NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

VULNERABILITY ANALYSIS OF THE PHYSICAL AND LOGICAL NETWORK TOPOLOGY ON THE U.S. VIRGIN ISLANDS

by

Cameron Jones

March 2022

Thesis Advisor:	Justin P. Rohrer
Co-Advisor:	David L. Alderson Jr.
Second Reader:	Daniel Eisenberg

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** March 2022	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE
VULNERABILITY ANALYSIS OF THE PHYSICAL AND LOGICAL NETWORK TOPOLOGY ON THE U.S. VIRGIN ISLANDS

5. FUNDING NUMBERS
1565443; RCP9G

6. AUTHOR(S) Cameron Jones

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Science Foundation, Alexandria, Virginia 22314; National Science Foundation, 2415 Eisenhower Avenue, Alexandria, Virginia 22314

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release. Distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT (maximum 200 words)

In 2017, two hurricanes, Irma and Maria, left the U.S. Virgin Islands with a destroyed telecommunications infrastructure, demolished homes and collapsed powerlines. Even though the communications system is broken into several sections (e.g., landline telephone, broadcast radio, and Internet service), the telecommunications network as a whole was severely impacted. Previous work has created a mapping and vulnerability analysis of the physical network infrastructure on the island of St. Croix, finding several single points of failure in the St. Croix network infrastructure. Data of the logical network infrastructure has been collected from the Center for Applied Internet Data Analysis (CAIDA) Ark Measurement Infrastructure, the Réseaux IP Européens (RIPE) Atlas Network, and the Naval Postgraduate School. This data is primarily traceroute data measuring the speed and route that messages take on their way to a specified destination. This thesis uses the traceroute data to create interface, router, and autonomous system-level network topologies of the U.S. Virgin Islands. We found that there are several nodes in the graph with high betweenness values, indicating that the network may be susceptible to congestion or disconnection during adverse events. To remedy this, we suggest adding redundancy to the important nodes or adding direct connections between distant nodes.

14. SUBJECT TERMS
internet, network, vulnerability, Center for Applied Internet Data Analysis, CAIDA, Réseaux IP Européens, RIPE

15. NUMBER OF PAGES
73

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

VULNERABILITY ANALYSIS OF THE PHYSICAL AND LOGICAL NETWORK TOPOLOGY ON THE U.S. VIRGIN ISLANDS

Cameron Jones
Civilian, Scholarship for Service
BS, California State University Monterey Bay, 2017

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
March 2022

Approved by: Justin P. Rohrer
Advisor

David L. Alderson Jr.
Co-Advisor

Daniel Eisenberg
Second Reader

Gurminder Singh
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

In 2017, two hurricanes, Irma and Maria, left the U.S. Virgin Islands with a destroyed telecommunications infrastructure, demolished homes and collapsed powerlines. Even though the communications system is broken into several sections (e.g., landline telephone, broadcast radio, and Internet service), the telecommunications network as a whole was severely impacted. Previous work has created a mapping and vulnerability analysis of the physical network infrastructure on the island of St. Croix, finding several single points of failure in the St. Croix network infrastructure. Data of the logical network infrastructure has been collected from the Center for Applied Internet Data Analysis (CAIDA) Ark Measurement Infrastructure, the Réseaux IP Européens (RIPE) Atlas Network, and the Naval Postgraduate School. This data is primarily traceroute data measuring the speed and route that messages take on their way to a specified destination. This thesis uses the traceroute data to create interface, router, and autonomous system-level network topologies of the U.S. Virgin Islands. We found that there are several nodes in the graph with high betweenness values, indicating that the network may be susceptible to congestion or disconnection during adverse events. To remedy this, we suggest adding redundancy to the important nodes or adding direct connections between distant nodes.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

AS	Autonomous System
ASN	Autonomous System Number
BGP	Border Gateway Protocol
CAIDA	Center for Applied Internet Data Analysis
DOD	Department of Defense
GDF	GUESS Data Format
GUESS	Graph Exploration System
ICMP	Internet Control Message Protocol
IGMP	Internet Group Management Protocol
IP	Internet Protocol
ISP	Internet Service Provider
IXP	Internet Exchange Point
MBT	Monotonic Bounds Test
MIDAR	Monotonic ID-Based Alias Resolution
NPS	Naval Postgraduate School
OSI	Open System Interconnection
SNMP	Simple Network Management Protocol
STX	St. Croix
TCP	Transmission Control Protocol

TTL Time to Live

UDP User Datagram Protocol

USN U.S. Navy

USVI U.S. Virgin Islands

viNGN Virgin Islands Next Generation Network

Acknowledgments

I would like to thank my team of advisors, Dr. Justin P. Rohrer, Dr. David Alderson, and Dr. Daniel Eisenberg, for their support and feedback. I couldn't have done it without you.

This material is based upon activities supported by the National Science Foundation under Agreement No 1565443. Any opinions, findings, and conclusions or recommendations expressed are those of the author and do not necessarily reflect the views of the National Science Foundation.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
Introduction

In 2017, two hurricanes, Irma and Maria, impacted the U.S. Virgin Islands (USVI), causing widespread damage. A report by the USVI Hurricane Recovery and Resilience Task Force documented the damage to the islands' infrastructures [1]. Upon the impact of the first storm, Irma, there was nearly a total loss of communications and power across the territory, and the infrastructures were further damaged during the impact of Hurricane Maria two weeks later.

The report describes the communications infrastructure on the USVI as consisting of both private and public communications services. Several private companies provide landline telephone, wireless network, cable television, broadcast radio and television, Internet, submarine cable connections, and power to the islands. On the other hand, public services use a government telephone network, microwave radio-based network and data storage, public safety radio network, middle-mile fiber-optic network, and provide public computer centers and WiFi hotspots.

Since the hurricanes, the infrastructure has been damaged on multiple occasions, causing disruptions in the islands' Internet services. In September 2019, a fiber-optic cable was cut by a USVI Department of Public Works contractor, which disrupted the customers of the Internet Service Provider (ISP) Viya [2], and in October 2020 and March 2021, Viya's fiber-optic cables were accidentally cut by their competitor, AT&T [3], [4]. Despite having "Call Before You Dig" legislation on the islands, where one can call to gain information on the area they plan to dig, these disruptions are still recurring.

1.1 Research Questions

As network disruptions are a persistent problem in the USVI, the question arises: What can be done about this? One possible strategy is to analyze the network topology to find where the weaknesses reside. With this, we ask the following research questions:

- Can we identify critical components of the telecommunications infrastructure based

on the network topology?

- What is the Layer 3 logical topology of the USVI?
- Can we correlate the IP addresses with the physical infrastructure on the USVI?

The motivation for this work is to understand the network topology of the USVI in order to find where the weaknesses reside and to make recommendations to mitigate these weaknesses and to create a more robust infrastructure. Also, in the event of a crisis, understanding the network topology may help with the recovery and repair of damaged infrastructure.

1.2 Network Topology

Despite what is known about the physical infrastructure and technologies that are used, little is known about the router-level network topology of the USVI. Network topology is how a network's nodes and links are arranged [5]. Typically, a network topology is represented as a graph, as shown in Figure 1.1. In a router-level network topology, the nodes in the graph represent the routers and end hosts in the network, and the edges in the graph represent the logical connections between them.

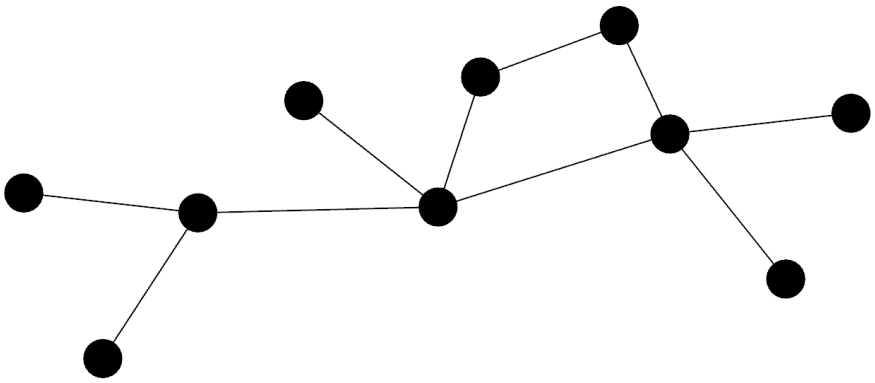

Figure 1.1. An example representation of a router-level network topology. The dots represent the nodes (routers and end hosts) connected to the network, and the lines represent the connections between the nodes.

Alderson et al. [6] state that the Internet as a whole "does not lend itself to direct inspection." This is because the Internet is made up of many smaller networks, each of which is owned and operated by separate organizations. The combination of the sheer size of the Internet and the lack of topology information from network owners make it difficult to model the topology. Alderson et al. [6] go on to say that the methods used for the discovery of network topology create an incomplete topology due to the complexity of network protocol suite.

Despite the difficulty in creating the topology of an unknown network, the knowledge of a network's topology is important from an administrative and engineering perspective [6]. A network's topology can greatly influence its performance. For example, if a network's traffic is funneled through a single router, the router may get inundated with data when the network gets busy. In addition, a network's topology can affect the detection of and reaction to security concerns, such as denial of service (DoS) attacks and malware.

1.3 Scope

This thesis is limited to analyzing previously collected traceroute data and Autonomous System IP subnet data. The main focus of this thesis is analyzing this data, so collecting it is not within scope. This means that measurements have not been taken, despite the increase in understanding that it would provide. However, some subnet data is aggregated from other sources. The network topology is created at the router level to find vulnerabilities in the logical infrastructure and correlate it with the physical infrastructure.

1.4 Summary of Findings

We create the interface-level graph of the USVI network infrastructure from the traceroute data, as well as a router-level graph using Center for Applied Internet Data Analysis (CAIDA)'s alias dataset. We also create Autonomous System (AS)-level graphs with this network topology. We find that the interface-level graphs and the router-level graph have high hop-count values and nodes with high betweenness values. These high values indicate that some nodes have a higher importance, and the network communications are at a higher risk of disruption should one of these nodes go down. The AS-level graphs show only four ASes that are registered in the USVI.

1.5 Thesis Structure

This thesis is structured as follows. Chapter 1 introduces the inspiration for the thesis, states the scope and research questions, summarizes the findings, and defines the thesis structure. Chapter 2 presents the necessary background information for understanding. Chapter 3 outlines the steps taken in the methodology. Chapter 4 presents the results and analysis. Chapter 5 concludes the thesis and presents ideas for future work.

CHAPTER 2:
Background

This chapter discusses the necessary background information for understanding networking topology and its generation. This includes the Internet structure, tools for gathering and interpreting Internet data, how vulnerabilities can be found in a network's topology, and how a network's topology affects its reliability.

2.1 Internet Structure

A starting point for understanding how the Internet works is the Open System Interconnection (OSI) model. The OSI model is "a seven-layer model of data communication with physical transport at the lower layer and application protocols at the upper layers" [7]. Each layer relies on the layer below it to provide a set of functions. These functions include reliable data transfer, packet delivery and routing, and the transmission of electrical signals. Figure 2.1, below, shows the seven layers of the OSI model and how they interact with each other.

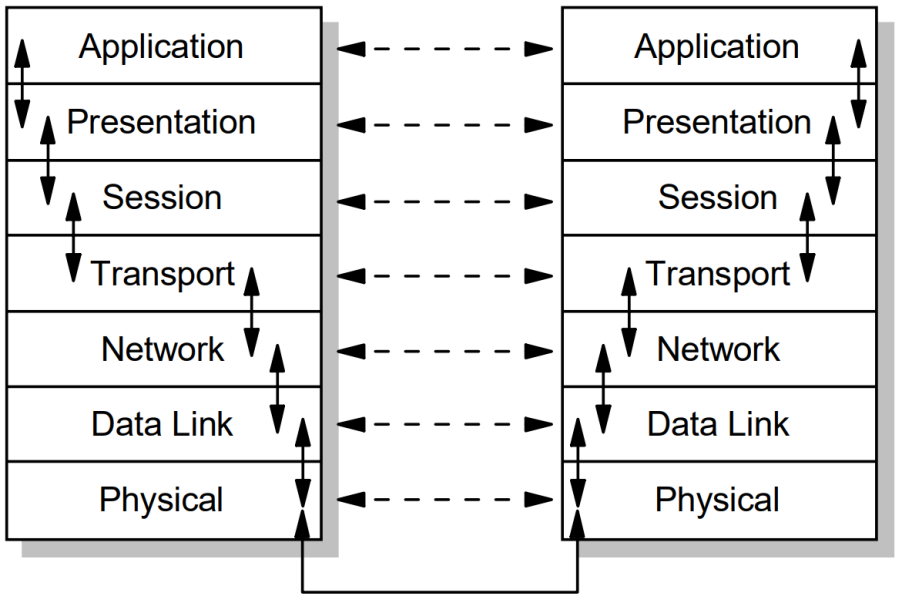

Figure 2.1. The OSI Reference Model. Adapted from Murhammer et al. [7].

Layer 3, or the Network Layer, is responsible for routing a packet from a source host to its intended destination host. This is done by using the Internet Protocol (IP), which is an abstraction layer above the physical network, creating a virtual network view. The Internet Protocol uses IP addresses to identify hosts on the Internet. These IP addresses are used by routers to pass packets along the correct path to the destination IP address. One tool that is used to view the IP addresses along the path to a destination IP address is Traceroute.

2.2 Traceroute

Traceroute is a software utility that is used to map the forward addresses between two hosts [8]. It does this by sending probe packets with increasing Time to Live (TTL) values starting at one and waiting for the Internet Control Message Protocol (ICMP) "time exceeded" replies. Each sequential reply is sent by the next node in the path, thus mapping the topology of the path to the target node.

Traceroute is limited in that it can only reveal a single IP address of each node in the path [9]. This is problematic because the nodes in the network infrastructure have multiple network interfaces, each with separate IP addresses. In order to reveal the IP addresses of all of the interfaces of a node, there must be multiple vantage points that launch traceroute probes.

Another limitation to traceroute is that inaccuracies can be created by load balancing routers [9]. Load balancing sends packets down different paths in order to increase resource utilization. This can create inaccuracies in the traceroute path when one probe takes one path and the next probe takes a different path. This creates a false link in the path between two nodes that are not connected. Per packet load balancing adjusts routes on a per packet basis, while per flow load balancing adjusts the route for each packet flow. The Paris traceroute was designed to resolve per flow load balancing by changing the ICMP headers in the probes so that they take the same path. It cannot, however, resolve per packet load balancing.

2.3 Alias Resolution

Routers typically have multiple network interfaces, each with separate IP addresses. Alias resolution is used to aggregate IP addresses that belong to a single router [10]. This is useful because it helps one gain a better understanding of the network topology and how it relates

to the physical infrastructure.

There are two main methods of creating a router-level topology: alias resolution and recursive router discovery [10]. Alias resolution uses aggregated traceroute data to infer the router-level topology from the interface-level topology. Recursive router discovery uses Simple Network Management Protocol (SNMP) and Internet Group Management Protocol (IGMP) to query a router about its neighbors. As recursive router discovery is an active form of router-level topology creation, it will not be used in this thesis.

There are several methods that are used for alias resolution [10]. Common Source Address sends a Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) probe to an unused port on the IP address in question. Then, it waits for an ICMP "port unreachable" message containing the router's shortest-path interface's IP address. Common IP-Identification Counter "assumes that a router has a single IP ID counter" and sends probes to high port numbers to two potential aliases. A third probe is sent to the IP that replies with an ICMP "Port Unreachable" response first. If the included IDs are within a small range and are sequential, then the addresses are assumed to be interfaces. Graph-Based Resolution Heuristics overlays individual traceroute paths to create a topology graph. The "common successor" heuristic says that if two IP addresses have the same following IP address, then they are aliases. However, this heuristic produces false positives in the presence of layer 2 switches and multiple-access clouds.

The alias resolution methods relevant to this thesis are Monotonic ID-Based Alias Resolution (MIDAR) and Speedtrap. MIDAR resolves aliases of nodes that use IPv4, and Speedtrap resolves aliases of nodes that use IPv6.

2.3.1 MIDAR

The MIDAR tool is the accepted best practice tool for alias resolution of network topology using traceroute data [11]. It is able to mitigate false positives by using IP ID values to construct time series and monotonic bounds tests to check the monotonicity requirements of the IP IDs. It also uses multiple probing methods, multiple vantage points, and a sliding window to make the large amount of necessary probing more scalable. MIDAR uses these components in the following four stages to infer aliases.

The first stage is the Estimation stage. This stage identifies each target's preferred probing method and estimated velocity. To determine the preferred probing method, each target is probed 30 times in a randomized probing order at an interval of about 7.8 seconds. The interval of about 7.8 seconds is short enough to estimate target velocities up to 2,520 ID/s.

The Discovery stage identifies potential address pairs that share an IP ID counter. It starts with a "sliding window probing schedule using the velocities found in the Estimation stage" [11]. The results of the probing are analyzed, and some pairs can be discarded after simple checks on the IP ID byte order and precision before moving on to the Monotonic Bounds Test. The Monotonic Bounds Test is the most important test to determine a shared counter.

The Elimination stage repeats the Monotonic Bounds Test on each potential shared counter address pair to eliminate false positives. The next set of probes use a graph structure to increase probing efficiency with addresses as nodes and potential pair relationships as edges. This graph is split into smaller subgraphs, and each subgraph is probed for 10 rounds.

The Corroboration stage applies the Monotonic Bounds Test on the remaining potential pairs from the Elimination stage, as well as on the pairs implied by the transitive closure of the remaining potential pairs. The probing is the same as in the Elimination stage, but with smaller sets of input that are broken into smaller subgraphs.

2.3.2 Speedtrap

Speedtrap is a router alias resolution tool for interfaces using IPv6 [12]. It uses the same Monotonic Bounds Test as MIDAR, but because there is no ID field in a standard IPv6 packet, it must obtain one using an ICMP Packet Too Big message. This makes the router fragment the IPv6 packet and add an extension header, which includes an ID field.

Luckie et al. [12] describe the Speedtrap algorithm in four steps. The first step determines the IP ID behavior of the interfaces. It does this by sending echo requests to each interface. If at least three responses are received and "the difference in ID values for adjacent responses is less than 65,535" [12] (the maximum size of a two byte integer), then it is inferred that the fragment ID values are derived from a counter.

The next step attempts to find groups of non-overlapping fragments. To do this, probes are

sent to each interface one at a time. Then, the interfaces are grouped together such that all probes that are sent before the reply from the first one is received is a group. This is visualized in Figure 2.2, where Group 2 begins with the first probe sent (probe G) after receiving the reply from the first probe of Group 1 (probe A). Groups that have any time overlap of probes cannot be probed at the same time.

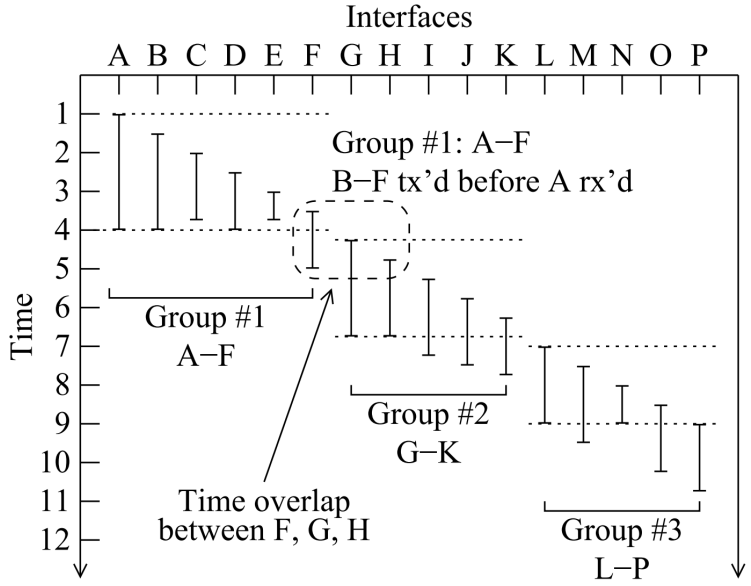

Figure 2.2. Probe grouping of Speedtrap's second step. Used with permission. Source: [12].

In the third step, transitive closure sets of all interface pairs that passed the Monotonic Bounds Test are created. Each closure that has more than three interfaces is then probed to try and force counters on different routers to diverge. After probing, smaller closures are formed where shared counters have been ruled out between interfaces.

The last step tests each pair of candidate interfaces from the third step. Each pair is alternatingly probed, then checked with a Monotonic Bounds Test (MBT). If the MBT suggests that they share a counter, then they are declared aliases. This final step outputs a set of routers with their interface addresses.

2.4 Synthetic Network Generation and Vulnerability Analysis in the U.S. Virgin Islands

Moeller [13] created a synthetic network topology of the Internet infrastructure of the USVI. Generating a synthetic network topology can be useful for measuring the performance of a topology when its ground truth is unknown. Using publicly available data, Moeller was able to create geospatial data sets for each node and edge in the last mile Internet connections and the middle mile network. The last mile infrastructure are the connections between Internet customers and the Fiber Access Points, and the middle mile infrastructure is the network between the Fiber Access Points. With these data sets, he was able to plot last mile Internet connections, middle mile network, and submarine fiber optic cables on a map. The last mile Internet connections included public and private Internet customers (shown in Figure 2.3), fiber access points, and the fiber optic cables connecting them on the island of St. Croix, and the middle mile network includes the fiber access points and the fiber optic cables between them. The submarine cables connect to three points on the island of St. Croix, two of which connect to the islands of St. John and St. Thomas, while the third connects to the network access points in both New York and Florida.

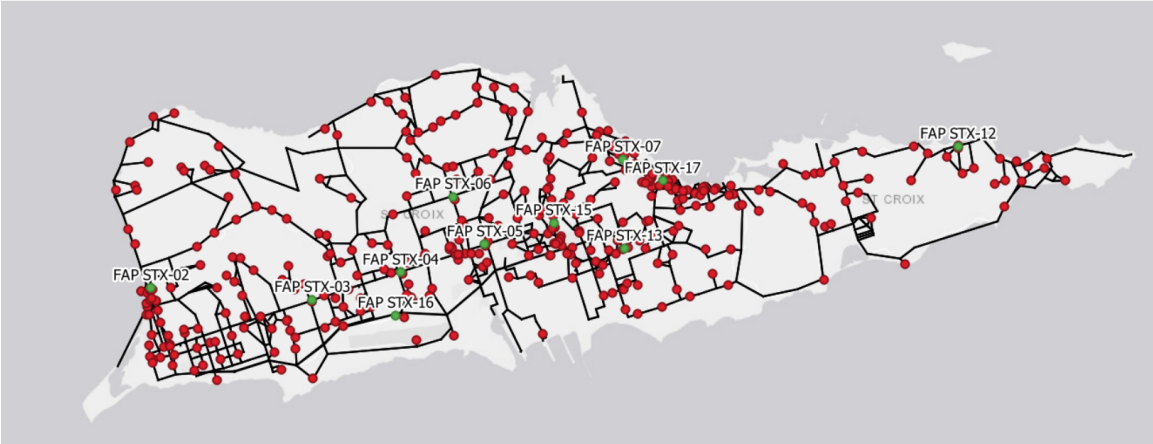

Figure 2.3. Moeller's representation of St. Croix's Last Mile Infrastructure [13]. The black lines show the locations of the fiber optic cables, the red circles show the Internet customers, and the green circles show the Fiber Access Points. Used with permission. Source: [13].

Moeller also developed a model to depict the Internet as a multi-layer network shown in

Figure 2.4. In this model, only layers one through four of the OSI model are used. Nodes are marked with a color corresponding to the highest OSI layer used in that node, as are connections between nodes. This model was used with the geospatial data sets to create Internet system models of the island of St. Croix at different OSI layers.

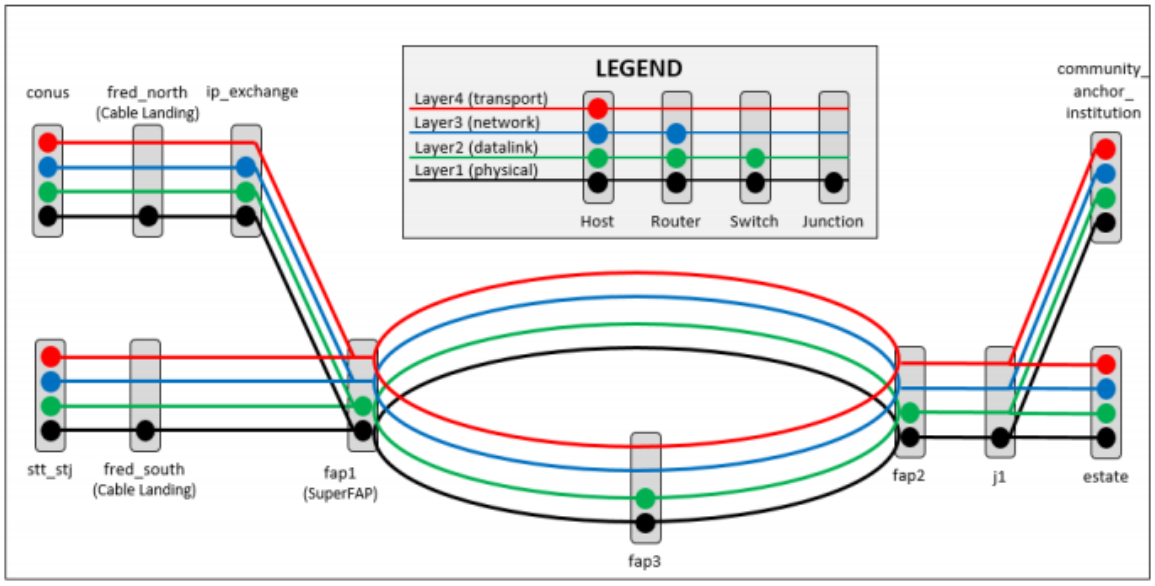

Figure 2.4. Moeller's Internet System model. Used with permission. Source: [13].

Combining the geospatial data and the Internet model, Moeller generated a synthetic network for St. Croix. This network was then analyzed using various Internet traffic scenarios including average and peak Internet demand, fully operational infrastructure, and one or more cable cuts in the fiber network. He found that the model was vulnerable to infrastructure failures and demand spikes. According to Moeller, the loss of key middle mile infrastructure could cause massive disconnection of the St. Croix population, and he found several single points of failure in layer-1 of the Virgin Islands Next Generation Network that, if lost, would disconnect customers. Lastly, he found that available bandwidth of the submarine links between islands and the link between the St. Croix Super fiber access point and Internet exchange point were limited during average demand scenarios, and the available bandwidth on St. Croix could not support surges in demand.

Moeller's work revealed the physical layer topology of the network infrastructure along

with the available bandwidth of the infrastructure. This thesis complements his work by revealing the network layer topology, using traceroute data. The traceroute data cannot be used to show the physical layer topology, nor can the physical layer topology be used to create the network layer topology. However, both topologies can be used together to show a more complete view of the Internet infrastructure on the USVI.

2.5 Network Resilience

The Internet, and the networks that make up the Internet, is depended upon by businesses and consumers, governments, and public services such as railways. In order for uninterrupted use of the Internet in the event of a system failure, resilience should be considered as a part of the network design [14]. Sterbenz et al. define network resilience "as the ability of the network to provide and maintain an acceptable level of service in the face of various faults and challenges to normal operation." This means that even if a part of the network is down (damaged, destroyed, disconnected, or disrupted), communication through the network should still continue by routing messages along alternate paths.

2.5.1 Resilience Disciplines

Sterbenz et al. base network resilience on a handful of disciplines [14]. The disciplines are initially separated into two categories: challenge tolerance and trustworthiness. These categories are depicted in Figure 2.5.

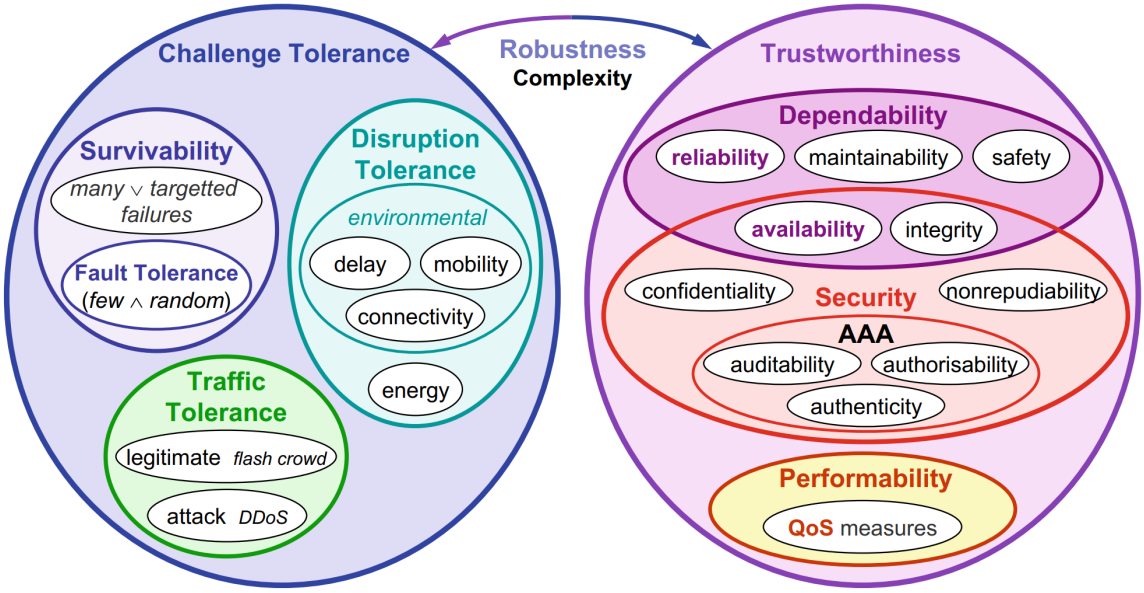

Figure 2.5. Resilience disciplines. Used with permission. Source: [14].

Challenge tolerance includes fault tolerance, survivability, disruption tolerance, and traffic tolerance [14]. Fault tolerance is a system's ability to tolerate faults. Historically, redundancy has been used to counteract failures caused by faults. Survivability is the ability for a system to continue operating while under attack or during a natural disaster. Survivability requires redundancy and diversity, which is the use of different types of components, to counteract failures caused by these threats. Disruption tolerance is the ability to tolerate weak connections or delayed messaging. Lastly, traffic tolerance is the ability to tolerate sudden increases in traffic from either legitimate or malicious sources.

Trustworthiness includes dependability, security, and performability [14]. Dependability is a combination of a service's reliability, maintainability, safety, availability, and integrity. Security shares availability and integrity with dependability and also includes confidentiality, nonrepudiability, auditability, authorizability, and authenticity. The last piece of trustworthiness is performability, which is a system's ability to operate as described by its specifications.

There are also two disciplines that reside outside of the two main categories: robustness and complexity [14]. Robustness relates a system's trustworthiness when faced with a challenge.

Complexity is the interaction between large numbers of systems and the emergent behavior from their interactions.

2.5.2 Resilience Strategies

Building on the resilience disciplines, Sterbenz et al. formalized a two-phase strategy for designing resilient networks [14]. The first phase loops a cycle of four steps: defend, detect, remediate, and recover. The defense step is the primary defense against failures and threats. It uses structural passive defenses such as trusted boundaries, redundancy, and diversity, as well as active defenses such as firewalls and the eventual connectivity paradigm. The detect step is for the network as a whole or for individual components to detect adverse events. This is done by detecting deviations from normal operations, detecting errors in a system, and detecting when a service fails. The remediate step reduces the effect of an adverse event using automated behavior. For example, using alternate network paths to route around a failed system. Lastly, the recover step restores systems back to normal operations.

The goal of the second phase is to improve the cycle of phase one. It is difficult to automate and, therefore, generally requires human intervention. First, the root cause of a fault must be diagnosed. Ideally, one can then remove the fault or increase fault tolerance by adding redundancy. Then, the phase one behavior is refined to increase the network's resilience.

2.5.3 Path Diversification

One way to increase network resilience is to use path diversification [15]. By choosing different paths for network packets, the network load is distributed across nodes. This leads to improved robustness when there is a node or link failure, as another path is chosen when a loss has been detected.

2.6 Our Contribution

This thesis creates network topology graphs at the interface, router, and AS levels. We use traceroute data to make the initial interface graphs. Then, CAIDA's alias dataset to consolidate interfaces for the router level graph. Lastly, we used AS registration and RouteViews' AS subnet data to create the AS level graph. We then analyze the graph for vulnerabilities, and compare this analysis to Moeller's [13] results.

CHAPTER 3:
Methodology

This chapter discusses the steps to create USVI network graphs from traceroute data. Shown in Figure 3.1, these steps include finding IP address subnets that are used in the USVI, filtering traceroute data for traces containing USVI IP addresses, and building a network graph of the USVI Internet infrastructure.

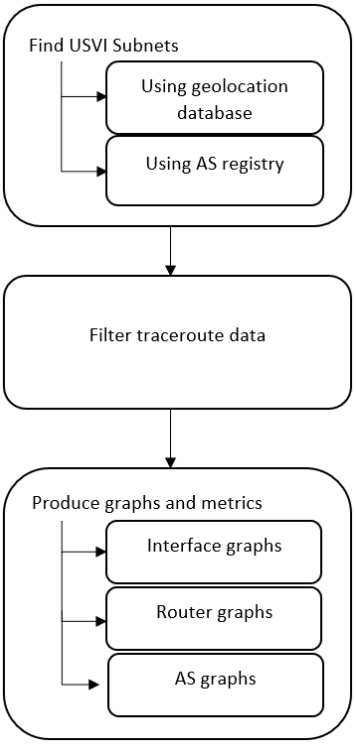

Figure 3.1. Methodology flowchart.

3.1 Finding USVI Subnets

Finding the physical location of a machine using a specific IP address, known as IP geolocation, is a difficult task that is known to have inaccuracies. The most well-known IP geolocation database is MaxMind [16]. Another way to find the physical location of a subnet is to use Autonomous System (AS) registrations. However, it is possible that the physical infrastructure of an AS resides outside of the registered zone.

3.1.1 MaxMind

MaxMind is an IP address geolocation database. This means that it can match a given IP address to a location. Despite IP geolocation being known to be imprecise, MaxMind claims 99.8% accuracy for geolocation on the country level [16]. However, not much is known about how MaxMind obtains this data.

We use MaxMind's locations file and blocks file to find IP address subnets in the USVI. First, the locations file is searched to find all entries that have 'U.S. Virgin Islands' as the entry's country name, and the entry's Geoname ID is saved. Then, the blocks file is searched for each of the saved Geoname IDs, as well as checked if the latitude is within -64.644065 and -65.04842 and the longitude is within 17.67281 and 18.38623. These latitudes and longitudes correspond to the location of the USVI. If either of these checks are true, then the entry is saved to another file. Using this method, we found 223 IP subnets that belong in the USVI.

3.1.2 AS Registry

Using an IP address geolocation database is not the only way to find IP address subnets from specific locations. Although there is not a way to directly search for subnets from specific locations, Autonomous System (AS) registrations can be checked in order to find AS numbers (ASN) registered to the USVI. To do this, we use Team Cymru's IP to ASN Mapping Service [17]. Their service allows us to do bulk requests where we enumerate every possible ASN. Unfortunately, the reply for each ASN request does not include IP address subnets for the registered AS, but it does include a country code that can be used to filter for ASes that are registered in the USVI. This leaves us with a list of ASNs that are registered in the USVI, as shown in Figure 3.2.

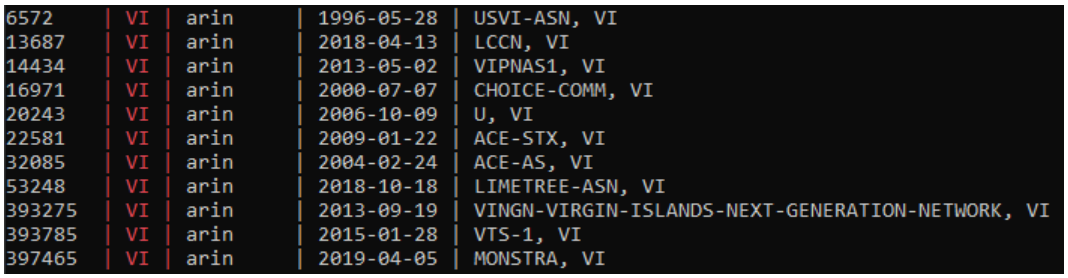

Figure 3.2. The output of Team Cymru's IP to ASN Mapping Service. The left column shows the ASN, followed by the country code (highlighted red), the registry, when the ASN was allocated, and the AS name.

Team Cymru is not the only whois server. Operated by Merit Network, Inc., RADb is another whois server [18]. Unlike Team Cymru's server, RADb's responses include IP address subnets used by the requested ASes. Using the ASN list from the previous step, we made a whois request for each of the ASNs and stripped the subnets from the responses. This leaves us with another list of subnets from the USVI. Using AS registrations, we found 484 IP subnets that belong in the USVI.

Typically, an IP subnet belongs to only a single AS. This is not the case with the list of subnets from RADb, however. The subnet list contains some conflicting data, in that some subnets appear in multiple ASes. For example, the node from the St. Croix (STX) vantage point is in the subnet 146.226.0.0/16, which, according to the RADb whois server, is part of AS 16971 and AS 20243. One possible reason for this conflicting data is that one or more of the AS entries in the whois server is outdated. Despite this, we still use the list of subnets from RADb for the filtering of the traceroute data because all of the ASes that we are looking at are registered in the USVI. In later steps (see Section 3.4), we use a different AS-subnet data set to create an AS-level graph.

3.1.3 Subnet Aggregation

The subnet lists from the MaxMind data and whois requests have multiple overlapping subnets, but there are many unique subnets between the two lists. We next consolidate the lists into a single list without any repeating subnets for use in the next step. This leaves us

with a total of 512 IP subnets that belong in the USVI.

3.2 IP Filtering

The traceroute data we are using is stored in a special binary file called a warts file [19]. These are created by CAIDA's Scamper utility. Scamper actively probes the Internet using techniques, such as ping, traceroute, and other techniques to collect this Internet data. The data we are using specifically used the traceroute technique and was collected in the year 2020. Scamper is not only used to collect data, but it also has tools for manipulating the data to make it easier to analyze. Of these tools, we use `sc_wartsfilter` and `sc_wartscat`.

In order to determine ingress and egress nodes, we first separate the traces that use a vantage point on the USVI from those that did not. The warts files with a USVI vantage point are prepended with "stx-vi", while the files with non-USVI vantage points are prepended with a different code affiliated with its location. Once these are separated, we then use `sc_wartscat` to concatenate the USVI and non-USVI warts. Next, we use `sc_wartfilter` to filter these files. We use the "check-hops" option and the "trace" record type to keep all traces that contain at least one IP address that is from a subnet from our list of USVI subnets. Unfortunately, `sc_wartfilter` can only filter for a single subnet at a time, so this must be repeated for each subnet in our list. The result of filtering is many small warts files, which are concatenated with `sc_wartscat`. Now, we are left with two warts files containing only traces that include at least one IP address from the USVI: one whose traces come from the vantage point on the USVI and one whose traces come from other vantage points across the world.

3.3 Building Network Graphs

Now that the traceroute data has been filtered for USVI IP addresses, we can create network graphs of the data by iterating through the data and tracking each connection. The individual nodes and their connections are then saved as a GUESS (Graph Exploration System) Data Format (GDF). A GDF file is a more specialized version of a .csv (comma separated value) file that is used by the Gephi software. It allows one to list nodes and their connections as well as attributes such as color, visibility, and weight [20]. We initially create an unlabeled graph, and use colors as labels in the next step.

Martineau [21] used the color attribute to track ingress (first hop into a network) and egress (first hop out of a network) nodes of mobile ASes. We use the same color coding as Martineau: green for ingress-only nodes, red for egress-only nodes, blue for nodes that are both ingress and egress, and black for nodes that are neither ingress nor egress. As our graph represents the network in the USVI as well as a few hops outside, we must also differentiate between the nodes that are outside of the network and the nodes that are inside the network. For this differentiation, we used black for nodes that are outside the network and orange for nodes that are inside the network.

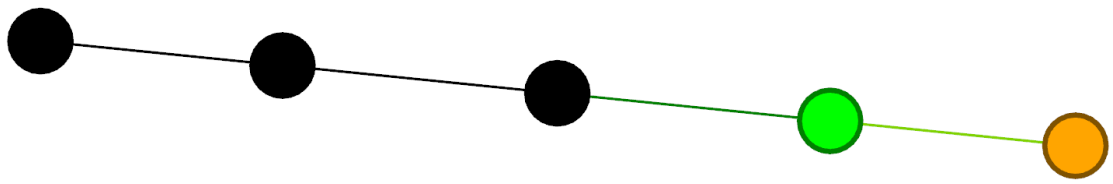

Figure 3.3. A graph showing a single trace. The black dots represent nodes outside of the USVI, while the green dot is an ingress node, and the orange dot is a non-entry node. Both the green and orange dots are part of the USVI network. As the trace originated from outside the USVI, there are no egress nodes (colored red).

Even though only traces that contain USVI IP addresses are kept, there are still relatively few USVI nodes. Non-USVI nodes that are far away (over two hops) from a USVI node are trimmed out of the GDF file, and, at the same time, our color labels are added. To add the color and trim the distant nodes, we iterate several times through the GDF file. The first iteration strips all the non-USVI nodes from the node section and sets the remaining nodes color to orange. The next several iterations strip the distant nodes out of the connections section of the file, until the only non-USVI nodes remaining are those that are within two hops from a USVI node. The last iteration checks each USVI node if it is an ingress or egress node (if it has a connection to a non-USVI node) and changes its color accordingly. This method takes advantage of Gephi's ability to recreate nodes that have been dropped from the node section of the GDF file, but still exist in the connection section.

3.4 Building AS Graphs

An AS-level graph is a graph that shows the connections between ASes. Because there is conflicting information about the subnets from the whois data, we used data from RouteViews to build an AS-level graph [22]. The RouteViews data is Border Gateway Protocol (BGP) Routing Information Base data from which we extracted subnets and the AS to which they belong to.

We build the graph by looking at the list of connections in the network graph. For each connection, we check the AS of each node. If the nodes are from the same AS or either node is within the private IP address space, then we ignore that connection. Otherwise, we save the connection between the two ASes. Some IP addresses that did not belong to an AS were found, along with some IP addresses that belong to ASes that were not in the RouteViews data. For these, we do a second pass: removing the addresses that do not belong to an AS and doing a whois request to find the AS they belong to for the rest. Lastly, the graph is color coded similar to the network graph: orange for USVI registered ASes, blue for non-USVI ASes, and black for IP addresses without and AS.

CHAPTER 4:
Results

This chapter discusses the graphs that we created, the statistics of each graph, the analysis of the resilience metrics of the graphs, and the vulnerabilities of the network topology.

Rohrer et al. [15] use a variety of metrics to analyze network topology graphs. Of these, we select node degree, clustering coefficient, network diameter, network radius, hop-count, closeness, and betweenness to characterize network graphs in this work. Of these metrics, we note that hop-count and betweenness are of special importance. Hop-count directly correlates to path failure probability and betweenness is an indicator that a node or link may be a central point-of-failure.

4.1 Interface-Level Graphs

The traceroute data was initially split into two parts: traces originating from the vantage point on the USVI and traces from vantage points outside of the USVI. From these parts, we created three graphs: one for each part and a graph combining them. These graphs are shown and analyzed below.

4.1.1 From UVSI Vantage Point

The interface-level graph of the traces from the vantage point on the USVI (shown in Figure 4.1) has a total of 669 nodes: 635 from locations off of the USVI and 33 from on the USVI. Of the 33 from the USVI, 11 are egress nodes and the remaining 22 are non-gateway nodes. A large majority of the nodes in this graph are located outside of the USVI. This is likely because most of the traces from the source on the USVI were destined for IP addresses outside of the USVI. The graph has an average node degree of 3.504, but a maximum node degree of 396. The maximum closeness centrality is 0.52 and the maximum betweenness centrality is 204134.96. The same node has the maximum node degree, the maximum closeness centrality, and the maximum betweenness centrality, which leads towards this node being of high importance in the graph. These statistics are shown in Table 4.1 below.

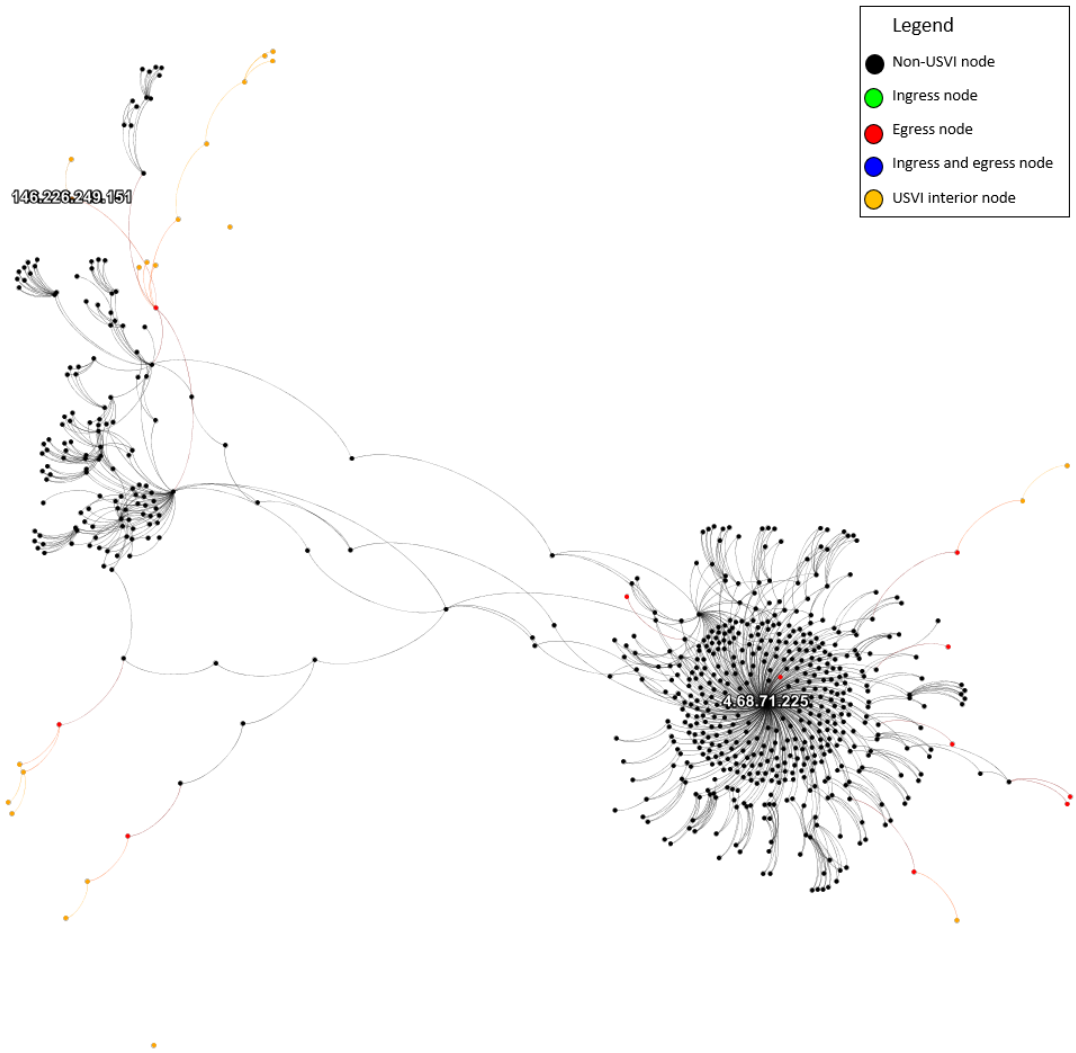

Figure 4.1. The graph from the USVI vantage point. Notable nodes are labeled with their IP address.

Metric	Value
Total Nodes	669
Total Edges	1172
Total USVI Nodes	33
Ingress (Green) Nodes	0
Egress (Red) Nodes	11
Ingress/Egress (Blue) Nodes	0
Inner (Orange) Nodes	22
Average Degree	3.504
Clustering Coefficient	0.449
Diameter	12
Radius	0
Average Hop-count	3.5697
Max Closeness	0.521944
Max Betweenness	204134.962446

Table 4.1. USVI vantage point graph statistics.

The node with the highest degree (the highest number of connections) is located at the IP address 4.68.71.225. In Figure 4.1, it is located in the center of the cluster on the right side of the image. It has a degree of 396, and also has the highest betweenness and closeness values at 204134.96 and 0.52, respectively. A whois lookup for this address shows that it belongs to Level 3 Parent, LLC and is within the 4.0.0.0/9 subnet in AS 3356 [23]. A reverse DNS lookup shows the address has the domain name ae-11.edge6.Miami1.Level3.net, indicating that its physical location is likely in Miami, Florida [24]. Although it is outside of the USVI, according to the geolocation database and AS registry, it is connected to the IP address 199.77.147.57, which is within the USVI. The IP address 199.77.147.57 is also owned by Level 3 Parent, LLC, but is in AS 3549 and does not have a domain name associated with it.

Another notable node on this graph is located at IP address 146.226.249.151. In Figure 4.1, it is located on the left side of the image. This node is a part of the University of the Virgin Islands network and, although it does not have particularly interesting metrics, it is

important because it is the source vantage point of the traceroute paths of this graph. It is in the AS 20243 and does not have a domain name associated with it.

4.1.2 From non-USVI Vantage Points

The interface-level graph from the non-USVI vantage points is shown in Figure 4.2. It has a total of 2800 nodes and 3325 links. There are 731 nodes outside of the USVI and 2069 nodes on the USVI. Of the nodes on the USVI, 1228 are ingress nodes and the remaining nodes are non-gateway nodes. Because the vantage points for each of the traces were outside of the USVI all the gateway nodes are ingress nodes with no egress nodes. The graph's average node degree is 2.375, with a clustering coefficient of 0.045. These statistics are shown in Table 4.2.

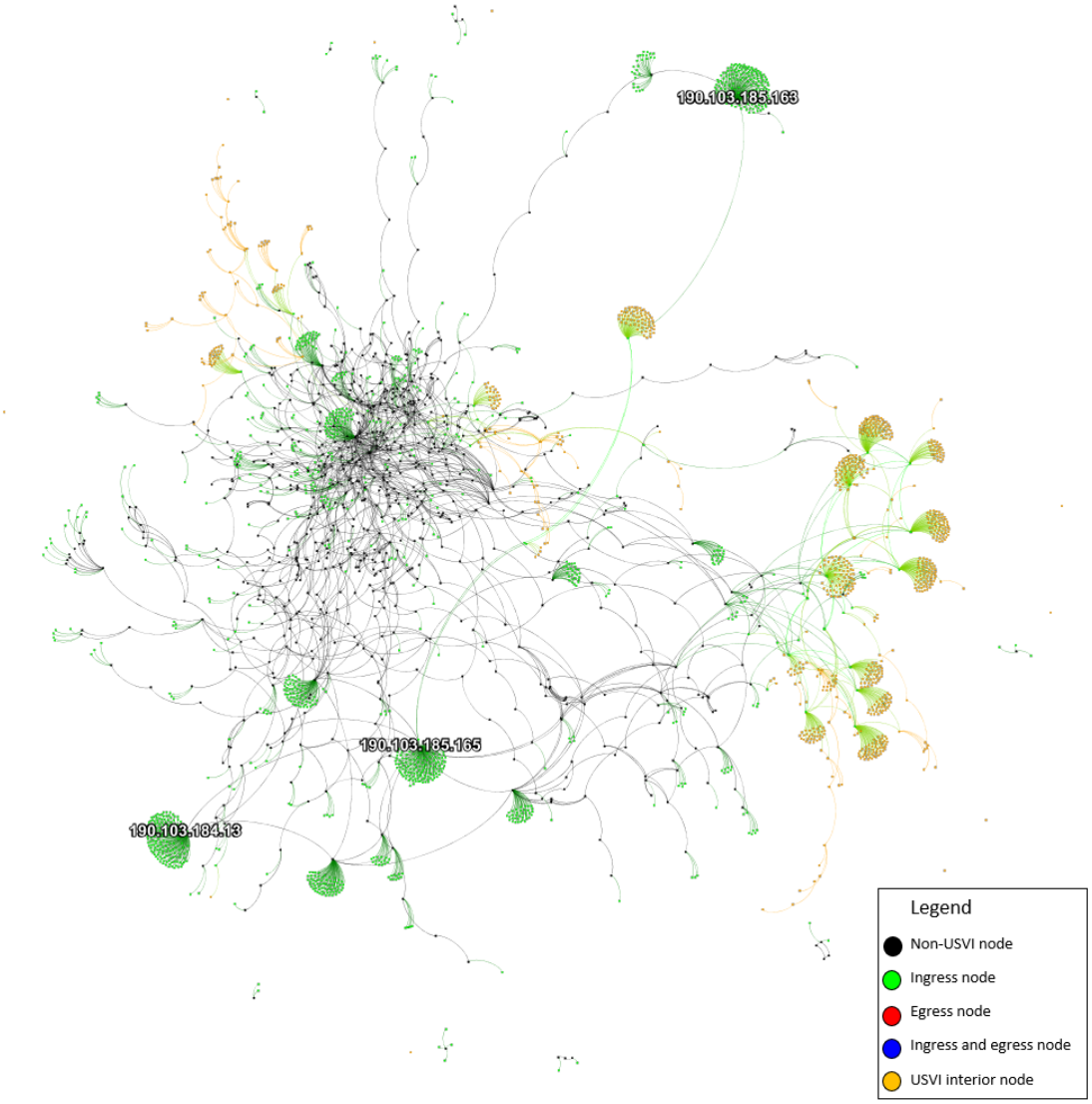

Figure 4.2. The graph from the non-USVI vantage points. Notable nodes are labeled with their IP address: 190.103.185.163, 190.103.184.13, and 190.103.185.165.

Metric	Value
Total Nodes	2800
Total Edges	3325
Total USVI Nodes	2069
Ingress (Green) Nodes	1228
Egress (Red) Nodes	0
Ingress/Egress (Blue) Nodes	0
Inner (Orange) Nodes	841
Average Degree	2.375
Clustering Coefficient	0.045
Diameter	20
Radius	0
Average Hop-count	7.3376
Max Closeness	1320992.77157
Max Betweenness	162.0

Table 4.2. Non-USVI vantage points graph statistics.

There are three nodes that have significantly higher degree than the rest of the nodes. They are located at the IP addresses 190.103.185.163, 190.103.184.13, and 190.103.185.165. These nodes are all part of the same subnet, 190.103.184.0/22, which belongs to AMPATH and is in AS 20080. Their associated domain names are ae0-952.rt04.ce.ampath.net, ae0-42.rt04.bb.ampath.net, and ae0-953.rt04.ce.ampath.net, respectively [24]. The node 190.103.185.165 also has the highest betweenness for the graph.

The node at 190.103.185.163 has a total of 162 connections: two that are outside of the USVI and 160 that are ingress nodes in the University of the Virgin Islands' 146.226.0.0/16 subnet in AS 20243. Of the ingress nodes, all but one are leaf nodes. The single non-leaf node ingress node is located at IP address 146.226.191.225 and is connected to another 65 USVI leaf nodes that are on the same subnet. It is also on the shortest path between 190.103.185.163 and the other two notable nodes.

The IP addresses 190.103.184.13 and 190.103.185.165 are similar in that they connect to

150 and 133 ingress leaf nodes, respectively. These leaf nodes are also in AS 20243 in the 146.226.0.0/16 subnet. They are also connected to several other nodes that are, in turn, connected to a cluster of ingress leaf nodes in AS 20243. In addition, they are both connected to a node at the IP address 129.250.200.114, which is owned by the telecommunications company NTT America, Inc [23]. It is in AS 2914 and has the domain name xe-0-0-23-2.a01.miamfl02.us.ce.gin.ntt.net, which indicates that it is likely located in Miami, Florida [24].

4.1.3 Combination Graph

The combined graph of the traces from both the USVI vantage point and the non-USVI vantage points has 3417 total nodes and 4482 links. Of these, 1323 nodes are outside of the USVI and 2,094 nodes on the USVI. Of the nodes on the USVI, 1224 are ingress nodes, 7 are egress nodes, 4 are both ingress and egress nodes, and the remaining 859 nodes are non-gateway nodes. It has an average node degree of 2.623 and a clustering coefficient of 0.222. As with the graph from the USVI vantage point, the same node has both the maximum closeness centrality and the maximum betweenness centrality. As the graph is very large and difficult to read, it has been omitted from the text. However, Table 4.3 shows the statistics for the graph.

Metric	Value
Total Nodes	3417
Total Edges	4482
Total USVI Nodes	2094
Ingress (Green) Nodes	1224
Egress (Red) Nodes	7
Ingress/Egress (Blue) Nodes	4
Inner (Orange) Nodes	859
Average Degree	2.623
Clustering Coefficient	0.222
Diameter	20
Radius	0
Average Hop-count	6.8282
Max Closeness	1931293.699866
Max Betweenness	396.0

Table 4.3. Combination graph statistics.

4.2 Router-Level Graph

The router-level graph uses the combination interface-level graph, along with CAIDA's alias dataset [25]. The alias dataset was created by using MIDAR on the traceroute data. It is comprised of a list of node IDs and the IP addresses of each interface believed to be on that node. In the resulting graph, shown in Figure 4.3, each node is a router as opposed to a single interface on a router.

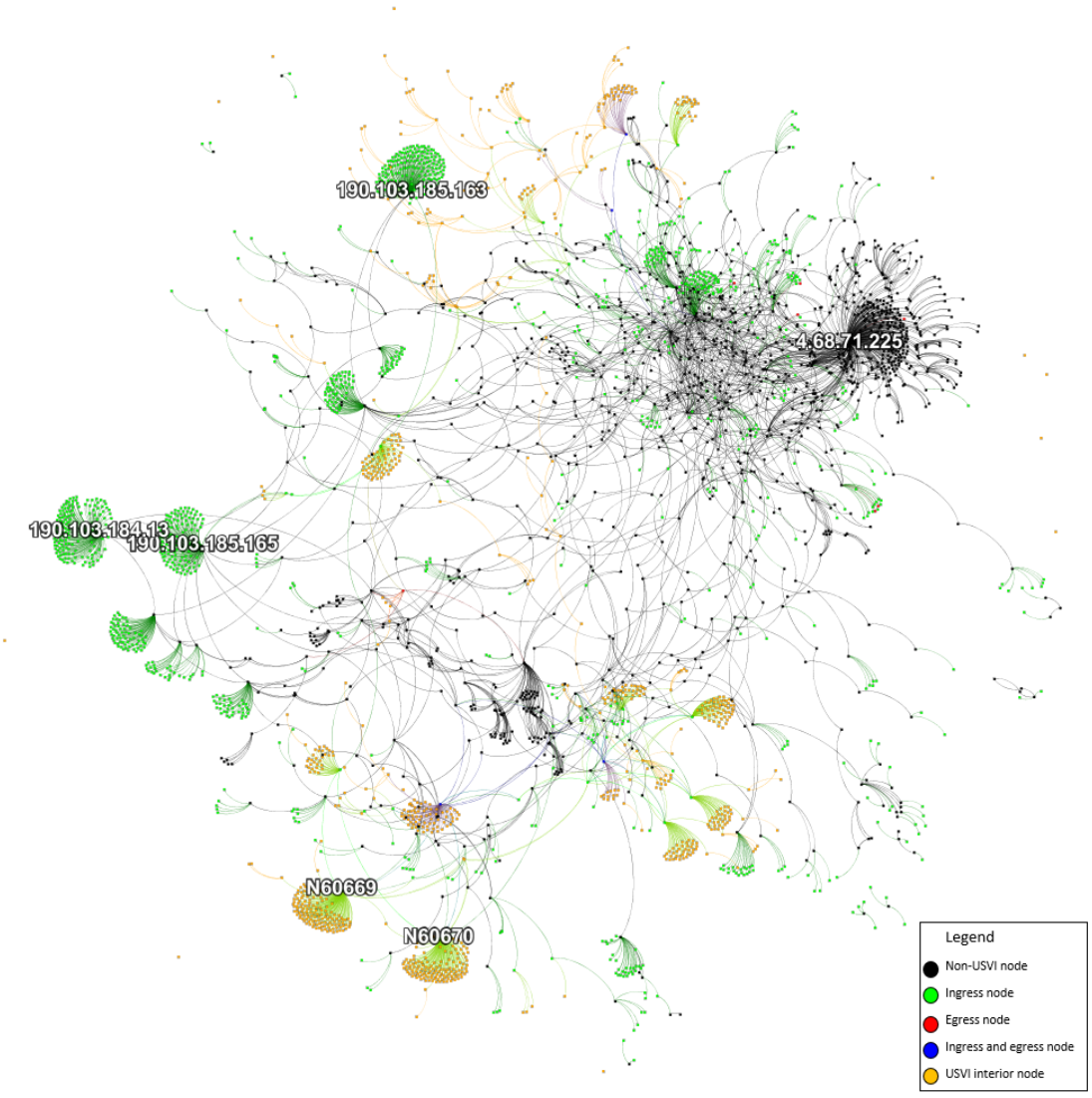

Figure 4.3. The router-level graph.

Overall, 879 nodes were found in CAIDA's alias dataset. However, a large majority of these nodes only have a single interface that was found in the interface-level graph. We found 71 aliases with more than one interface, which leaves the router-level graph with 91 fewer nodes than the combination interface-level graph and nine fewer USVI nodes. This decrease

in node count is due to the consolidation of some nodes from the interface-level graph into a single node. In addition to a decrease in node count, there is also a decrease in edge count: from 4482 edges to 4316 edges. This is also due to the consolidation of nodes. Table 4.4, below, shows the statistics of the graph.

Metric	**Value**
Total Nodes	3326
Total Edges	4316
Total USVI Nodes	2085
Ingress (Green) Nodes	1220
Egress (Red) Nodes	7
Ingress/Egress (Blue) Nodes	4
Inner (Orange) Nodes	854
Average Degree	2.595
Clustering Coefficient	0.235
Diameter	20
Radius	0
Average Hop-count	6.4973
Max Closeness	1.0
Max Betweenness	1678730.923904

Table 4.4. Router-level graph statistics.

The statistics for the router-level graph are very similar to the combination interface-level graph. The average degree and the average hop-count both decrease slightly, which is to be expected when a graph is altered to be slightly smaller. The clustering coefficient increased slightly, which means that when the interfaces were consolidated into their aliases, some clusters were created.

None of the notable nodes from the graphs above were consolidated with other nodes. The nodes at IP addresses 4.68.71.225, 190.103.185.163, 190.103.184.13, and 190.103.185.165 still each have the highest node degrees: 371, 162, 154, and 140, respectively. However, two consolidated nodes, with node IDs N60669 and N60670, also have very high node

degrees, at 140 and 138, respectively. Node N60669 is comprised of IPs 65.112.145.133 and 66.248.175.60. IP address 65.112.145.133 is owned by CenturyLink Communications, LLC according to a whois lookup and 66.248.175.60 is a USVI IP address owned by VI Powernet, LLC [23]. Node N60670 is comprised of 65.112.145.134 and 66.248.175.59. IP address 65.112.145.134 is also owned by CenturyLink Communications, LLC and 66.248.175.59 is owned by VI Powernet, LLC. Interestingly, all the interfaces on N60669 and N60670 are a part of VI Powernet, LLC's AS14434.

4.3 AS-Level Graphs

Figure 4.4 shows the AS-level graph. It was created by mapping each node in the combined interface-level graph to an AS and saving the connections. Then, nodes that are not part of an AS were filtered out. It has a total of 118 nodes and 203 links. Only four of the nodes represent ASes in the USVI and there are two non-USVI ASes that are completely disconnected with two more that are only connected to each other (due to the intermediate links being a part of private address space or not being a part of an AS). The average node degree is 3.441 and the clustering coefficient is 0.463. Because this is an AS-level graph, the node degree indicates the number of peers an AS is connected to.

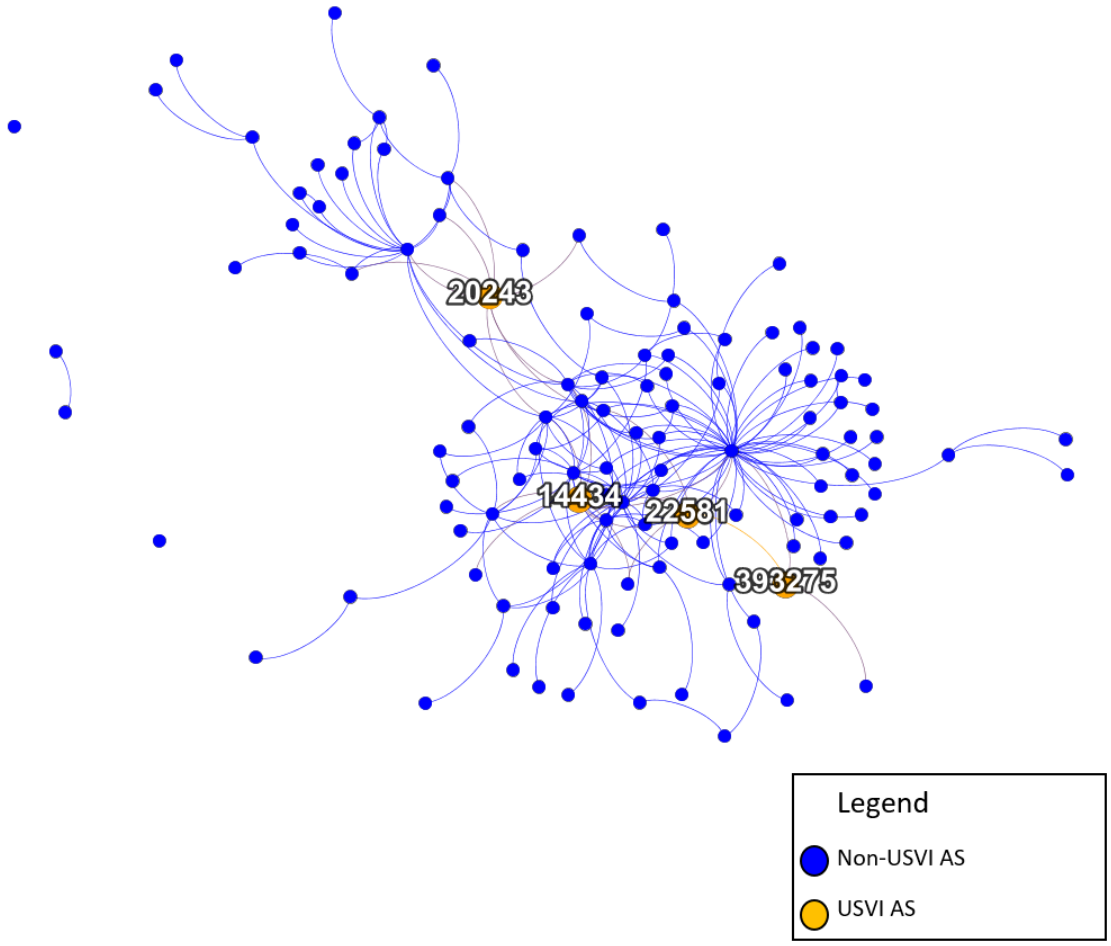

Figure 4.4. The AS-level graph of the USVI networks.

The four ASes that are registered in the USVI are AS14434, AS20243, AS22581, and AS393275. According to Internet Corporation for Assigned Names and Numbers' lookup service, AS14434 is registered by the internet service provider VI Powernet, LLC, AS20243 is registered by the University of the Virgin Islands, AS22581 is registered by the ISP Broad-

band VI, LLC, and, lastly, AS393275 is registered by the Virgin Islands Next Generation Network [26]. Table 4.5 shows the number of nodes and number of peers for each of these four ASes.

ASN	Organization	Number of Nodes	Number of Peers
AS14434	VI Powernet, LLC	649	9
AS20243	UVI	769	8
AS22581	Broadband VI, LLC	151	5
AS393275	viNGN	107	3

Table 4.5. USVI AS statistics.

4.3.1 VI Powernet, LLC (AS14434)

VI Powernet, LLC is a former name of the company Viya, an ISP on the USVI that owns AS14434 [27]. Of the four USVI ASes, it is the second largest graph and the largest of the three graphs of ISPs.

The AS14434 graph shown in Figure 4.5 (along with the other AS graphs) is not completely connected. This is because it is made up of only the AS's nodes and their direct connections. These disconnected portions are likely to be connected through a path that is outside of the AS, but do not have a path connecting them through the AS in the traceroute data.

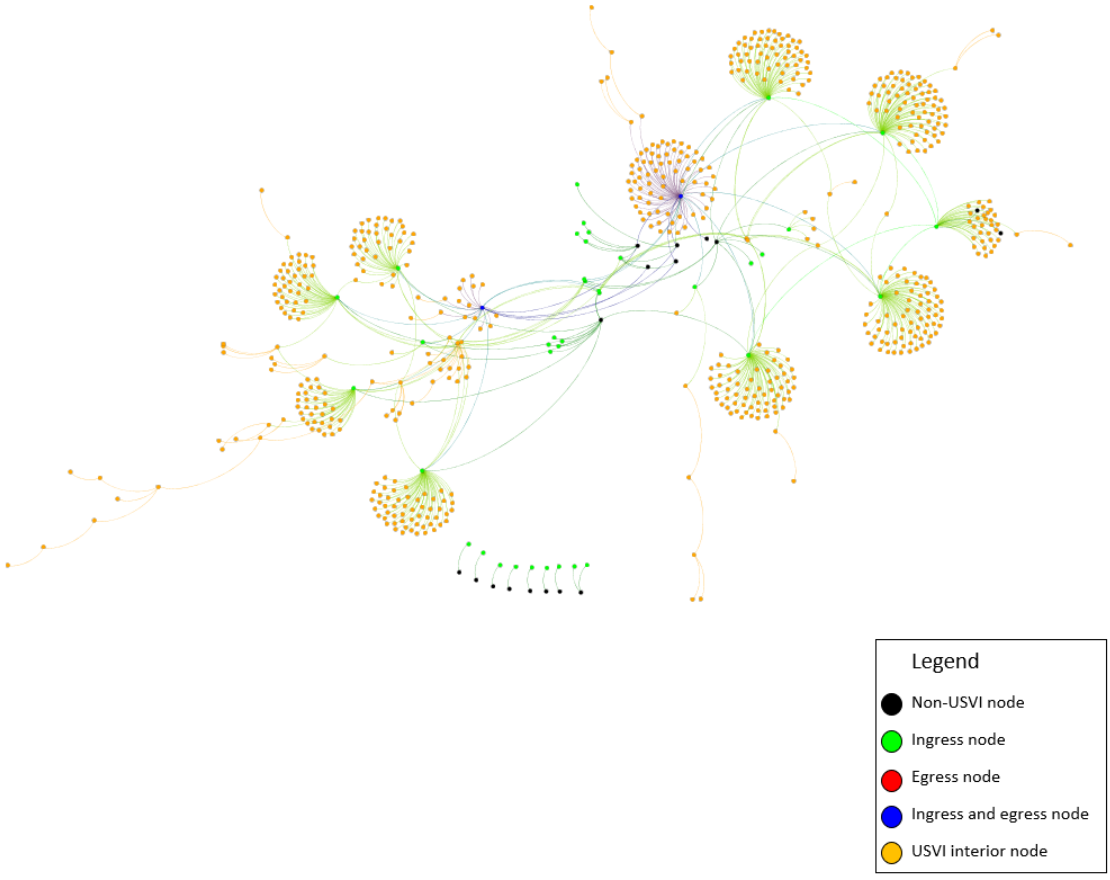

Figure 4.5. The combined graph of AS14434.

Table 4.6 shows the peers and connections of AS14434. It has a total of nine AS peers, none of which are registered in the USVI. All of them, however, are network service providers. In addition to the AS peers, there are also three private IP addresses that are connected to the AS. These are likely customers of Viya's internet service whose private IP addresses

are a part of the traceroute data.

Organization	ASN	**Number of connections**
Cogent Communications, Inc.	AS174	3
Lumen Technologies Inc.	AS209	1
Lumen Technologies Inc.	AS3356	2
Lumen Technologies Inc.	AS3549	1
Aussie Broadband Limited	AS4764	2
Telecom Italia Sparkle S.p.A.	AS6762	1
Hurricane Electric	AS6939	1
Beanfield Technologies	AS21949	2
Liberty Global	AS23520	1
Private IP		3
Total		**17**

Table 4.6. AS14434 connections.

Figure 4.6 shows the degree distribution of the AS14434 graph. It is a typical distribution for a graph, with few nodes with a high degree and many nodes with a low degree. There are over 600 nodes with a degree of one. These are most likely internet service customers with a single IP address.

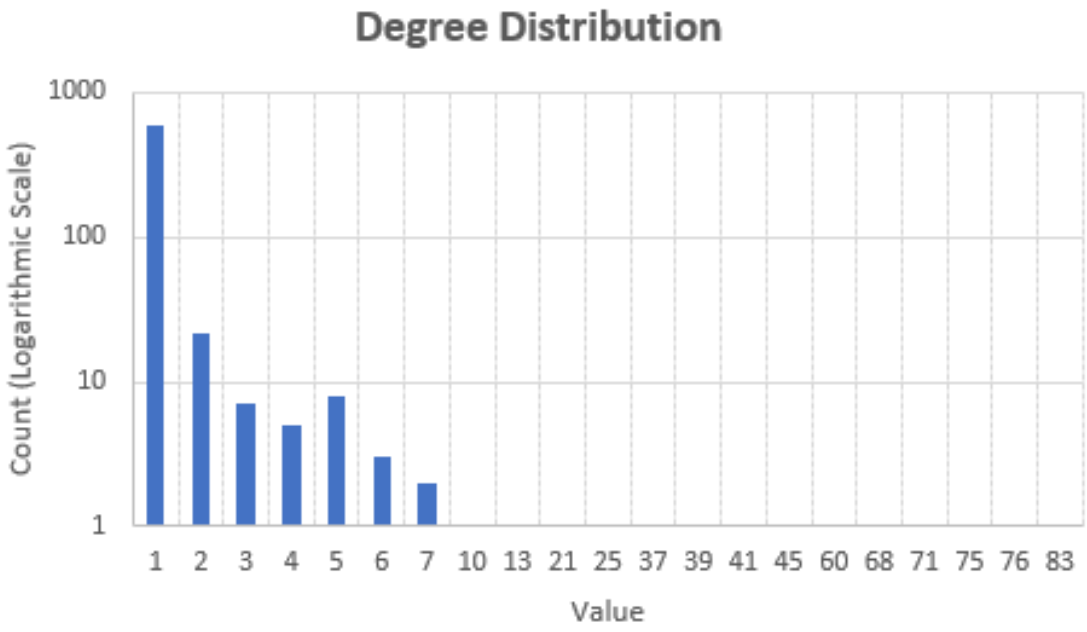

Figure 4.6. The degree distribution of AS14434.

4.3.2 University of the Virgin Islands (AS20243)

The University of the Virgin Islands owns AS20243, shown in Figure 4.7. It is the largest of the four USVI AS graphs with 790 nodes and 773 edges. At first glance, it seems odd that a university would have a larger network than an ISP, but given the limited size and infrastructure of the USVI it is understandable.

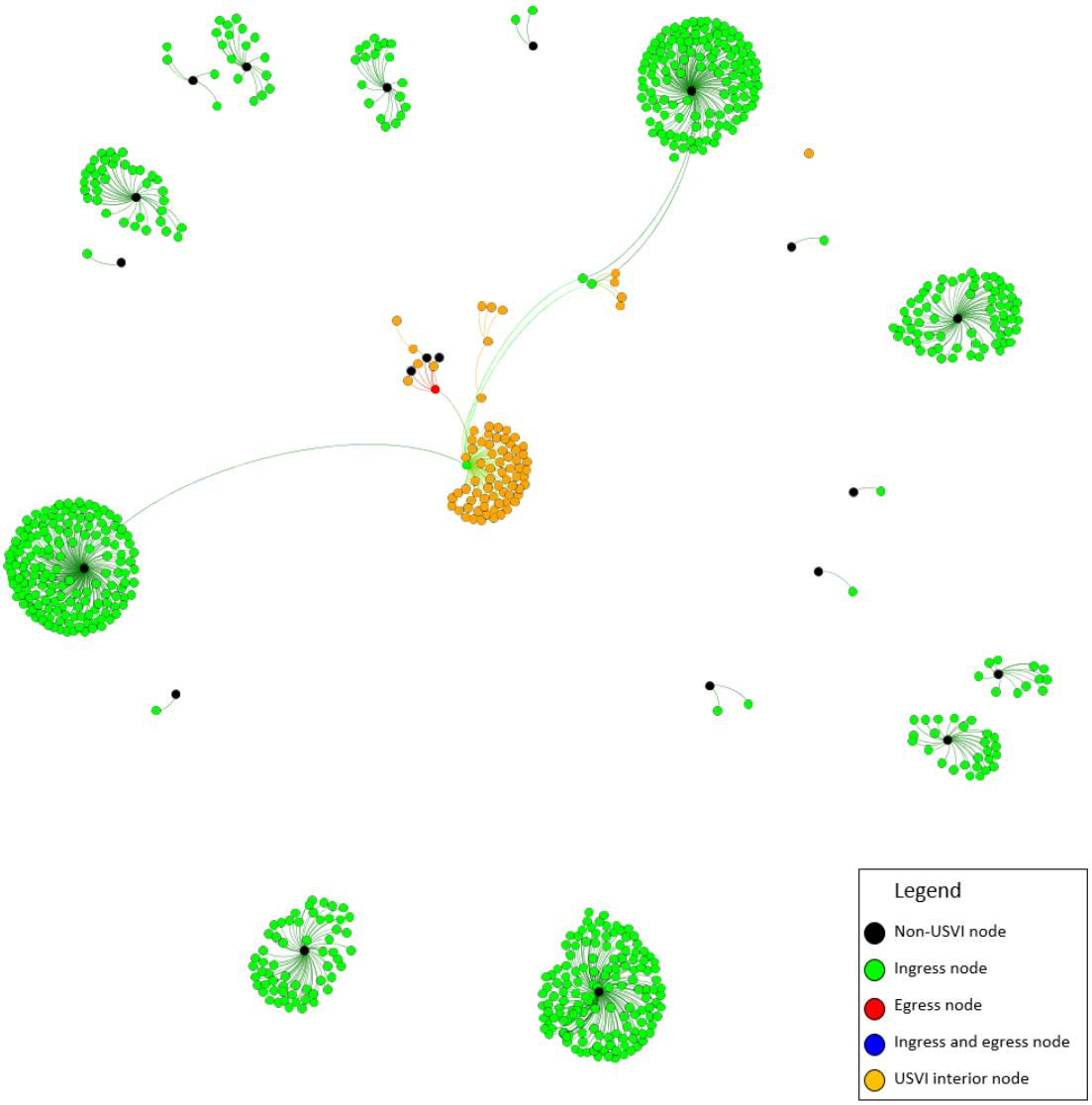

Figure 4.7. The combined graph of AS20243.

This AS is peered with seven other ASes as shown in Table 4.7. It is connected to several network service providers, education and research organizations, and an ISP. AMPATH's AS20080 has the most peering connections. Of these nine connections, three of them are the notable nodes from Section 4.1.2: 190.103.185.163, 190.103.184.13, and 190.103.185.165.

The two IPs that are not in an AS appear to be Internet Exchange Point (IXP)s in Florida and Chicago according to a whois lookup [23].

Organization	ASN	**Number of connections**
CyrusOne	AS62	1
Telstra (International)	AS4637	1
Telecom Italia Sparkle S.p.A.	AS6762	1
Hurricane Electric	AS6939	2
Florida LambdaRail LLC	AS11096	2
UCAID	AS11537	1
AMPATH	AS20080	9
Vodafone Libertel B.V.	AS33915	2
IP not in an AS		2
Total		**21**

Table 4.7. AS20243 connections.

The degree distribution (shown in Figure 4.8) follows a similar trend as before, with a large number of nodes with a low degree and a few nodes with a high degree. Being a university, it is likely that a majority of the nodes with a degree of one are workstations for students and faculty.

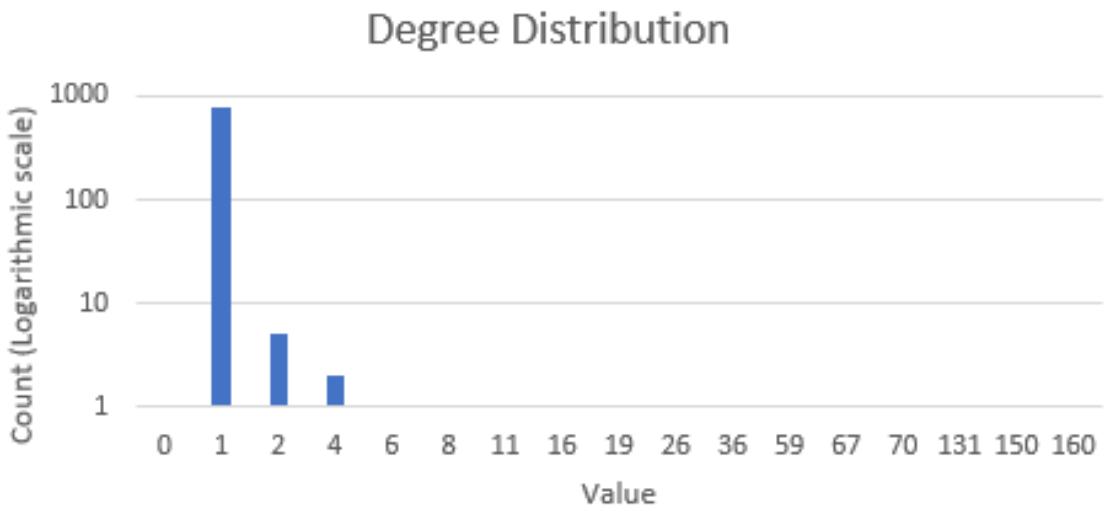

Figure 4.8. The degree distribution of AS20243.

4.3.3 Broadband VI, LLC (AS22581)

Broadband VI, LLC is another ISP on the USVI, and owns AS22581 [28]. It is the smallest USVI AS with 110 nodes and 122 edges. Interestingly, the only two USVI ASes that are connected to each other are AS22581 and AS393275. Figure 4.9 shows the graph of AS22581.

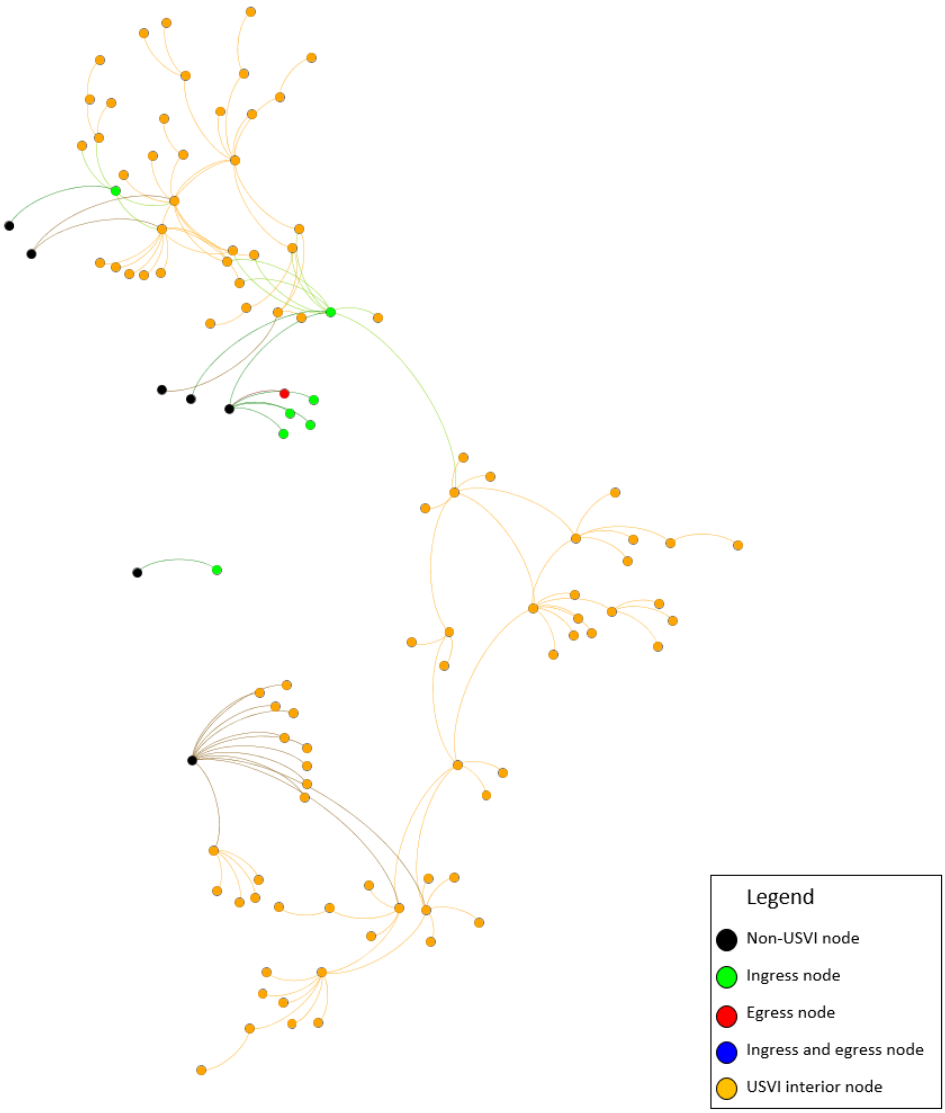

Figure 4.9. The combined graph of AS22581.

Table 4.8 shows AS22581's peers and connections. It is peered with three of the same network service provider ASes as AS14434, so even though they aren't directly peered, they are still close. It is also peered with AS33915 which is a common peer with AS20243.

Organization	ASN	**Number of connections**
Lumen Technologies, Inc.	AS3356	1
Lumen Technologies Inc.	AS3549	1
Aussie Broadband Limited	AS4764	1
Vodafone Lebertel B.V.	AS33915	1
viNGN	AS393275	4
Total		**8**

Table 4.8. AS22581 connections.

Having the smallest AS graph, AS22581's degree distribution is the easiest to read of the four AS graphs, as it has fewer data points. Shown in Figure 4.10, it, again, follows the same pattern of the previous graphs, but has a slightly higher average degree due to there being fewer leaf nodes.

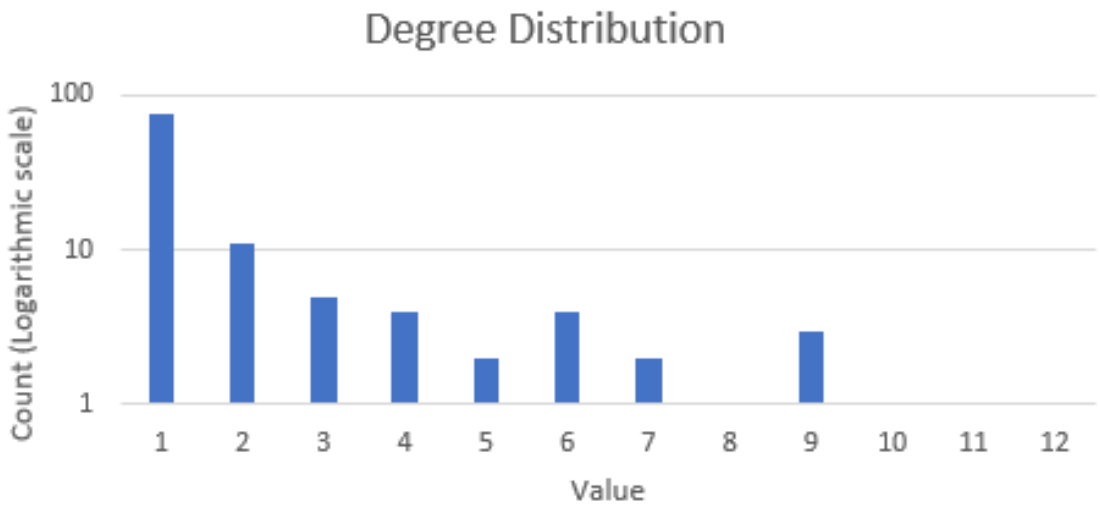

Figure 4.10. The degree distribution of AS22581.

4.3.4 Virgin Islands Next Generation Network (AS393275)

Virgin Islands Next Generation Network (viNGN) is a public broadband bandwidth provider that owns AS393275 [29]. Shown in Figure 4.11, it is the second smallest of the USVI ASes with 127 nodes and 125 edges. As mentioned above, AS393275 is connected to AS22581. According to the viNGN's website, Broadband VI, LLC is an ISP partner with viNGN.

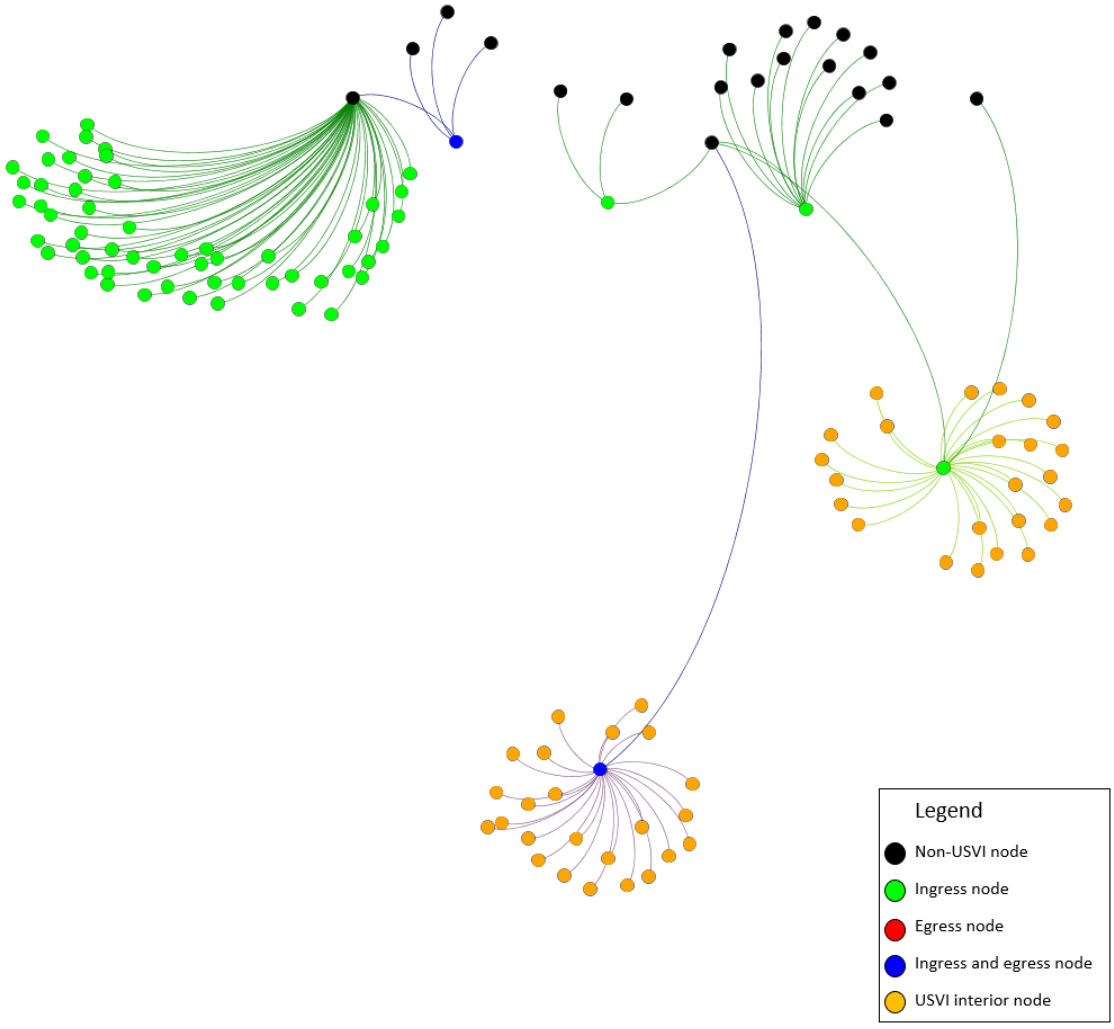

Figure 4.11. The combined graph of AS393275.

This AS has the fewest peers of the USVI ASes, despite having the second fewest nodes of the USVI ASes. It has one unique peer in AS16929, which is registered by the Florida, based ISP, Surge Communications, LLC. Surge's assets and operations were acquired by Broadband VI in February 2019 [30]. Table 4.9 shows the peers and connections of AS393275.

Organization	ASN	**Number of connections**
Lumen Technologies, Inc.	AS3356	2
Surge Communications, LLC	AS16929	3
Broadband VI, LLC	AS22581	15
Total		**20**

Table 4.9. AS393275 connections.

Like the other degree distributions above, AS393275 follows the same pattern of many low-degree nodes and few high-degree nodes as shown in Figure 4.12. However, this graph has only seven nodes with a degree greater than one, with large gaps in value between them. The remaining 120 nodes with a degree of one are likely broadband service customers.

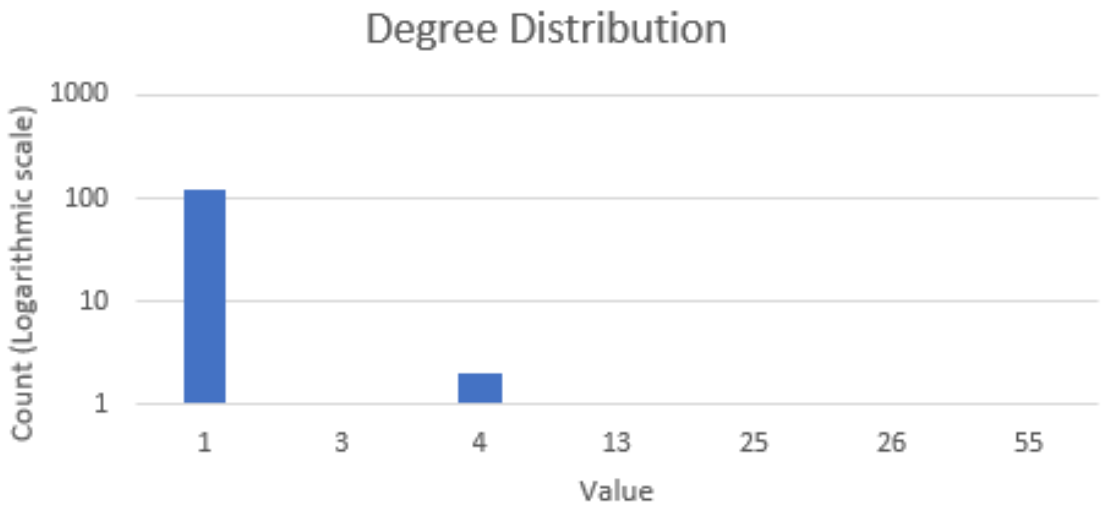

Figure 4.12. The degree distribution of AS393275.

4.4 Vulnerability Analysis

Having these graphs allows us to analyze for potential network vulnerabilities. Network disruptions, such as high traffic and lost nodes, can cause congestion in the network or disconnections. The average hop-counts and node betweenness values of the graphs show that there are some weaknesses in the network topology.

The average hop-counts for the three interface-level graphs are 3.5697, 7.3376, and 6.8282 for the USVI vantage point, non-USVI vantage points, and the combination graphs, respectively, and the average hop count for the router-level graph is 6.4973. For perspective, the average hop-count of a 20 node full-mesh graph is 1, the average hop-count of a 25 node star graph is 1.92, and the average hop-count of a 25 node ring graph is 6.5 [15]. With this perspective, we can see that the interface-level graphs have a relatively high average hop-count, especially the non-USVI vantage point graph and the combination graph, as well as the router-level graph. This means that the failure probability for the non-USVI vantage point graph and the combination graph are higher than those of the full-mesh, star, and ring graphs, while the router-level graph is nearly the same as the ring graph. They are all,

however, comparable to the failure probability of a ring graph.

A commonality between the three interface-level graphs and the router-level graph is that they have few nodes with a high betweenness value and many nodes with a low betweenness value. The router-level graph's betweenness centrality distribution is shown below in Figure 4.13. Similar to the degree distributions of the AS graphs, this is the typical betweenness distribution of a graph.

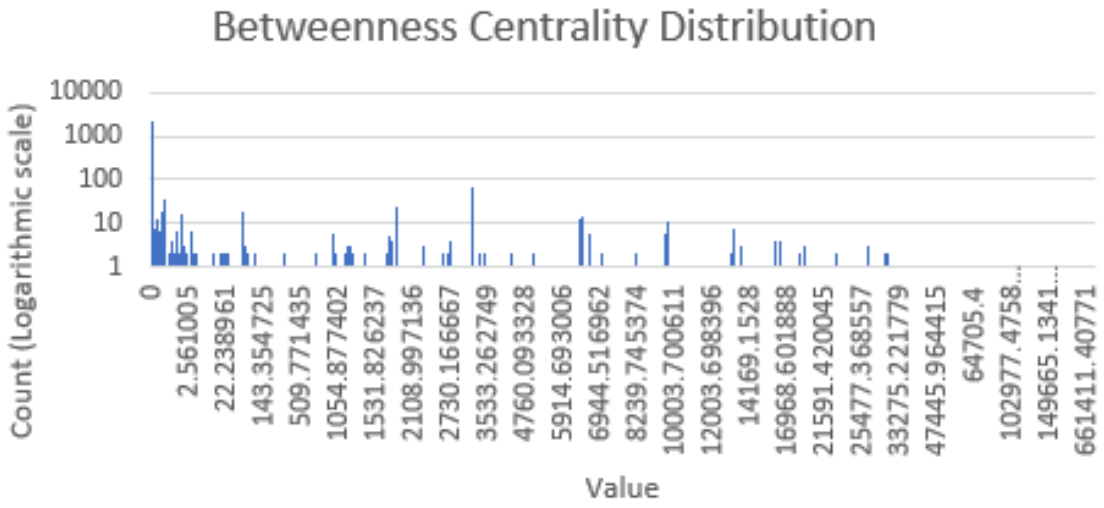

Figure 4.13. The betweenness centrality distribution of the router-level graph.

These graphs have a high average hop-count and nodes with high betweenness values. This indicates that there are several nodes that have a higher importance than the others. Having few nodes with high importance means that the graph is susceptible to congestion during high traffic events and disconnection or congestion of alternate routes if a high importance node goes down. During a high traffic event, the congestion would occur at the high importance nodes, causing slowdowns in the network traffic. If one or more of these nodes were to go down, then there would few, if any, alternate paths which could lead to network traffic slowdowns if the alternate routes are unable to handle the increase in traffic.

The metrics for only the ingress and egress nodes mirror the same patterns as the graphs.

Figure 4.14, below, shows the degree distribution, betweenness distribution, and the clustering coefficient of only the ingress and egress nodes of the router-level graph. Each of these graphs follow the pattern of having a few nodes with a high value and many nodes with a low value. This means that even among only the ingress and egress nodes, there are some nodes with a higher importance than others.

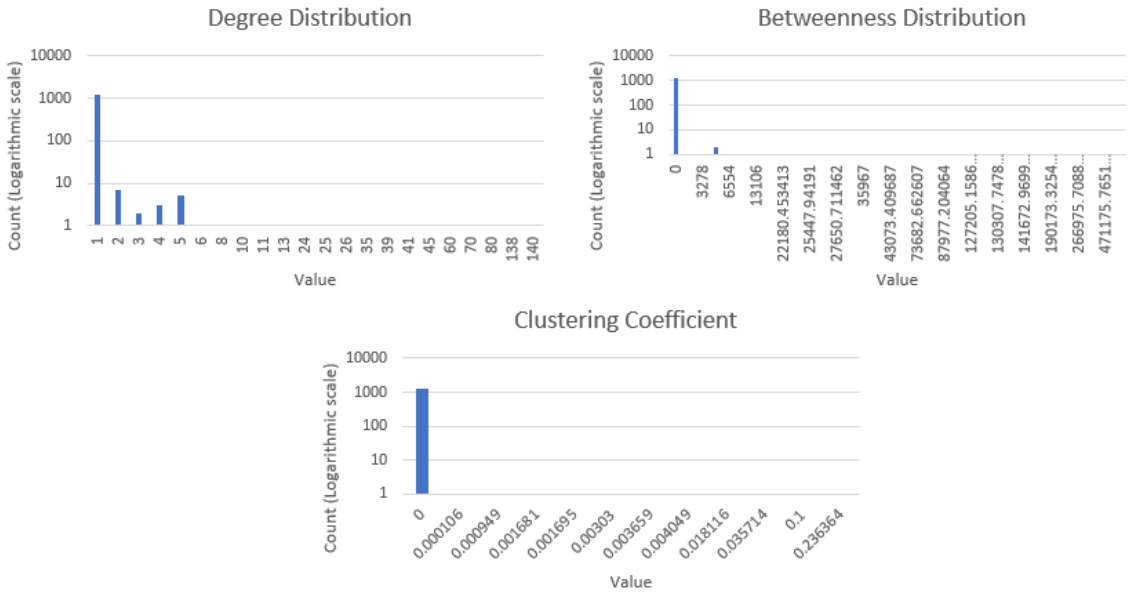

Figure 4.14. The degree distribution, betweenness distribution, and clustering coefficients of the ingress and egress nodes of the router-level graph.

Moeller [13] created a map of the physical network infrastructure and ran simulations with various cuts in the network. He found that cutting links in the network created congestion and disconnected portions of the network. This supports our findings that some nodes have a higher importance than others, and the disconnection of these nodes (or a single link of these nodes) can cause disruptions in the network. Finding ground truth IP addresses for the physical infrastructure would help to solidify these findings.

There are several limitations with the data that we used for this thesis. First, the traceroute data is limited in that it does not include the IP address of every host or router in the USVI.

This is apparent in the AS graphs, as an AS is expected to be a connected network and each acAS graph has disconnected portions of the network. Second, IP address geolocation is very difficult and prone to errors. Knowledge of the IP addresses of key network infrastructure, such as the Network Access Point that connects to Florida and New York, would help to mitigate these errors.

4.5 Recommendation

There are a few steps that can be taken in order to increase the survivability of the network. Adding redundancy to the high importance nodes will help mitigate congestion during high traffic events, as well as provide a backup in the case that one of the nodes goes down. Another way to increase survivability is to add connections between nodes that have a long path between them. This will decrease both the average hop-count of the graphs and the high node betweenness values by providing alternate routes for the network traffic. Decreasing these values means that the importance of individual nodes also decreases, so there will be less of an impact if a node goes down.

Network service providers and organizations, such as the University of the Virgin Islands, can help decrease the average hop-count by adding more AS peering connections to other USVI ASes. Looking at the AS graphs, only AS22581 (Broadband VI, LLC) and AS393275 (viNGN) had a peering connection out of the USVI ASes. If the USVI ASes had more peering relationships with each other, it would provide more alternate paths in the event of a network disruption.

4.6 Summary

The network graphs show that there are weak points in the network topology because they have high average hop-counts and nodes with high betweenness values. If these nodes were to go down or there was some other network disruption, this could cause congestion in network traffic or disconnections. To help prevent this, redundancy can be added to the high importance nodes and links can be added between distant nodes.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5:
Conclusion and Future Work

This chapter discusses the conclusion of this thesis as well as future work that can be done to continue this research.

5.1 Conclusion

This thesis used traceroute data in order to create graphs of the USVI network infrastructure at the interface, router, and AS levels. Analyzing these graphs found that they have relatively high hop-counts and that there are several nodes in each graph of high importance. This means that in the event of a node failure or a high amount of network traffic, the network is susceptible to disconnections or slow downs. In order to increase survivability, we recommend to add redundancy to the high importance nodes and to add connections between nodes with a high path distance between them to reduce the impact of node failures and high traffic events.

5.2 Future Work

There is still much that can be done to continue this work. Physical mapping of the USVI network infrastructure can be compared with the graphs we created. Also, IP address geolocation still has a margin of error. Future improvements to IP address geolocation can help ensure that USVI IP addresses are accurately selected from the traceroute data. In addition to ensuring that the USVI IP addresses are accurate, future work can use IP geolocation to differentiate USVI IP addresses between which island they are on.

We used an alias resolution dataset to create a router-level graph from the interface-level graphs. Future work can use this graph to match and compare the nodes to the physical infrastructure. Matching the nodes to the physical infrastructure will be easier if there are any known ground truth IP addresses of the infrastructure.

As discussed before, IP geolocation can be imprecise. For example, there are several nodes that are identified as non-USVI nodes, but directly connect USVI leaf nodes to the rest of

the network. As these techniques improve, so too will the accuracy of these graphs. We also discussed the descrepancies in the AS registration. Any changes made to the AS registration process may also change the accuracy of future work.

As there aren't any existing commensurable studies, future work can also apply this analysis to other regions. This will help to clarify exactly how much more vulnerable theUSVI network topology is than other, larger network topologies.

List of References

[1] USVI Hurricane Recovery and Resilience Task Force, "Final report," St. Thomas, USVI, Tech. Rep., 2018 [Online], available: https://www.usvihurricanetaskforce.org/.

[2] Viya, "Viya network update," *Facebook*, 2019, retrieved 26 August 2021, https://www.facebook.com/ViyaUSVI/posts/10157700261291385.

[3] A. RAO, "Viya fiber cut linked to AT&T trenching," *The Virgin Islands Daily News*, Oct. 10, 2020 [Online]. Available: http://www.virginislandsdailynews.com/news/viya-fiber-cut-linked-to-at-t-trenching/article_bf4b47a5-ba1a-52f8-ae93-00ce8810cf9c.html

[4] "Viya restores service after fiber cut," *The Virgin Islands Daily News*, Mar. 29, 2021 [Online]. Available: http://www.virginislandsdailynews.com/news/viya-restores-service-after-fiber-cut/article_59004343-db7f-5cda-9f21-0fd45ef1be89.html

[5] OmniSci, "Network Topology," Accessed Aug. 9, 2021 [Online]. Available: https://www.omnisci.com/technical-glossary/network-topology

[6] D. Alderson, W. Willinger, L. Li, and J. Doyle, "An optimization-based approach to modeling internet topology," in *Telecommunications Planning: Innovations in Pricing, Network Design and Management*, S. Raghavan and G. Anandalingam, Eds. Boston, MA: Springer US, 2006, pp. 101–136.

[7] M. W. Murhammer, O. Atakan, S. Bretz, L. R. Pugh, K. Suzuki, and D. H. Wood, *TCP/IP Tutorial and Technical Overview*. Upper Saddle River, NJ Prentice Hall, 1998.

[8] *traceroute*, 2006 [Online]. Available: https://man7.org/linux/man-pages/man8/traceroute.8.html

[9] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira, "Avoiding traceroute anomalies with Paris traceroute," in *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement* (IMC '06). New York, NY, USA: Association for Computing Machinery, 2006, p. 153–158. Available: https://doi.org/10.1145/1177080.1177100

[10] R. Motamedi, R. Rejaie, and W. Willinger, "A survey of techniques for internet topology discovery," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1044–1065, 2014 [Online]. doi: https://doi.org/10.1109/COMST.2014.2376520.

[11] K. Keys, Y. Hyun, M. Luckie, and K. Claffy, "Internet-scale IPv4 alias resolution with MIDAR," *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 383–399, 2012.

[12] M. Luckie, R. Beverly, W. Brinkmeyer, and k. claffy, "Speedtrap: Internet-Scale IPv6 Alias Resolution," in *ACM Internet Measurement Conference (IMC)*, 10 2013, pp. 119–126.

[13] B. T. Moeller, "Synthetic network generation and vulnerability analysis of internet infrastructure systems in the U.S. Virgin Islands," M.S. thesis, Dept. of Operations Research, NPS, Monterey, CA, USA, 2020 [Online]. Available: https://calhoun.nps.edu/handle/10945/66112

[14] J. P. G. Sterbenz, D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, and P. Smith, "Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines," *Computer Networks: Special Issue on Resilient and Survivable Networks (COMNET)*, vol. 54, no. 8, pp. 1245–1265, June 2010. Available: https://cdn.jprohrer.org/documents/publications/Sterbenz-Hutchison-Cetinkaya-Jabbar-Rohrer-Scholler-Smith-2010.pdf

[15] J. P. Rohrer, A. Jabbar, and J. P. Sterbenz, "Path diversification for future internet end-to-end resilience and survivability," *Telecommunication Systems*, vol. 56, no. 1, pp. 49–67, May 2014. Available: https://cdn.jprohrer.org/documents/publications/Rohrer-Jabbar-Sterbenz-2012.pdf

[16] MaxMind, [Online], GeoIP2 Country Database - Auguest 8, 2021. Available: https://www.maxmind.com/en/geoip2-country-database

[17] Team Cymru, "IP to ASN mapping," Accessed Jul. 14, 2021 [Online]. Available: https://team-cymru.com/community-services/ip-asn-mapping/

[18] RADb, "RADb query help," Accessed Jul. 14, 2021 [Online]. Available: https://www.radb.net/query/help

[19] M. Luckie, "Scamper: A scalable and extensible packet prober for active measurement of the Internet," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement* (IMC '10). New York, NY, USA: ACM, 2010, pp. 239–245. Available: http://doi.acm.org/10.1145/1879141.1879171

[20] Gephi, "GDF format," [Online], Accessed July 20, 2020. Available: https://gephi.org/users/supported-graph-formats/gdf-format/

[21] J. Martineau, "Mapping mobile ipv6 providers," M.S. thesis, Dept. of Computer Science, NPS, Monterey, CA, USA, 2020 [Online]. Available: https://calhoun.nps.edu/handle/10945/66104

[22] RouteViews, [Online], MRT format RIBs and UPDATEs - September 3, 2020. Available: http://archive.routeviews.org/bgpdata/

[23] “Ip whois lookup,” Dec 2021. Available: https://www.whatismyip.com/ip-whois-lookup/

[24] “Network tools: Dns,ip,email.” Available: https://mxtoolbox.com/SuperTool.aspx

[25] “Macroscopic internet topology data kit (itdk),” Oct 2020. Available: https://www.caida.org/catalog/datasets/internet-topology-data-kit/release-2020-01/

[26] “Registration data lookup tool.” Available: https://lookup.icann.org/lookup

[27] “Viya: High speed internet,” Jan 2022. Available: https://broadbandnow.com/Viya

[28] “Fast, reliable internet service in the virgin islands.” Available: https://broadband.vi/

[29] “Connecting the vi to the world.” Available: https://www.vingn.com/

[30] “Surge with broadband vi.” Available: https://broadband.vi/surge/

THIS PAGE INTENTIONALLY LEFT BLANK

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California