SeeStar: An Efficient Starlink Asset Detection Framework

Linkang Zhang[1,2,3], Yunyang Qin[1,2,3], Yujia Zhu[1,2,3(✉)], Yifei Cheng[1,2,3], Zhen Jie[1,2], and Qingyun Liu[1,2,3]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
zhuyujia@iie.ac.cn
[2] National Engineering Research Center of Information Security, Beijing, China
[3] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

Abstract. Starlink is a new communication network architecture that uses thousands of low-orbiting satellites to provide high-speed, low-latency Internet services. However, there is still much information about Starlink that has not been disclosed to the public. The details of Starlink network architecture, and key nodes which are important to deeply understand and evaluate the performance, security, and impact of Starlink, etc. are still not known. In this paper, we propose an efficient Starlink asset detection framework based on active detection, passive detection, and non-intrusive search engine-based detection methods for the effective discovery and identification of Starlink assets. Based on this framework, this paper implements SeeStar, a Starlink asset mapping system, and provides a detailed analysis of Starlink ground stations and key nodes, exploring their roles and characteristics in the network. Finally, this paper provides an aggregated analysis of Starlink assets in terms of device and service dimensions, and attempts to evaluate their security. The work in this paper provides a powerful methodology and system to unravel the mystery of Starlink network.

Keywords: Starlink · Satellite Internet · Network Asset · LEO · Detection · Mapping

1 Introduction

Starlink [1] is a LEO (Low Earth Orbit) satellite Internet communication system built by SpaceX (Space Exploration Technologies Corp.) in recent years, aiming to provide high-speed, low-latency, and highly stable Internet services through the deployment of low-orbit broadband satellites with global coverage. As of March 2023, Starlink has more than 3,800 satellites in orbit [2], providing Internet services to 50 countries and regions, and currently has more than 1 million subscribers [3,4], with plans to deploy 12,000 satellites and eventually expand to 42,000 [5].

As a project to provide Internet services using space technology, Starlink has attracted global attention since its launch in 2015. Starlink uses an innovative

M. Yung et al. (Eds.): SciSec 2023, LNCS 14299, pp. 139–156, 2023.
https://doi.org/10.1007/978-3-031-45933-7_9

network system that assigns dedicated IP addresses and protocols to satellites and hardware devices. In addition, Starlink has installed laser cross-link technology on the satellites, changing the traditional satellite communication model. This technology allows satellites to transmit and forward data to each other, reducing reliance on ground stations. Starlink also uses a new P2P network protocol and end-to-end hardware encryption technology that outperforms conventional Internet technologies in terms of security and prevents data theft or cracking.

Although Starlink has launched a large number of satellites and provided services to some countries and regions, the status of its assets is not well known. Starlink assets are familiarly known as satellite operations, and known TLE files [6] and websites publish the status of space segment assets. However, as an Internet service provider, little is known about its ground segment network assets. In order to deeply analyze and understand the characteristics and advantages of the Starlink network, its ground segment network assets need to be effectively detected and identified. At present, some cyberspace search engines [7–11] have started to include Starlink network assets, but because their search scope is too broad and lacks targeting, and there is no unified Starlink network asset identification standard and dynamic update mechanism, the data they provide often have poor timeliness, much noise, and low accuracy. To address these problems, this paper proposes a Starlink asset detection method that integrates active detection, passive detection, and non-intrusive search engine-based detection methods, aiming to improve the efficiency and accuracy of Starlink asset detection.

Specifically, the main contributions of this paper are as follows:

1. Propose **the first efficient and targeted Starlink asset detection framework** based on the Starlink network architecture and open data sources, and integrate a heuristic algorithm for Starlink asset detection.
2. Deploy **Starlink Asset Mapping System called SeeStar**, which enables continuous dynamic detection, Starlink ground station discovery and critical node classification and mapping of Starlink IPv4 and some IPv6 assets.
3. Obtain Starlink asset data and perform aggregation analysis to **refine the characteristics of Starlink assets**, and can **infer Starlink key assets** based on Starlink asset attributes and **analyze its security** based on the characteristics of assets.

The rest of the paper is organized as follows: Sect. 2 reviews related work. Section 3 introduces the starlink network architecture and definition of starlink asset. Section 4 describes in detail the starlink asset detection framework. Evaluations are presented in Sect. 5. Finally, we make final remarks in Sect. 6 and Sect. 7.

2 Related Works

Satellite Internet mainly relies on the space satellite constellation to achieve seamless global Internet connection and provide broadband Internet access to

users anytime and anywhere, which is the new generation of Internet infrastructure and is the inevitable trend of future network infrastructure development. The academic and industrial communities are also increasingly interested in studying the satellite Internet represented by the Starlink network.

In the study of Starlink, Michel et al. [12] summarized the analysis of Starlink performance conducted by researchers using active measurement methods, as well as their evaluation of performance under load and packet loss using QUIC. M. M. Kassem et al. [13] utilized a web browser extension to measure Starlink connectivity performance, aiming to answer questions such as how Starlink connectivity compares to other ISPs in the same geographic region, whether connection quality changes over time, and whether weather affects performance. Stock et al. [14] discusses the use of distributed on-demand routing for LEO mega-constellations, using the Starlink case study as an example. Ma et al. [15] uses experiments and observations to study the network characteristics and performance of Starlink, the largest LSN constellation, with a focus on end-to-end user experiences.

Another research aspect related to our work is asset detection. Feng et al. [16] proposes a scalable framework for profiling physical devices on the Internet, using network reconnaissance and banner grabbing to extract device information. Meidan et al. [17] describes the development of a multi-stage meta classifier that utilizes machine learning algorithms to accurately identify and classify IoT devices based on network traffic data, which was collected and labeled from a heterogeneous set of devices. Leonard et al. [18] summarizes the authors' use of the IRLscanner tool to perform 21 Internet-wide experiments for service discovery, and analyzes feedback generated while suggesting novel approaches for reducing blowback.

Currently, mainstream search engines such as ZoomEye [7] have scanned and collected some Starlink asset data, but their exploration is too broad to be comprehensive or dynamic. Due to Starlink's continuous development and changing IP allocation and assets, identifying its assets is inconsistent, leading to suboptimal data accuracy and timeliness. We propose a combination of exploration methods based on active detection, passive detection and search engines, along with continuous exploration using publicly available data sources, for comprehensive and timely identification of Starlink assets with high accuracy.

3 Starlink Architecture and Assets Definition

As a satellite constellation system, Starlink uses a constellation of LEO satellites designed to provide high-speed Internet service to rural and remote areas where Internet connectivity is unreliable or non-existent, ultimately achieving global Internet coverage. The Starlink system is divided into three main parts: the user segment, the space segment, and the ground segment. The Starlink system architecture is shown in Fig. 1.

The user segment belongs to the user intranet. This segment contains mainly user devices and Starlink terminals. The wireless router is used to provide Wi-Fi

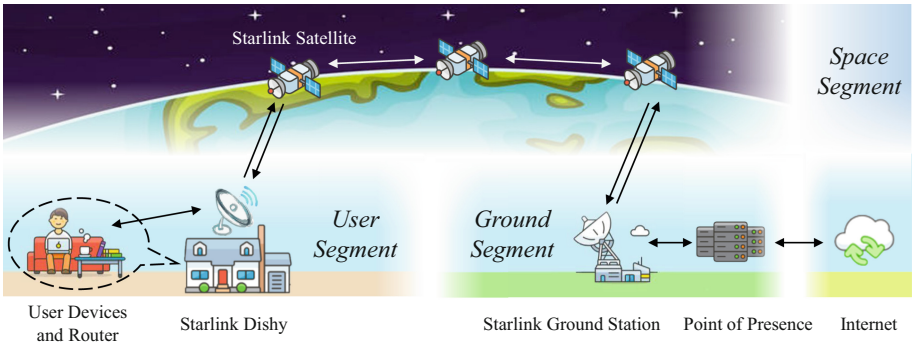

Fig. 1. Starlink System Architecture

signals for user devices to access the local LAN empowered by Starlink. The satellite receiver with phased-array antennas is used to track Starlink satellite signals in near-Earth orbit.

The space segment is part of the inter-satellite network. In this segment, the components communicate and relay via a proprietary, new hardware encryption technology that has not been officially disclosed by Starlink, rather than the traditional TCP/IP protocol stack. Because of the extremely limited public data available for this segment, the lack of clarity on its specific technical details, and the insufficient work at the protocol level, it is currently not possible to probe this segment through existing mapping techniques, and electromagnetic or signal analysis may serve as an entry point for subsequent research.

The ground segment contains the ground gateway station, the point of presence, and the data center network. The ground station is used to communicate with the satellite and send data read from the satellite to the point of presence. The data center is where Internet service providers aggregate their networks and share bandwidth. In the Starlink network, the ground gateway station does not directly access the Internet but instead accesses a nearby point of presence via fiber optics to reach the data center, which provides Internet access.

In short, in the Starlink network, user devices send data packets to Starlink user terminals through the local LAN, and the user terminals encode the packets and send the data to the LEO satellite through the uplink. After the inter-satellite transmission, the LEO satellite sends the data to a specific ground gateway station through the downlink according to the pre-designed routing algorithm, and the ground gateway station further processes the data and transmits the data to the point of presence through the optical fiber. Finally, the data enter the Internet through the data center.

In this paper, according to the Starlink network architecture, Starlink assets are divided into three parts according to function and location: user segment assets, space segment assets and ground segment assets. User segment assets mainly include Starlink terminals and other devices used by users. Space segment assets mainly include Starlink satellite network, sensors and other devices, and

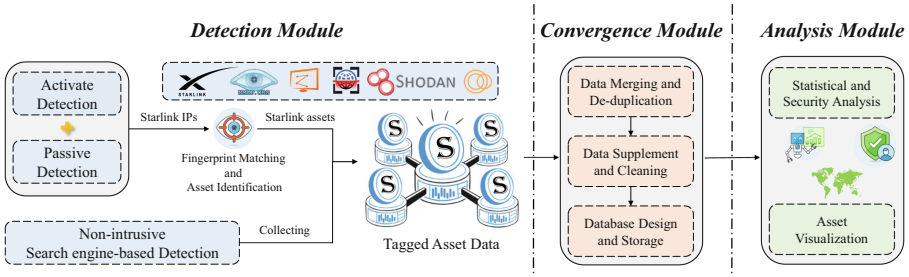

Fig. 2. Starlink Asset Detection System Architecture

ground segment assets mainly include Starlink ground stations and PoP(point of presence) and other key nodes. This paper argues that because the Starlink network uses a new communication protocol and data transmission method different from the traditional TCP/IP protocol stack, as well as its fewer public data sources, it is difficult for user segment and space segment assets to be effectively detected and mapped. Therefore, the research in this paper **focuses on the detection of assets in ground segment**. The public information of Starlink's key nodes and the locations and parameters of ground stations are obtained by collecting and collating public data sources. Then, active detection, passive detection, and non-intrusive search engine-based detection are carried out on top of this to obtain more information about Starlink assets. Finally, this paper constructs the Starlink asset detection framework and deploys the Starlink asset mapping system to achieve the detection and mapping of Starlink ground segment assets.

4 Starlink Asset Detection Framework

In this paper, we propose an efficient Starlink asset detection framework. The framework is divided into three modules: detection, convergence and analysis. The architecture of Starlink asset detection framework is shown in Fig. 2.

4.1 Deteciton

Detection Criteria. At present, there exists a huge amount of network assets on the Internet, and it is a very significant challenge to accurately obtain the IPs associated with Starlink and identify their assets from the large-scale network assets. In order to improve the accuracy and efficiency of Starlink asset identification, we develop the following Starlink asset identification criteria to achieve accurate identification by using Starlink official data [1], Whois database [19], BGP database [20] and other diversified intelligence. For the collected network assets, we verify whether the relevant attributes in the whois database satisfy the rules shown in Table 1, and any one of them will be considered as a candidate Starlink-associated asset.

whois和bgp用于验证，但是验证规则已经稍微过时

Table 1. Starlink Asset Detection and Identification Criteria

Attribute	Value
Organization	SpaceX/SpaceX Services, Inc/Space Exploration Technologies Corporation/Space Exploration Holdings LLC
ISP	Space Exploration Technologies Corporation/SpaceX Services,Inc/Spacex.com
Hostname	SpaceX/starlinkisp/.pop.starlinkisp.net./.mc.starlinkisp.net
Banner	SpaceX/Starlink/Starlink GW AP FTP server/Starlink Exporter/gRPC connection state to Starlink/dishy.starlink.com
AS	AS14593/AS27277/AS397763

Detection Methods. This section focuses on the introduction of Starlink asset detection methods. The detection section mainly utilizes active detection, passive detection and non-intrusive search engine based detection methods, aiming to enhance the ability of Starlink asset data acquisition and improve the validity and accuracy of Starlink asset data by using multi-source data and multi-dimensional methods.

Active Detection. Active Detection enables the acquisition of Starlink asset information by performing alive scanning and port scanning to the destination host. The input of the active detection is a list of network IP addresses and the output is the IP addresses associated with Starlink. The main function of this module is to extract the IP addresses associated with Starlink from the Internet IP address space for subsequent acquisition of Starlink-associated asset data. Specifically, the network IP addresses are probed in an active way using a combination of ICMP, TCP, and other types of probing packets, combined with Zmap [21] and Nmap [22]. The Whois information corresponding to the IP addresses contains information about the segment name, the organization to which it belongs, etc. Therefore, the network IP addresses are scanned to collect the data. By scanning the IP addresses of the whole network and collecting the surviving IP addresses, combined with the information in the Whois database, the IP addresses associated with Starlink can be mined based on the asset identification criteria above accurately and efficiently. Then the Starlink-associated IP addresses are labeled to build the Starlink asset-associated IP dataset. Finally, active detection identifiers are added to each piece of data. In addition, as a supplement, combining DNS data, autonomous system information, and public information, it is also possible to mine Starlink-associated IP addresses with certain characteristics among them. The rDNS [23] can be obtained by scanning the global IP address space for rDNS probes. If the domain name text in the detection result contains Starlink-associated keywords such as starlinkisp, the corresponding IP is considered a candidate Starlink-associated IP address. Since the current global IP address space is unusually large and inefficient scanning cannot accomplish the goal of fast scanning, a distributed scanning method is used to efficiently scan global IP addresses based on global scanning nodes and using a distributed architecture.

XMAP-IPV6

Passive Detection. Passive Detection uses network sniffing tools to get the traffic, analyze the IP and other data in it, and combine the IP corresponding

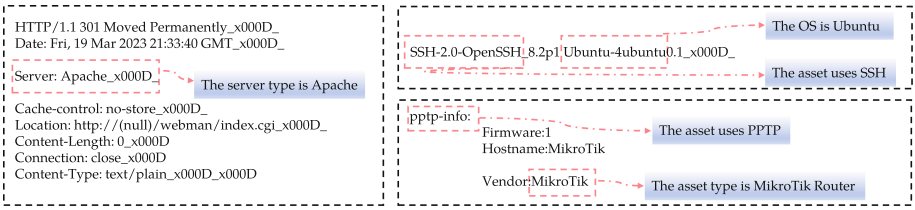

Fig. 3. Identifying assets based on banner

domain name information and Whois information to get Starlink assets. The input of the Passive Detection is the traffic on the Internet (which may involve Starlink-associated IP addresses) obtained by using network sniffing tools, and the output is the Starlink-associated IP addresses. For the IPs in the collected traffic, we first query their Whois information to determine whether their domain names, organization names, etc. match the characteristics of Starlink assets, such as the domain name is customer.*.starlinkisp.net or the organization name is SpaceX, etc. The Starlink-associated IP addresses are also labeled and incorporated into the Starlink asset-associated IP dataset, and a passive probe identifier is added to each piece of data.

Non-intrusive Search Engine-Based Detection. At present, the mainstream cyberspace search engines have included a portion of Starlink asset data, however, due to the detection cycle and other factors, these data may have poor timeliness and other problems. In order to get a more comprehensive overview of Starlink assets, a non-intrusive detection method based on cyberspace search engines is used to obtain Starlink asset data to expand the data obtained by active and passive detection. The input of the search engine-based non-intrusive detection module is a search statement designed based on asset identification criteria, and the output is a list of Starlink-associated assets. Relying on the current relatively mature cyberspace search engines Zoomeye [7], Shodan [8], Quake [9], fofa [10], and Censys [11], such cyberspace search engines are earlier developed and more mature in technology, and usually provide information such as IP, port number, geographic location information, service type, device type, and product type. Therefore, the Starlink asset detection and identification criteria developed above are used to construct search statements that conform to the syntax of each search engine. The Starlink-associated asset data is retrieved by the APIs provided by each search engine, and each piece of data is included in the Starlink asset-associated IP data set, adding a non-intrusive search engine-based detection marker and tagging its data source.

能否搞到一个端口列表？

Asset Identification Methods. For the list of Starlink-associated IP addresses obtained, a list of candidate scanning ports is designed based on the search engine results, combined with experience. We use TCP protocol to connect to the specified port of the target IP and print out the information returned from the port to get the banner shown in Fig. 3.

The asset's corresponding device, operating system, product and service are then identified by means of fingerprint matching and open-source tools. The identification results are merged with the asset data obtained from non-intrusive search engine-based detection and stored in the original Starlink asset dataset.

This section further processes the data tables by updating the original three data tables. For each Starlink asset data, it contains information such as detection type and data source, asset identification (IP and port number), IP address, port number, hostname, autonomous system number, service, device, operating system, product, latitude, longitude, country and banner and other raw information.

Algorithm 1:

Input: I, T, R, P
Output: Starlink Assets Table A

IPs← Randomly Rearrange(IPs);
for IP in I **do**
  if IsAlive(IP) **then**
    for r in R **do**
      if RuleTest(IP, r) **then**
        Store((Active, IP), IP_{SL});
      end
    end
  end
end
for t in T **do**
  IP←ExtractIP(t) **for** r in R **do**
    if RuleTest(IP, r) **then**
      Store((Passive, IP), IP_{SL});
    end
  end
end
for IP in IP_{SL} **do**
  for p in P **do**
    $b \leftarrow$ ExtractBanner(IP, p);
    if FingerprintMatch(b) **then**
      Store((Type, IP, p, Details), A);
    else
      run Open Source Tools and store the results into A;
    end
  end
end
for r in R **do**
  $Query$←GenerateQuery(r);
  Res←SearchUsingEngine($Query$);
  Store(Res, A);
end

Heuristic Algorithm. Based on the description of the Starlink asset detection framework, we have summarized a related heuristic algorithm shown in Algorithm 1.

The algorithm is mainly used for Starlink asset detection. The input of this algorithm is the list of IP addresses to be probed I (because the IPv6 address space is too large, the list of IP addresses to be probed in this algorithm description is a subset of the IPv4 address space), the passive traffic data T, the Starlink asset discrimination rules R and the list of port numbers to be scanned P. The output is Starlink asset data table A.

In order to avoid intensive scanning of the same network segment, which may cause the defense mechanism of the target segment and thus affect the detection process, the list of IPs to be scanned is first randomly rearranged. If the IP is alive, the IP is identified as a Starlink-associated IP based on a predefined Starlink asset identification criteria (e.g., domain name information, autonomous system information, organization information, and other keywords), and if it passes the identification policy, the IP address is stored in the $IPSL$ list with an active detection identifier. If the identification policy is passed, the IP address is stored in the $IPSL$ list and the active detection mark is attached.

For the collected passive traffic data T, extract the IP address from each piece of information and determine whether it is a Starlink-associated IP as described above, and if it passes the identification policy, store the IP address in the $IPSL$ list and attach a passive detection mark. The type of service/request of Starlink-associated IPs and the port number and protocol they use may be present in the passive traffic, and this part of data is saved to help in the subsequent acquisition of asset data.

For each IP in $IPSL$, each port in the port list P to be scanned is scanned, the banner returned by the port is extracted and fingerprinted for matching, and if there is matching information, the asset data is stored in the asset data table A in the form of (probe type (active/passive), IP address, port number, asset details).

Next, Starlink asset discrimination rules are used to generate a series of cyberspace search engine query statements and retrieve them, and store each piece of information in the returned results in the form of (probe type (search engine), IP address, port number, asset details) in the asset data table A.

4.2 Convergence

This section focuses on merging and de-duplicating Starlink asset data obtained through active detection, passive detection, and non-intrusive search engine-based detection methods. The data is extracted, verified, and cleaned according to Starlink asset characteristics. And then we implement database design and data storage. Data extraction and verification make the same object data from different sources conform to a unified form and provide guarantees for the fusion of multiple heterogeneous data. Data cleaning can improve the credibility and validity of the subsequent data fusion analysis results.

Data Merging and De-duplication. Starlink assets obtained through active and passive detection and non-intrusive search engine-based detection methods may have duplicate data, so this part of the assets needs to be merged and de-duplicated to form the final Starlink asset dataset.

Data Supplementation. For Starlink asset data from search engines, there may be missing asset data, or there may be cases where the original alive assets no longer exist. Therefore, this data needs to be supplemented. Firstly, we use alive scanning to screen the alive assets, and then we use a combination of Nmap and traditional fingerprint matching methods to identify the services and operating systems of assets, etc.

Data Cleaning. Preliminary analysis of the acquired asset data reveals that there are cases of different descriptions of the same operating system, service, product, and device information in Starlink asset information due to user-defined or vendor differences. In addition, there are cases of inconsistent country names (e.g. using country Chinese names, country English names, country codes, etc.) and inconsistent default values, so the data cleaning rules were developed to facilitate statistics and analysis.

- Attribute naming unification (eg. Ubuntu/ubuntu/UBUNTU)
- Empty value replacement (eg. convert unknown/UNKNOWN to null value)
- Attribute case conversion (eg. convert UBUNTU/Ubuntu to ubuntu)
- Conversion of country names (eg. convert US to United States)
- IP asset expansion (eg. unroll the list of IP addresses)
- Data correction (eg. modify the incorrect data such as ASN)
- Same asset data conflict resolution(eg. Assets change over time, and different conflicting methods may yield different results depending on the time of data collection).

4.3 Analysis

This section focuses on the in-depth analysis of Starlink asset data, which mainly involves the following three aspects: screening of Starlink asset features, aggregation and security assessment of asset data, and visual display of asset data.

Screening of Starlink Asset Features. This aspect mainly involves extracting Starlink asset data from the Starlink asset database according to certain criteria such as operating system, product, device type, and so on, and extracting Starlink asset data that meet the conditions. For example, if you need to analyze the data of all assets whose product type is camera, the system will return a table containing the data of all assets whose product type is camera for subsequent analysis.

Asset Data Aggregation and Security Assessment. This aspect is mainly to aggregate and analyze the asset data obtained according to specific asset characteristics in different dimensions such as open ports, operating systems, device types, etc., and generate corresponding statistical charts. At the same

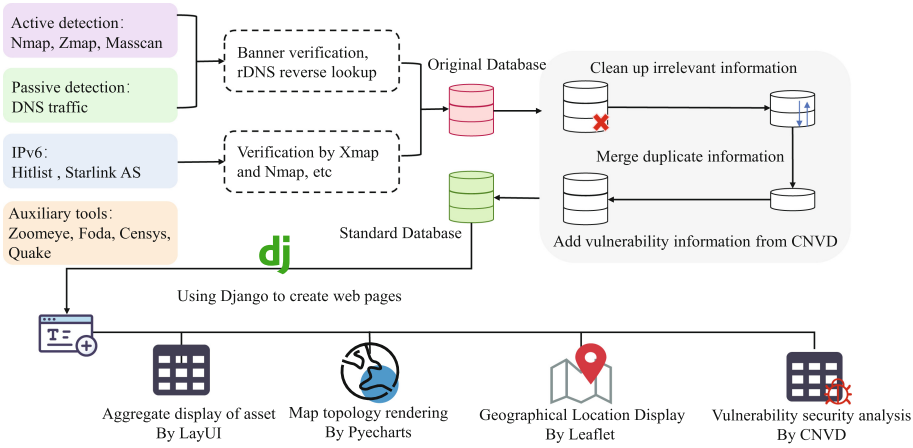

Fig. 4. Starlink System Implementation

time, the vulnerability knowledge base such as CNVD (National Information Security Vulnerability Sharing Platform) [24] is used for aggregation analysis, and the historical vulnerabilities and current unpatched vulnerabilities of asset data are queried according to their operating systems and product types to achieve security assessment of asset data.

Visualization Display of Asset Data. This aspect mainly utilizes some visualization libraries such as Echarts and Leaflet to present the results of convergence analysis and security assessment in a more intuitive and visual way, such as through charts and maps to show the overall situation and distribution characteristics of Starlink's asset data.

5 Evaluation

5.1 Implementation

This paper proposes a Starlink asset detection framework and implements a mapping system SeeStar that can automatically detect, identify and analyze Starlink assets based on this framework. The system mainly consists of detection, aggregation and analysis from the bottom up.

As mentioned above, the overall system operation is divided into three parts, namely detection, aggregation and analysis: in the detection stage, it is mainly active and passive detection to obtain asset information and use tools for verification; in the aggregation stage, it is mainly data processing and integration; in the analysis stage, it is mainly statistics and presentation of assets from various dimensions. The main technologies and the overall architecture flow used in each segment are shown in Fig. 4. In this paper, we adopt an active detection method, based on public data sources and open source tools, to send detection

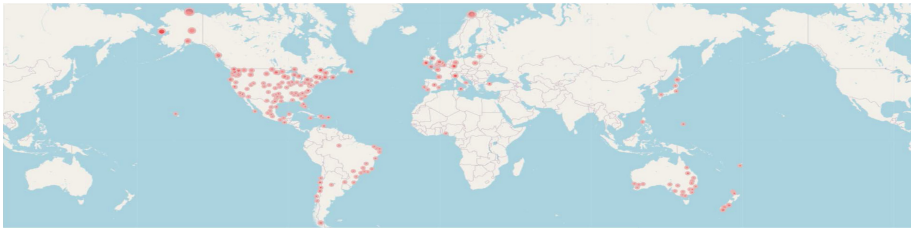

Fig. 5. Distribution of Starlink Ground Station

packages and extract Starlink-related IP. then, we combine data from multiple search engines to collect Starlink's asset information. Then, we fuses the data with CNVD database and uses Django framework to build our Starlink asset mapping system SeeStar. Finally, utilizing Echarts and Leaflet for visual presentation of the data.

5.2 Application

Ground Station Discovery and Critical Node Classification. Based on the public data from FCC [25], Google Map [26], etc., a total of 211 ground stations are collected and analyzed, which are distributed in 27 countries. We mapped the geographic locations of the ground stations onto the map and finally obtained their global distribution which is shown in Fig. 5.

We conducted an in-depth analysis of the IP addresses of the Starlink network, revealing the structure and characteristics of the Starlink network in terms of autonomous systems, host names and network architecture. In this paper, we first count the autonomous systems to which Starlink-associated IP addresses belong, and find that these IP addresses are mainly distributed in three autonomous systems, AS14593, AS27277 and AS397763 [27]. Then, this paper performs reverse domain name resolution on these IP addresses to obtain their host names, and classifies and identifies the nodes of Starlink network according to the naming rules of host names. This paper finds that there are mainly the following types of nodes in the Starlink network:

- **PoP node:** This is an important node in the Starlink network, which is located near the ground station, communicates with the satellite, and connects to the Internet core to provide Internet access services for Starlink users. the PoP node can also interconnect with other PoP nodes or MC nodes to achieve interconnection and redundancy of the network.
- **MC node:** This is another important node in the Starlink network, which is responsible for coordinating data transmission between satellites, earth stations and user terminals, as well as assigning IP addresses and subnet masks, etc. The MC node is the core component of the Starlink user access network.

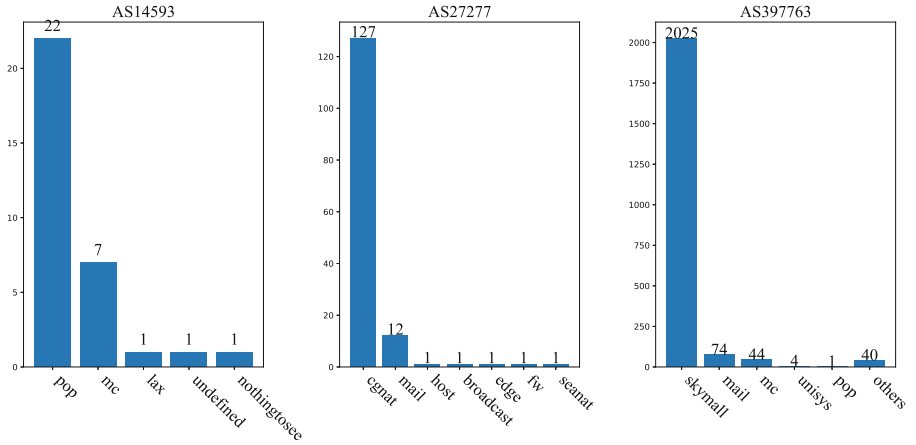

Fig. 6. Distribution of Hostname in AS14593, AS27277, AS397763

– **CGNAT node:** This is a node in the Starlink network used to solve the IPv4 address exhaustion problem and protect the security of user terminals. CGNAT node can realize the conversion between private IP addresses and public IP addresses, thus supporting two-way communication.

In this paper, we analyzed the types of host names in the three autonomous systems, as shown in Fig. 6, and came to the following conclusions:

– **AS14593** is the most important autonomous system in the Starlink network, which contains a large number of PoP nodes and MC nodes, which are key components of Starlink in providing satellite Internet access services and managing network devices. Among these nodes, the number of PoP nodes is significantly more than the number of MC nodes. The highest percentage of assets is also found in AS14593 in the detection results.
– **AS27277** consists mainly of CGNAT nodes, which are used to solve the IPv4 address exhaustion problem and to protect user terminal security. In addition, this autonomous system contains a small number of mail services and special nodes, which may be related to other functions of Starlink.
– **AS397763** has a large number of hostnames that are not obviously associated with Starlink, such as hostnames containing keywords like skymall, mail, unisys, etc. These hostnames may be historical legacies of this autonomous system or domain names planned for future use. In addition, there are a certain number of MC nodes and a small number of PoP nodes in the autonomous system, so it is presumed that the autonomous system has not been fully utilized by Starlink and may be in the construction or testing stage.

Finally, this paper also found that, in addition to AS14593, AS27277 and AS397763 provided by public data sources, there are cases that Starlink associated IPs are attached to other autonomous systems (e.g., AT&T's AS7018,

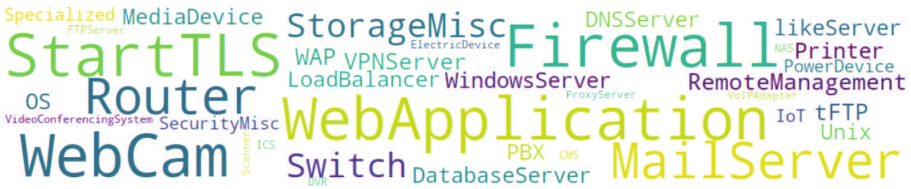

Fig. 7. Starlink Asset Device Type Distribution

Google's AS39462, etc.) or belong to independent network prefixes. These cases may reflect cooperation or competition between Starlink networks and other networks.

Starlink Asset Data Aggregation Analytics. Based on our deployed SeeStar, we have accumulated 10,188 Starlink-associated IPs and 23,132 asset data. We analyzed Starlink asset data specifically and found that all asset data involved 2073 open ports, 219 open services, 22 operating systems. Then we provide an in-depth analysis of the device types in Starlink assets and finds that Starlink assets contain 40 different device types, with firewalls and webcams as the main ones. These device types reflect the current trend of network security and IoT, and also reveal the main uses and functions of Starlink assets. Figure 7 shows the specific distribution of device types.

This paper identifies and counts the device products in Starlink assets, and finds that Starlink assets mainly use products from some well-known companies, such as Amerest, Hikvision, foscam, Dahua and other webcam products, and Sonic WALL, Fortinet, pfSense, and FortiGate and other firewall products. The features and performance of these products can help us further understand the management and configuration strategies of Starlink assets, as well as their potential attack surfaces and vulnerabilities. In order to understand the vulnerability of Starlink assets, this paper integrates CNVD based on the device and product data of Starlink assets to obtain the vulnerabilities that have existed in Starlink assets and those that have not been fixed.

In addition, this paper also counts the number of open ports in the IP addresses of Starlink assets and finds that there are large differences in the number of open ports in the IP addresses of Starlink assets, which may be related to their roles and importance in the Starlink network architecture. We believe that the higher the number of open ports, the more functional and utilized the IP address is, so the importance of Starlink IP addresses can be assessed based on the number of open ports.

Taking an IP shown in Fig. 8 with the highest number of open ports as an example, the IP 129.222.251.*** is located in the United States and its host name is customer.*.pop.starlinkisp.net, so we judge that this host exists as a PoP node by collecting data coming from Internet satellites from ground stations and importing it into the terrestrial Internet. The high number of open ports

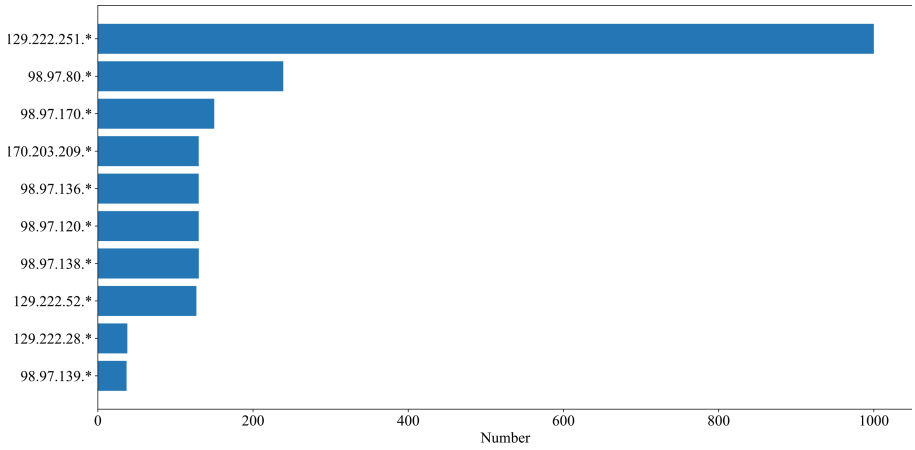

Fig. 8. The top 10 IPs of the detected port openings

indicates that it has more types of services, reflecting that this IP is in a more important position in Starlink's network architecture.

System Accuracy and Proof of Advantage. The purpose of this section is to evaluate the advantages of the system in terms of both accuracy and breadth of data and timeliness of data.

In terms of data breadth and accuracy, this paper adopts three methods: active detection, passive detection and non-intrusive detection based on search engines, and conducts in-depth analysis for Starlink IP allocation, which significantly improves the data breadth through multi-source data fusion. At the same time, this paper uses the Whois database and other Starlink asset features to conduct a comprehensive detection and identification of the entire network IP address space, which also improves the accuracy of the data. In order to verify the detection effect of this paper, the detection results and data volume are compared with the Starlink asset data included in major search engines in this paper, as shown in the table. From the table, we can see that the asset detection framework in this paper can obtain more Starlink asset data more effectively and accurately.

In terms of data timeliness, SeeStar built in this paper achieves real-time updates of Starlink assets across the network by continuously detecting and using Starlink IP to assign public data sources, Whois information and other Starlink asset features. In contrast, the conventional cyberspace search engine is a broad detection of all kinds of assets on the whole network and does not focus on Starlink assets, so it is more lenient in the development of detection criteria and will miss a large number of Starlink assets, and because the search scope of the cyberspace search engine is larger and the operation cycle is longer, the timeliness of the assets it obtains is lower compared with this system, and the system includes the proportion of surviving IPs is higher.

Table 2. Comparison of SeeStar and search engines

System	IPs	Surviving IPs	Survivability	Assets	Surviving Assets	Survivability	Service	Device	OS	Vulnerabilities	Whois
ZoomEye	7388	442	5.98%	13777	924	6.70%	•	•	•	•	•
FOFA	3279	455	13.8%	8045	1325	16.4%	•	•	•	○	○
Quake	3255	388	11.92%	9172	782	8.52%	•	•	•	○	○
CenSys	1696	354	20.87%	3506	680	19.39%	•	•	•	○	•
SeeStar	**10188**	**3015**	**29.59%**	**23132**	**7317**	**31.63%**	•	•	•	•	•

We performed survivability probes on Starlink IPs and assets provided by various search engines and compared our system (SeeStar) results with those provided by these search engines to obtain the results in Table 2. It is proved that our system provides more IP and asset data in terms of data accuracy and breadth. Both IP survivability (29.59%) and asset survivability (31.63%) detected by our system are higher compared to existing search engines in terms of timeliness.

6 Discussion

This paper proposes an efficient Starlink asset detection framework that combines active detection, passive detection and non-intrusive search engine based detection techniques, aiming to acquire Starlink asset data efficiently and accurately.

In this paper, a Starlink asset mapping system SeeStar is implemented using this framework to map Starlink ground stations and key nodes, achieve ground station discovery and key node classification, and perform aggregation and analysis of Starlink asset data. The detection scheme in this paper has high accuracy because it combines publicly available data sources assigned by Starlink IP, uses multiple detection methods, and uses a combination of fingerprint matching and open source tools.

The work in this paper also has some limitations. On the one hand, this paper has a relatively limited understanding of Starlink IP assignments. Since there is a time lag between the release of Starlink public data sources and the allocation and use of IPs, and the Starlink IP allocation is more complex, there are some IPs attached to other autonomous systems or independent network prefixes in addition to the three publicly available autonomous systems. On the other hand, due to the oversized IPv6 address space, this paper does not explore IPv6 assets in depth, although the public data sources and existing studies indicate the existence of IPv6 assets in Starlink.

7 Conclusion

In this paper, we propose a Starlink asset detection framework that combines active detection, passive detection, and non-intrusive search engine based detection methods, and implement a Starlink asset mapping system SeeStar based on this framework.

The system implements ground station discovery and critical node classification. In terms of ground stations, the study finds that the distribution of ground stations matches the countries and regions where Starlink has officially released its services. By analyzing Starlink IP, we found that Starlink IP distribution is not only limited to AS14593, AS27277 and AS397763, but also a small portion is attached to other autonomous systems or affiliated with independent network prefixes. In terms of key nodes, the study found that Starlink key nodes are mainly classified into three categories: PoP, MC, and CGNAT, in addition to other special nodes and unknown nodes. In addition, we specifically analyzed Starlink asset data, aggregating and analyzing Starlink assets from different dimensions including service, device, operating system, and product. Finally, we compare the performance difference between the proposed system and existing search engines and demonstrate that the system outperforms them in terms of accuracy, breadth and effectiveness. In the future, we will further fuse data from multiple sources, expand Starlink IP allocation intelligence to increase the amount of Starlink asset data, and combine IPv6 address space prediction algorithms to more completely probe Starlink IPv6 assets. We will also further optimize the detection strategy by adjusting the detection range and detection period to achieve more effective and efficient detection.

Acknowledgments. This work is supported by the Scaling Program of Institute of Information Engineering, CAS (Grant No. E3Z0191101) and the Strategic Priority Research Program of the Chinese Academy of Sciences with No. XDC02030400.

References

1. Starlink. https://www.starlink.com/
2. McDowell, J.: Starlink Launch Statistics. Planet4589 (2022). https://planet4589.org/space/con/star/stats.html. Accessed 18 Dec 2022
3. SpaceX [@SpaceX]. Starlink now has more than 1,000,000 active subscribers (Tweet) (2022). https://twitter.com/SpaceX/status/1604872936976154624. Accessed 13 Mar 2023
4. Starlink Internet Review 2023: Plans, Pricing, and Speeds. https://www.satelliteinternet.com/providers/starlink/
5. Launches. https://www.spacex.com/launches/
6. CelesTrak: NORAD Two-Line Element Set Format. https://celestrak.org/NORAD/documentation/tle-fmt.php
7. Zoomeye. https://www.zoomeye.org/
8. What is Shodan? - Shodan Help Center. Shodan. https://help.shodan.io/the-basics/what-is-shodan. Accessed 11 Nov 2021
9. Quake. https://quake.360.net/quake/#/index
10. FOFA. https://fofa.info/
11. Censys. https://search.censys.io/
12. Michel, F., Trevisan, M., Giordano, D., Bonaventure, O.: A first look at starlink performance. In: 22nd ACM Internet Measurement Conference (IMC 2022), pp. 130–136. Association for Computing Machinery, New York (2022)

13. Kassem, M.M., Raman, A., Perino, D.: A browser-side view of starlink connectivity. In: 22nd ACM Internet Measurement Conference (IMC 2022), pp. 151–158. Association for Computing Machinery, New York (2022)
14. Stock, G., Fraire, J.A., Hermanns, H.: Distributed on-demand routing for LEO mega-constellations: a starlink case study. In: 2022 11th Advanced Satellite Multimedia Systems Conference and the 17th Signal Processing for Space Communications Workshop (ASMS/SPSC), Graz, Austria, pp. 1–8. IEEE (2022). https://doi.org/10.1109/ASMS/SPSC55670.2022.9914716
15. Ma, S., Chou, Y.C., Zhao, H., Chen, L., Ma, X., Liu, J.: Network characteristics of LEO satellite constellations: a starlink-based measurement from end users. arXiv (2022). http://arxiv.org/abs/2212.13697. Accessed 22 Apr 2023
16. Feng, X., et al.: Active profiling of physical devices at internet scale. In: 2016 25th International Conference on Computer Communication and Networks (ICCCN), Waikoloa, HI, USA, pp. 1–9. IEEE (2016). https://doi.org/10.1109/ICCCN.2016.7568486
17. Meidan, Y., et al.: ProfilIoT: a machine learning approach for IoT device identification based on network traffic analysis. In: Proceedings of the Symposium on Applied Computing, Marrakech Morocco, pp. 506–509. ACM (2017) https://doi.org/10.1145/3019612.3019878
18. Leonard, D., Loguinov, D.: Demystifying internet-wide service discovery. IEEE/ACM Trans. Netw. **21**(6), 1760–1773 (2013). https://doi.org/10.1109/TNET.2012.2231434
19. ASN/IP Whois Query–IPIP.NET. https://whois.ipip.net/
20. BGP.Tools. https://bgp.tools/
21. Durumeric, Z., Wustrow, E., Halderman, J.A.: ZMap: fast internet-wide scanning and its security applications (2013)
22. Lyon, G.F.: NMap Network Scanning: The Official NMap Project Guide to Network Discovery and Security Scanning (2009)
23. rDNS. https://en.wikipedia.org/wiki/Reverse_DNS_lookup
24. National Information Security Vulnerability Sharing Platform. https://www.cnvd.org.cn/
25. Satellite Earth Station: License. https://fcc.report/IBFS/Filing-List/SES-LIC
26. Starlink Global Gateways & PoPs. https://www.google.com/maps/d/viewer?mid=1805q6rlePY4WZd8QMOaNe2BqAgFkYBY&hl=en_US&ll=47.6144489%2C-122.33867770000002&z=8
27. Starlink AS. https://whois.ipip.net/search/SPACEX