
Conclusion: This research provides a comprehensive guide for mapping internet topology in specific regional areas, consisting of nine sequential steps grouped into four major steps. This guide can be used to assist similar research efforts in other regional areas as well as provide further knowledge regarding studies in this field. This research is different from previous studies, because it provides a comprehensive guide to the internet topology mapping process, which has not been available in previous studies.

Keywords: Internet Topology Mapping, Regional Internet Topology, Autonomous System, Border Gateway Protocol

Article history: Received 10 January 2024, first decision 19 March 2024, accepted 13 May 2024, available online 28 June 2024

I. INTRODUCTION

The internet is involved in various aspects of people's lives, so research aimed at improving internet performance is very important. Internet topology research is one area of study that relates to enhancing internet performance. The connecting structure of the many parts that make up an internet network, such as the router, host, and autonomous system, is known as internet topology. The rapid expansion of the internet and its various issues are driving forces behind research in the field of internet topology. Researchers are interested in this topic of study because it offers fresh insights into the internet through research on internet topology. The topology of the internet is usually divided into several different levels, including autonomous system (AS), point-of-presence (POP), router, and interface levels. Because the amount of granularity at the AS level is the highest compared to other levels and because AS data collection is relatively simpler, the most notable study on internet topology occurs at this level [1].

ASes can be described as a growing "network of networks" managed independently by each AS administrator. Interconnection between ASes occurs autonomously, namely by direct communication between administrators of

* Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.10.2.191-205

each AS that wants to be connected, and is not managed by an institution centrally. From a scientific point of view, the internet is a self-organizing complex network. The development of the internet which runs independently can give rise to new problems that are not visible now, therefore researchers have carried out various studies to study the structure and development of the internet [2]. Generally speaking, there are four types of reasons why people study network topology: scientific (to comprehend a network's general properties), managerial (to get comprehensive details on an owned network), informational (to obtain network information from the service provider), and adversarial (to get information about competitors' networks) [3]. In particular, internet topology research is driven by several motivations: to obtain information about certain topological properties of the internet, to review what internet topology looks like, to produce internet topologies for simulation purposes, and to measure topological developments internet in the future [4].

The AS level internet topology is at the highest granularity, so problems that occur at the AS level will affect the levels below. Issues with internet topology at the AS level typically have an international effect. Identification and handling of problems at the AS level requires internet topology information, but there is no complete internet topology map owned by any particular party. Information about AS is relatively easy to obtain from Regional Internet Registry (RIR), the organization that manages AS Number (ASN) allocation; but what is difficult is to obtain information about relations between ASes. Relations between ASes in the internet topology are the result of business agreements between ASes' administrators, therefore relations between ASes are generally private and confidential. AS level research is the most remarkable because it facilitates the collection of research dataset and has the highest granularity [4, 1]. More and more researchers are studying global internet topology, and less research is being conducted on internet topology in specific countries or regions. Nonetheless, several studies concentrate primarily on assessing the internet topology inside certain nations or areas. The performance of a nation or region's internet can be impacted by the structure of its internet topology, and study on internet topology can aid in this. Research into the internet topology can also predict the future of the internet in some countries and regions. The results of the study also provide a representation of the topology of the internet and provide useful information for internet stakeholders in certain countries and regions [5, 2].

The structure of an internet topology may affect a nation's or region's internet performance, and study on internet topology can help that region's internet topology get better. This research chose to study internet topology mapping at the AS level in a selected regional area, because AS represents the highest level of granularity of the internet, so that internet topology at the AS level can provide a picture of the internet economy of the selected area. This research initially mapped internet topology at the AS level utilizing datasets from border gateway protocol (BGP) table dumps, with a case study on the national internet topology of Indonesia. This research then utilizes the knowledge gained from each stage of internet topology mapping to produce a comprehensive guide that outlines the steps involved in internet topology mapping at the AS level in a specific regional area.

This research contributes to the scientific field of internet topology by presenting a comprehensive guide for mapping internet topology, focusing on the AS level and specific regional areas. It utilizes passive mapping and public datasets, such as BGP table dumps from RV and RIS. While some studies concentrate on mapping within a particular country [6, 7], this research offers guidelines that can be applied more broadly, not limited to a particular country. Other studies have explored broader mapping areas, such as continents [8, 9, 10], or have compared and studied multiple countries [11, 12]. This research offers a unique contribution by presenting a comprehensive guide to the process of mapping internet topology, filling a gap that has not been addressed in previous studies, and serves as a point of reference for similar research endeavors in other regions.

This paper is further organized as follows: Section II discusses an overview of the topic of internet topology mapping, Section III explains the data and methodology used to carry out regional mapping of internet topology, Section IV discusses a comprehensive guide to implementing regional internet topology mapping as a result of this research and its evaluation, Section V discusses the contribution of this research among similar studies that conducted regional internet topology mapping along with its limitations and future research, while Section VI provides the conclusions of this research.

II. Literature Review

In the early days of the internet, through the presence of The Advanced Research Projects Agency Network (ARPANET) around 1969, researchers were able to directly draw detailed and accurate maps of the internet network. This could be done because all parts of the internet network at that time were still very expensive and installation was also very complicated, plus there were only a few people involved in operating the network. As the internet network continues to develop and its complexity also increases, now no one can draw maps of the internet like researchers did at the beginning of the internet. Around 1995 the internet shifted from a large, monolithic network to a very diverse "network of networks" called an autonomous system (AS). Nowadays the internet has become public internet, because it is formed by various ASes owned by various organizations and companies, namely: internet service provider (ISP), transit provider, network service providers, giant companies to small businesses, academic and research organizations, content providers, content distribution network, web hosting companies, and cloud providers. This transition draws the attention of the research community to the structure of internet connectivity which is predominantly based on economics (economics driven) and is usually referred to as the AS graph of the internet. AS graph is a logical internet topology where nodes represent ASes and edges reflect the relationships between ASes. Even though AS graph does not provide a physical picture of internet connections nor does it provide an actual picture of data traffic on the internet, interest in AS graph has generated numerous research publications since 1995. The resulting publications are related to various aspects, including measuring, modeling and analyzing internet topology at the AS level and its evolution over time [3].

Internet topology research is conducted mainly at AS level due to its highest granularity and data set availability. The AS relationship on the internet topology is divided into three types: peer-to-peer (P2P), client-to-providers (C2P), and siblings-to-siblings (S2S) [1, 13]. The majority of internet topology researchers at the AS level conduct research with a global scope, but only a few conduct research limited to countries or regions. Below are the researchers who focus on specific countries and regions in their research. Wahlisch et al. analysed the AS structure in Germany, likewise Berenguer and Pintor studied AS level internet topology in Latin America and the Caribbean [6, 8]. Similarly, research on AS level internet topology analysis is also conducted in Southeast Asia, including the internet topology of Thailand by Siritana et al. [7]. Furthermore, Gregori et al. and Fanou et al. have studied internet topology in Africa [9, 10]. Cakmak and Aydin also studied internet topology in 25 countries, while Leyba and colleagues concentrated on the internet borders of eight selected countries [12, 11].

Seeing the importance of the availability of internet topology maps at the AS level, researchers have made various efforts to estimate relations between ASes through evaluating public dataset such as BGP table dump [14, 15, 16]. Despite numerous attempts to collect internet topology data at the AS level, obtaining a complete internet topology is extremely difficult and even almost impossible. Even though getting complete AS data is quite easy, the most difficult part is getting information on relations between ASes. There are two main factors that hinder efforts to obtain comprehensive internet topology data: the nature of internet administration is decentralized, and network management is exclusively held by the commercial ISP. For these reasons, obtaining an accurate internet topology remains an ongoing and essential research challenge in this sector. A worse outcome may result from a poor depiction of the internet network topology, considering the large number of advanced studies that use this topology as a basis for experiments or simulations [17].

Researchers used BGP table dump to create a representative diagram of AS level internet topology. In the representation of graphs, nodes represent AS with a particular AS Number (ASN), edges represent connections between two ASes [3]. ASes use BGP to communicate with each other and exchange routing information. Regarding BGP, there are two long-running projects that continue to collect BGP table dumps on many Route Collectors (RCs) around the world, namely: University of Oregon's Route Views project (RV) and RIPE Network Coordination Center's Route Information Services (RIS). Researchers used BGP table dumps from RV and RIS as their source of experimental results, but there are also other sources, such as trace-route probes, Internet Routing Registry (IRR), and Internet Exchange Point (IXP) looking glass [18, 19].

III. METHODS

This research initially mapped internet topology at the AS level using the BGP table dump dataset, with a focus on Indonesia's national internet topology as a selected case study. The knowledge gathered from each stage of internet topology mapping is then used to create a complete guide that covers the stages required in internet topology mapping at the AS level in a given regional area. The presence of a thorough guide to internet topology mapping processes, particularly for specific regional areas, distinguishes this research from earlier studies that have also undertaken internet topology research in specific areas.

The internet topology mapping case study was carried out in a limited area, namely the Indonesia's national internet topology, by inferring relationships between ASes. Relations between ASes are generally private and confidential, so efforts are needed to infer relations between ASes so that the internet topology can be mapped. Efforts to infer relationships, or also called relationship inference method, between ASes are carried out by carrying out an inference process against BGP table dump. BGP table dump as the source of inference will be taken from BGP public dataset which is up to date. The knowledge gained from each stage in conducting internet topology mapping is then used to build a comprehensive guide on internet topology mapping methods at the AS level in a particular regional area, explained in Section IV.

The internet topology mapping method chosen is passive, namely by using BGP table dump provided by two public dataset, namely Routing Information Services (RIS) by RIPE Network Coordination Center and Route Views project (RV) by University of Oregon. Both organizations have a repository of BGP table dump data that is up to date and collected historically. The BGP table dump data is collected by their Route Collector (RC) devices spread across various geographic areas in the world. Each RC is connected to a BGP router which becomes a Vantage Point (VP), so that the RC can obtain BGP table information from the partial internet topology from the point of view of each VP. This provides an opportunity to carry out fairly comprehensive passive mapping of internet topology. Meanwhile, active internet topology mapping requires ownership of access to many VP points in various geographic areas in the internet topology. Due to the enormous need for resource to collect information independently via VP in various geographic locations, active mapping was not chosen in this research.

A. Data Sources

The main research data source is BGP table dump taken from Routing Information Service (RIS) [20] and from Route Views project (RV) [21]. Additional data sources include a list of AS Number from Asia Pacific Network Information Center (APNIC) [22], information on AS Organization [23] and AS Relationship [24] from Center for Applied Internet Data Analysis (CAIDA), as well as AS' related information from PeeringDB (PDB) [25].

The main data source for the research is BGP table dump provided by public dataset RIS and RV. BGP table dump will provide BGP paths data which can be processed by relationship inference method to become AS type-of-relationship thereby forming an internet topology map at the AS level. RIS dataset accessed via the following repository [20], and RV dataset accessed via the following repository [21]. Information regarding delegated AS Number is taken from APNIC. Data from APNIC accessed from the following page [22]. Information on AS Organization and AS Relationship is taken from CAIDA, data from CAIDA are accessed via the following repositories [23, 24]. Other AS related information is taken from PDB, data items from the PDB are accessed from the following CAIDA repository page [26].

B. Research Steps

This section describes the research procedures used to map internet topology at the AS level using BGP table dump dataset, with a case study on the Indonesia's national internet topology. The research step begins by collecting datasets from the main data source and additional data sources. Datasets that have been collected are then read, sanitized and processed so that they can be mapped into Indonesia's national internet topology along with the results of analysis of its characteristics. The sequence of research steps is: acquire datasets from both primary and secondary data sources, reading and cleaning datasets, mapping a national internet topology from datasets, and finally analyzing the characteristics of the national internet topology.

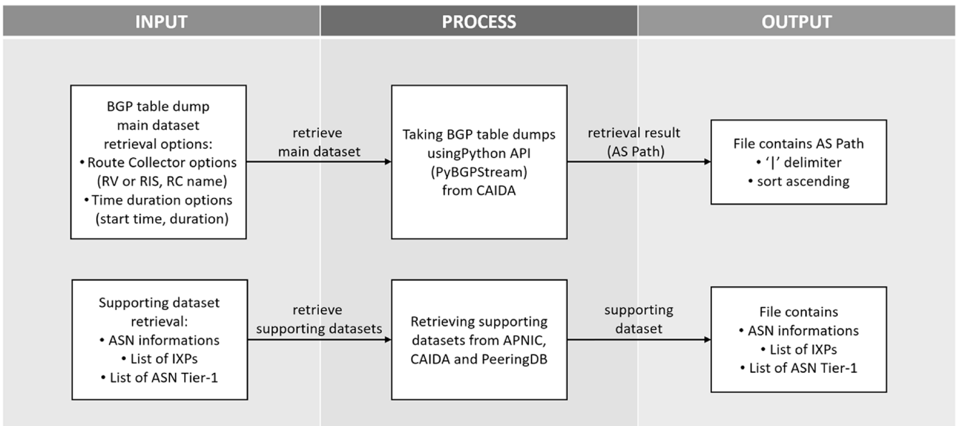


Fig. 1 Acquire datasets from both primary and secondary data sources

The first research step, namely acquiring datasets from both primary and secondary data sources, is depicted in Fig. 1. The main datasets in the form of BGP table dump are collected from public dataset provided by RIS and RV according to the specified time period. Supporting dataset, such as: Tier-1 ASN lists, IXP lists, and other supporting data, are also collected as needed, utilizing data sources from public datasets belonging to APNIC, CAIDA, and PDB. The process of retrieving the main dataset in the form of Routing Information Base (RIB) from BGP table dump belonging to RV or RIS is processed using PyBGPStream [27] from CAIDA. The data taken specifically using PyBGPStream from selected RCs is AS Path, the results of taking AS Path are stored in the `.txt` file. AS Path is marked with delimiter `'|'` as a separator between ASN, and has been sorted according to the ASN order per line in ascending.

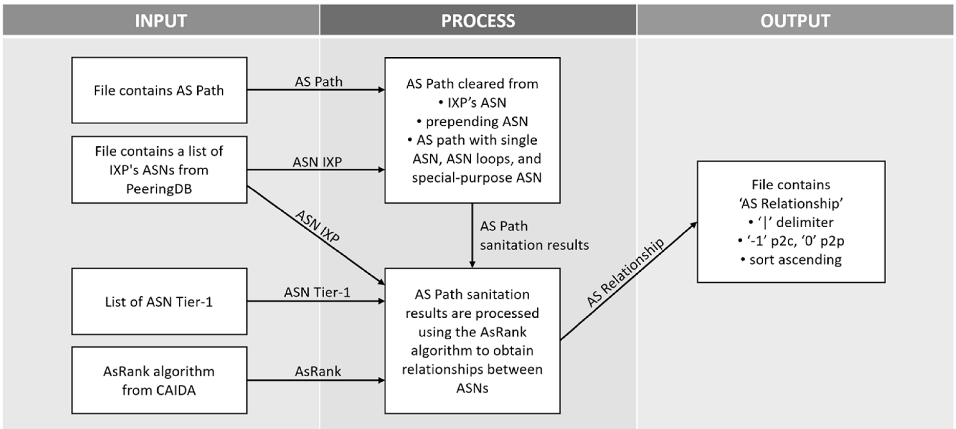


Fig. 2 Reading and cleaning datasets

Fig. 2 depicts the flow of the second research step, namely reading and cleaning datasets. AS Path that has been obtained needs to be cleaned first from unnecessary data. Some data that needs to be cleaned, for example, include ASNs that appear more than once in AS path or ASNs that are reserved. AS Path also needs to be combined with additional dataset from PeeringDB that has been collected by CAIDA on the repository page [26]. The sanitization process is carried out to clean the collected AS Path dataset. The sanitation carried out to clean is: remove IXP's ASN, remove prepend ASN, pass single ASN, pass ASN loop (non unique ASN), and pass special-purpose ASN. The `.txt` file containing the sanitized AS Path is then sorted in ascending order. The process of obtaining relationships between ASNs is carried out using the AS Rank [14] algorithm from CAIDA. The results of the relationship between ASNs are sorted based on the order of ASNs per line in ascending order in the `.txt` file. The file contains Peer-to-Peer (P2P) and Customer-to-Provider (C2P) information, with the following format [24]: `<Provider-ASN>|<Customer-ASN>|-1` or `<Peer-ASN>|<Peer-ASN>|0`.

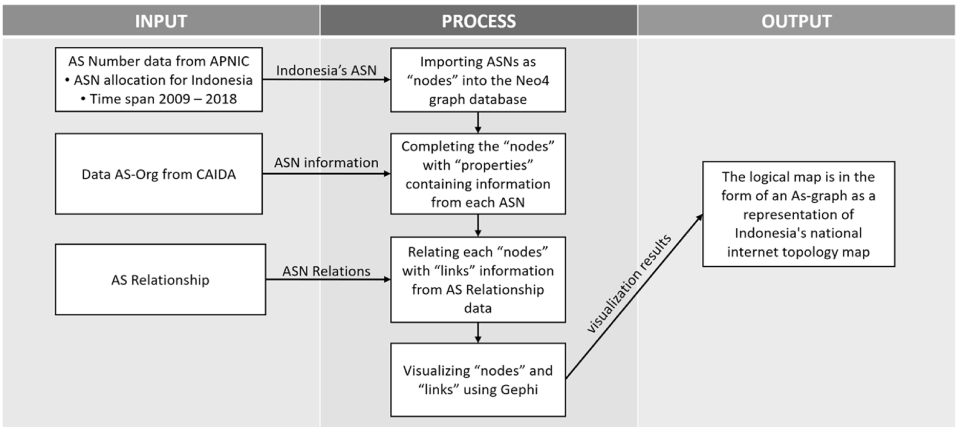


Fig. 3 Mapping a national internet topology from datasets

The next step as the third research step, namely mapping a national internet topology from datasets, is explained in Fig. 3. Topology mapping process is carried out by adding relationships between Indonesia's ASes based on ASNs from APNIC, using the relationship between AS obtained from the previous process. The mapping process yielded a logical map of Indonesia's national internet topology in the form of an AS graph representation, which at the AS level, can offer a logical description of the form of Indonesia's national internet topology. This research took Indonesia's ASN from APNIC [22]. Until 2018, Indonesia has allocated 1452 ASNs, the first ASN being allocated in 1994. This research took the AS Organizations (AS Org) [23] dataset from CAIDA. ASN datasets from APNIC are transformed to comma separated value (CSV) and put into Neo4j as a graph database. From 2009 to 2018, ASN is imported into Neo4j as a node each year, and every node has `"label:AsYY"` and `"property:ASN"` (YY: year). Once all the nodes `"label:AsYY"` with `"property:ASN"` are created, another property is added to every associated ASN nodes, based on the AS Organizations of CAIDA data set. Following this step, the node `"label:AsYY"` has the following properties of additional properties: `"ASN, ASname, ORGID, ORGName, Country, Source"`. The next step is to construct a link between two connected nodes using the relationship information from from the previous process. Based on the information contained in the dataset, the data for each ASN pair in the dataset is matched with the data in the graph database using Cypher as the graph query language in Neo4j as the graph database used. The results of nodes and links from Indonesia's ASNs stored in Neo4j are exported to Gephi, as a graph analysis and visualization tool used, to be displayed as a logical map of Indonesia's internet topology as shown in Fig. 5.

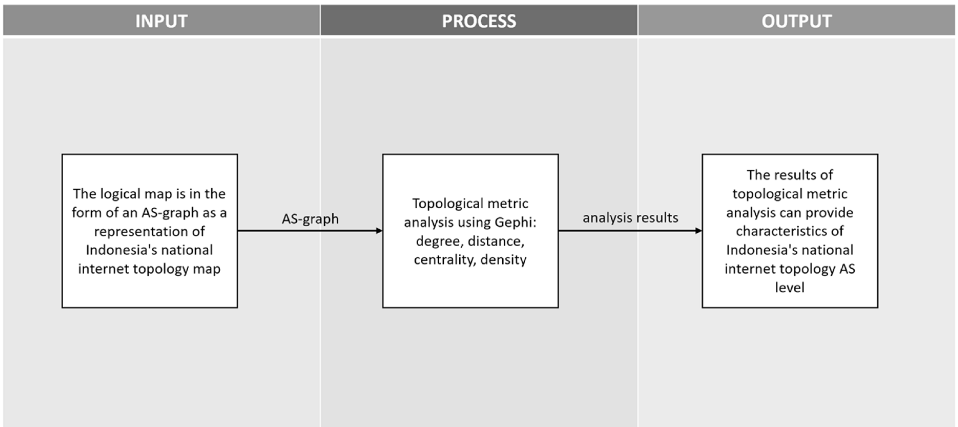
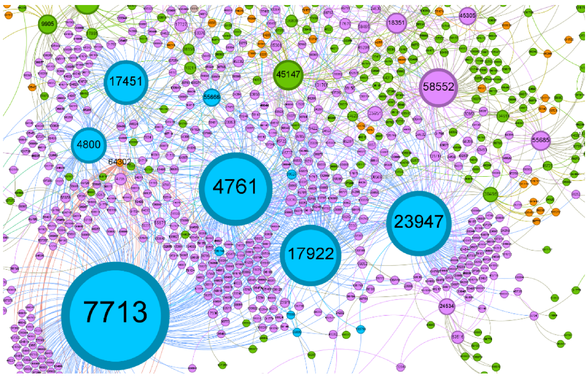


Fig. 4 Analyzing the characteristics of the national internet topology

Fig. 4 presents the process flow for the fourth research stage, which is analyzing the characteristics of the national internet topology. The resulting AS graph is then analyzed using several topological metrics, such as: degree, distance, centrality, and density. The topological metric analysis results can give AS level properties of Indonesia's national internet topology. This study used Gephi as a tool for analysis and the creation of a representation of Indonesia's national internet topology at the AS level using simple undirected graphs. The data table processed by Gephi for the analysis and visualization process looks like in Fig. 5. The discussion regarding the results of the visualization and analysis of the AS graph of Indonesia's national internet topology is fully explained by the author of this paper in another paper [13] and is not the focus of the discussion in this paper.



year	country	asn	Degree	Eccentricity	Betweenness Centrality	Authority	Hub
2018	ID	7713	279	5.0	0.279864	0.345067	0.34497
2018	ID	4761	197	5.0	0.16114	0.355047	0.354966
2018	ID	23947	195	4.0	0.222869	0.268955	0.268919
2018	ID	17922	169	5.0	0.066424	0.292854	0.292774
2018	ID	58552	120	4.0	0.124925	0.144228	0.144211
2018	ID	17451	104	5.0	0.107539	0.145236	0.145234
2018	ID	58463	104	5.0	0.042247	0.115211	0.115194
2018	ID	56258	103	5.0	0.094552	0.158301	0.158294
2018	ID	4800	91	5.0	0.051565	0.07492	0.074902
2018	ID	45147	87	5.0	0.0679	0.13184	0.131828
2018	ID	64302	58	6.0	0.002692	0.059769	0.059746
2018	ID	4787	50	6.0	0.030748	0.023859	0.023855
2018	ID	18351	46	5.0	0.036843	0.051196	0.051198
2018	ID	55685	45	5.0	0.032053	0.048569	0.048567
2018	ID	24534	38	5.0	0.007666	0.028127	0.028122
2018	ID	7717	36	5.0	0.022932	0.069998	0.070001
2018	ID	9905	35	6.0	0.016373	0.01678	0.016777

Fig. 5 Left-side: Visualization of AS graph using Gephi as a logical map representation of Indonesia's national internet topology at the AS level; Right-side: Results of topological metric analysis of the AS graph using Gephi

IV. RESULTS

In accordance with the authors' experience in mapping Indonesia's national internet topology, where the outline of the method is written in Section III, in this article the authors have created a comprehensive guide to each process in mapping internet topology at a regional level. This detailed guide can be useful for researchers to further understand the mapping process and can be a useful instrument for conducting similar research in different regional areas of the internet topology. This comprehensive guide has been tested in one selected country to evaluate the results of the study, accompanied by assessments from experts which are also used as an evaluation of the research results.

A. Research Result

The results of the documentation of each process carried out by the author in carrying out regional-based internet topology mapping, allows conclusions to be drawn regarding steps that can serve as guidelines for mapping internet topology in certain regional areas. The comprehensive flow of the mapping procedure is depicted in the flow diagram in Fig. 6, providing a visual representation of the entire sequence of actions involved in this mapping process. This visualization helps in understanding the complexity of internet topology mapping and acts as a resource aid to perform similar procedures in other cases of regional topology mapping.

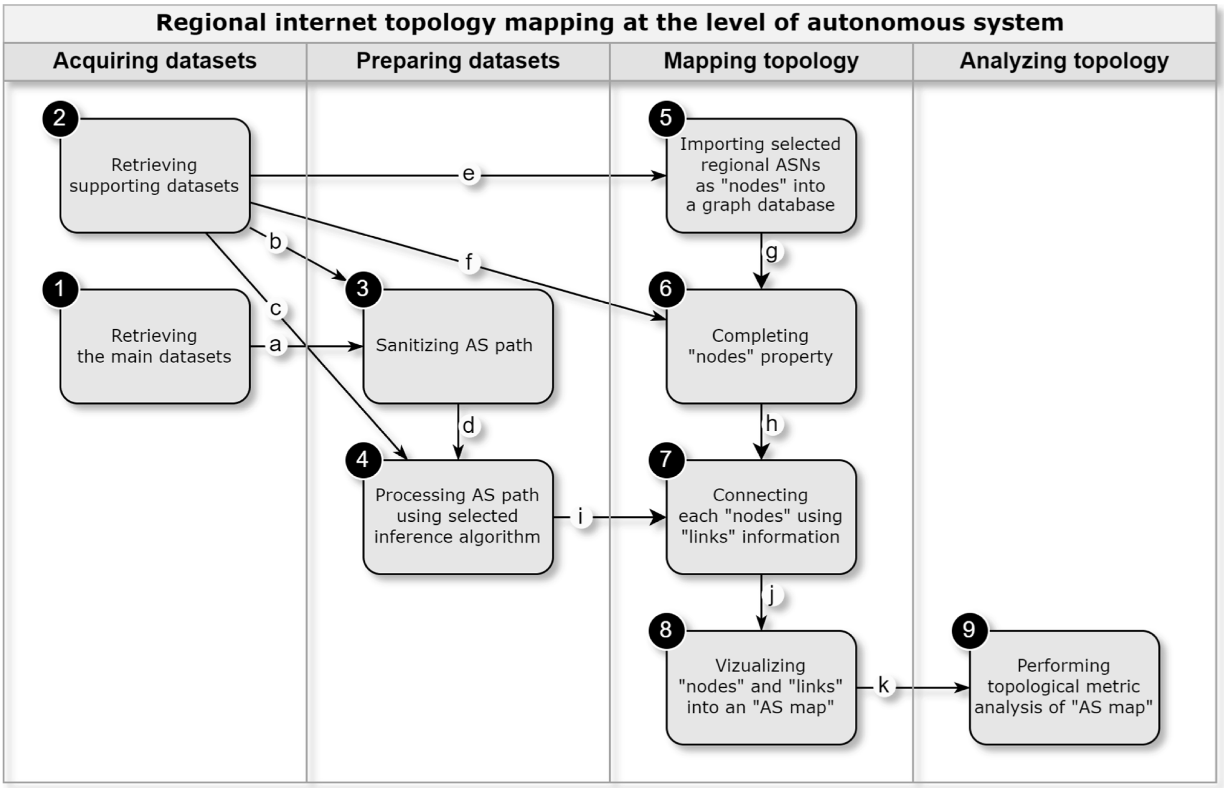


Fig. 6 Methodological guidelines for internet topology mapping in certain regional areas

The nine steps involved in mapping internet topology are represented visually in Fig. 6 and are divided into four groups, each of which focuses on separate aspects of the whole process. The first category, acquiring datasets, emphasizes the significance of compiling the primary dataset as well as auxiliary datasets, which serve as the basis for the mapping procedure and offer crucial data for further studies. In order to assure correctness, the major goal of the second category, preparing datasets, is to sanitize the AS paths derived from the main dataset by eliminating any unnecessary or incorrect data. To facilitate the construction of connections between various ASes, a suitable inference technique is also utilized to process the AS paths and produce "links" information between ASes. In the third category, mapping topology, AS Numbers are imported into a graph database as "nodes", with each AS Number having specific properties attributed to it. A comprehensive visualization of the internet layout is produced by connecting these "nodes" with each other utilizing "links" information between ASes to produce an "AS map" visual representation. The fourth category, analyzing topology, finally describes how the generated "AS map" can be examined according to certain needs. This enables researchers and analysts to derive important insights by using different analysis methods to obtain a deeper comprehension of the internet topology. Table 1 provides a detailed description of each step mentioned in Fig. 6, acting as a reference point for understanding the internet topology mapping process. These resources provide useful information that can aid research in the field of internet topological mapping and provide further knowledge in this field.

TABLE 1
DETAILED EXPLANATION OF EACH STEP FROM FIGURE 6

<table>
<tr><th>Step</th><th>Stage</th><th>Explanation of the step</th></tr>
<tr><td>(1) Retrieving the main datasets</td><td>Process</td><td>BGP table dumps are fetched using PyBGPStream from CAIDA with specified Route Collector and Time Duration options.
Router Collector (RC) can be selected from:
• Route Views' RC
o https://www.routeviews.org/routeviews/index.php/collectors/
o BGP table dumps on Route Views' RC are collected every 2 hours
• RIPE RIS' RC
o https://ris.ripe.net/docs/route-collectors/
o BGP table dumps on RIPE RIS' RC are collected every 8 hours</td></tr>
<tr><td></td><td>Output</td><td>(a) BGP table dump (with selected Route Collectors and Time Duration)</td></tr>
<tr><td>(2) Retrieving supporting datasets</td><td>Process</td><td>Supporting datasets are drawn from APNIC, CAIDA, and PeeringDB
• AS informations
o Selected regional ASNs from APNIC or other RIRs (https://ftp.apnic.net/stats/apnic/)
o AS Organizations from CAIDA (https://www.caida.org/catalog/datasets/as-organizations/)
• List of IXPs from Peering DB (https://publicdata.caida.org/datasets/peeringdb/)
• List of ASN Tier-1 from CAIDA's AS Rank (https://www.caida.org/catalog/datasets/as-relationships/)</td></tr>
<tr><td></td><td>Output</td><td>(b) Supporting dataset: ASN informations, list of IXPs, list of ASN Tier-1
(c) File contains a list of IXP's ASNs from PeeringDB, list of ASN Tier-1, AS Rank algorithm from CAIDA
(e) AS Number data from APNIC (ASN allocation for selected regional and time span)
(f) AS Org from CAIDA</td></tr>
<tr><td>(3) Sanitizing AS path</td><td>Input</td><td>(a) BGP table dump (with selected Route Collectors and Time Duration)
(b) Supporting dataset: ASN informations, list of IXPs, list of ASN Tier-1</td></tr>
<tr><td></td><td>Process</td><td>AS path are cleared from
• IXP's ASN (IXP's ASN resulting from multilateral peering connection at IXP)
• Prepend ASN (multiple copies ASN for traffic engineering purposes)
• Single ASN (single ASN in an AS path)
• ASN loops (same ASN repeats itself in an AS path)
• Special-purpose ASN (https://www.iana.org/assignments/iana-as-numbers-special-registry/iana-as-numbers-special-registry.txt)</td></tr>
<tr><td></td><td>Output</td><td>(d) AS path</td></tr>
<tr><td>(4) Processing AS path using selected inference algorithm</td><td>Input</td><td>(c) File contains a list of IXP's ASNs from PeeringDB, list of ASN Tier-1, AS Rank algorithm from CAIDA
(d) AS path</td></tr>
<tr><td></td><td>Process</td><td>AS path sanitation results are then processed using AS Rank or other inference algorithms to infer "links" between AS.
AS Rank result format:
• <Provider-ASN>|<Customer-ASN>|-1
• <Peer-ASN>|<Peer-ASN>|0</td></tr>
<tr><td></td><td>Output</td><td>(i) ASN relationship data</td></tr>
<tr><td>(5) Importing selected regional ASNs as "nodes" into a graph database</td><td>Input</td><td>(e) AS Number data from APNIC (ASN allocation for selected regional and time span)</td></tr>
<tr><td></td><td>Process</td><td>Selected regional ASNs from APNIC or other RIRs are imported as "nodes" into the Neo4j or other graph databases.
• AS Number information is taken from RIR that carry out ASN assignments in selected regions, for example APNIC (https://ftp.apnic.net/stats/apnic/)</td></tr>
<tr><td></td><td>Output</td><td>(g) Regional ASN as nodes</td></tr>
<tr><td>(6) Completing "nodes" property</td><td>Input</td><td>(f) AS Org from CAIDA
(g) Regional ASN as nodes</td></tr>
<tr><td></td><td>Process</td><td>ASNs "nodes" are completed with "properties" containing information from each ASN.
• Properties are added to the AS Number as nodes, one source of which could be AS Org information from CAIDA (https://www.caida.org/catalog/datasets/as-organizations/)</td></tr>
<tr><td></td><td>Output</td><td>(h) Regional ASN as nodes with properties</td></tr>
</table>

TABLE 1 (CONTINUED)

Step	Stage	Explanation of the step
(7) Connecting each "nodes" using "links" information	Input	(h) Regional ASN as nodes with properties (i) ASN relationship data
	Process	ASNs "nodes" are related each other with "links" information from AS relationship data. • The relationships between AS Numbers as nodes can be taken from the results of process step (4), or can directly use relationships from public repositories such as AS Rel from CAIDA (https://www.caida.org/catalog/datasets/as-relationships/)
	Output	(j) Regional ANS as nodes with properties and links connected
(8) Vizualising "nodes" and "links" into an "AS map"	Input	(j) Regional ASN as nodes with properties and links connected
	Process	ASNs "nodes" and "links" are visualized using Gephi or other graph visualization tools into an AS Map. • Visualizations can be selected using a layout algorithm that suits research needs, one of which can be using the layout feature from Gephi (https://gephi.org/features/)
	Output	(k) Complete regional AS graph
(9) Performing topological metric analysis of "AS map"	Input	(k) Complete regional AS graph
	Process	AS map is analyzed based on the required topological metrics using Gephi or other graph analysis tools. • The resulting AS map can be analyzed using several topological metrics, such as: "degree, distance, centrality, and density"

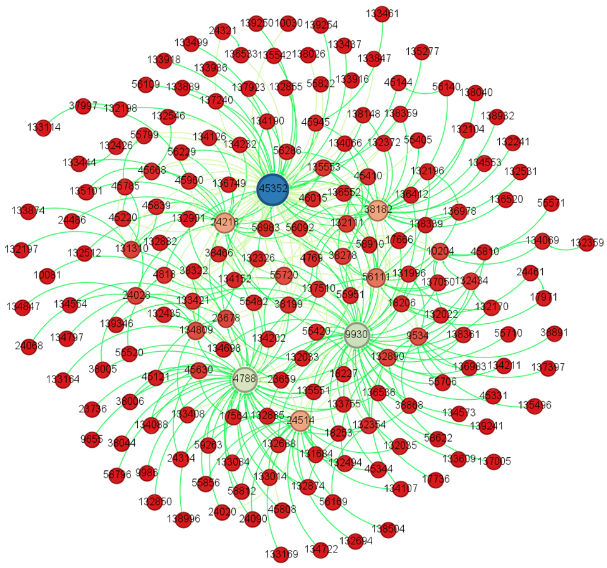
B. Research Evaluation

This research produces a comprehensive guide to internet topology mapping at the AS level with regional coverage, and can be applied to help internet stakeholders understand and carry out the internet mapping process in certain regional areas. As an evaluation of the research results, the authors uses this mapping guide to map the internet topology at the AS level in one selected country. Malaysia's internet topology at the AS level, as a selected regional area, was chosen as a means of implementing the comprehensive guidelines resulting from this research. The implementation of each step of the guide are further explained in Table 2.

TABLE 2
IMPLEMENTATION OF EACH STEP OF THE GUIDE

Steps of the guide	Notes from the implementation of each step of the guide
(1) Retrieving the main datasets	Taking the main dataset only on selected RCs
(2) Retrieving supporting datasets	Retrieve supporting datasets that match the timestamp of the main dataset
(3) Sanitizing AS path	AS path sanitation is carried out according to predetermined scenarios
(4) Processing AS path using selected inference algorithm	The inference algorithm used is AS Rank [14]
(5) Importing selected regional ASNs as "nodes" into a graph database	The graph database used is Neo4j version 5.11.0 community
(6) Completing "nodes" property	Additional property information for each node is taken from the CAIDA AS Organization [23] data source
(7) Connecting each "nodes" using "links" information	The total number of "nodes" is 275 and connected by a total of 391 "links", 284 links are C2P and the other 107 are P2P
(8) Vizualising "nodes" and "links" into an "AS map"	AS Map visualization was carried out using Gephi version 0.10.1 with the Fruchterman Reingold layout algorithm
(9) Performing topological metric analysis of "AS map"	Analysis was carried out using Gephi version 0.10.1 with several standard topological metrics, such as degree, distance, centrality, and density

Every step of the comprehensive guide as a result of this research has been successfully implemented. Table 2 lists each step taken during the implementation, while Fig. 7 shows the final results of the implementation in the form of a map and several metrics of internet topology at the AS level for the country of Malaysia with data sources in 2019.



ASN	Degree	Eccentricity	Betweenness Centrality
4769	5	4.0	1.70119
4788	57	4.0	4135.204257
4818	6	4.0	215.811111
9534	13	3.0	704.401476
9655	1	5.0	0.0
9930	59	3.0	4220.894908
9986	1	5.0	0.0
10030	1	4.0	0.0
10081	1	5.0	0.0
10204	10	3.0	459.134149
17564	3	4.0	3.396429
17666	2	4.0	0.0
17736	1	4.0	0.0
17971	2	5.0	0.0
18206	4	4.0	364.0
18227	2	4.0	0.0
18253	2	4.0	0.0
23659	2	4.0	0.0
23678	10	4.0	307.707179
23736	1	5.0	0.0
24020	2	5.0	0.0
24028	9	4.0	417.676984
24068	1	5.0	0.0

Fig. 7 Implementation results of the internet topology mapping guide at the AS level for the regional area of Malaysia, with data sources in 2019 (left side: topological map visualization results, right side: snippet of calculation results of several topological metrics)

Evaluation was also carried out by asking for assessments from experts in the field of computer networks and the internet regarding the results of this research. A total of nine experts provided assessments as an evaluation of the comprehensive guidelines for internet topology mapping produced through this research. Table 3 presents the background of each expert who carried out the assessment. On average, these experts have ten years of experience in the field of computer networks and the internet. Meanwhile, in terms of professional background, these experts are divided into: researchers, lecturers, practitioners, and administrators of computer networks and the internet.

TABLE 3
THE IDENTITY AND BACKGROUND OF EACH EXPERT PROVIDING FEEDBACK FOR THE EVALUATION

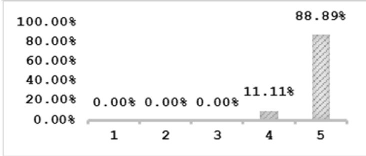
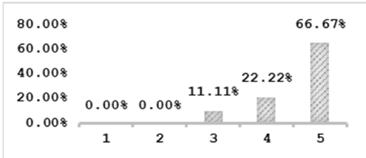
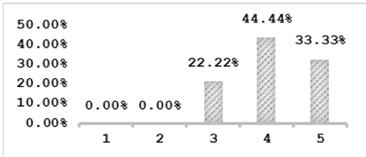
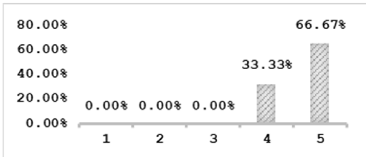
Name initials	Length of experience (years)	Profession in the field of computer networks and the internet			
		Researcher	Lecturer	Practitioner	Administrator
DJS	7		✓		✓
AF	20			✓	
CEHP	13		✓	✓	✓
SFS	12	✓	✓		✓
OML	7	✓	✓	✓	✓
RAN	3				✓
MTK	8	✓	✓		
DWW	6			✓	✓
TAS	18			✓	✓

* Average: 10 years

As an evaluation of the results of this research, experts were asked to assess four statements, with detailed assessment results written in Table 4. The assessment is carried out by giving a score on a scale of ‘1 to 5’, with a score of ‘1’ means strongly disagree while a score of ‘5’ means strongly agree. For statements number (1), (2), and (4), the majority of experts gave a score of ‘5’, which means they strongly agree with each statement. Meanwhile, for statement number (3), the answers are spread across a score of ‘3 to 5’, with a score of ‘4’ slightly higher. In general, it can be concluded from the evaluation assessment that the experts agreed with the statements related to the research results.

The Cronbach's alpha calculation yielded a value of 0.75269, indicating acceptable reliability for the assessment instrument. Each statement item ranged from 0.45582 to 0.81929, surpassing the general minimum threshold of 0.3. Comprehensive results of the reliability analysis can be found in Table 5. The outcome highlights the resilience of the assessment instrument and offers important information about how well it measures the targeted objectives.

TABLE 4
EXPERTS' ASSESSMENT RESULTS ON STATEMENTS REGARDING THE RESULTS OF THIS RESEARCH

List of statements assessed by experts regarding research results	Number of experts who chose the rating scale *					Percentage graph of assessment results (according to rating scale 1 to 5)
	1	2	3	4	5	
(1) The existence of network topology map documentation is important for network administrators or internet stakeholders.	0	0	0	1	8	
(2) The research results in the form of visualization of a logical map of Indonesia's national internet topology at the autonomous aystem level are useful in providing an overview of Indonesia's internet topology.	0	0	1	2	6	
(3) The flow of steps from the research results in the form of "a comprehensive guide to regional internet topology mapping at the autonomous system level" can be understood.	0	0	2	4	3	
(4) The results of the research in the form of "a comprehensive guide to regional internet topology mapping at the autonomous system level" can provide guidance for people who need to carry out internet topology mapping.	0	0	0	3	6	

* Scale 1 to 5: 1 strongly disagree, 5 strongly agree

TABLE 5
RESULTS OF RELIABILITY ANALYSIS OF THE ASSESSMENT INSTRUMENT

Statement	Item-rest correlation
S1	0.52764
S2	0.60698
S3	0.45582
S4	0.81929

* Scale Reliability Statistics: Cronbach's α = 0.75269; McDonald's ω = 0.81165

V. DISCUSSION

This research produces an internet topology mapping method with regional coverage, and can be applied to map and analyze internet topology in certain regional areas. This method can be used in similar research that will carry out internet topology mapping in certain areas, by following the sequence of steps in the guide that has been produced by this research. This research differs from earlier studies in that it offers a thorough manual for the process of mapping the internet topology, something that earlier studies did not offer.

A. Research Contribution

This research provides a contribution in the form of a comprehensive guide to internet topology mapping, as well as being a reference for similar research projects in other regional areas. The majority of AS level internet topology researchers conduct global study, while just a few focus on a single nation or area. Table 6 shows the position of this research when compared with other studies that have carried out internet topology mapping in limited areas.

The unique position of this research is that it is the only research that focuses on discussing a complete guide to the internet topology mapping process specifically for selected regional areas. Meanwhile, for data sources, all research agreed to carry out passive mapping with the majority using public datasets in the form of BGP table dumps from RV and RIS. Regarding regional coverage in topological mapping, this research and several other studies both focus on a particular country [6, 7]. Nevertheless, this research produces guidelines that can be used generally and not only apply to the case study country of Indonesia. Several other studies have taken a wider topological mapping area, for example within the scope of a continent [8, 9, 10]. Meanwhile, there are also several studies that take several selected countries as comparisons or case studies [11, 12].

TABLE 6
LIST OF STUDIES THAT CARRY OUT INTERNET TOPOLOGY MAPPING IN CERTAIN REGIONAL AREAS

Selected Regional Areas	References	Data Sources	Study Results
Indonesia	this research	BGP table dump (RIS, RV), AS Number (APNIC), AS Organization and AS Relationship (CAIDA), AS' related information (PDB)	This research produces a comprehensive flow of internet topology mapping procedures that can provide a visual representation of the entire series of actions involved in the mapping process. The flow visualization along with a detailed description of the methodology helps in understanding the complexity of internet topology mapping and acts as an excellent resource for carrying out similar mapping procedures in specific regional contexts.
8 selected countries	[11]	CAIDA AS relationship, Team Cymru ASN lookup service, Freedom House's Freedom On The Net (FOTN) report	The research findings suggest that over the last decade, some countries have tightened their control over internet access, while others have shown signs of increased openness due to the development of new connections to the global internet. Additionally, the study has identified a correlation between national chokepoint potential and measures of openness such as Freedom on the Internet and Freedom of the Press, with statistically significant associations found in both cases.
Latin America and the Caribbean	[8]	Team Cymru's WHOIS service, IANA's web page, RIPEstat project API, IXP web pages, The World Bank's web page, CIA's World Factbook, CAIDA's AS Relationship Inference algorithm, Routing data from RIS, RV, PCH and LGs	This study focuses on regional topology analysis in Latin America and the Caribbean (LAC region) by enhancing AS-level graphs using different connectivity metrics. It reveals that basic metric comparisons may not be enough to identify connectivity issues. Country-level studies reveal correlations between graph characteristics and socioeconomic indicators, identifying countries for Internet Exchange Point deployment. This helps in assessing the interconnection level and regional internet robustness, highlighting the need for improved connectivity in the region.
Africa	[9]	BGP data by Isolario, RIS and Route Views, AFRINIC registry, Maxmind GeoLite2 Country4, Inter-AS economic inferences through a hygiene phase on BGP data	The AS-level ecosystem in Africa reflects a range of cultural, economic, and developmental characteristics. South Africa stands out as the most developed, boasting a peering ecosystem akin to that of European countries. Progressive nations such as Egypt, Kenya, Nigeria, and Tanzania are witnessing an expansion in internet accessibility, while many others are still in the nascent stages of development. Nevertheless, the limitations of BGP route collectors in uncovering peering connectivity among ASes have impacted graph analysis. The introduction of ten new full feeders could potentially address this issue in most African countries.
25 selected countries	[12]	RIB entries that are collected from RIS and Route Views projects, each AS is later assigned to a country by using datasets from the Regional Internet Registries, CAIDA's AS Relationships dataset	The study analyzes internet interconnections in a country-specific manner, considering factors like IP prefix announcement behavior, transit links, multi-homing inclinations, and the presence of active peering communities (IXPs). It reveals that countries with well-established IXPs have a competitive transit market, more evenly distributed betweenness centrality among ASes, and increased multi-home tendency for stub networks, enhancing the resilience of national interconnection constructs. IXPs are crucial parts of peering fabrics and can provide complementary benefits for the interconnection ecosystem.
Africa	[10]	RIPE Atlas, OpenIPMap (OIM), Reverse DNS lookups (RDNS), MaxMind GeoIP2City (MM), Team Cymru (TC), RIR delegated files, RIR Databases (WHOIS)	The African Internet's AS level topology, particularly its IXP substrate, is understudied due to a lack of vantage points. A study from 2013 to 2016 improved the RIPE Atlas measurement infrastructure, revealing IPv4 and IPv6 topologies interconnecting local ISPs. The study found a variety of ISP transit habits and their dependence on socio-economic factors. It also revealed the dominance of ISPs outside Africa for intra-continental paths and traffic localization efforts. The study highlights the launch of new IXPs and their impact on connectivity.

TABLE 6 (CONTINUED)

Selected Regional Areas	References	Data Sources	Study Results
Germany	[6]	RIPE database (DB) IP-blocks, Team Cymru, RRC12 of the RIPE RIS	The paper introduces a methodology for identifying and categorizing pertinent ASes within a country, offering a comprehensive examination of crucial internet stakeholders. The study focused on Germany and demonstrated an identification of 25% more ASes compared to prefix-based techniques. The analysis uncovered that entities within the same public or business sector interconnect through both national and international ASes. Furthermore, it identified Deutsche Telekom, Level 3, Lambdanet, Colt, and Versatel as the primary transit ASes for intra-Germany communication.
Thailand	[7]	BGP routing tables from Internet Thailand Public Company Limited (INET), Whois (APNIC)	This paper examines the relationships between Autonomous Systems (AS) in Thailand, focusing on the internet topology and inference type. It reveals that CS LoxInfo Public Co., Ltd (AS4750) is the largest AS in Thailand, followed by CAT telecom (AS9931) and Asia Infonet Co., Ltd (AS7470). These ASes have no direct links and some are connected to multiple providers. Most relationships are C2P, allowing AS administrators to adapt policies based on Thailand's internet map.

B. Limitations and Future Works

The limitation of this study lies in the fact that the guidance provided in Section IV was developed solely based on the experience of conducting research in a specific area, namely Indonesia's national internet topology. To enhance the comprehensiveness of the guide, it would be beneficial to incorporate insights gained from conducting regional topology research in other geographic areas. Although this guide has been validated through its implementation on internet topology mapping in a selected country, further refinement can be achieved by conducting experiments on a larger scale, such as mapping the internet topology across an entire continent or several neighboring countries.

One of the further work that can be done based on this research is to provide an automatic tool that can carry out the flow of each step from the research results in Figure 6. If this automatic tool is available, stakeholders are expected to be able to generate an internet topology map by simply entering time options and regional area options according to their own needs.

VI. CONCLUSIONS

This research proposes a comprehensive guide to map internet topology at the AS level in selected regional areas. This research is different from previous studies, because this research proposes a comprehensive guide for the internet topology mapping process, something that was not done in previous studies. This guide was developed based on the internet topology mapping process at the AS level, with case studies focused on regional areas of Indonesia. In general, this guide consists of the following nine sequential steps: retrieving the main datasets, retrieving supporting datasets, sanitizing AS path, processing AS path using selected inference algorithm, importing selected regional ASNs as “nodes” into a graph database, completing “nodes” property, connecting each “nodes” using “links” information, vizualising “nodes” and “links” into an “AS map”, and performing topological metric analysis of “AS map”.

Evaluation of the comprehensive guide as the result of this research has been carried out, either by testing the application of the comprehensive guide to a mapping case in one selected country, or by asking for assessments from experts in the related field regarding the comprehensive guide. Evaluation from both sides shows that the results of this research are acceptable. The work of this research contributes to the scientific field of internet topology by proposing a comprehensive guide to internet topology mapping specifically at the AS level and in limited regional areas. This comprehensive guide can also serve as a resource to help researchers who need guidance to carry out topological mapping according to their individual cases.

The results of this study are designed to be used effectively in similar research efforts to map internet topology in specific regional areas, particularly at the AS level. Although this approach has been confirmed by its execution on internet topology mapping in a specific area, additional refinement may be performed through larger-scale studies, such as mapping the internet topology over an entire continent or several adjacent countries.

Author Contributions: *Timotius Witono*: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing. *Setiadi Yazid*: Review, Supervision. *Yudho Giri Sucahyo*: Review, Supervision.

All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from Maranatha Christian University, Bandung, Indonesia.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: The data supporting the findings of this study are openly available in public repositories, with source details listed in references numbers [20-26]. However, certain processed data from this study are available from the corresponding author, upon reasonable request.

Institutional Review Board Statement: Not applicable.

Informed Consent: There were no human subjects.

Animal Subjects: There were no animal subjects.

ORCID:
Timotius Witono: https://orcid.org/0000-0002-3495-1002
Setiadi Yazid: https://orcid.org/0000-0002-5291-5144
Yudho Giri Sucahyo: https://orcid.org/0009-0009-5146-8352

REFERENCES

[1] R. Motamedi, R. Rejaie and W. Willinger, "A Survey of Techniques for Internet Topology Discovery," *IEEE Communications Surveys & Tutorials,* vol. 17, p. 1044–1065, 2015.

[2] G. Accongiagioco, "Modeling and Analysis of the Internet Topology," IMT Institute for Advanced Studies, Lucca, 2014.

[3] W. Willinger and M. Roughan, "Internet Topology Research Redux," in *Recent Advances in Networking*, ACM SIGCOMM, 2013, p. 1–59.

[4] Y. He, G. Siganos and M. Faloutsos, "Internet Topology," in *Computational Complexity*, R. A. Meyers, Ed., New, York: Springer, 2012, p. 1663–1680.

[5] M. E. Tozal, "Policy-Preferred Paths in AS-level Internet Topology Graphs," *Theory and Applications of Graphs,* vol. 5, March 2018.

[6] M. Wählisch, T. C. Schmidt, M. de Brün and T. Häberlen, "Exposing a Nation-Centric View on the German Internet – A Change in Perspective on AS-Level," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* vol. 7192 LNCS, p. 200–210, 2012.

[7] H. Siritana, S. Kitisin and S. Gertphol, "A study of autonomous system relationships within Thailand," in *2011 Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2011.

[8] S. Silva Berenguer and F. Valera Pintor, "Radiography of internet autonomous systems interconnection in Latin America and the Caribbean," *Computer Communications,* vol. 119, p. 15–28, April 2018.

[9] E. Gregori, A. Improta and L. Sani, "On the African Peering Connectivity Revealable via BGP Route Collectors," *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST,* vol. 250, p. 368–376, 2018.

[10] R. Fanou, P. Francois, E. Aben, M. Mwangi, N. Goburdhan and F. Valera, "Four years tracking unrevealed topological changes in the african interdomain," *Computer Communications,* vol. 106, p. 117–135, July 2017.

[11] K. G. Leyba, B. Edwards, C. Freeman, J. R. Crandall and S. Forrest, "Borders and Gateways: Measuring and Analyzing National AS Chokepoints," in *Proceedings of the Conference on Computing & Sustainable Societies - COMPASS 19*, New York, New York, USA, 2019.

[12] G. Cakmak and M. N. Aydin, "A country-specific analysis on internet interconnection ecosystems," in *2017 9th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2017.

[13] T. Witono and S. Yazid, "Portrait of Indonesia's Internet Topology at the Autonomous System Level," in *Computational Science and Technology*, Singapore, 2020.

[14] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas and K. Claffy, "AS relationships, customer cones, and validation," in *Proceedings of the 2013 conference on Internet measurement conference - IMC '13*, New York, New York, USA, 2013.

[15] Y. Jin, C. Scott, A. Dhamdhere, V. Giotsas, A. Krishnamurthy and S. Shenker, "Stable and practical as relationship inference with problink," *Proceedings of the 16th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2019,* p. 581–597, 2019.

[16] G. Feng, S. Seshan and P. Steenkiste, "UNARI: an uncertainty-aware approach to AS relationships inference," in *Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies*, New York, NY, USA, 2019.

[17] R. Oliveira, D. Pei, W. Willinger, B. Zhang and L. Zhang, "The (In)Completeness of the Observed Internet AS-level Structure," *IEEE/ACM Transactions on Networking,* vol. 18, p. 109–122, February 2010.

[18] B. Huffaker, M. Fomenkov and k. claffy, "Internet Topology Data Comparison," Cooperative Association for Internet Data Analysis (CAIDA), 2012.

[19] V. Giotsas, A. Dhamdhere and K. C. Claffy, "Periscope: Unifying Looking Glass Querying," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* vol. 9631, p. 177–189, 2016.

[20] RIPE NCC, "Routing Information Service (RIS)," Accessed: August 2023. [Online]. Available: https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/routing-information-service-ris.

[21] University of Oregon, "University of Oregon Route Views Archive Project,", Accessed: August 2023. [Online]. Available: https://archive.routeviews.org/.

[22] Asia Pacific Network Information Centre, "Apnic Allocation and Assignment Reports," Accessed: August 2023. [Online]. Available: https://ftp.apnic.net/stats/apnic/.

[23] Center for Applied Internet Data Analysis, "Inferred AS to Organization Mapping Dataset," Accessed: August 2023. [Online]. Available: https://www.caida.org/catalog/datasets/as-organizations/.

[24] Center for Applied Internet Data Analysis, "AS Relationships," Accessed: August 2023. [Online]. Available: https://www.caida.org/catalog/datasets/as-relationships/.

[25] PeeringDB, "The Interconnection Database," PeeringDB, Accessed: August 2023. [Online]. Available: https://www.peeringdb.com/.

[26] Center for Applied Internet Data Analysis, "Repository of daily snapshots of historic PeeringDB data," Accessed: August 2023. [Online]. Available: https://publicdata.caida.org/datasets/peeringdb/.

[27] Center for Applied Internet Data Analysis, "Python API (PyBGPStream)," Accessed: August 2023. [Online]. Available: https://bgpstream.caida.org/docs/api/pybgpstream.