Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.064523

ARTICLE

Determination of Favorable Factors for Cloud IP Recognition Technology

Yuanyuan Ma[1], Cunzhi Hou[1], Ang Chen[1], Jinghui Zhang[1], Ruixia Jin[2] and Ruixiang Li[3,*]

[1]College of Computer and Information Engineering, Henan Normal University, Xinxiang, 453007, China
[2]Intelligent Medical Engineering, SanQuan Medical College, Xinxiang, 453003, China
[3]Information Engineering University & Key Laboratory of Cyberspace Situation Awareness of Henan Province, Zhengzhou, 450001, China

*Corresponding Author: Ruixiang Li. Email: ruixiang_li@yeah.net

Received: 18 February 2025; Accepted: 07 April 2025; Published: 09 June 2025

ABSTRACT: Identifying cloud IP usage scenarios is critical for cybersecurity applications, yet existing machine learning methods rely heavily on numerous features, resulting in high complexity and low interpretability. To address these issues, this paper proposes an approach to identify cloud IPs from the perspective of network attributes. We employ data mining and crowdsourced collection strategies to gather IP addresses from various usage scenarios, which including cloud IPs and non-cloud IPs. On this basis, we establish a cloud IP identification feature set that includes attributes such as Autonomous System Number (ASN) and organization information. By analyzing the differences in the properties of different IP usage scenarios in the detection results, we can find out the factors that are conducive to cloud IP identification. Experimental evaluation demonstrates that the proposed method achieves a high identification accuracy of 96.67%, surpassing the performance of traditional machine learning models such as CNN, MLP, XGBoost, KNN, SVM, and Decision Tree, whose accuracies range between 81% and 92%. Furthermore, this study reveals that latency and port information exhibit insufficient discrimination power for distinguishing cloud IP from non-cloud IP scenarios, highlighting ASN as a simpler, more interpretable, and resource-efficient criterion. To facilitate reproducible research, datasets and codes are publicly released.

KEYWORDS: Cloud IP identification; organization information; network attributes; IP usage scenario

1 Introduction

Cloud IP identification aims to determine whether an IP belongs to a cloud service provider (such as Alibaba Cloud, Tencent Cloud, AWS, Google Cloud, etc.) and identify its specific service type by analyzing the IP address and its related features. Cloud IP identification has a wide range of applications, especially in the fields of network security [1–4], access control [5,6], and data analysis [7,8].

In recent years, the rapid development and widespread adoption of cloud computing technologies have led to an ever-expanding variety and scale of cloud services, thereby rendering the network environment increasingly complex and dynamic. Traditional IP identification methods often fall short in managing this complexity, creating an urgent need for advanced data processing and intelligent analysis techniques to enhance both accuracy and efficiency. Consequently, both academic and industrial research communities have gradually incorporated machine learning, deep learning, and big data approaches into the study of cloud IP identification. These efforts focus on extracting richer and more nuanced IP feature information from multiple dimensions, with the goal of achieving rapid and precise identification even when confronted with massive datasets and diverse service scenarios. Such endeavors not only offer a theoretical foundation

for refining subsequent methodologies but also underpin practical applications in security defense and resource allocation.

Existing studies rely heavily on high-dimensional features combined with deep neural networks, often overlooking the fundamental question: "Which network attributes actually contribute to the effective and interpretable identification of cloud IPs?" This gap has hindered the development of simpler, computationally-efficient methods that offer clear insights into their decision processes. Moreover, few existing methods systematically examine the discriminatory power of different network features—such as latency, ports, and ASN (Autonomous System Number is a unique identifier assigned to a network operator or organization (e.g., Internet Service Providers, cloud providers). Each ASN distinctly identifies a network entity responsible for managing IP address blocks and routing policies on the Internet)—for IP usage scenario recognition.

To bridge this gap, this study proposes an approach that explicitly investigates network attribute differences between cloud and non-cloud IP scenarios, focusing on identifying favorable discriminative features. By systematically analyzing attributes like latency, port availability, and ASN organizational information across diverse IP scenarios, our work seeks to uncover clear and interpretable factors that enhance cloud IP identification accuracy. Our research is motivated by the practical need for simpler and interpretable methods capable of effectively distinguishing cloud IPs from other IP usage scenarios.

Traditional IP identification methods struggle to effectively manage this complexity, prompting researchers and practitioners to adopt machine learning (ML), deep learning (DL), and big data analytics for cloud IP identification. These approaches focus on extracting richer IP features from multiple dimensions to enhance identification accuracy even in large-scale datasets. Zhou et al. [9] proposed a comprehensive feature extraction method, focusing on geographic locations, path, ports, and WHOIS information, but primarily relied on deep neural networks, resulting in limited interpretability and high computational cost. Wang et al. [10] proposed a deep ensemble learning approach to classify IP blocks into typical usage scenarios, improving accuracy but maintaining high complexity. Li et al. [11] introduced a continuous neural tree model capable of handling complex feature interactions but similarly lacked transparency in decision-making processes. Furthermore, Liu et al. [12] utilized a graph-based framework (GraphCyber) that improved computational efficiency but still depended on deep neural network structures that were difficult to interpret clearly.

Despite these advancements, a clear research gap remains. Specifically, previous studies primarily employ complex DL approaches and large-scale, high-dimensional feature spaces, yet seldom explicitly analyze which network attributes truly contribute to the effective and interpretable identification of cloud IP scenarios [13]. This gap limits our understanding of IP usage scenarios and restricts the development of simpler, interpretable, and more resource-efficient methods. Moreover, few studies have systematically evaluated the discriminatory power of different network attributes—such as latency, port openness, and ASN—in differentiating cloud IPs from non-cloud IPs [14,15].

To address these gaps, our research focuses explicitly on analyzing network attributes to identify the favorable factors for cloud IP identification. We emphasize computational efficiency, interpretability, and resource efficiency, aspects which are frequently overlooked in existing approaches.

The specific objectives of this research are:

- To systematically analyze different network attributes (including latency, port openness, and ASN organization information) to determine their effectiveness in distinguishing cloud IPs from non-cloud IP usage scenarios.

- To explicitly identify favorable network attributes that significantly enhance the accuracy and interpretability of cloud IP recognition.
- To propose a simplified, computationally efficient cloud IP identification method primarily based on ASN organizational attributes, overcoming the complexity and opacity associated with traditional deep learning approaches.
- To provide publicly accessible datasets and reproducible experimental methods, thereby facilitating further research in cloud IP recognition within the research community.

The main contributions of this work are as follows:

- **We propose a cloud IP identification method based on ASN organization information.** The latency connectivity, port opening information, and ASN organization information in the cloud IP network attributes are different from those in the non-cloud IP network attributes. Based on these properties, we use network differences to capture richer correlation feature information. Therefore, we obtain experimental results with higher robustness and generalization.
- **We have achieved effective identification of cloud IP.** We verified that the cloud IP usage scenario and the locally hosted non-cloud IP usage scenario have similar characteristics such as latency and ports. However, these characteristics cannot be used to distinguish between cloud IP and non-cloud IP. On the other hand, IP ranges under the same ASN have similar IP usage scenarios. Based on the experimental results, we summarize the obstacles and advantages of cloud IP usage scenario identification. This provides theoretical support for more accurate identification of cloud IP usage scenarios.
- **Community Contribution.** In order to enable more scholars to study IP usage scenarios, we provide real data to the community and open up our research methods. The open-source repository is available at: https://gitee.com/henan-normal-university_4_0/determination-of-favorable-factors-for-cloud-ip-recognition-technology.git (accessed on 6 April 2025).

The structure of the paper is organized as follows. Section 1 introduces the background and motivation for cloud IP identification. Section 2 presents related work, highlighting the limitations of existing approaches. Section 3 discusses the key challenges and the rationale behind our research. In Section 4, we describe our proposed cloud IP recognition methodology, including data collection and feature design. Section 5 provides a detailed analysis of the network attributes–latency, port, and ASN and evaluates their discriminative power. Section 6 presents the experimental evaluation, including setup, dataset construction, comparative model results, and justification for selecting ASN as the core feature. Section 7 summarizes the findings and outlines directions for future research. Section 8 discusses the limitations of the proposed approach.

2 Related Work

Cloud IP identification technology has recently garnered significant attention, especially with the rapid expansion and adoption of cloud computing infrastructure. Existing methods primarily leverage machine learning and deep learning techniques to address IP usage scenario identification.

Zhou et al. (2022) [9] proposed a comprehensive feature extraction framework that uses various IP attributes, including geographic location, path information, ports, domain name services (DNS), and WHOIS information, achieving promising results in distinguishing IP usage scenarios. Wang et al. (2022) introduced a deep ensemble learning model designed to classify IP blocks into four typical usage scenarios: home broadband, private enterprise networks, mobile networks, and data centers. They improved decision tree models to achieve higher accuracy within specific regions. Similarly, Li et al. (2024) introduced a deep

continuous neural tree model capable of effectively handling complex feature interactions and transfer learning capabilities across different regions.

Additionally, Liu et al. (2024) [12] developed the GraphCyber framework based on graph neural networks (GNN) to improve computational efficiency. Their approach divides IP nodes into regional blocks, enhancing computational efficiency while identifying IP usage scenarios at a granular level. However, despite the performance gains, the intrinsic "black-box" nature of deep learning and the requirement for high computational resources remain critical limitations.

Previous studies have also extensively relied on IP geolocation databases. However, these databases' internal methodologies for cloud IP identification are not publicly accessible, contributing further to challenges related to computational overhead and interpretability.

To address these limitations, our study adopts an alternative approach, emphasizing interpretability and efficiency by systematically analyzing network attributes such as latency, port openness, and ASNs. Our method specifically investigates which network attributes significantly contribute to accurately distinguishing cloud IP usage scenarios from non-cloud IPs. In contrast to previous methods, our approach highlights interpretability, systematically examining the contribution of each network attribute to the identification accuracy and efficiency of cloud IP recognition.

In summary, while prior studies focus on enhancing accuracy through complex feature sets and deep learning techniques, our research seeks to identify simple, interpretable, and computationally efficient features, significantly contributing to practical and effective cloud IP identification.

3 Challenges and Motivations

Cloud IP identification is increasingly important for maintaining network security and stability, yet current methodologies face several critical challenges that hinder practical deployment and limit their effectiveness.

Firstly, existing cloud IP recognition methods are highly complex and lack interpretability. Most current approaches, including deep learning and ensemble models, utilize large feature sets and complex model architectures. Although these methods achieve relatively high classification accuracy, their complexity imposes significant computational overhead and makes the decision-making process opaque. Consequently, network administrators often find it challenging to trust and implement these models in practical security applications.

Secondly, a significant obstacle in cloud IP recognition research is the lack of publicly available and reliable datasets. Existing IP datasets are often privately held, incomplete, or inaccessible due to national or organizational restrictions. This situation severely limits reproducibility, independent verification of results, and fair comparative analysis of different IP identification techniques. Thus, developing publicly accessible datasets is critical for enabling standardized evaluations and community-wide progress in this area.

Thirdly, the current literature has not adequately addressed the fundamental differences in IP traffic characteristics across different usage scenarios. Most studies focus on simply classifying IPs without deeply analyzing or interpreting the network-level attributes (such as latency patterns, open ports, or ASN distributions) that distinguish cloud IP from non-cloud IP scenarios. Without understanding these underlying characteristics, it remains difficult to establish meaningful theoretical foundations or clear guidelines for IP scenario differentiation.

Lastly, there is a noticeable absence of clearly defined criteria or baselines for identifying IP usage scenarios. The field lacks standardized benchmarks or criteria for systematically distinguishing cloud IP addresses from others, causing inconsistent methodologies and conclusions across studies. Establishing

transparent and interpretable criteria or bases for classification is therefore necessary to advance rigorous and comparable research efforts.

Motivated by these key challenges, this study aims to systematically address the gaps highlighted above. Specifically, we focus on clearly identifying discriminative network attributes (e.g., ASN, latency, ports) that effectively separate cloud IPs from other usage scenarios. By emphasizing interpretability and resource efficiency, we aim to develop a simplified yet robust approach that overcomes existing limitations. Additionally, we provide openly accessible datasets and reproducible methods, facilitating community-wide validation and further research in cloud IP recognition.

4 Coud IP Recognition Method

Based on this motivation and goal, this paper uses data mining methods to obtain IP addresses in different IP usage scenarios and detect the network attributes of IP addresses. By analyzing the differences in network attribute values, the favorable factors and obstacles for identifying cloud IPs are obtained. Then, according to these favorable factors, we can accurately identify cloud IP usage scenarios, as shown in Fig. 1.

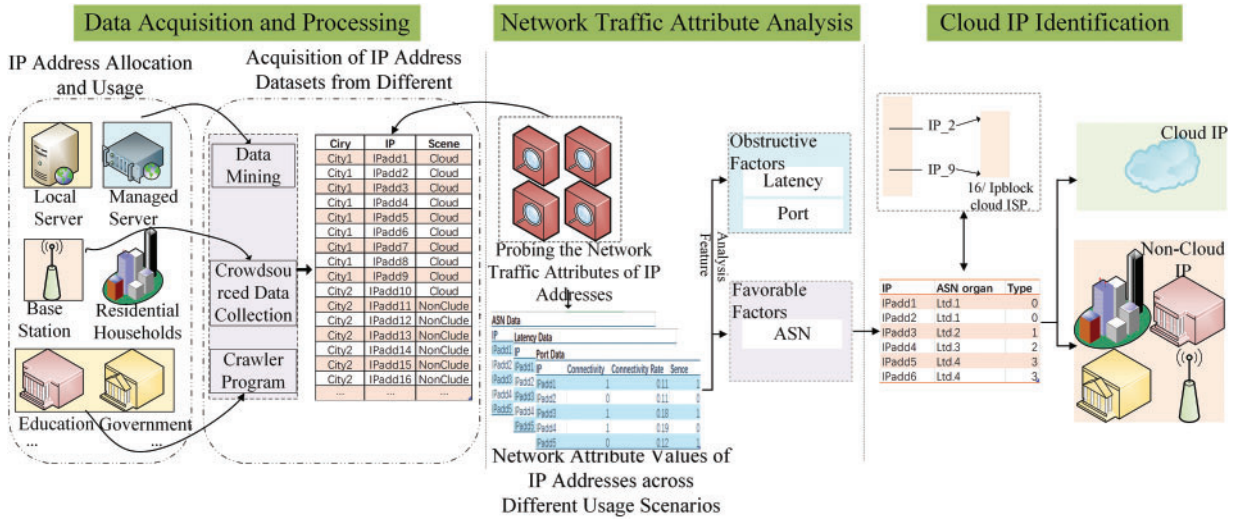

Figure 1: Framework for identifying cloud IP usage scenarios based on favorable factors

Step 1: Data Collection and Preprocessing

We begin by gathering IP address data from multiple usage scenarios to ensure comprehensive coverage. Specifically, we collect:

Cloud IP addresses from well-known providers, including Alibaba Cloud, Tencent Cloud, Amazon Web Services (AWS), and Google Cloud.

Non-cloud IP addresses from various real-world scenarios: university campus networks, mobile base stations, dedicated line users, and home broadband users.

All raw data undergoes a rigorous cleaning process to remove duplicates, resolve inconsistencies, and eliminate outliers. We also cross-verify IP addresses with existing databases (e.g., IP2Location, IPIP, and Chun Zhen) to ensure reliability.

Step 2: Network Attributes Measurement

Once the dataset is established, we employ standardized measurement tools to probe the network attributes of each IP address:

Latency (Round-Trip Time): We use the Ping command (ICMP protocol) to measure RTT from our probing server to the target IP. This value reflects the network latency.

Port Scanning: We leverage the MASSCAN tool to scan ports ranging from 1 to 10,000, capturing which ports are open on the target IP. This data reveals potential services and provides insight into the IP's usage scenario.

ASN: Using a GeoLite2 ASN Database (by MaxMind) lookup, we retrieve the ASN for each IP, thereby identifying its organizational affiliation. This step is crucial for analyzing whether an IP is managed by a cloud provider, a national telecom operator, or an independent enterprise network.

In this study, we utilized the GeoLite2 ASN database, which is a freely available product provided by MaxMind. The GeoLite2 ASN database offers comprehensive global IP-to-ASN mappings, is updated regularly monthly, and is widely recognized and adopted by the research community. While commercial databases such as MaxMind's GeoIP2 ASN provide more frequent updates (weekly) and higher granularity, GeoLite2 ASN was specifically selected for this research due to its open-access nature, making our experiments easily reproducible and verifiable by other researchers without incurring additional cost or access restrictions. The broad usage and community acceptance of GeoLite2 also enable straightforward comparisons with other studies in cloud IP identification.

Step 3: Feature Analysis and Discriminative Attribute Selection

Next, we systematically analyze and compare the measured network attributes across different IP usage scenarios:

Latency: We investigate whether latency alone can reliably distinguish cloud IPs from non-cloud IPs. Our findings indicate that simple RTT measurements often fail to discriminate effectively, largely due to varied user configurations and potential ICMP filtering.

Port Information: We examine port scanning results for both cloud and non-cloud IPs. Although certain ports (e.g., 80, 443, 22) appear frequently in cloud environments, they also show up in many self-hosted or dedicated-line scenarios, reducing their discriminative power.

ASN Organization Information: We then focus on ASN-based attributes. Experimental results strongly suggest that IP addresses within the same ASN tend to share similar usage scenarios. Cloud providers typically manage dedicated ASNs, whereas non-cloud IPs are associated with ISPs offering general broadband, base station services, or enterprise networks. Therefore, ASN emerges as a particularly strong and interpretable feature for distinguishing cloud IP usage scenarios from others.

Step 4: Cloud IP Identification and Validation

After isolating ASN as the most effective factor:

Model Building and Comparison: We construct a cloud IP identification model based primarily on ASN organization attributes. For a thorough performance comparison, we also train multiple machine learning models (e.g., MLP, CNN, XGBoost, KNN, SVM, Decision Tree) on various network feature subsets (including latency and port information).

Evaluation Metrics: We evaluate each model's classification accuracy and AUC (Area Under the ROC Curve). Our proposed ASN-based method achieves an accuracy of 96.67%, surpassing the performance of traditional machine learning models, which exhibit accuracies between 81% and 92%.

Interpretability and Overhead Analysis: Since our approach relies on a single, well-defined feature ASN, it significantly reduces computational overhead and increases transparency in decision-making. By contrast, deep learning methods often operate as "black-box" solutions and require extensive computational resources.

Experimental Validation: We verify the robustness of our approach using IP addresses from different regions (China and Europe) and a diverse range of usage scenarios (e.g., base station IPs, home broadband, data center hosting). This confirms that focusing on ASN effectively generalizes across multiple contexts.

5 Feature Analysis

In order to reveal the essential differences in different IP usage scenarios, this paper conducts network detection on IP addresses from the perspective of network attributes. It is well known that each IP address has its own unique network attribute characteristics. Next, we analyze the obstacles and advantages of identifying cloud IP from the perspectives of latency, port, and ASN.

5.1 Latency

Latency refers to the time required from sending a request to receiving a response, and it is one of the key indicators for measuring computer network performance. For many applications (especially real-time applications, such as video conferencing, online gaming, and remote control), latency is critical to their performance. The main components of latency include sending latency, transmission latency, processing latency and queuing latency. The latency used in our paper is the sum of them.

This paper uses the Ping command to measure latency. The round-trip time (RTT) is calculated by calculating the time difference between sending an ICMP Echo Request to the destination address and receiving the Echo Reply. The measurement and analysis of latency is of great significance for identifying the organizational entities corresponding to different types of IP addresses.

In network communication, different IP usage scenarios have varying requirements for connectivity. However, when an IP address serves as a web server, the host associated with the IP address must provide web services to the user. This requires mutual connectivity between the server host and the user host. Therefore, whether it is a cloud IP or a non-cloud IP, the connectivity of IP addresses in different usage scenarios needs to be connected. This may cause the ICMP protocol to respond.

In the experiment of studying IP address connectivity in different usage scenarios, we discovered a problem. When a host has an ICMP filtering policy, it will not respond to ICMP requests. Even if the host remains connected.

5.2 Port

Ports are used to identify different applications or services. Each data packet contains a destination port number. And the receiver uses this port number to forward the packet to the corresponding application or service.

The characteristic of port value is that it has extremely strong application service differences. The port value range is 0~65,535. It means that a host can open ports for up to 65,536 services at the same time and distinguish application differences of them. Ports are divided into well-known ports, registered ports and dynamic or private ports. The well-known ports range from 0 to 1023. They have the strongest application service differentiation, and each port number is fixed to the corresponding service. The range of registered ports is from 1024 to 49,151. These ports are also bound to certain services, but the port numbers are not fixed. For example, port 3306 is commonly associated with MySQL databases. However, the MySQL database does not force users to use port 3306. Additionally, in order to protect the database from attacks, users often

avoid exposing their database port and even changing the database port number on their own initiative. Other ports are dynamic or private ports. Other ports are dynamic or private ports. They are usually used for short-lived connections or temporary services and range from 49,152 to 65,535.

It can be seen that analyzing the port values between 0–1023 is the most valuable for research. By analyzing the differences in port values in different usage scenarios, we can identify IP usage scenarios. It is the focus of our paper.

During the port detection process, the detection result of the port value is usually inaccurate. The main reasons for this inaccuracy are as follows. First, the target host may use firewalls or network filters to restrict access to certain ports. Second, In the case of an unstable network environment or a target host response timeout, the detection tool may not be able to accurately obtain the response information of the port status. Finally, Technologies such as port masquerading or dynamic ports prevent us from obtaining port information.

It is necessary to study the port feedback rate within the network segment. By filtering the protocol, the host no longer responds to protocol requests. Users can prevent network attacks and protect host security. On the other hand, the feedback rate of the port in the network segment can be used to reflect the network attribute characteristics of the IP usage scenario corresponding to the host. Other researchers have proposed that the geographical location and latency of /24 network segments are very similar [16]. In this paper, the port feedback rate within the /24 network segment is used.

5.3 Autonomous System Number

An ASN is a number that used to uniquely identify different Autonomous Systems (AS) on the Internet. The allocation of ASNs is managed by the Internet Assigned Numbers Authority (IANA) and regional internet registries (such as ARIN in the US, RIPE NCC in Europe, APNIC in the Asia-Pacific region, etc.) [17,18]. ASN not only identifies an Autonomous System but also contains information (about the range of IP addresses) managed by it. The number of IP addresses that an Autonomous System can manage ranges from a few to hundreds of millions.

In communication protocols, ASN is primarily associated with the Border Gateway Protocol (BGP). BGP uses ASN to identify different Autonomous Systems and plays a crucial role in routing decisions. When a BGP router receives routing information from other autonomous systems, it uses the ASN to determine the validity and priority of the route. Additionally, BGP can prevent routing loops by checking the list of ASNs traversed in routing updates. This ensures that duplicate paths do not occur. Multiprotocol BGP (MP-BGP) also relies on ASN to support routing propagation for IPv6 and other protocols. Therefore, ASN plays a vital role in BGP. It ensures the effective exchange of routing information between different Autonomous Systems, and maintains the stability and reachability of the internet.

Based on ASN information, we can effectively distinguish IP addresses assigned to cloud providers from those allocated to traditional network operators. Each ASN belongs exclusively to one organization, making it a reliable feature for classifying cloud vs. non-cloud IP usage. Unlike large telecom carriers that manage a diverse array of residential and enterprise IP allocations, cloud providers typically operate specialized ASNs dedicated to data centers and virtual machines. This clear organizational boundary captured by ASN significantly enhances the interpretability and accuracy of identifying cloud IP addresses. Furthermore, while mobile data usage scenarios often span multiple ASNs, cloud IP addresses tend to be concentrated under a smaller set of well-defined ASNs, further simplifying detection.

Based on the above analysis, the ASN information for mobile data usage scenarios is relatively complex. The ASN information for cloud IP usage scenarios is simple and centralized.

6 Experimental Evaluation

In this section, we will study the network attribute characteristics of cloud IPs and analyze the network attributes of IP addresses from different scenarios. We aim to identify network attributes that are advantageous for recognizing cloud IPs and those that are disadvantageous.

6.1 Experimental Setup

In this paper, we conducted network probing on a large ground truth dataset from the perspective of network attributes. The probing source was an Alibaba Cloud server host with the following configuration, dual-core processor, 2 GB RAM, 100 Mbps peak bandwidth, running Linux CentOS 7.6. The geographic locations of the server were in Hangzhou, China; Shanghai, China; London, UK; Frankfurt, Germany; and Silicon Valley, USA. This paper used the network measurement tool Scamper (developed by the Cooperative Association for Internet Data Analysis (CAIDA)) to probe IP addresses. The tool includes commands such as Ping and Traceroute, and supports protocols such as ICMP, UDP, TCP, etc. For port scanning, we used the MASSCAN tool, which can quickly perform batch scans on target IP ports. The scan was configured to target ports within the range of 1 to 10,000. For querying the ASN, our paper used the ASN database provided by Max-Mind, which is used to look up the ASN corresponding to each IP address.

6.2 Dataset

We have made the dataset and related experimental results of this article public. This section will describe the dataset collection and cleaning process to ensure the accuracy of IP addresses in IP usage scenarios.

6.2.1 Cloud IP Addresses

Our paper first collected the IP addresses of Tencent Cloud, Alibaba Cloud, Amazon Cloud, and Google Cloud.

Cloud service providers such as Tencent Cloud and Alibaba Cloud do not disclose relevant cloud IP address documents. Therefore, this paper first screens out IP addresses whose Internet service providers (ISPs) are Tencent Cloud and Alibaba Cloud from three major databases (IPIP [19], Chun Zhen [20], and IP2location [21]). Then, we searched for city-level location results for these IP addresses and organized the ones belonging to the same city based on the results.

Two cloud service providers, Amazon Cloud and Google Cloud, provide cloud IP documentation on their official websites. The documentation clearly specifies the IP ranges and the regions associated with those IP ranges. In this paper, we select IP addresses from major cities in the United States and Europe, then include the IP addresses within the specified ranges in the dataset. The dataset size for each city is shown in Table 1.

Table 1: Ground truth IP address dataset for different IP usage scenarios

IP	Quantity	Distribution regions	IP	Quantity	Distribution regions
Alibaba cloud	2,243,850	Shanghai, Beijing, Hangzhou, Shenzhen, Qingdao	Google Cloud	4,820,736	Las Vegas, Salt Lake City, Council Bluffs, The Dalles, Dallas, Berlin, Frankfurt, Turin, Paris

(Continued)

Table 1 (continued)

IP	Quantity	Distribution regions	IP	Quantity	Distribution regions
Tencent cloud	2,162,937	Shanghai, Beijing, Nanjing, Tianjin, Guangzhou, Chengdu, Jinan, Shenzhen, Chongqing	Base Station	37,864	Zhengzhou, Jinan, Dezhou, Luoyang, Beijing, Weifang, Zaozhuang, Xi'an, Hefei, Xinxiang
Amazon cloud	2,5191,115	London, Paris, Ireland, Frankfurt, Oregon, California	Non-Cloud	184,275	China, Europe

Building on this foundation, we further processed the raw data to ensure its reliability and suitability for subsequent analysis. Initially, a rigorous data cleaning procedure was implemented to remove duplicate entries, resolve inconsistencies, and eliminate any outliers that could skew the analysis. Each IP address was cross validated across multiple sources to confirm its legitimacy and correct geographic assignment. Following this, we enriched the dataset by integrating additional network attributes such as latency measurements, port usage patterns, and self-assigned system identifiers, which are critical for distinguishing between cloud and non-cloud IP scenarios.

Subsequently, feature extraction was carried out using a combination of statistical analysis and domain-specific heuristics to derive meaningful indicators from the raw network data. These features were then normalized and categorized to facilitate an accurate comparison across different regions and service providers. Our methodology emphasizes not only computational efficiency but also the interpretability of the resulting model, addressing the common challenges posed by black-box deep learning approaches.

6.2.2 Base Station IP Addresses

This paper adopts crowdsourcing method to obtain base station IP. Before data was collected, users were clearly instructed not to use Wi-Fi for data connections. And they can only submit their IP addresses over cellular data connections. After users submitted their IP addresses, the city corresponding to the IP address will be recorded. Then, we queried the size of the subnet (the current IP address belongs), and designated this subnet as the base station IP subnet. The IP addresses within this subnet were treated as base station IPs. The dataset is displayed in Table 1.

6.3 Feature Effectiveness Validation Experiment

In Section 5, three unique network properties of IP addresses were introduced: latency, port, and ASN. We conducted experiments to verify whether these attributes can be used to distinguish cloud IPs from non-cloud IPs. This section will present the experimental results and analyze the advantages and disadvantages of identifying cloud IP usage scenarios.

6.3.1 Latency Effectiveness Validation Experiment

As analyzed in Section 5.1, latency and connectivity represent the ability of the server and user to communicate with each other. To ensure users can access server resources, connectivity is typically maintained. However, different usage scenarios have varying connectivity requirements. Therefore, we probed the true IP datasets of the four major cloud service providers and non-cloud IP addresses. As shown in Table 2.

In Table 2, the connectivity test results for each IP are shown. The results indicate that most of the IPs are not reachable. The reason for this is that the host resources have not yet been allocated to the user, or the user has configured security policies to prevent the host from being detected. We also tested the connectivity of individual cloud IPs and non-cloud Ips. The results (line 5) show that the cloud IPs do not exhibit any significant difference in characteristics compared to other usage scenarios in this network metric. Therefore, we conclude that using latency as a feature to identify cloud IPs is ineffective.

Table 2: Connectivity test for different IP usage scenarios

IP usage scenario	Total IP count	Number of inaccessible IPs	Inaccessible IP percentage
Alibaba cloud	2,243,850	1,990,945	88.73%
Tencent cloud	2,162,937	1,725,727	79.79%
Google cloud	4,820,736	3,356,207	69.62%
Amazon cloud	2,5191,115	2,353,663	93.43%
Non-cloud	184,275	161,210	87.48%

6.3.2 Port Effectiveness Validation Experiment

Ports are the unique identifiers that distinguish different applications or services on a network. When a host needs to provide services to users, the corresponding port must be opened for communication. The purpose of studying the port openness of different IP usage scenarios is to distinguish cloud IPs from non-cloud IPs by port values. First, port probing was performed on the reachable IPs of the four cloud service providers (Alibaba Cloud, Tencent Cloud, Google Cloud and Amazon Cloud). Then, we scanned all ports within the range of 1–10,000 and identified the ten most frequently occurring ports for each cloud service provider. The proportion of each port in the top ten was calculated and illustrated as shown in Fig. 2.

Fig. 2 shows the port scan results of the cloud IP real data set. We can see that ports 22, 80, 443, and 3389 appear most frequently. As shown in Fig. 2, ports 22, 80, 443, and 3389 appear most frequently. Ports 80 (HTTP) and 443 (HTTPS) are particularly dominant due to their ubiquity in hosting web-based services in the cloud.

However, relying solely on single-IP scans might not capture the broader network context. Different users or hosts under the same provider could open or block various ports based on specific needs or security policies.

Fig. 2 presents the port scan results for individual cloud IP addresses from Alibaba Cloud, Tencent Cloud, Google Cloud, and Amazon Cloud (in the range 1–10,000). As illustrated, ports 22, 80, 443, and 3389 appear most frequently across multiple cloud service providers. However, relying solely on single-IP scans might overlook broader network-level patterns and could be affected by a particular user's or host's service configurations. To investigate whether these port distributions differ significantly across the entire network segment, we extended our scanning from a single IP to the full /24 subnet that the IP belongs to. This broader approach, detailed in Table 3, helps capture the variety of active hosts within that subnet and reveals whether certain ports are consistently tied to cloud usage scenarios or are also found in common ISP services, mobile data networks, and other non-cloud environments. If the same ports remain prominent across both cloud and non-cloud subnets, it suggests that port-based indicators alone may be insufficient to reliably identify cloud IP usage.

Based on these considerations, we present the extended scanning results in Table 3, which summarizes the top-five open ports in the /24 subnet of each IP address in the dataset. This approach amplifies the

correlation between port values and their IP usage scenarios. We set the scan range to be limited to ports 1–1000 to comply with network security regulations. In addition, a small port scan range helps reduce the network load on the scanning host. The probing results are shown in Table 3.

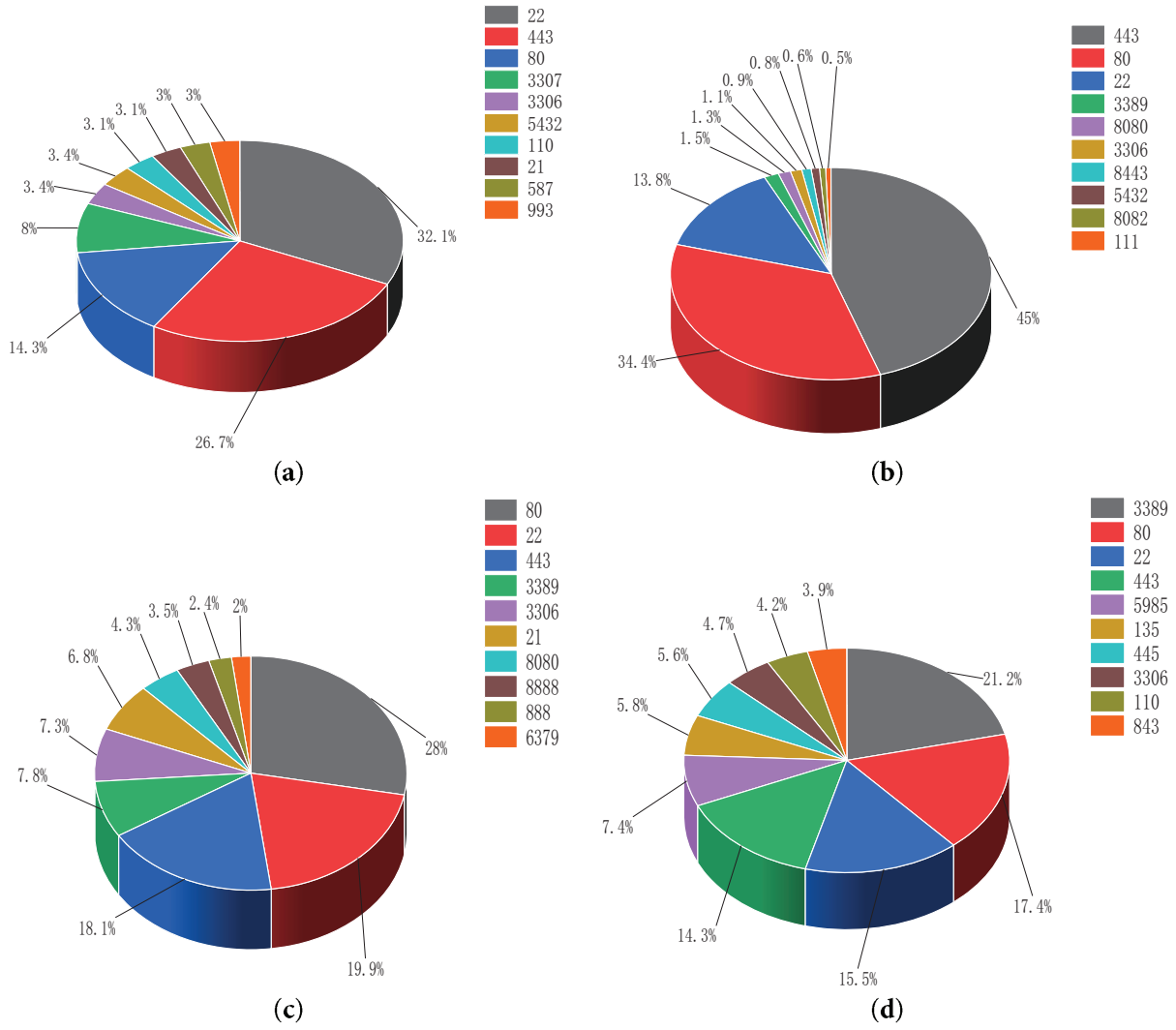

Figure 2: Top 10 most frequent ports for different cloud service providers. (a) Top 10 most frequently used ports in Amazon Cloud; (b) Top 10 most frequently used ports in Google Cloud; (c) Top 10 most frequently used ports in Tencent Cloud; (d) Top 10 most frequently used ports in Alibaba Cloud

After expanding a single IP to a /24 subnet, ports 80 and 443 still account for the largest share, which is similar to the port information opened by the cloud IP. In addition to ports 80 and 443, port 22, which is opened for remote login protocol, should have a strong correlation with cloud IP. However, port 22 also appears in mobile data scenarios and ordinary broadband scenarios (accounting for 0.12% and 0.23%, respectively). Port 22 appears in the detection results of both cloud IP and non-cloud IP. Obviously, remote login cannot be regarded as a unique feature of cloud IP. On the other hand, the port opening status is determined by the user's business needs. Therefore, we believe that using port-related information to identify cloud IP is a disadvantage.

Table 3: Top 5 most frequent ports in subnet and their proportions

IP usage scenario	Detailed usage scenario	Top five ports and their proportions
Non-cloud IP	Mobile data	443(1.11%), 80(1.10%), 587(0.79%), 465(0.79%), 21(0.79%)
	Dedicated connections	443(25.30%), 80(24.16%), 22(11.52%), 21(6.55%), 587(3.68%)
	General broadband	80(0.68%), 443(0.62%), 21(0.30%), 53(0.26%), 465(0.26%)

After expanding a single IP to a /24 subnet, ports 80 and 443 still account for the largest share, which is similar to the port information opened by the cloud IP. In addition to ports 80 and 443, port 22, which is opened for remote login protocol, should have a strong correlation with cloud IP. However, port 22 also appears in mobile data scenarios and ordinary broadband scenarios (accounting for 0.12% and 0.23% respectively). Port 22 appears in the detection results of both cloud IP and non-cloud IP. Obviously, remote login cannot be regarded as a unique feature of cloud IP. On the other hand, the port opening status is determined by the user's business needs. Therefore, we believe that using port-related information to identify cloud IP is a disadvantage.

6.3.3 Effectiveness of ASN for Cloud IP Recognition

The uniqueness of ASN was introduced in Section 5.3. To verify its utility in identifying cloud IP scenarios, we queried the ASN information of four cloud service providers. Table 4 presents the results.

Table 4: Statistics of ASN for cloud companies

Cloud service provider	AS number	Number of IP addresses in the ASN	Percentage of the company's IP addresses
Alibaba cloud	AS37963	2,243,850	100.00%
	AS45090	2,140,459	98.98%
	AS132203	4064	0.19%
	AS139341	1016	0.05%
	0	762	0.04%
Tencent cloud	AS23724	508	0.02%
	AS4837	8636	0.40%
	AS136958	4318	0.20%
	AS38283	2032	0.09%
	AS4538	508	0.02%
Google cloud	AS396982	4,512,256	93.60%
	AS15169	308,480	6.40%
Amazon cloud	AS16509	21,454,628	85.17%

Table 4 lists the principal ASNs for Alibaba Cloud, Tencent Cloud, Google Cloud, and Amazon Web Services (AWS), along with the number of IP addresses each ASN contains and the corresponding percentage of the total. Notably, some cloud providers register multiple ASNs but tend to concentrate the majority of their

IP addresses under one or two main ASNs. For instance, Tencent Cloud owns 13 ASNs; however, AS45090 alone accounts for nearly 99% of its IP resources. Alibaba Cloud, by contrast, exclusively uses AS37963, unifying all IP addresses under a single ASN. One reason for this pattern is that major cloud providers typically manage routing and infrastructure across a few core ASNs to streamline network administration and maintain consistent routing policies. Although a provider may occasionally acquire or register additional ASNs (due to mergers, new data centers, or experimental services), most production traffic and customer VMs remain clustered in its primary ASN(s).

Google Cloud and AWS similarly appear in multiple geographic regions—spanning Europe and North America—but still belong to a limited set of ASNs. This centralized ASN usage simplifies the identification of their IP resources, since administrators only need to track a handful of global ASN entries to categorize a significant portion of cloud IP addresses. In rare cases, minor discrepancies may arise when smaller, less-frequented ASNs are not fully documented or have incomplete data in external databases.

To further investigate whether different network service providers might share ASNs—or if one ASN covers multiple locations—we examined base station IPs from various provinces. The results are summarized in Table 5.

Table 5: ASN statistics for base station IPs

IP range	Province	City	ISP	AS number
39.144.188.0/24	Henan	Xinxiang	China Mobile	AS24445
117.136.106.0/23	Henan	Zhengzhou	China Mobile	AS24445
113.121.196.0/23	Shandong	Dezhou	China Telecom	AS4134
106.33.160.0/20	Henan	Zhengzhou	China Telecom	AS4134
171.9.166.0/19	Henan	Zhengzhou	China Telecom	AS4134
182.33.128.0/20	Shandong	Weifang	China Telecom	AS4134
182.39.128.0/18	Shandong	Zaozhuang	China Telecom	AS4134
219.144.0.0/22	Shaanxi	Xi'an	China Telecom	AS4134
61.54.108.0/23	Henan	Xinxiang	China Unicom	AS4837
112.224.141.0/22	Shandong	Yantai	China Unicom	AS4837
220.205.249.0/24	Beijing		China Unicom	AS4837
124.64.16.0/22	Beijing		China Unicom	AS4808

Table 5 shows how base station IP addresses (non-cloud) are distributed among distinct ASNs belonging to various ISPs (e.g., China Mobile, China Telecom, China Unicom). For instance, China Telecom uses AS4134 across provinces such as Henan, Shandong, and Shaanxi, while China Mobile relies on AS24445, and China Unicom predominantly employs AS4837 or AS4808. Even though these ASNs cover multiple cities, each ASN remains tied to a single ISP, indicating that non-cloud IPs—despite their wide geographic reach—still map uniquely back to their operator. This organizational consistency confirms that ASN effectively differentiates cloud IPs (centrally managed by cloud providers) from non-cloud IPs (allocated by traditional network operators for mobile data, broadband, etc.).

Overall, Tables 4 and 5 demonstrate that while IP addresses under the same ASN can be geographically diverse, they share a single organizational affiliation. In practice, IP usage scenarios typically fall into two major categories: cloud IP (managed by cloud providers) and non-cloud IP (managed by ISPs). Thus, ASN proves to be a viable and interpretable feature for identifying these distinct usage contexts—an insight crucial for security, resource allocation, and network policy enforcement.

6.4 Experiment on Identifying Cloud IP Using Multiple Network Attribute Factors

By detecting the network attributes of cloud IP and non-cloud IP, we obtained the latency value, port information, and ASN organization information corresponding to the IP address. We constructed a dataset containing these network attributes for machine learning experiments and to verify the method proposed in this Organizations with strict compliance standards (e.g., in finance or healthcare) often maintain private network policies that block or sandbox traffic from external cloud services. To ensure geographic diversity, the dataset includes IP addresses collected from major cities in China—Beijing, Shanghai, Guangzhou, and Shenzhen—as well as multiple European metropolitan areas—Paris, Turin, Berlin, and London—covering a variety of cloud providers. This broad regional coverage allows us to assess the robustness of our approach across different latency conditions, port patterns, and ASN distributions. The experimental results are shown in Table 6.

Table 6: AUC of different models for cloud IP recognition

Model	Implementation	Hyperparameters	Accuracy (Ping + Port + ASN)	Accuracy (ASN Only)
Decision tree	scikit-learn	criterion = 'gini', max_depth = None, random_state = 42	0.92	0.89
XGBoost	XGBoost	max_depth = 6, learning_rate = 0.3, n_estimators = 100, random_state = 42	0.86	0.86
KNN	scikit-learn	k = 5, metric = 'euclidean'	0.86	0.87
SVM	scikit-learn	kernel = 'rbf', C = 1.0, gamma = 'scale', random_state = 42	0.81	0.81
MLP	scikit-learn	hidden_layers = (100,50), activation = 'relu', max_iter = 500, solver = 'adam', random_state = 42	0.82	0.81
CNN	TensorFlow/Keras	Conv1D(filters = 64, kernel_size = 2, activation = 'relu'), epochs = 10, batch_size = 32, optimizer = 'adam', loss = 'categorical_crossentropy'	0.81	–
Ours	ASN		–	0.96

To clearly demonstrate and validate our proposed method, we specifically sampled the dataset to contain 500 non-cloud IP addresses (TYPE = 0) and 2000 cloud IP addresses (TYPE = 1), maintaining a ratio of approximately 1:4. Although this intentionally introduces class imbalance, this proportion mirrors realistic scenarios commonly observed in practice, where cloud-hosted IPs are generally more prevalent. Such design decisions allow our experiments to closely reflect real-world conditions. Furthermore, the use of diverse mainstream machine learning models (MLP, CNN, XGBoost, KNN, SVM, and Decision Tree) helps demonstrate that our ASN-centric approach consistently performs well despite class imbalance.

In Table 6, we construct datasets with multiple network attributes (including latency, port, and ASN information) and datasets using only ASN attributes. Then, we use mainstream machine learning methods

(MLP, CNN, XGBoost, KNN, SVM, Decision Tree) to classify cloud IP usage scenarios. The results show that the recognition accuracy of machine learning methods under various network attribute data sets is slightly different. When only ASN features were used for experiments, the machine learning classification results did not improve or decrease significantly compared to using multiple network attributes as feature data sets for classification. In the dataset using only ASN, the average recognition accuracy of multiple machine learning methods is 0.84, while our method is 0.96, which is significantly higher than the recognition accuracy of machine learning.

This ASN-focused approach to cloud IP identification provides notable advantages in scalability, resource efficiency, and geographic adaptability. By reducing reliance on multiple, high-dimensional features, it is especially suitable for large-scale monitoring contexts, while its explicit reliance on ASN records offers straightforward traceability for network administrators. Moreover, because ASNs are assigned globally through Regional Internet Registries (RIRs), the method generalizes effectively across diverse regions and provider ecosystems.

Although we compared our proposed method to a variety of mainstream machine learning models (Section 6.4), it is equally important to evaluate its relative performance and utility against the latest published methods in this domain. Table 7 summarizes our method's performance alongside selected state-of-the-art cloud IP identification approaches proposed by Zhou et al. [9], Wang et al. [10], Li et al. [11], and Liu et al. [12]. The experimental results are shown in Table 7.

Table 7: Comparison of cloud IP identification methods and their key characteristics

Method	Feature dimensionality	Key technique	Reported accuracy	Interpretability	Computational overhead
Zhou et al. [9]	More features	–	–	Low	High (due to large feature set)
Wang et al. [10]	–	DT & Deep learning	98%	Medium	High
Li et al. [11]	46 features	Deep continuous neural tree	94%	Medium	Moderate
Liu et al. [12]	Graph-based approach	GNN-based GraphCyber Framework	97%	Low	High (building graph structures)
Ours	**1 primary feature (ASN)**	ASN-based classification	96.67%	**High**	**Low**

Observations from the Comparative Analysis:

1. Higher Accuracy with Fewer Features

While prior works often employ 40+ features (e.g., geographic location, ports, DNS, WHOIS) combined with deep learning architectures, our method focuses on ASN organization attributes supplemented by minimal supporting data (latency, port). Despite this lean approach, we achieve a higher reported accuracy of 96.67%.

2. Improved Interpretability

Most prior SOTA models (e.g., [9,10,12]) rely on deep or ensemble learning techniques that are difficult to interpret. By contrast, our ASN-based classification is inherently more transparent: network administrators can clearly see whether an IP belongs to a known cloud service provider's ASN or a general ISP. This clarity can be especially valuable for real-time security decision-making. Unlike feature importance methods (e.g., SHAP, permutation importance) employed by prior studies, which merely indicate numerical significance of attributes without clear practical meaning, our approach provides straightforward, actionable explanations tied directly to real-world entities.

3. Reduced Computational Overhead

Methods like DNN or GNN [9,12] often require large-scale feature extraction and extensive training, which can be computationally intensive. By trimming the feature set drastically, our approach minimizes resource usage and speeds up deployment, making it suitable for high-volume, real-time environments.

Overall, these comparisons emphasize the superior accuracy, interpretability, and lower overhead of our proposed method relative to existing state-of-the-art solutions. The use of ASN as a primary discriminative feature underscores the potential for simpler, more transparent systems that maintain or surpass the performance of more complex deep learning frameworks.

The method proposed in this paper eliminates unfavorable factors for identifying cloud IP usage scenarios and focuses on the ASN organization affiliation information of the IP address. Experimental results show that our method improves the accuracy and interpretability of identifying cloud IP usage scenarios without considering the uncertainty of network properties. That is to say, even if there is a lot of noise in the network attributes, our model can still achieve high-precision cloud IP identification.

6.5 Rationale for Choosing ASN as a Key Discriminative Feature

In this subsection, we provide a detailed justification for selecting ASN as the primary factor in identifying cloud IP usage scenarios. By highlighting ASN's uniqueness, interpretability, and organizational consistency, we demonstrate why it outperforms latency- or port-based approaches in terms of both accuracy and explainability. Specifically, the following points clarify the key advantages of ASN in our classification framework.

- **Uniqueness and Non-repetition.** An ASN is a globally unique identifier that is assigned to network entities (e.g., Internet Service Providers, data center operators, cloud service providers). Because ASNs cannot overlap—each corresponds to a distinct organization—this provides a clear-cut way to group IP addresses by their operational or organizational affiliation. In contrast, latency and port usage are not strictly tied to any single organization and can be influenced by factors such as user configurations or network conditions, making them less reliable for classification.
- **Clear Organizational Boundaries Each.** ASN typically belongs to a specific network operator or service provider. For instance, cloud service providers (e.g., AWS, Alibaba Cloud, Tencent Cloud, and Google Cloud) often control well-defined ASNs used predominantly for hosting virtual machines, storage, and other cloud-based resources. On the other hand, non-cloud IP addresses are typically associated with ASNs owned by national telecom operators, universities, or smaller ISPs. These clear boundaries naturally lend themselves to an interpretable classification system.
- **Reduced Ambiguity and Greater Interpretability Port.** usage and latency are highly context-dependent and subject to individual host configurations and security policies (e.g., firewall settings, port forwarding). As a result, they can vary substantially even within the same IP usage scenario, leading to ambiguous or conflicting signals when classifying cloud vs. non-cloud IPs. Conversely, ASN is assigned

and managed at the organizational level and remains consistent for all IP addresses under the same network entity. This consistency translates into improved interpretability—one can readily see that if a given IP belongs to an ASN known to be managed by a cloud provider, that IP is highly likely to be a cloud IP.

- **Low Overhead and High Scalability.** Extracting ASN from an IP address via a lookup in a reputable database (e.g., MaxMind) is computationally lightweight and easily scalable for large IP datasets. In contrast, high-dimensional machine learning approaches that rely on numerous features (e.g., 40+ attributes involving port scans, latency measurements across different protocols, DNS records, WHOIS lookups) can be computationally expensive and more difficult to maintain. ASN-based identification thus offers a resource-friendly alternative while still achieving high classification accuracy.
- **Robustness across Different Regions.** Our empirical results indicate that relying on ASN provides robust classification performance across multiple geographic regions (China, North America, Europe) and diverse IP usage scenarios (base station IPs, university networks, home broadband, and data center environments). Since ASN allocation is managed by regional internet registries under global standards, it remains relatively stable across countries and continents, making it suitable for large-scale, cross-regional IP classification tasks.

Combined with the above factors, we choose ASN as the key factor.

7 Conclusions and Future Work

At present, the identification of cloud IP usage scenarios is mainly carried out through database query and machine learning methods. These methods generally have problems such as inaccurate identification, black box, and complexity. This paper starts from the perspective of network attributes and proposes a method based on ASN organization information to identify cloud IP. The new method can explain the favorable factors for identifying cloud IP scenarios, discard useless features in network attributes, and realize cloud IP identification only by relying on the attribution information of IP addresses. Explainability perfectly solves the black box problem. Experimenting with only the ASN feature also greatly reduces overhead. Most importantly, the experimental results show that our method can achieve a recognition accuracy of 96.67%, which is higher than the recognition method based on machine learning. In the community contribution module, we also make the experimental code and data public to facilitate discussions with other scholars.

Our method depends heavily on timely and accurate ASN databases; outdated or incomplete records can lead to misclassification. Sub-leased or multi-tenant ASNs also pose a challenge when "cloud" and "non-cloud" traffic coexist under the same organizational umbrella. As IPv6 adoption grows, extending our focus beyond IPv4-based ASNs will be critical for comprehensive coverage. Finally, dynamic IP allocation within cloud providers necessitates periodic re-checking of IP usage scenarios, ensuring sustained accuracy in real-world, rapidly changing environments.

In future work, we plan to expand upon threshold determination for latency and port openness. Specifically, we will explore quantile-based methods or machine learning-driven optimization to systematically identify boundary values that separate cloud IP from non-cloud IP usage scenarios. This approach can be particularly valuable for resolving ambiguities in shared ASN subnets or dynamic IP allocation, where small differences in latency or port usage can yield crucial insights. Additionally, we aim to integrate real-time ASN database synchronization or a hybrid ASN-DNS validation strategy, ensuring precise automated responses in rapidly evolving network environments. By adopting these enhanced thresholding techniques, our ASN-centric identification method can achieve finer-grained decision-making and maintain robust performance, even in borderline or ambiguous classification cases.

Cloud IP ranges often change dynamically, and multiple providers may share or sub-lease the same ASN, creating classification ambiguities. Future research could address these issues by regularly updating ASN-IP mappings, incorporating secondary verification (e.g., DNS or WHOIS data), or combining active and passive traffic analysis methods. These strategies can effectively manage the complexity of dynamic and overlapping ASN scenarios.

8 Limitations

Despite the promising results of our ASN-based cloud IP identification approach, several limitations warrant attention:

1. Dependence on ASN Database Accuracy. Our method heavily relies on external ASN databases (e.g., MaxMind). If these databases contain outdated or inaccurate records, the classification may mislabel certain IPs. Nevertheless, as these databases are extensively used and regularly updated within both academic and industry communities, inaccuracies are generally rare and typically have minimal impact on large-scale analytical results. Regular database updates and cross-verification with alternative sources can mitigate, but not fully eliminate, this risk.
2. IPv6 Adoption. While our experiments focused primarily on IPv4 addresses, the continued growth of IPv6 may reveal different patterns or require separate ASN databases. In high-IPv6 environments, additional research and validation will be necessary to confirm the approach's robustness.

Acknowledgement: The authors wish to express their appreciation to the reviewers for their helpful suggestions which greatly improved the presentation of this paper.

Funding Statement: This research was funded by the Henan Province Science Foundation for Youths No. 222300420058, Henan Province Science and Technology Research Project No. 232102321064, Teacher Education Curriculum Reform Research Priority Project No. 2023-JSJYZD-011, Key Project of Henan Provincial Higher Education Teaching Reform (Graduate Education) No. 2023SJGLX062Y.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Yuanyuan Ma, Cunzhi Hou, and Ruixiang Li; data collection: Cunzhi Hou, Ruixia Jin and Ang Chen; analysis and interpretation of results: Yuanyuan Ma, Cunzhi Hou, and Jinghui Zhang; draft manuscript preparation: Cunzhi Hou and Ruixiang Li. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: In this study, we used a public dataset, which can be downloaded from the website if needed (https://gitee.com/henan-normal-university_4_0/determination-of-favorable-factors-for-cloud-ip-recognition-technology.git, accessed on 6 April 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Mukhtarov M, Miloslavskaya N, Tolstoy A. Network security threats and cloud infrastructure services monitoring. In: Proceedings of the Seventh International Conference on Networking and Services (ICNS); 2011 May 22–27; Venice, Italy. p. 141–5.
2. Yang J, Wang C, Liu C, Yu L. Cloud computing for network security intrusion detection system. J Netw. 2013;8(1):140–7. doi:10.4304/jnw.8.1.140-147.
3. Chen Z, Han F, Cao J, Jiang X. Cloud computing-based forensic analysis for collaborative network security management system. Tsinghua Sci Technol. 2013;18(1):40–50. doi:10.1109/TST.2013.6449406.

4. Ma Y, Chen A, Hou C, Jin R, Zhang J, Li R. DC-FIPD: fraudulent IP identification method based on homology detection. Comput Mater Contin. 2024;81(2):3301–23. doi:10.32604/cmc.2024.056854.
5. Ali I, Sabir S, Ullah Z. Internet of things security, device authentication and access control. arXiv:1901.07309. 2019.
6. Ghaffari F, Bertin E, Hatin J, Crespi N. Authentication and access control based on distributed ledger technology: a survey. In: Proceedings of the 2020 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS); 2020 Sep 28–30; Paris, France. p. 79–86.
7. Nirmala MB. Wan optimization tools, techniques and research issues for cloud-based big data analytics. In: Proceedings of the 2014 World Congress on Computing and Communication Technologies; 2014 Feb 27–Mar 1; Trichirappalli, India. p. 280–5.
8. Wani A, Mohite S, Gonge S, Joshi R, Vora D, Kotecha K. Data analytics technique of different ip address utilized for cloud services. In: Proceedings of the 2023 9th International Conference on Signal Processing and Communication (ICSC); 2023 Dec 21–23; Noida, India. p. 274–8.
9. Zhou F, Zhang W, Wang Y, Zhong T, Trajcevski G, Khokhar A. Identifying IP usage scenarios: problems, data, and benchmarks. IEEE Netw. 2022;36(3):152–8. doi:10.1109/MNET.012.2100293.
10. Wang Z, Zhou F, Zhang K, Wang Y. Large-scale IP usage identification via deep ensemble learning (student abstract). Proc AAAI Conf Artif Intell. 2022;36(11):13077–8. doi:10.1609/aaai.v36i11.21675.
11. Li Z, Zhou F, Wang Z, Xu X, Liu L, Yin G. Measuring and classifying IP usage scenarios: a continuous neural trees approach. Sci Rep. 2024;14(1):5144. doi:10.1038/s41598-024-55750-x.
12. Liu L, He Y, Zhang L, Liu L, Li B. GraphCyber: identifying IP usage scenarios for cyberspace mapping. In: Proceedings of the ICC 2024-IEEE International Conference on Communications; 2024 Jun 9–13; Denver, CO, USA. p. 3310–5.
13. Nazarenko E, Varkentin V, Polyakova T. Features of application of machine learning methods for classification of network traffic (features, advantages, disadvantages). In: Proceedings of the 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon); 2019 Oct 1–4; Vladivostok, Russia. p. 1–5.
14. Lakhina A, Papagiannaki K, Crovella M, Diot C, Kolaczyk ED, Taft N. Structural analysis of network traffic flows. In: Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems; 2004 Jun 10–14; New York, NY, USA. p. 61–72.
15. Iglesias F, Zseby T. Analysis of network traffic features for anomaly detection. Mach Learn. 2015;101:59–84. doi:10.1007/s10994-014-5473-9.
16. Dan O, Parikh V, Davison BD. IP geolocation using traceroute location propagation and IP range location interpolation. In: Proceedings of the Companion Proceedings of the Web Conference; 2021 Apr 19–23; Ljubljana, Slovenia. p. 332–8.
17. Magoni D, Pansiot JJ. Analysis of the autonomous system network topology. ACM SIGCOMM Comput Commun Rev. 2001;31(3):26–37. doi:10.1145/505659.505663.
18. Di BG, Patrignani M, Pizzonia M. Computing the types of the relationships between autonomous systems. In: Proceedings of the IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies; 2003 Mar 30–Apr 3; San Francisco, CA, USA. p. 156–65.
19. IPIP [Online]. [cited 2025 Feb 18]. Available from: https://www.ipip.net.
20. ChunZhen [Online]. [cited 2025 Feb 18]. Available from: https://cz88.net/.
21. IP2location [Online]. [cited 2025 Feb 18]. Available from: https://www.ip2location.com.