CHAPTER 1

INTRODUCTION

1.1 ABOUT VEHERE

Founded in 2006, Vehere [1] is a leader in Situational Intelligence with its Unified Platform encompassing Cyber Intelligence and Cyber Security, transgressing all technology and industry verticals. Vehere's Situational Intelligence Platform harnesses the power of advanced big data analytics, machine learning and artificial intelligence to intercept, interactively visualize and analyse data from diverse and disparate Sensors. Currently our Development Centres are in Kolkata & Armenia and Sales Offices in New Delhi, Mumbai, Singapore & Cape Town.

1.1.1 INTEGRITY

We have more than a decade of experience with deployments ranging from tactical solutions to large nationwide strategic monitoring centres across governments, businesses and continents. Vehere's solutions have been successfully protecting privacy and security of sensitive customer data for more than a decade.

1.1.2 A DECADE OF INNOVATION

We were the first to introduce integrated link analysis, geospatial mapping tools and speech recognition systems against the background of a 'dumb receiver' system way back in 2006.

1.1.3 RESEARCH DRIVEN

We publish cutting-edge research in prestigious institutions including IEEE, Springer, and IJCN and on our blog, 'The Chatter'. Vehere publishes new research on an average every quarter.

1.1.4 ENHANCING HUMINT

Our scientists have been working for close to a decade to develop technologies that augment human intelligence. Our scientists believe that machines are meant to augment and not replace human intelligence.

1.2 VEHERE PRODUCT DOMAINS

1.2.1 CYBER INTELLIGENCE

1.2.2

Governments [1] strive to stay ahead of disruptive threats. Vehere CommuSAS is deployed at critical government institutions for advanced cyber intelligence. Vehere's solutions play a vital role in National Security by providing a mission critical data acquisition and analytics platform to defend countries and save lives. Our solution harnesses data from multi-sensors viz. satellite, submarine cable, web

1.2.2 CYBER SECURITY

The risk of cyber-attack has never been so high! Every connected asset is a potential target and the impact and disruption of cyber-attack is incalculable. Vehere Packet [1] worker protects businesses, government entities and nations. Protect your critical assets using Vehere’s cutting-edge Network Situational Awareness solutions.

1.3 PURPOSE

The main purpose of our project is to design and develop as an automated device fingerprinting system which will provide real time updates regarding incoming & outgoing data packets & process in a network. It will also provide live monitoring & analysis of data packets with detection and tracking features. This system will reduce capital, efficiency, dependency on 3rd party firewall, risk of intrusion & privacy overall maintains security standards.

1.4 OBJECTIVE

To develop a system which maintain monitoring, tracking and detection of anomaly signatures or sources. Also, to keep a record of all the incoming & outgoing packets. Basically, a fingerprinting device to counter security risks.

1.5 MOTIVATION

This project idea being chosen by industries self-experiencing, some issues which could be easily handled by a smart automated technology. The solution we provide will not only help the clients but also help the service or admin authorities to save lot of resources. Our system will manage authorization & validation and will also maintain security standards of airport facility. This automated system will reduce the manual work load, dependency and increase security, performance and efficiency.

1.6 DEFINITION AND OVERVIEW

Our system "**Anomaly Device Fingerprinting**" is a fingerprinting device used but in the specific context of anomaly, providing signatures & pattern analysis of incoming & outgoing processes of a network of an organisation to maintain security standards. Device Fingerprinting is a new way of differentiating between a valuable client, employee and professional fraudster online. Online identity verification and authentication is a significant challenge and concern to all business owners to safeguards their organisation interest.

Device Fingerprinting is part of a broader class of technologies called Device Intelligence used to determine whether the computer or a network you are doing business with should be trusted or not. Device Fingerprinting is a new way of differentiating between a valuable client, service and administrator and a professional fraudster online. Today, the prevalence of identity theft and hackers has meant that it is much harder to verify that the person you are doing business with is who they say they are. That new client or customer could be a compromised computer transacting on behalf of a sophisticated outsider.

So, the topic "Anomaly Device Fingerprinting" is chosen to develop an automated system which can validates internal security from inside and outside world. It will not only act as a self-propelled firewall from an organisation but as a monitoring, tracking and detection system for an organisation having full control of it. This automated system will reduce the risk as having own monitoring system than trusting other 3rd party firewall. Here an automated programmed & algorithm will be developed as per datasets generated after pre-processing & transformation of raw data which will help in having an automated device fingerprinting system for network analysis.

1.6.1 TYPES OF DEVICE FINGER PRINTING:

- Client-based methods
- Server-based methods

1.6.1.1 CLIENT-BASED METHODS

Client-based methods require installing a software executable on an end computer. The advantage of client-based methods is that they have access to otherwise hidden operating system information such as the hard drive serial number and MAC address of the network card. This information is highly unique, persistent and harder to tamper with. The major disadvantage of this method that excludes it from practical use in most ecommerce transactions is that the process requires some action or permission on behalf of the user. This may be ok if you are a bank, but definitely not if you are an ecommerce, media or retail financial services business. The other issue is of course that most corporate computers won't allow anything to be installed from an external website.

1.6.1.2 SERVER-BASED METHODS

Transparent Device Profiling methods on the other hand rely on information that can be measured remotely via a profiling server. This information is based on anonymous attributes that can be measured or derived about the user's browser, operating system and connection. Because this method has zero impact to the user's customer experience and their privacy and does not require registration, this is often the only practical method available to ecommerce, online media and retail financial businesses. The trade-off with transparent profiling is that protected device attributes such as MAC address or hard-drive serial number are not available. However, recent advances in TCP protocol and Operating System profiling now enable a device to be uniquely identified beyond more obvious browser characteristics such as browser type and version.

1.7 COMBAT STRATEGY

Device Fingerprinting is a valuable [6] system because it enables a fraudster's device to be recognized even when they change their identity through the use of proxies and stolen credit information or cloning data.

Typically, four main strategies are used to leverage Device Fingerprinting to combat fraud.

- Anomaly detection
- Device velocity
- Transaction linking
- Account linking
- Device reputation

1.7.1 ANOMALY DETECTION

Detecting anomalies related to the device fingerprint is a powerful way of providing real time fraud detection. An example of anomaly detection would be determining that a device was connecting through a proxy to hide its real location, or determining that a device is currently under the control of a botnet.

1.7.2 DEVICE VELOCITY

When fraudsters find a hole in your defences, they will try to extract the maximum value as fast as they can. Creating velocity filters based on a Device Fingerprint will enable you to minimize fraud costs even when names, credit card details and IP Addresses are changed.

1.7.3 TRANSACTION LINKING

A Device Fingerprint is a powerful tool for finding related transactions either as an identifier in itself or as a means of finding transactions with related characteristics e.g. finding related transactions performed from the same ISP and location.

1.7.4 ACCOUNT LINKING

A Device Fingerprint is a valuable tool to be able to detect when accounts or subscriptions are being accessed or shared illegally.

1.7.5 DEVICE REPUTATION

If a device has been involved with fraud, adding that device to a blacklist will enable you to protect customers that share the same device reputation network.

There are several ways of implementing and using the device fingerprinting technique. So, the effectiveness to gather data and identify a single source may vary considerably from one method to another. Either on its own or when combined with other methods, device fingerprinting can be an effective tool for tracking behaviour and identifying users. As such, this powerful technique can be used for both legit and shady activities, so being informed about their basic mechanisms is certainly a good starting point.

CHAPTER 2

OVERALL DESCRIPTION

2.1 METHODLOGY

2.1.1 INITIALIZING

At the beginning, it is ready to collect the data passing over the network, with some initial values or parameters.

2.1.2 CALIBRATING

In this stage the collected data are calibrated as no gaps are present, any broken or missing links.

2.1.3 ANALYSING

Here all the calibrated data are structed and filter as per requirement and data sets are created.

2.1.4 TRAINING

In this stage models are designed and the data sets generated on bases of them and are evaluated.

2.1.5 BUFFERING

Here expected distribution of values are build and wait for enough data for being monitored.

2.1.6 MONITORING

At this stage it compares new data to the excepted distribution of values generated. Monitoring is continuously performed on new data till it found any anomaly, then alert is generated.

2.2 PROJECT PERSPECTIVE

2.2.1 USER VIEW

In terms of user perspective our project can be used by the users for monitoring incoming & outgoing packets. also, it can be used for tracking and detection, where they will be scan incoming data packets and can analyse them. And can used to keep a record of the transfer of data. Which can be used in further reference if any kind of similar pattern or abnormal pattern arises with fraud indication an alert will be generated and if possible, it will try to stabilized the current situation with a solution or passes it to a response team for further enquiry or action. After being an automated system, it will be helpful for the users to access all the services provided by the system to analyse the network.

2.2.2 DEVELOPER VIEW

In terms of developer's perspective our project will help the admin authorities to manage and maintain all the process on network, from where they will be connected with each devices of an entire organisation. Developers can change, develop or define algorithm or parameters as per the needs or requirements of the client or the service organisation. Admin can have full control of upgrading and modifying the system.

User View = Client Side

Developer View = Service Side

2.3 PROJECT PARAMETERS

2.3.1 UNIQUENESS

Uniqueness is the measure of how accurately and confidently you can identify a return computer and differentiate it from other computers on the internet. It depends on the amount of entropy, or information, that is contained in the fingerprint [8]. For example, screen-resolution by itself is does not represent a unique fingerprint while the MAC address of a computer is generally considered unique.

2.3.2 PERSISTENCE

Persistence is the measure of how long you can expect to uniquely identify a device based on the fingerprint technique used. For example, the operating system would be a persistent fingerprint attribute, while JavaScript version used by the browser would change more frequently.

2.3.3 RESISTANCE

Resistance is a measure of how well the Device Fingerprinting technique stands up to tampering by a hacker or fraudster. For example, a browser cookie may be unique, but it is easy to delete or copy.

2.3.4 FIT

Fit is the measure of how seamlessly the Device Fingerprinting technology integrates with your business and technology requirements. For example, requiring a user to have a hardware token or to download software in order to uniquely identify them is inconvenience and not practical for most online businesses. Ideally the Device Fingerprinting method should be transparent to the end user.

2.4 PROJECT FUNCTIONALITY

2.4.1 MONITORING & ANALYSIS

Our system monitors the network and analyses the real time data captured and provides us with live updates of the present scenario. Users or clients[6] can use our system to stay connected with the network and can analyse live feeds. Here during analysis phase data are structured and filter as per requirement needed for training, analysis and testing. Our system is designed in such a way that it gives access to modify and upgrade as per requirements. It provides greater accuracy with full support as uses a trained dataset for getting result, which increases it capability and efficiency.

2.4.2 TRACKING & DETECTION

Our project can perform detection and tracking on bases of anomaly patterns and signatures. Here if any suspicious packet is present or enters into our network, then we can start tracking on bases of our previous records and parameters. We can prevent and detect from fraud as safeguards our interests by taking precautionary methods before on hand. Our system can combat from intrusion by continuously monitoring our network. And make a reference sets of signatures to detect anomaly on devices over a network.

2.4.3 TRAINING & AUTOMATION

In our project we will combine trained & test datasets then operate with effective algorithm and model to get a higher accuracy & performance. Further the train data will be used for automation, where it will automatically respond accordingly as per previous track record. An automated system will make it simpler, faster and more effective to operate and maintain and safeguard an organization interest.

CHAPTER 3

FIGURES & DIAGRAMS

Fig. 1 Methodlogy Stages

At the beginning, it is ready to collect the data passing over the network, with some initial values or parameters. Then the collected data are calibrated as no gaps are present, any broken or missing links. After that the calibrated data are structed and filter as per requirement and data sets are created. Then data sets generated on bases of them and are evaluated. Here expected distribution of values are build and wait for enough data for being monitored. Where it compares new data to the excepted distribution of values generated. Monitoring is continuously performed on new data till it found any anomaly, then alert is generated.

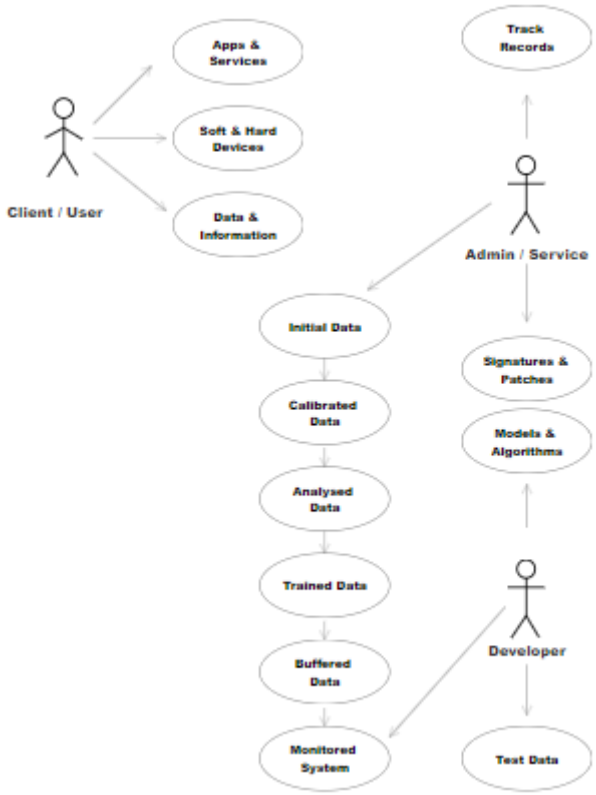


Fig. 2 Activity Diagram

Fig. 2 Activity Diagram shows the relationship between client / user with Admin / Service / Developer. Here the client / user is connected with apps, service, devices, data & info. Whereas admin / service & developer are connected with test data, models, algo, track records. This as a whole then passes through methodology stages.

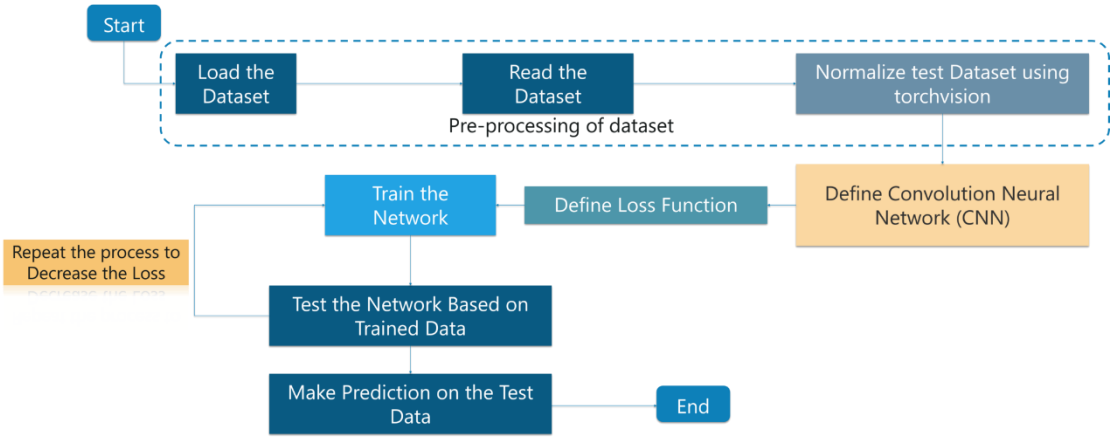


Fig. 3 Flow Diagram

Fig. 3 denotes flow of data which takes places during testing & training using CNN & TensorFlow. Here the process begins with loading datasets where the datasets are then pre-processed & normalized then passes through CNN model [3] where loss functions are defined then trained in a network. Then finally data are tested and prediction are processed.

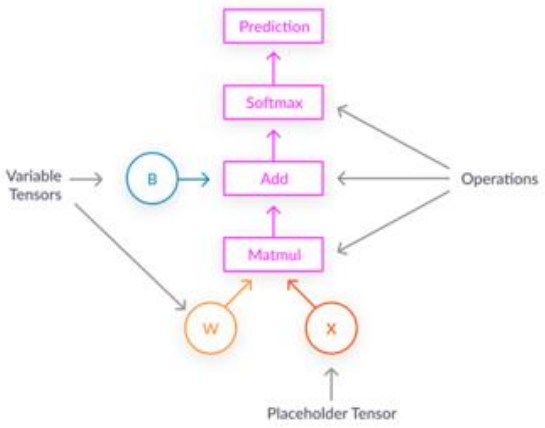


Fig. 4 Flow Diagram

Fig 4 denotes way in which weights & variables are to be used in training.

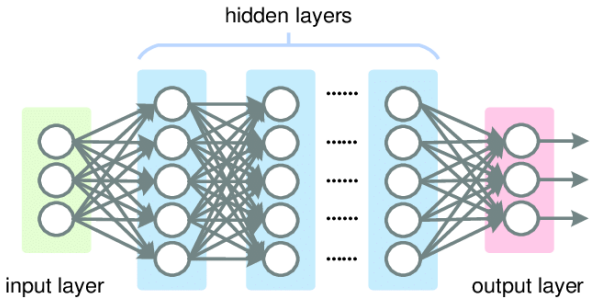


Fig. 5 layers structure

Fig 5 & Fig 6 represent cnn network architecture[3] and layers consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of a series of convolutional layers that *convolve* with a multiplication or other dot product. The activation function is commonly a RELU layer, and is subsequently followed by additional convolutions such as pooling layers, fully connected layers and normalization layers, referred to as hidden layers because their inputs and outputs are masked by the activation function and final convolution.

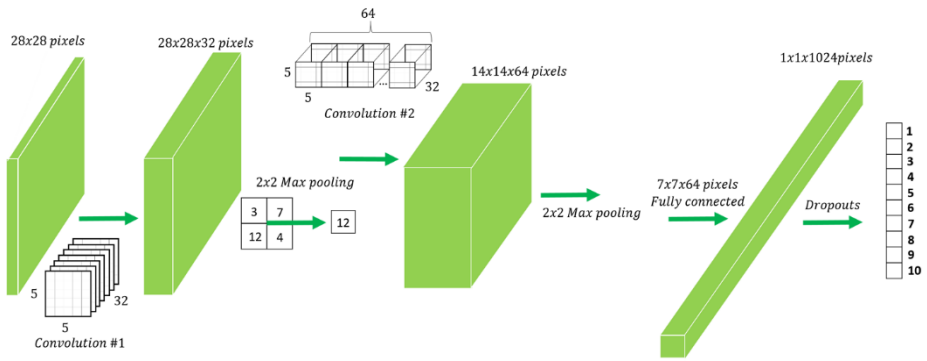


Fig. 6 CNN Architecture

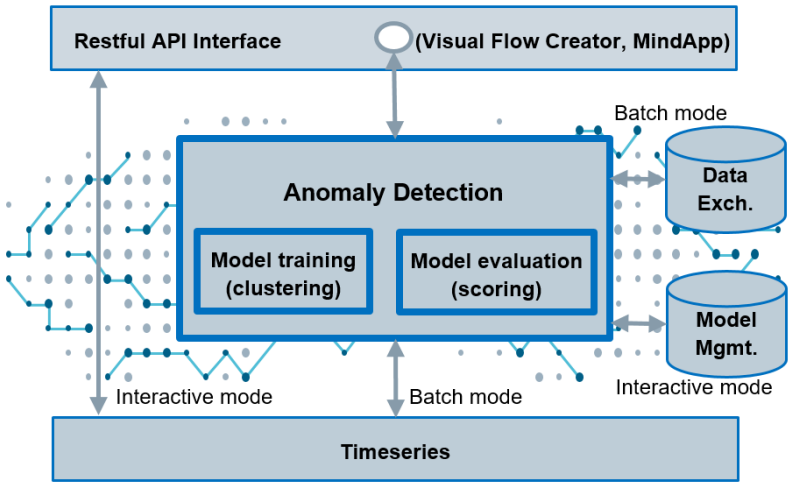


Fig. 7 Model Flow

Here in Fig. 8 & Fig 7 denotes flow of models used for anomaly detection, where cluster & classified training data are passes through tensor flow and hankel tensor[3] secifications, where the datasets are processed with different stages to achieve desire outputs.

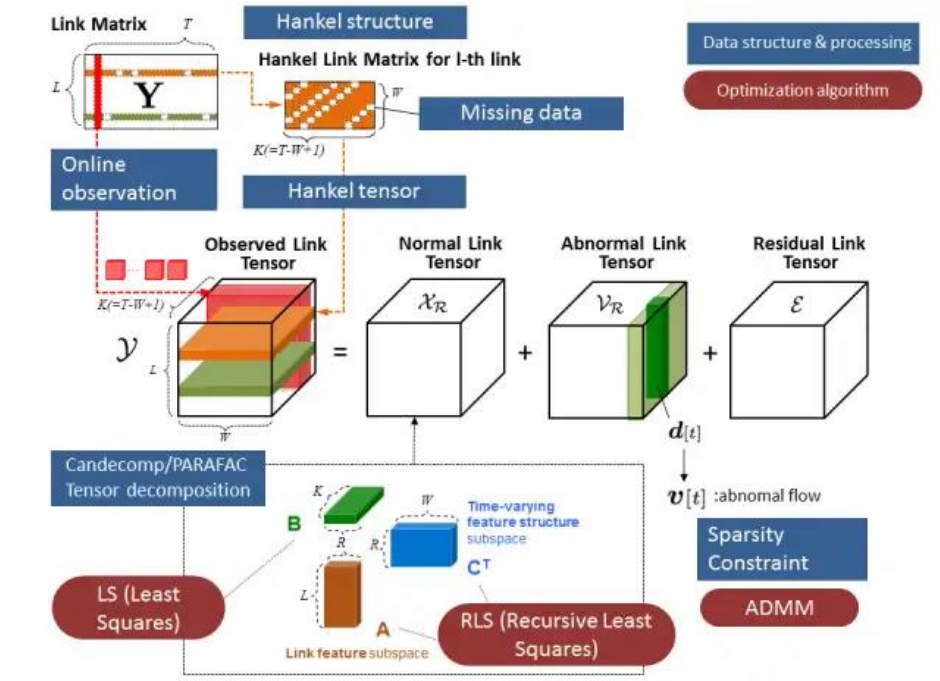


Fig. 8 Anomaly Detection Architecture

CHAPTER 4

INTERFACES

4.1 **DEVELOPER USER INTERFACES**

Img. 1 Raw Data

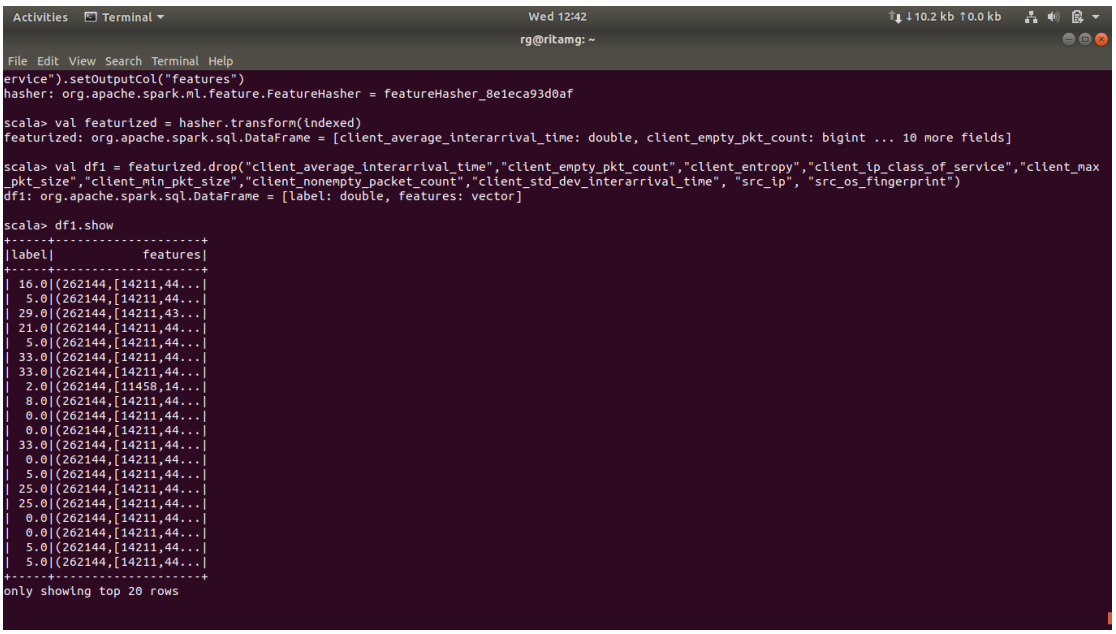
In Img. 1 displays [4] collected raw data that have been fetched and save into json or libsvm format where will be then analysed and proceesed as per requirements.

Data are loaded and displayed into spark by using spark scala commands as:

val df = spark.read.json("/home/rg/Project/eno.json")

df.show

df.printSchema



Img. 2 Featured Data Frame

In Img. 2 displays featured data with labels.

Data are featured into spark by using spark scala commands as:

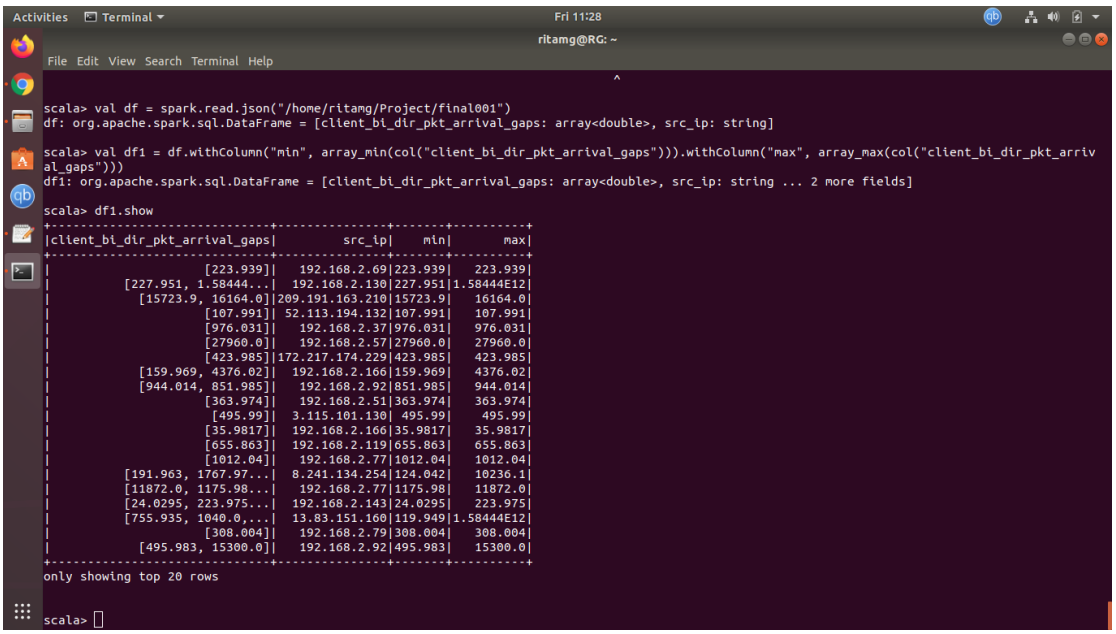
val df = spark.read.json("/home/rg/Project/sample005.json")

val indexer = new StringIndexer().setInputCol("src_ip").setOutputCol("label")

val indexed = indexer.fit(df).transform(df)

val hasher = new FeatureHasher().setInputCols("client_entropy", "src_ip", "client_average_interarrival_time", "client_empty_pkt_count", "client_max_pkt_size", "client_min_pkt_size", "client_nonempty_packet_count", "client_std_dev_interarrival_time", "src_os_fingerprint", "client_ip_class_of_service").setOutputCol("features")

val featurized = hasher.transform(indexed)



Img. 3 Min Max Data Frame

In Img. 3 displays [5] featured data with min max values.

Data are transform into spark by using spark scala commands as:

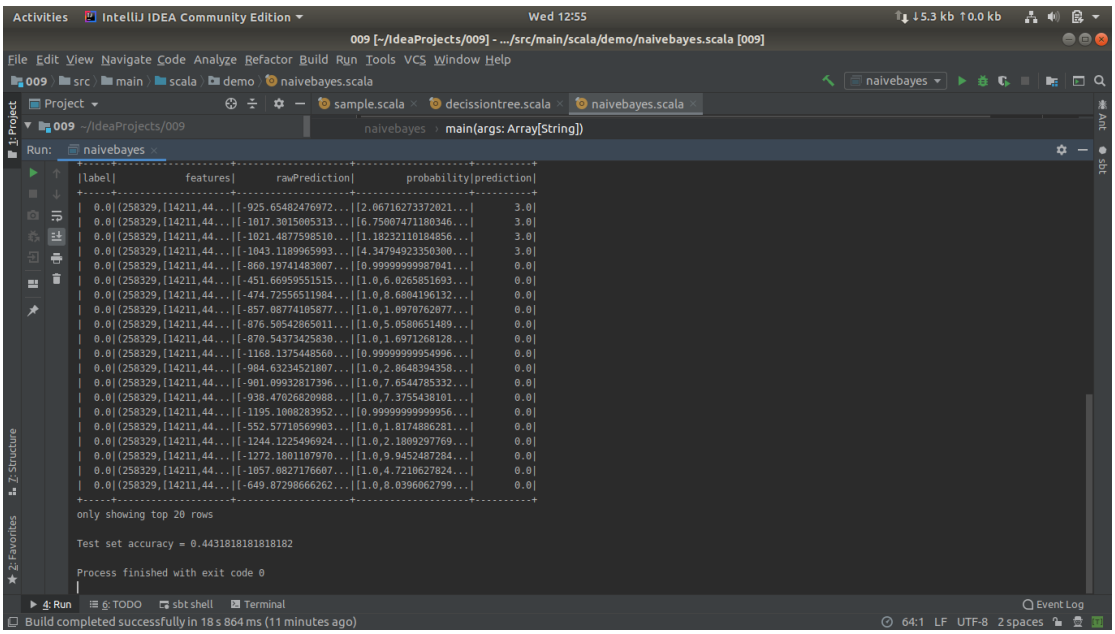
val min_max = df1.agg(min("min"), max("max")).head()

#val global_min = min_max(0)

#val global_max = min_max(1)

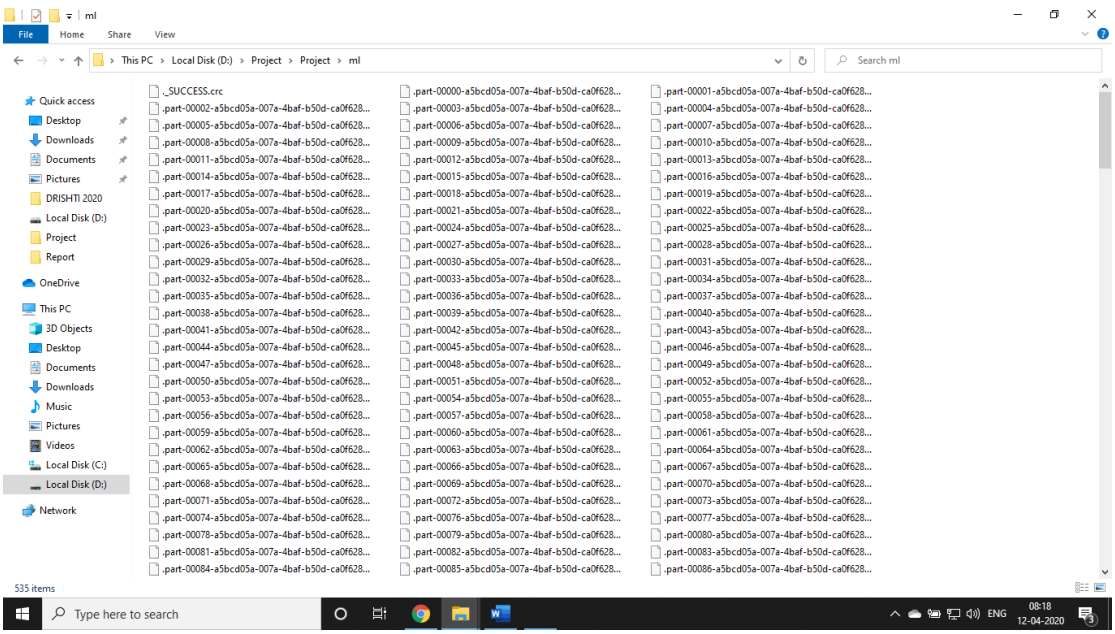
#val generate_bucketing = udf((gap_seq: Seq[Double], min: Double, max: Double) = bucketing_udf(gap_seq, min, max))

#val df4 = df3.withColumn("feature", generate_bucketing(col("client_bi_dir_pkt_arrival_gaps"), col("min"), col("max")))



Img. 4 Model Accuracy

In Img. 4 displays accuracy 0.4431818 with predication and probability.



Img. 5 Packet Data

4.2 HARDWARE INTERFACES

- PC Requirements:
  - Intel core i9 or i7 5th Generation
  - 3.3 or 4.5 GHz processor
  - 32 or 64 GB RAM
  - 512 GB SSD or 2 TB HDD
  - 4 or 8 GB VRAM

- Routers (3-5)
- Modems (2-3)
- Repeater
- Bridge
- Switch
- Server
- Firewall
- Ethernet Cables
- Wireless Devices

Initially we need to setup a small internal network connected with an admin or master pc. Master pc must have a high configuration as training, gathering, monitoring, tracking & detection as all process will be continue simultaneously.

4.3 SOFTWARE INTERFACES

- SPARK
  Version 2.4.4
  Version 2.4.5
- Hadoop
  Version 2.7
- JAVA
  Version 1.8 (Specifically)

- PYTHON 3.6
- WINDOWS 10
- UBUNTU 18.04

- Packet Tracer (Generic)
- Network Monitor (Generic)
- Firewall (Generic)

- Tensor Flow
- Anaconda
- INTEL J

CNN & RNN are applied to the datasets to generate desire outputs with clustering & classification algorithms. Vehere generic support & network tools are used in continue. Scala is the preferred language used during our process with python. Various algorithms as Naïve Bayes, Tensor Flow and other self-designed algorithms are used.

CHAPTER 5

CONCLUSION AND SCOPE OF WORK

5.1 CONCLUSION

As industry modernizes and revenues grow, fraud inevitably follows. The absence of a physical client or users at the time of the transaction with a new era of technology-driven identity theft means that there is a need for a new alternative to online identity verification. Device Fingerprinting,[2] while not a panacea, is a valuable Device Intelligence tool capable of monitoring, detecting and stopping fraud that currently flies under the radar of outdated and outmatched fraud detection systems.

Device fingerprinting has grown a lot over the past few years. As this technique is closely tied to network technology, its evolution is hard to predict but its usage is currently shifting. What we once thought could replace cookies as the ultimate tracking technique is simply not true. Recent studies show that, while it can be used to identify some devices. Instead, fingerprinting [2] is now being used to improve security. More and more companies find value in it to go beyond traditional IP analysis. They analyse the content of fingerprints to identify bots or attackers and block unwanted access to online systems and accounts.

Finally, device fingerprinting will stay in the near future at a decade. This technique is so rooted in mechanisms that exist since the beginning of the web that it is very complex to get rid of it. It is one thing to remove differences between users as much as possible. It is a completely different one to remove device-specific information altogether. Only time will tell how fingerprinting will change in the coming years but its evolution is something to watch closely as the frantic pace of development will surely bring a lot of surprises along the way.

5.2 SCOPE OF WORK

The development of the Internet along with progress in network technology brought a booming diversity of devices at the turn of the century. This diversity gave birth to browser fingerprinting, a simple technique [7] that consists in collecting information about the configuration and the composition of a user's device. Its fascinating aspect is that it is at a crossroads between companies, academic research groups, law makers and privacy advocates. As it got out of the research lab, it has a concrete impact on the web as it is now used in real-world scenarios.

Devices fingerprinting represents an alternative to current methods of tracking and identification at a time where the ad landscape is undergoing tremendous changes with the rise of ad blockers. For law makers, device fingerprinting [7] represents an additional tracking mechanism that must be regulated so that the control is given back in users' hands. For journalists, activists, businesses or members of the military that rely on the confidentiality and privacy of their communications, they must now take this technique into account to protect their activities.

All in all, device fingerprinting is still a fairly new technique. Yet, it already had a lot of impact in its short time of existence. Our effort to systematize existing knowledge proves there are still many open challenges and problems to be solved as researchers and developers are coming to grasp with its intricacies. We hope that our paper will provide the necessary basis for researchers to analyse even further the inner-workings of fingerprinting as novel solutions based on it have the potential to provide real-world benefits to millions by improving online security.

5.3 REFERENCES

1. https://vehere.com/
2. http://www.cse.iitd.ernet.in/~siy117527/sil765/readings/Device-Fingerprinting-and-Online-Fraud-Protection-Whitepaper.pdf
3. https://www.tensorflow.org/tutorials/images/cnn
4. https://spark.apache.org/docs/latest/
5. https://docs.scala-lang.org/?_ga=2.98505370.461464815.1586724030-863908889.1586724029
6. https://core.ac.uk/download/pdf/157587918.pdf
7. https://ieeexplore.ieee.org/document/8533867
8. https://hal.inria.fr/hal-01431014/document
9. https://www.hindawi.com/journals/scn/2018/6124160/
10. https://www.oreilly.com/
11. https://www.google.co.in/
12. https://www.wikipedia.org/
13. https://www.youtube.com/
14. https://blog.mozilla.org/internetcitizen/2018/07/26/this-is-your-digital-fingerprint/
15. https://pixelprivacy.com/resources/browser-fingerprinting/
16. https://branch.io/glossary/device-fingerprint/
17. http://www.m2sys.com/blog/comments-on-recent-biometric-news-stories/device-fingerprinting-care/
18. https://blog.jscrambler.com/introduction-to-device-fingerprinting/

19. https://www.eff.org/deeplinks/2018/06/gdpr-and-browser-fingerprinting-how-it-changes-game-sneakiest-web-trackers
20. https://blog.getsocial.im/device-fingerprinting-for-mobile-attribution/
21. https://www.cybage.com/sites/default/files/whitepaper/2016-03/Mobile%20Device%20Fingerprinting%20and%20Unique%20User%20Identification.pdf
22. https://www.adjust.com/glossary/fingerprinting/
23. https://www.scitepress.org/Papers/2017/63757/63757.pdf