National College of Ireland

Title

MSc Research Project
Programme Name

Jasmine Killedar
Student ID: x19204663

School of Computing
National College of Ireland

Supervisor: Liam McCabe

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: …. ……Jasmine Killedar …

Student ID: ……x19204663……..……

Programme: ……Cybersecurity…… ……

Year: ……2021-2022…… ……..

Module: …… ……..……

Supervisor: ……Liam McCabe…… ……..……

Submission Due Date: …… ……..……

Project Title: ……Detecting border gateway protocol connectivity monitoring using machine learning and security configurations …… …..……

Word Count: …… **Page Count**…….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……Jasmine Killedar …… ……

Date: …16-12-2021…… ……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Detecting Border gateway protocol (BGP) connectivity monitoring using machine learning and using security configurations

Jasmine Killedar
x19204663

Abstract

Border Gateway Protocol (BGP) has been the inter domain routing protocol which was first described in 1989, the current version of BGP is version 4 which was published in 2006. Because of its destination-based routing, it is unable to select a precise end-to-end AS-level route. Border gateway protocol (BGP) still has on going issues related to security such as IP prefix hijacking attacks , Ddos attacks , man-in-the-middle attack , sniffing , routing to end points in malicious networks , creation of route instabilities and this is the only protocol that is connected to internet.

This article proposes a technique for monitoring the connectivity of suspected autonomous systems obtained via a software tracing IP prefix hijacking signature, which uses anomaly detection method, and I have also configured the devices using security commands which prevents any malicious attacker from hijacking the device which uses a comprehensive cross validation test to examine the approach's correctness.

Random forest method is used to deal hijacking of IP prefix. The primary characteristics are derived from the autonomous system path properties of autonomous systems that are possibly suspicious. The characteristics are a combination of the behavioural aspects of router connection

BGP in turn uses transmission control protocol (TCP) to establish the connection, there are many security flaws in the TCP standard such as it relies on IP source address for authentication and minimal or no authentication in network control mechanism example routing protocol , congestion control, flow control , ICMP messages, as well as additional flaws in some of its implementations. These flaws might allow an intruder to “uattack” TCP-based systems, allowing him or her to “hijack” a TCP connection or deny legitimate users service.

Table of Contents

1 Introduction 2
2 Related Work 4
2.1 Evaluating Machine Learning Approaches 4
2.2 Features based on GNS3 lab 5
3 Research Methodology 5
3.1 Business understanding 6
3.2 Data preparation 6
3.3 Modelling 7
3.3.1 Choosing anomaly detection 7
4 Design Specification 8
4.1 Project Requirement 9
4.2 Solution Architecture 10
4.3 Train and testing the model 10
5 Implementation 11
6 Evaluation 11
6.1 Experiment / Case Study 1 11
6.2 Experiment / Case Study 2 14
6.3 Experiment / Case Study 3 16
6.4 Experiment / Case Study 4 16
6.5 Below diagram mentions the workflow in the study of random forest 17
6.5 Discussion 17
7 Conclusion and Future Work 18
8 References 18

1 Introduction

How can we avoid BGP vulnerabilities in a network, by detecting BGP connectivity monitoring using machine learning and by securing the network using TTL-security method .

The flaw in BGP related to security has caused significant network instability and could be exploited by black-hole traffic attackers, spammers, and other malicious attackers such as man-in-the-middle attacks and distributed denial-of-service attacks. Border gateway protocol (BGP) is the only routing protocol that is connected to the internet and the only protocol that is connected to autonomous systems. There have been security concerns as a result of many attacks. Hijackers may also utilise Border gateway protocol (BGP) traffic redirection for hijacking purposes. Many safe procedures were implemented to prevent fraud or hijacking of cryptocurrency transactions in the Google hijacking incident in 2018. A thorough or broad deployment is unlikely since each organisation instals its own solution individually. [1]

Many research were carried out based on the various techniques and an approach to make browsing or connectivity more reliable but the research was carried on and focused solely on border gateway protocol , however this research paper tries to do the same but with new technology like machine learning approach , however it was already used earlier but this approach focuses on securing network connectivity using TTL-secuirty mechanism , I have come up with the idea to secure border gateway protocol using as machine learning approach . Cyber assaults are uninvited acts taken by network fraudsters and intruders with the goal of destroying, stealing, or manipulating critical data. Anomaly is a term used to describe

unexpected behaviour that differs from typical network activity. Individuals and businesses alike suffer significant economic and societal consequences when network data is breached.

BGP is the sole way to connect to the internet; it has been used for decades and will continue to be used in the future; yet, it has a number of security problems, including a susceptibility to prefix attacks. This vulnerability causes significant network instability and might be used by black-hole traffic attackers, spammers, and other malicious attackers. Man-in-the-middle and distributed denial-of-service attacks are two types of assaults. Hijackers may also utilise BGP traffic redirection for hijacking purposes. [1]

BGP hijacking occurs when an attacker maliciously redirects Internet traffic. Attackers do this by claiming fake ownership of IP prefixes, or groups of IP addresses, that they do not actually own, control, or route to. When a domain name server expires, domain names are reregistered and stored in regional internet registries. Currently, 40% of IPv4 address space is assigned but not publicly stated; this void is the perfect environment for malicious BGP hijacking. In a study, Goldberg claimed that the main reason for the long time it takes for BGP to be secured is because, aside from deployment challenges, the infrastructure is security deficient. There is no central authority because each organisation deploys separately. As a result, a large-scale or broad deployment is unlikely to occur. [1]

The Internet is made up of thousands of administrative domains known as Autonomous Systems (ASes), each of which uses the Border Gateway Protocol to share IP address space reachability. (BGP). Given the widespread usage of BGP today and the unpredictability of ASes' transmitted information, any misconfiguration or breakdown of the protocol would be disastrous. The Internet's stability may be jeopardised by protocol. Regardless of the motivation for these oddities, whether they are intentional or not, are malevolent, such as malware or targeted assaults, or are caused by misconfiguration There has been an increasing interest in identifying and repairing network or connection problems monitoring BGP traffic to mitigate BGP abnormalities, without RPKI, for example, is a large-scale deployment option. [7]

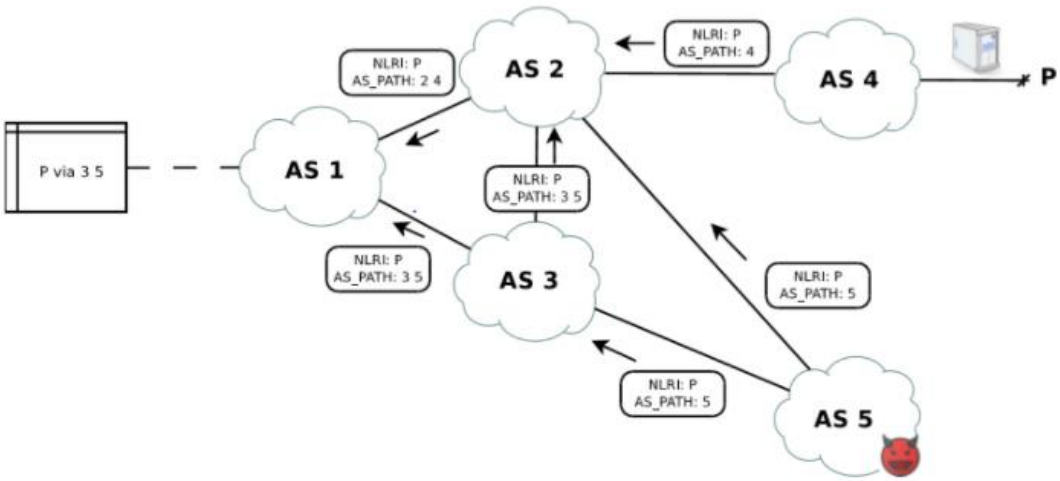


Figure 1: BGP Attack into autonomous system (AS)

2 Related Work

The ASes connections are inferred from studies that maximise the number of valley-free routes on the internet, if all routes that violate the valley-free criteria are leaks. This assumption may not hold true for some of the more intricate AS connections, such as siblings and mutual transit.

Attacks in BGP has been around since the time it was first described in 1989 in RFC 1105, and it has been in use on the internet since 1994 , The idea behind BGP hijacking is to find an ISP that isn't screening ads (intentionally or unintentionally) or that has an internal or ISP-to-ISP BGP session that is vulnerable to a man-in-the-middle attack. Once discovered, an attacker can advertise whatever prefix they choose, diverting part or all traffic away from the true source and towards the attacker. This can be done to either overwhelm the infiltrated ISP or to launch a DoS or impersonation attack against the company whose prefix is being broadcast. It's not unusual for an attacker to create major disruptions, up to and including full connection loss. Ever since various research have been going on to mitigate it , there were various methods used to secure BGP. It is the sole way to connect to the internet; it has been used for decades and will continue to be used in the future; yet, it has a number of security problems such as Ddos attack , man-in-the-middle attack , including a susceptibility to prefix attacks. This vulnerability causes significant network instability and might be used by black-hole traffic attackers, spammers, and other malicious attackers. Man-in-the-middle and distributed denial-of-service attacks are two types of assaults. Hijackers may also utilise BGP traffic redirection for hijacking purposes. [2]

Other options, advocate altering the control plane by adding flags or colouring systems to distinguish between ASes based on their business interactions. ASes can use this information to accept or deny updates based on the kind of incoming connection. To help protect the colours, these approaches require additional signature verification procedures; otherwise, they may be vulnerable to exploitation. It appears that improvements to the control plane will be difficult to implement in the near future in the global internet community.

Other methods include studying route announcements through the worldwide internet via vantage points that illustrate how route modifications may be compared and tracked. These systems are capable of detecting route leaks, but they need a significant amount of computing power and resources. Because of the reliance on information from vantage points, blind corners may occur, and certain pathways may not be explored.

The route leaks problem that meets the five requirements below after examining the problem : 1) be able to remediate the leak event quickly enough before the leaked prefixes spread throughout the network. 2) Keep the computational overhead to a minimum so that it may be used with current systems. 3) Do not rely on third parties to speed up and improve the decision-making process. 4) Do not rely on the ASes' business contacts because they are private; and, finally, 5) Aim for high accuracy and minimal false alarms. [6]

2.1 Evaluating Machine Learning Approaches

Methods to prevent vulnerability in BGP using machine learning, various methods have been carried out for the same. Machine learning techniques for BGP anomaly detection include support vector machine (SVM) , long short term memory (LSTM) , BGP anomaly prediction based on ensemble learning , random forest , naïve bayes classifier (NB) , decision trees (DT) , multi layer perceptron (MLP) [5]

2.1 Analysis various methods

There are different ways to conduct vulnerability research. Traditional ways include Blacklist-, Heuristic-, Visual similarity etc. and machine learning use different algorithms like SVM, Random Forest, Naïve Bayes, etc. approaches to detect phishing. It is very important to identify the best approach to start detection Several studies have looked on securing BGP using historical and statistically based behavioural models. In previous studies, the top five supervised classifiers were utilised, and the features of the dataset will be used in this work to evaluate the detection technique. The detection approach is able to detect hijacks with 81 percent accuracy when using different learning algorithms, such as Random Forest and J48 classifiers.

Zhang et al. emphasised the importance of signature-based and anomaly-based intrusion detection in today's intrusion detection, as well as their inherent drawbacks – ambiguity for signature-based approaches and inability to recognise fresh attacks for anomaly-based analysis. The connection model is a new method for tracing the behaviour of opportunistic networks. A paradigm shift from mobility to connection models, according to Kathiravelu, is essential. To validate the detection technique, I have used numerous supervised machine learning classifiers based on a comprehensive cross-validation test strategy. [1]

2.2 Features based on GNS3 lab

Features Based on lab		
Routers	c7200-adventerprisek9-mz.152-4.M7.image	image description
Switch	CiscoIOSvL215.2(20170321:233949)-1	
NetworkAutomation-1	1024*768	VNC console resolution

Table 1 : Features based on GNS3 lab

3 Research Methodology

Research methodology used in this project is CRISP-DM methodology . It stands for Cross industry standard process for data mining .

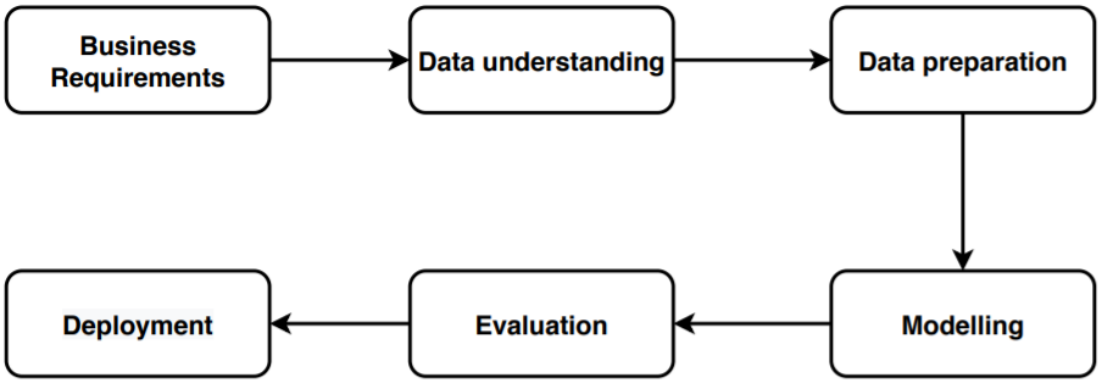


Figure 2: CRISP – DM

3.1 Business understanding

When BGP was invented in 1989 , security was a major concern for protocols used on the internet in the late 1980s and early 1990s, when BGP was invented. Since then, I have learned that BGP may have two security issues: you might be talking to the incorrect device, or the appropriate device could be saying the wrong thing. I choose three configuration features MD5 , IPSec and GTSM for the security of BGP , I configure these using gns3 software and also implement it using python coding , along with python based machine learning technique used to detect the prefixes , I have used anomaly detection method.

Someone rerouting the cable towards your BGP peer to a different router, then having that router imitate your BGP neighbour, may be as dramatic as talking BGP to the wrong device. TCP reset assaults are a more everyday threat. TCP sessions can be active indefinitely, or they can be inactive over extended periods of time. So it's not out of the question that if systems A and B are in a TCP connection and system A reboots, A will receive TCP messages from B. As a result, A's TCP session is no longer operational. As a result, it sends B a "TCP reset" message, causing B to terminate the session.

A third party can send a faked TCP reset to A that appears to come from B since TCP has no strong security protections. A will then terminate its TCP session, as well as the BGP session that runs over it, halting traffic flow between A and B until a new session can be established.

The main objectives to enhance the model and application included:
 Improve vulnerability detection by including anomaly detection using machine learning and including the security features while configuring the devices.
 Obtain features_ importance to identify features that contribute to the unidentified IP addresses. [7]

3.2 Data preparation

In this phase, the data has been prepared using GNS3 software in which the IP addresses has been pre-configured , the configuration has been made as such it can directly be used for implementation , I have used python libraries to implement the configuration so as to

automate the configuration , I have implemented machine learning method as well for the anomaly detection using random forest which gives 82% accuracy on test data .

Any anomaly detection algorithm, whether supervised or unsupervised, must be assessed to determine its effectiveness. I can't use accuracy as an assessment metric since the number of occurrences of anomalies is so small compared to normal data points. If a model predicts everything as non-anomalous, the accuracy will be more than 99.9%, and I wouldn't have caught any abnormalities. The dataset used for learning is supposed to include all non-anomalous training instances, which is one of the most critical assumptions for an unsupervised anomaly detection system (or very very small fraction of anomalous examples).

3.3 Modelling

I have concentrated on using various modelling approaches on the provided variables to develop models that may potentially produce the required output in this phase. Detecting anomalies, cluster based approach for anomaly detection , anomaly detection as a classification problem , anomaly detection in time series data , deep learning based method , The IP addresses are typically easy to locate, and port 179 is one of them (the BGP port number). Two the attacker transmits packets with all conceivable leftover port and sequence numbers in the hopes of hitting the proper combination within a minute or so. RFC 2385, issued in 1998, describes the BGP TCP MD5 password mechanism. [8]

3.3.1 Choosing anomaly detection

Detecting anomaly by just seeing

The development of anomalies in data is inextricably linked to the generation of data points. The method is good enough to proceed in simulating this. Let's look at some fundamental statistics in the form of a boxplot (such minimum and maximum values, first quartile values, and so on). Boxplot, since I have obtained all of the following information in one graphic location [9]

Clustering based approach for anomaly detection

In this method, I have begun by grouping items that are similar in nature. Distance measurement functions such as Euclidean distance, Manhattan distance, and others are used to calculate this similarity mathematically. When deciding between several distance measuring functions, Euclidean distance is a common option.

Anomaly detection in time series data

Any data that is related to time is referred to as a time series (daily, hourly, monthly etc). For example, daily income at a store is a time series data at the day level. Many application cases, such as demand estimates and sales forecasting, are common time series forecasting problems that may be handled using algorithms like as SARIMA, LSTM, and Holtwinters, among others. By projecting future demands using present data, time series forecasting assists us in preparing for them. When I get the prediction, I can compare it to the actuals to see if there are any abnormalities.

Deep learning method for anomaly detection

There are advanced Neural Network designs (such as Autoencoders) that may be used to successfully simulate an anomaly detection situation.

Decision tree classifier

Decision Trees are the most easily understood classifiers because they are visually presented in the same way that the human brain works, which includes a feature selection process by nature. However, it may be prone to overfitting and sensitive to outliers. The experiment is carried out using the "Breiman" [10] algorithm with the following parameters: Minimum split=2, Maximum depth=5, and the "Entropy" based splitting scheme.

Random forest classifier

Random forest classifier is regarded as a more sophisticated variant of the Decision Tree classifier. It overcomes the overfitting problem in Decision Trees by merging the ideas of "randomness in dividing features selection" and "forest, which refers to employing more than one decision tree." The final predictor is decided by a majority vote. The experiment is carried out using Breiman's random forest method [11], with a forest of 500 random trees, a minimum split of 2, a maximum depth of 5, and a "Entropy" based splitting strategy for each tree.

SVM Classifier (SVM)

Another form of classifier is the Support Vector Machines classifier, which aims to divide data into subsets in a way that optimises the margins around the dividing line or function. There are a number of splitting kernel functions available. Functions like "polynomial," "radial," and "sigmoid" are examples. The kernel function employed is a degree 2 polynomial with Gamma =1/15 and Cost=1. [9]

3.4.2 Improving the model

The network topology has been configured with MD5 hashing feature , since it is not much in use as it doesn’t provide complete security I configure using IPsec and generalized TTL security mechanism (GTSM) , I have also used anomaly detection using machine learning approach.

4 Design Specification

This document contains the application's technical requirements, URL feature extraction, model training and testing, and assessment of essential characteristics.

4.1 Project Requirement

Software Configuration: Google Browser V87: This browser is used in Developer Mode to add Chrome-Extensions. This Extension will serve as a user interface for the Anti-Phishing programme.
Python2.7 is used in the development of this project . As a result, Python2.7 is required to run the application successfully.
Visual Studio Code: This IDE is used to test the response by running the.py files locally.
Gns3 : This software is used to run the network topology

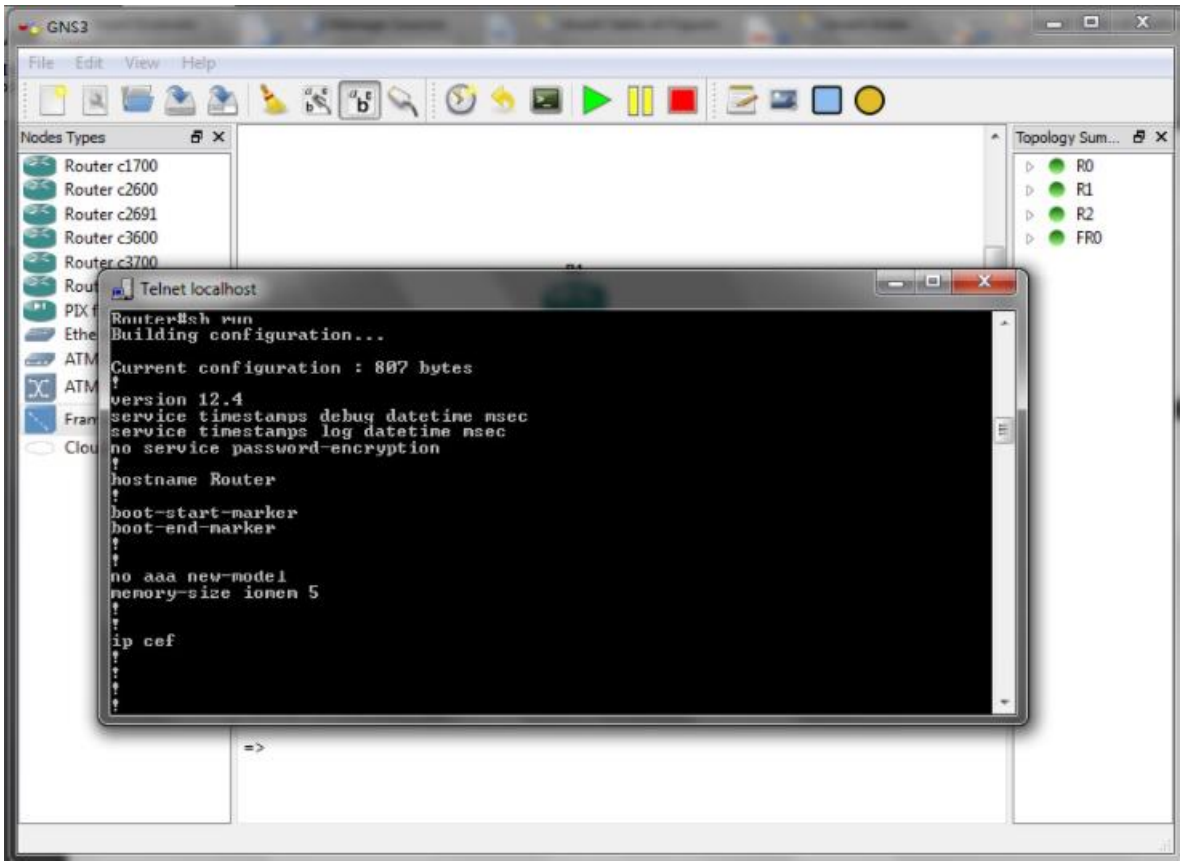


Figure 3: GNS software

4.2 Solution Architecture

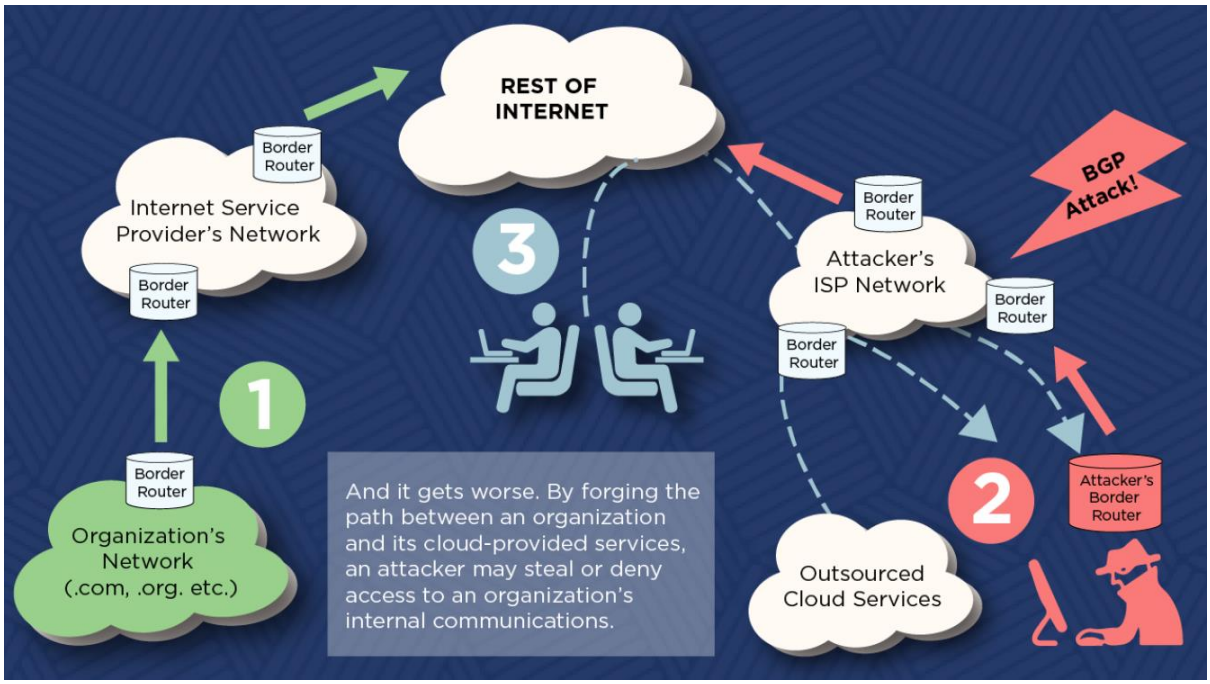
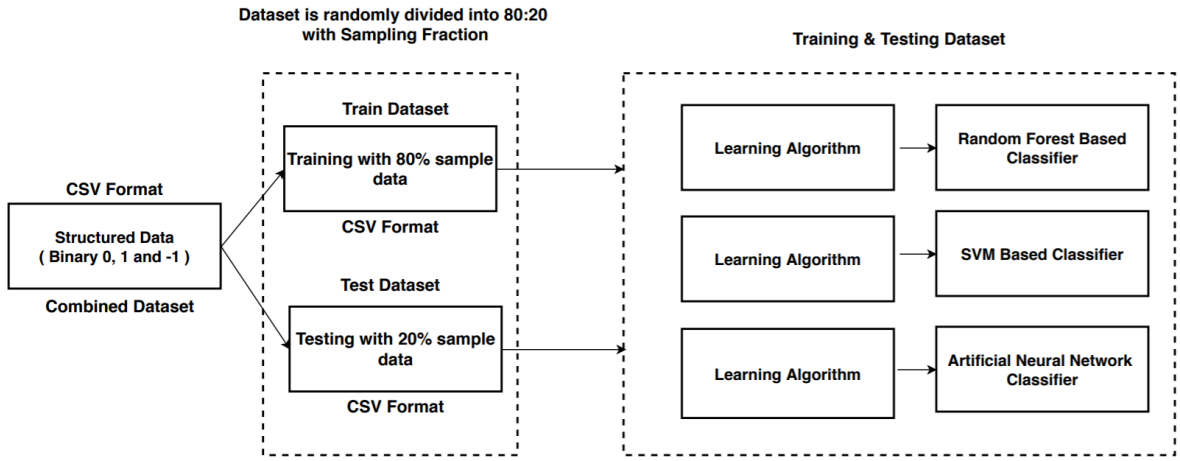


Figure 4 : Solution architecture

4.3 Train and testing the model

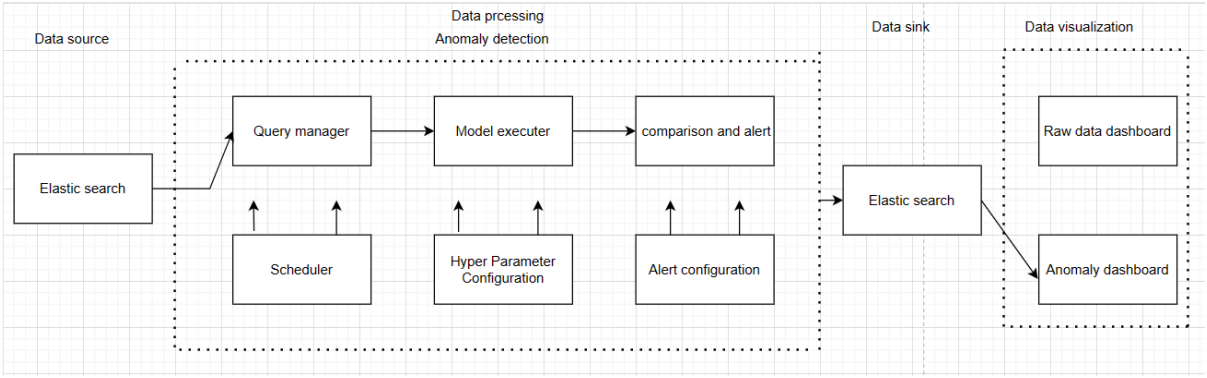
The combined structure dataset feature extraction and the result (class labels) marked as 1 for all benign URLs is prepared for model training and testing. By splitting the combined dataset into an 80:20 ratio, the train and test dataset files are formed. This is accomplished through the use of the sampling fraction sample(frac=1)[12]. Pandas.DataFrame.iloc [13] is used to choose columns from the train and test CSV files in the Dataframe. The graphic in Fig. 4 shows how the training and testing files were used to test each model.



Model training and testing data

5 Implementation

This section focuses on the finished product. The finished product's Application Journey and Process workflow are shown below. The project uses GNS3 for the implementation of security features , also it uses random forest anomaly detection which gives a model accuracy of 82% . The distinction between Paramiko and Netmiko is that the Netmiko module is used to connect to switches more easily using ConnectHandler, which also uses SSH in the backend. In addition, while utilising Netmiko, I must indicate the device type for which the script will be used.



Anomaly detection

6 Evaluation

During this phase, the model is built and ready for review. Before sending the model t deployment, the model is evaluated to review its performance. For this project security update and random forest classifier was evaluated using the below ways.

6.1 Experiment / Case Study 1

For my findings I have conducted the experiment in the lab using GNS3 , also I shall be using python code for configuring BGP , I would be using python code for running machine learning using anomaly based which would be running on visual code this is used for the detection of test data .

Figure 7 illustrates the simulator that I have used to build a network topology , left side of the column depicts routers and switches that are used in configuration.

Using the gns3 simulator

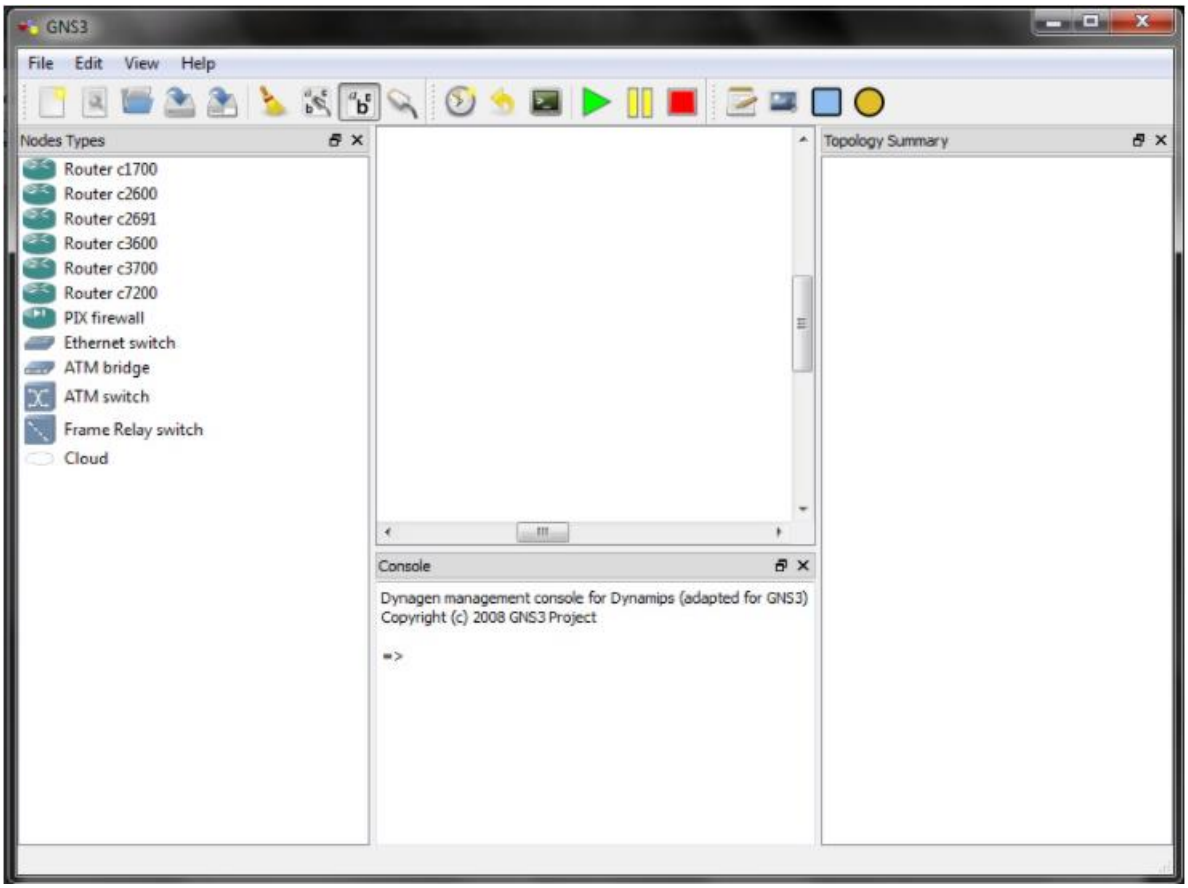
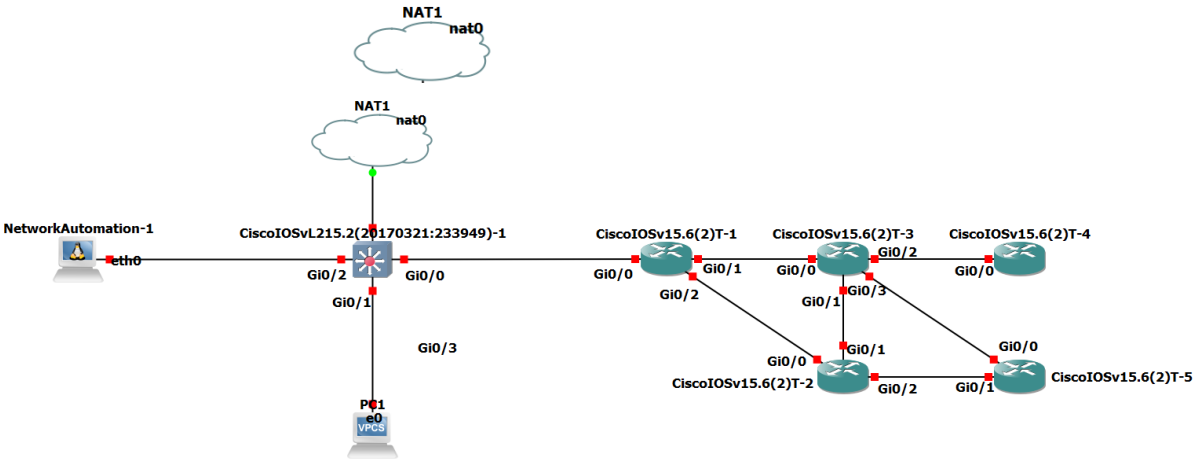


Figure 5 : GNS3 simulator

Four routers have been taken, each router in different autonomous system and working independently, each router has been configured with border gateway the fifth router(R5) is the attacker router with the attacker trying to send packets to router R2 and R3.

This case study is done using four routers running BGP individually I have used python scripting to configure the devices however rest of the configuration has been directly implemented into it.

The figure shows network automation icon where I shall be running the script on, then there is a NAT (network address translation) gateway which is connected to the multilayer switch , there is a PC (system/personal computer) connected which is used to check the configuration is something goes wrong , rest of the routers are connected in the topology diagram . Gi0/1 , Gi0/2 …. Identifies the interface that it is connected to



A well-defined network diagram for better clarity has been shown in the below figure.
R1, R2, R3, R4: are the routers
S1/0, s1/1, s1/2 ……. Are the serial interfaces that are used to connect different routers together and these interfaces are used for the flow of packets.

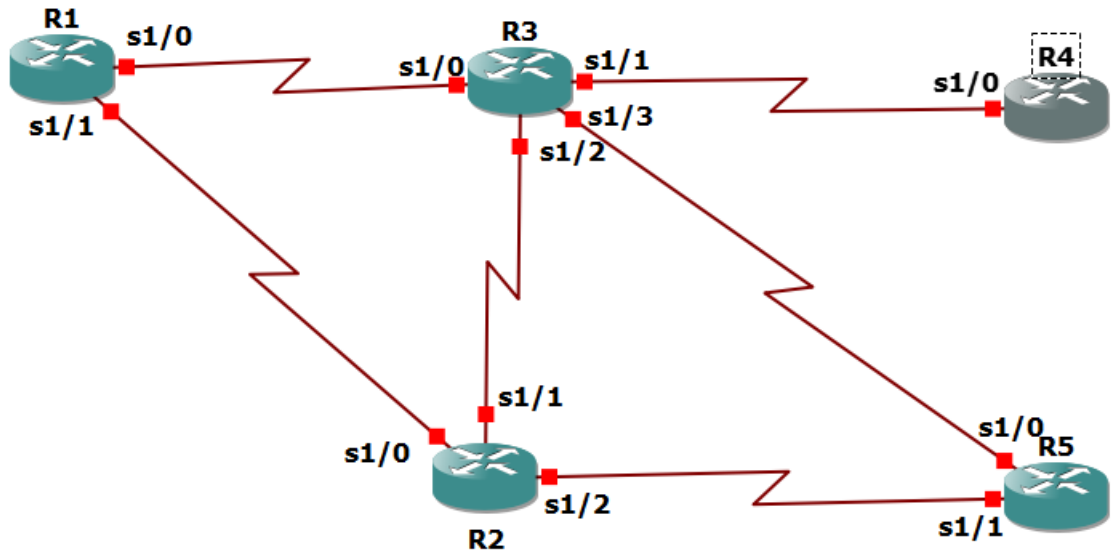
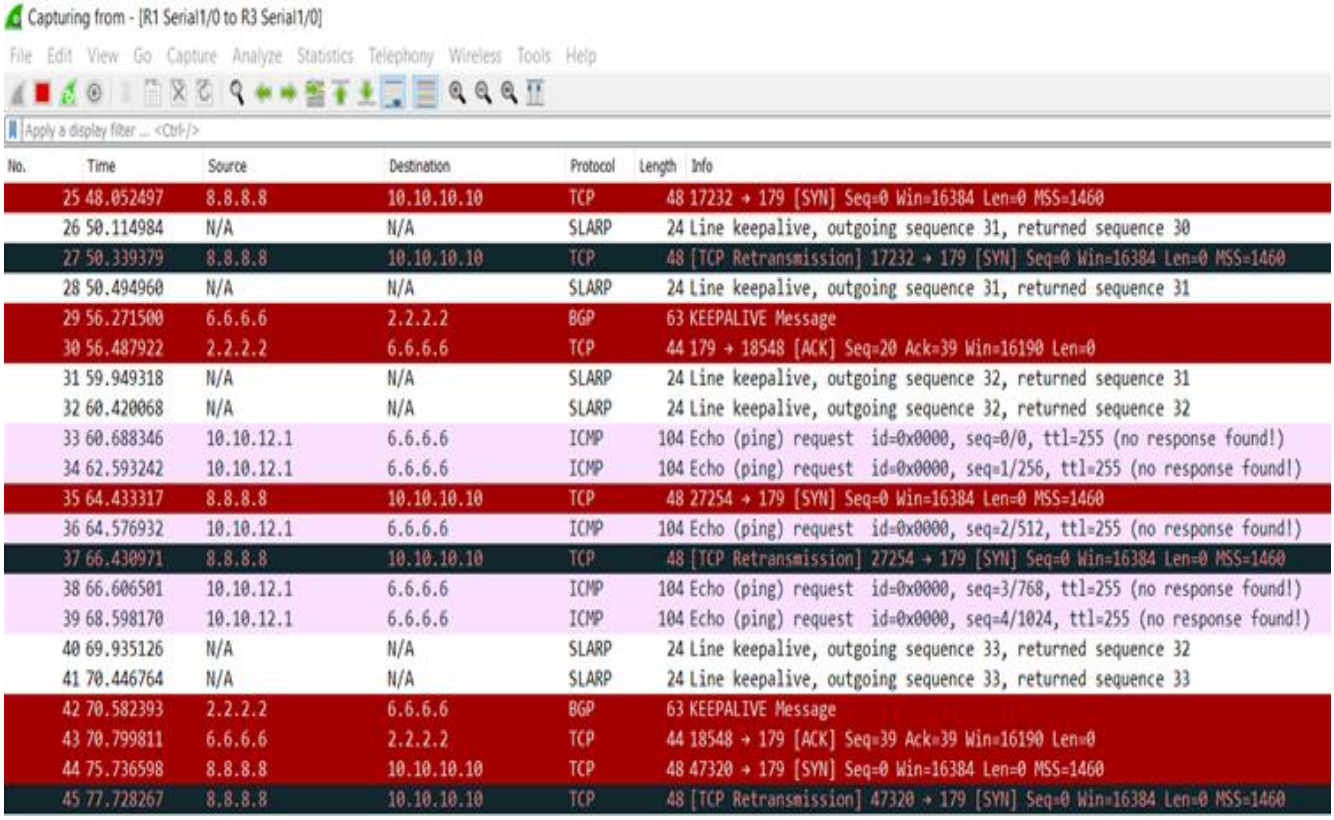


Figure 7 : Network Diagram

6.2 Experiment / Case Study 2

The below image shows the packet capture in between two routers where BGP is configured, the packets are captured using Wireshark, the image shows the TCP connection , keepalive messages going from one IP address to another .



No.	Time	Source	Destination	Protocol	Length	Info
25	48.052497	8.8.8.8	10.10.10.10	TCP	48	17232 → 179 [SYN] Seq=0 Win=16384 Len=0 MSS=1460
26	50.114984	N/A	N/A	SLARP	24	Line keepalive, outgoing sequence 31, returned sequence 30
27	50.339379	8.8.8.8	10.10.10.10	TCP	48	[TCP Retransmission] 17232 → 179 [SYN] Seq=0 Win=16384 Len=0 MSS=1460
28	50.494960	N/A	N/A	SLARP	24	Line keepalive, outgoing sequence 31, returned sequence 31
29	56.271500	6.6.6.6	2.2.2.2	BGP	63	KEEPALIVE Message
30	56.487922	2.2.2.2	6.6.6.6	TCP	44	179 → 18548 [ACK] Seq=20 Ack=39 Win=16190 Len=0
31	59.949318	N/A	N/A	SLARP	24	Line keepalive, outgoing sequence 32, returned sequence 31
32	60.420068	N/A	N/A	SLARP	24	Line keepalive, outgoing sequence 32, returned sequence 32
33	60.688346	10.10.12.1	6.6.6.6	ICMP	104	Echo (ping) request id=0x0000, seq=0/0, ttl=255 (no response found!)
34	62.593242	10.10.12.1	6.6.6.6	ICMP	104	Echo (ping) request id=0x0000, seq=1/256, ttl=255 (no response found!)
35	64.433317	8.8.8.8	10.10.10.10	TCP	48	27254 → 179 [SYN] Seq=0 Win=16384 Len=0 MSS=1460
36	64.576932	10.10.12.1	6.6.6.6	ICMP	104	Echo (ping) request id=0x0000, seq=2/512, ttl=255 (no response found!)
37	66.430971	8.8.8.8	10.10.10.10	TCP	48	[TCP Retransmission] 27254 → 179 [SYN] Seq=0 Win=16384 Len=0 MSS=1460
38	66.606501	10.10.12.1	6.6.6.6	ICMP	104	Echo (ping) request id=0x0000, seq=3/768, ttl=255 (no response found!)
39	68.598170	10.10.12.1	6.6.6.6	ICMP	104	Echo (ping) request id=0x0000, seq=4/1024, ttl=255 (no response found!)
40	69.935126	N/A	N/A	SLARP	24	Line keepalive, outgoing sequence 33, returned sequence 32
41	70.446764	N/A	N/A	SLARP	24	Line keepalive, outgoing sequence 33, returned sequence 33
42	70.582393	2.2.2.2	6.6.6.6	BGP	63	KEEPALIVE Message
43	70.799811	6.6.6.6	2.2.2.2	TCP	44	18548 → 179 [ACK] Seq=39 Ack=39 Win=16190 Len=0
44	75.736598	8.8.8.8	10.10.10.10	TCP	48	47320 → 179 [SYN] Seq=0 Win=16384 Len=0 MSS=1460
45	77.728267	8.8.8.8	10.10.10.10	TCP	48	[TCP Retransmission] 47320 → 179 [SYN] Seq=0 Win=16384 Len=0 MSS=1460

Figure 8 : Packet capture using wireshark

The below mentioned graph describes the input data in graph format

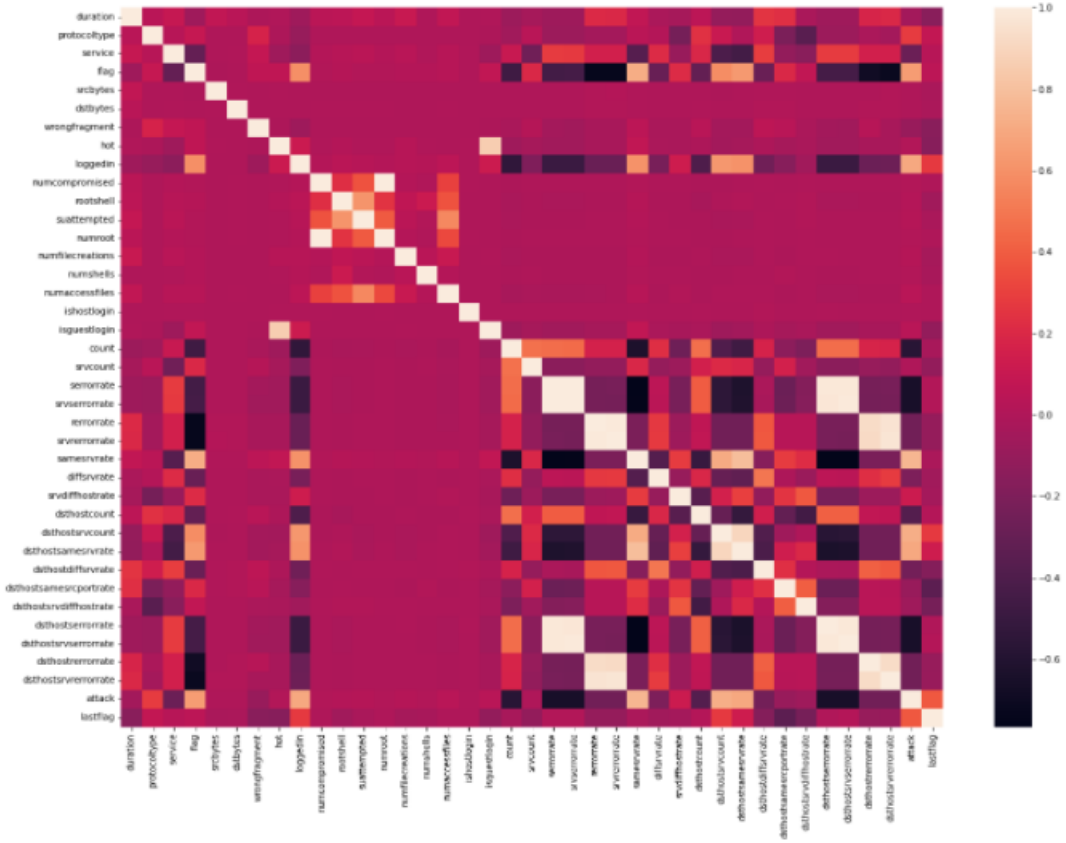


Figure 9 : Input data in graph format

6.3 Experiment / Case Study 3

The below graph depicts the attack vs the count that has been caused due to hijacking BGP

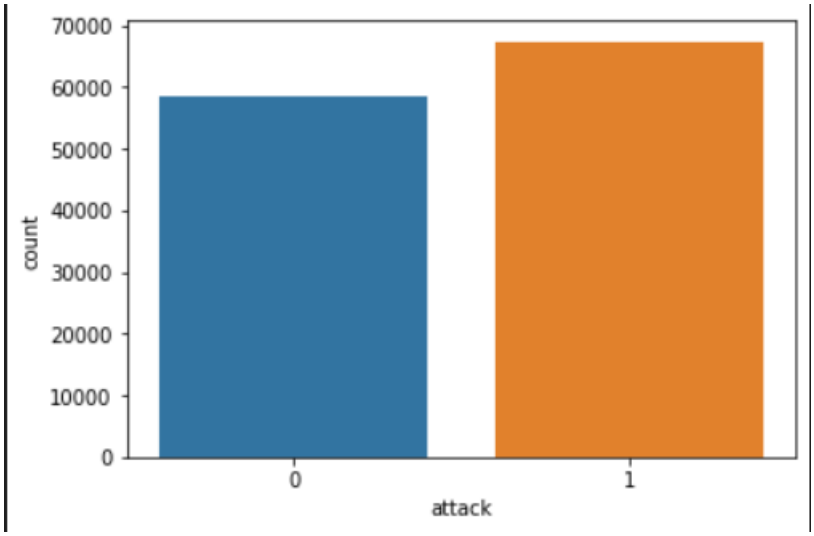


Figure 10 : Attack vs count

6.4 Experiment / Case Study 4

Figure 13 illustrates the route leaks detection system , BGP update file or the input file is the dataset used for the analysis of routes , features are extracted from the file and passed on the classifier (random forest) through which route leaks are detected .

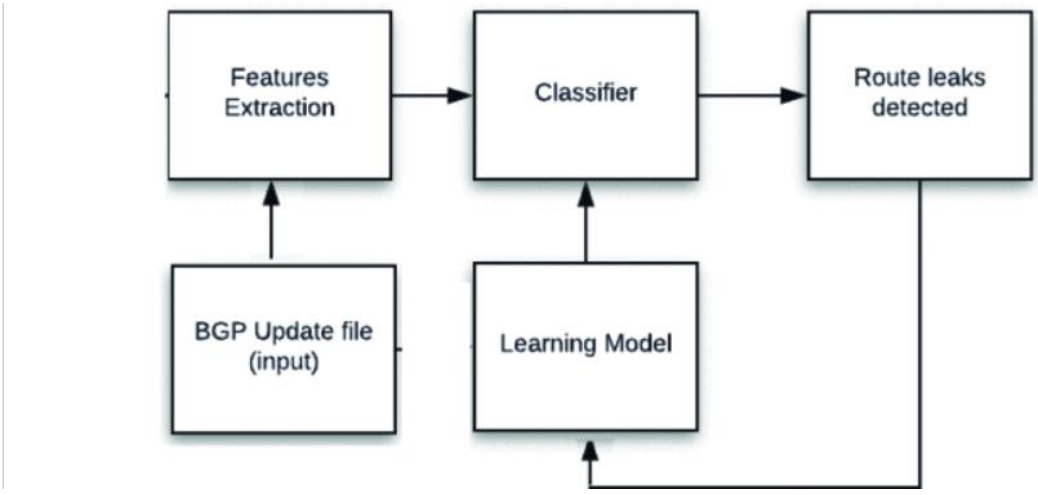


Figure 11 : System model for the route leaks detection system

6.5 Below diagram mentions the workflow in the study of random forest

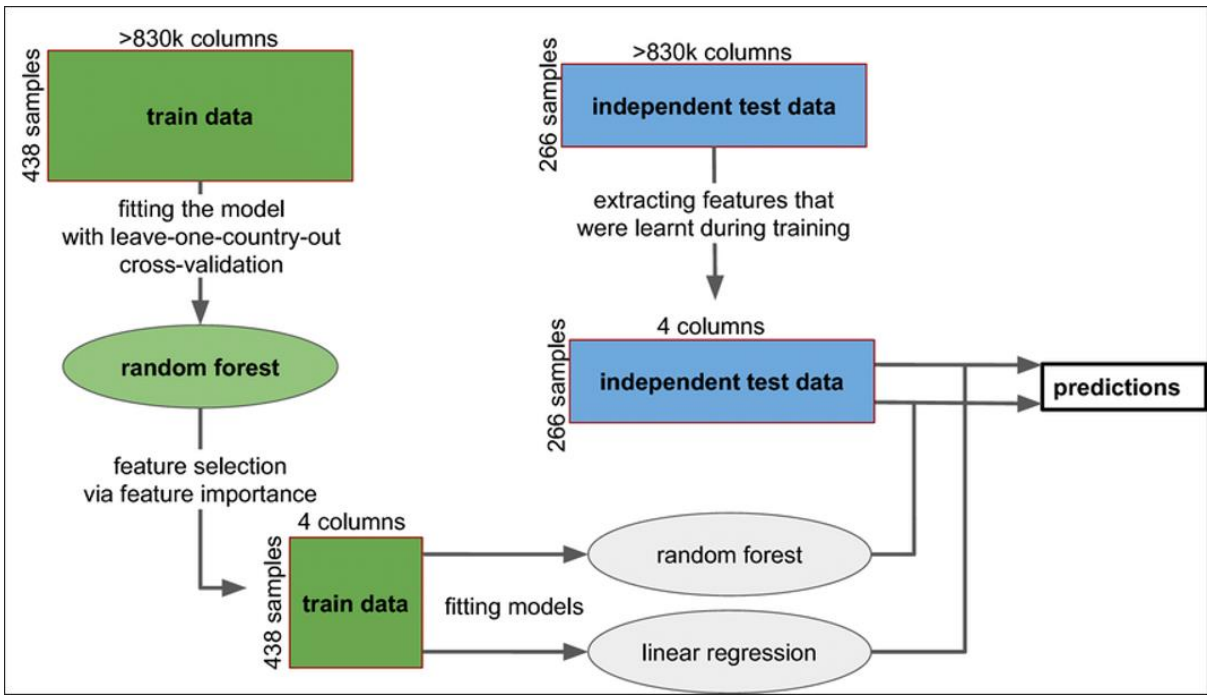


Figure 11 : Workflow of random forest classifier

6.5 Discussion

This study investigated the long-standing issue of BGP route leaks in depth. According to this research, a full solution to the problem should address all three parts of the problem: the secrecy of ASes interactions, the lack of publicly available datasets, and the system's real-world applicability. Based on the suggested taxonomy, a new taxonomy for distinguishing different types of route leaks is proposed, which is then utilised to introduce a list of the most effective characteristics that are retrieved solely from BGP Update messages. The initial dataset for genuine route leakage occurrences is obtained and annotated, and then classification algorithms are utilised to create a real-time detection system that can be deployed on any border router without the need for further data. With an average accuracy of 87 percent, a balanced accuracy of 64 percent, a precision of 86 percent, recall of 87 percent, and an F1 Score of 85 percent, the suggested system achieves the best accuracy. The worst time complexity of our system is O(NM), where N is the number of prefixes and M is the length of each prefix. In our testing setup, the system can run in real-time with each BGP update file taking less than 2 seconds on average.

The security features in BGP have also been highlighted along with the approach of detection of attacks using anomaly detection technique , when BGP was first put to use in the internet in 1994 there were no security features with it since then there were many research going on in order to secure BGP , although many attempts were made , there are many attacks related to BGP like the Amazon attack in 2018 , Google attack in November 2018, European telecommunication networks in June 2019 , companies who use the internet have to work on securing their network.

Many companies escape the security configuration irrespective of the consequences , hence this paper suggest the security configuration along with the detecting the attacks in BGP using machine learning approach using unsupervised learning method , because a model may be randomly right in identifying an abnormality, effectiveness and consistency are extremely crucial in this aspect. I have ensured that the model regularly performs effectively when it comes to detecting abnormalities.

7 Conclusion and Future Work

The suggested route leaks detection method includes a number of potential success features, including a short execution time, high efficiency, independence from third-party information, and cheap calculation cost. These parameters may be improved with more data in order to make the system more universally applicable and help put an end to the long-standing problem of BGP route leaks. Instead of depending solely on ipv4 messages, ipv6 update messages might be included to the dataset in the future. Increasing the dataset size during real-world operations can assist to apply deep learning methods like as Neural Networks, which can help to improve classification accuracy and recall.

Including security features during border gateway protocol configuration can avoid attacks and it avoids any malicious router from being advertised into the network , machine learning method can detect the attacks using unsupervised detection , further to this there can be an implementation and configuration using machine learning which would give the liberty to automate the network and monitor the attacks or any malicious activity

8 References

[1] H. K. Thakkar,, 2020. *Machine Learning Techniques for detecting BGP anomalies*, s.l.: School of Engineering Science faculty of applied science.

[2] Alshamrani, H. & Ghita, B., 2016. IP Prefix Hijack Detection Using BGP Connectivity Monitoring. *IEEE 17th International Conference on High Performance Switching and Routing.*

[3] B. Guha & B. Mukherjee, 2015. *Network Security Via Reverse Engineering of TCP Code*, India: IEEE.

[4] Shaun Nicholas , 2018 . BGP hijacking committee ‘grand theft internet’ https://www.theregister.com/2018/11/13/google_russia_routing/

[5] Dai, X. & Wang, N., 2019. *Application of machine learning*, s.l.: Journal of Physics: Conference Series 1176, 3.

[6] https://support.huawei.com/enterprise/en/doc/EDOC1, 2019. *NE40E V800R010C10SPC500 Configuration Guide - Security 01*, s.l.: HUAWEI.

[7] https://www.noction.com/, 2015. *BGP Security: the MD5 password and GTSM*, US: Noction .

Improta , A. & Sani, L., 2019. *Vulnerabilities of BGP*, s.l.: Catchpoint Systems, Inc..

[8] Systems, c., 1998. *Protection of BGP Sessions via the TCP MD5 Signature Option*, s.l.: https://datatracker.ietf.org/doc/html/rfc2385.

[9] b, t., 2020. *Machine Learning and BGP Anomaly Detection*, s.l.: https://www.bizety.com/2020/06/18/machine-learning-and-bgp-anomaly-detection/.

[10] R. Naik, "Malicious-Web-Content-Detection-Using-Machine-Learning", GitHub, 2017. [Online]. Available: https://github.com/philomathic-guy/Malicious-Web-ContentDetection-Using-Machine-Learning.

[11] "Developer Information", PhishTank. [Online]. Available: https://www.phishtank.com/developer_info.php

[12] "Sampling fraction", En.wikipedia.org. [Online]. Available: https://en.wikipedia.org/wiki/Sampling_fraction

[13] "pandas.DataFrame.iloc — pandas 1.2.1 documentation", Pandas.pydata.org. [Online]. Available: https://pandas.pydata.org/pandasdocs/stable/reference/api/pandas.DataFrame.iloc.html