Cape Peninsula
University of Technology

FRAUD DETECTION IN A HYBRID CLOUD NETWORK UTILIZING SOFTWARE-DEFINED NETWORKING

by

ELISHA INDARJIT

Thesis submitted in fulfilment of the requirements for the degree Doctor of Engineering in Electrical Engineering

in the Faculty of Electrical, Electronic, and Computer Engineering

at the Cape Peninsula University of Technology

Supervisor: Prof. Vipin Balyan
Co-supervisor: Prof. Marco Adonis

Bellville Campus
Date submitted 06 February 2025

DECLARATION

I, Elisha Indarjit, declare that the contents of this dissertation/thesis represent my own unaided work, and that the dissertation/thesis has not previously been submitted for academic examination toward
s any qualification. Furthermore, it represents my own opinions and not necessarily those of the Cape Peninsula University of Technology.

e.indarjit

19 Jun 25

Signed **Date**

ABSTRACT

The increase of fraudsters and fraud attacks on the communication network plays a major role in the loss of revenue, network abuse, and degradation of services. Communication and cloud networks belong to separate companies and consist of handover points from one network to another. The scope is to bring the communication and cloud networks closer and understand network traffic profiles using *Software-Defined* Networking (SDN) concept.

The SDN controller serves to route the extracted tapped data to a central server instead of making use of the core network to route traffic. To apply policies to the traffic and identify which user traffic is a fraud case, and further to send a block signal to the network charging element on the communication network.

The flow process is automated, and to protect the network from fraud attacks, using the SDN controller in a seamless approach and maintaining the performance of the network.

Currently, there are outdated fraud detection systems but no automated blocking, previous work shows high expense and uses hierarchical layers of the infrastructure. Also, SDN is a concept that is designed on a vertical layer, which oversees the environment, the uses SDN on an in-line setup.

The study defines a new framework for Communication and Cloud Providers to enable the detection and blocking of fraud. The study presents two scenarios, Smart energy abuse, and Service application abuse. The Smart Grid leverages intelligent communication technologies to modernize the electric infrastructure and raises new vectors for cyber risks and energy frauds. While services on the network require a secure platform that serves to block fraudsters.

The research solution proposes three network architectures for the identification of fraud and integration within communication and cloud networks.

The aim of the study included:

- Real-time data classification;
- Reporting of fraudulent subscribers, applications, and protocols;
- Real-time analytics; and
- Real-time blocking of data is classified as a fraudulent activity within the mobile data network.

The main purpose was to achieve a higher energy/service fraud detection rate than similar work in the field, to provide a solution that integrates within the traditional network, and further blocks the fraud within the network.

ACKNOWLEDGEMENTS

I wish to thank:

- Prof Vipin Balyan for his continuous support, expertise knowledge in leadership, domain, and skill set.
- Prof Marco Adonis for his timeously support, recommendations and dedication.

TABLE OF CONTENTS

	Declaration	ii
	Abstract	iii
	Acknowledgements	iv
	Glossary	ix
1	Introduction	1
1.1	The impact of fraud	2
1.2	Software-Defined Network in the communication network	4
1.3	Introduction to Fraud Case Detection	4
1.4	Introduction of Architecture Connectivity for Fraud Detection	5
1.5	SDN Placement and Fraud Detection	7
1.6	Communication Flow for Fraud Detection	8
1.7	Compliance Specifications for Fraud Detection and Blocking Solution	9
2	Literature Review	12
2.1	Introduction to Fraud Detection within Communication Network	12
2.2	Fraud discussion on methods	13
2.2.1	Random Sampling and Graph Neural Network	13
2.2.2	Broad Learning System	15
2.2.3	Label propagation community detection algorithm and Rule-based method	15
2.2.4	Call Detail Record	15
2.2.5	Vector Machine and Random Forest	16
2.3	Impact of Fraud Case	19
2.3.1	Impact of Fraud on Communication Network – Revenue loss	19
2.3.2	Impact of Fraud on Communication Network – Detection Features	21
2.3.3	Impact of Fraud on Communication Network – Fraud technology solutions	23
2.3.4	Impact of Fraud Case on Smart Energy and Service Application	26
2.3.5	Hybrid Cloud Network	28
2.4	Challenges faced with Cloud network	29
2.4.1	Software as a Service	30
2.4.2	Platform as a Service	30
2.4.3	Infrastructure as a Service	31
2.5	Software-Defined Network	32
2.5.1	Integration of Communication network and SDN controller	33
2.5.2	Summary of Software-Defined Network Challenges	37
2.5.3	Summary of Software-Defined Network Benefits	37
2.6	Fraud case detection	38
3	Architecture and Connectivity	44
3.1	Architecture and Connectivity	44
3.2	Fraud Detection Selected Architecture	46
3.3	Key Factors on the Architecture	48
3.3.1	Hybrid Cloud	48
3.3.2	Aggregator deployment	52
3.3.3	SDN Placement	58
3.4	Energy Abuse fraud	62
3.5	PCRF/PCF	63
3.6	Compliance Specifications	68
3.7	Introduction to 5G Standalone	73
4	Characteristics and Solution Scoping	76
4.1	The large internet pool of users	77
4.2	Correlation of the Billing system	78
4.3	Characteristics of the User/Traffic Profiles	78
4.4	Sim registration and swop	81
4.5	Aggregation/interface points in the communication network	82
4.6	SDN location and specification	83
4.7	Energy Fraud identification	87
4.8	Packet de-encryption	89
4.9	Blocking Integration and Rules	94
4.10	Test Case(s) proposal	95
4.10.1	Test Case 1: Fraud detection with an unpaid token on the Communication Network	95

4.10.2 Test Case 2: Fraud detection utilizing the network 96
4.10.3 Test Case 3: Fraudulent Access on Cloud Network 100
4.10.3.1 Test Case 3a 101
4.10.3.2 Test Case 3b 104
4.10.4 Test Case 4: DDoS attack 106
5 Design and Test Cases 1,2,3,4 111
5.1 Background of design 112
5.2 Test Case 1: Fraud detection with unpaid token on the Communication Network 114
5.3 Test Case 2: Fraud detection utilizing the network 116
5.4 Test Case 3: Fraudulent Access on Cloud Network 125
5.5 Test Case 4: DDoS attack 134
5.6 Summary of Test Case(s) 141
6 Discussion 143
6.1 Test Case 1: Fraud detection with unpaid token on the Communication Network 146
6.2 Test Case 2: Fraud detection utilizing the network 148
6.3 Test Case 3: Fraudulent Access on Cloud Network 151
6.4 Test Case 4: DDoS attack 153
6.5 Cross References 156
6.6 Future Recommendations 157
7 Conclusion 159
7.1 Characteristics 159
7.2 SDN Controller Placement 161
7.3 Blocking Integration and Rules 162
7.4 Test Case Proposals 162
7.5 Summary discussion on Fraud Detection Algorithms 167
7.6 Challenges and Considerations 168
7.7 Future considerations for researchers 169
8 References 171

LIST OF FIGURES

Figure 2.1 Equipment for mapping data in a communication network to detect phantom subscribers 14
Figure 2.2 Accuracy analysis from existing methods 17
Figure 2.3 Number of reported fraud cases from Romance 21
Figure 2.4 Social media platforms ranked by usage 24
Figure 2.5 Distribution of incorrect reading counts 28
Figure 2.6 x as a Service on Infrastructure, Platform, Software 30
Figure 2.7 SDN Architecture 35
Figure 2.8 Imbalance issue 40
Figure 2.9 Flow process of Competitive Swarm Optimization Deep Convolutional Neural Network 42
Figure 3.1 Blocker-First Architecture 45
Figure 3.2 Interconnect-Server Architecture 45
Figure 3.3 Software-Defined Network Controller Architecture 46
Figure 3.4 Hybrid cloud network and traditional Spine/Leaf network based on physical connectivity 47
Figure 3.5 Logical network packet flow on tapping/aggregating and probing 48
Figure 3.6 Layer 2 General Routing Encapsulation tunnel 52
Figure 3.7 Application Server query 54
Figure 3.8 Pcap files 54
Figure 3.9 Pcap file – Before De-Duplication 55
Figure 3.10 Pcap file – After De-Duplication 55
Figure 3.11 5G Core control and userplane Network 57
Figure 3.12 VXLAN Encapsulated network packet 59
Figure 3.13 GRE Encapsulated network packet 59
Figure 3.14 Flow between Central Processing Unit(CPU) and Application Specific Integrated Circuit(ASIC) of the SDN controller 60
Figure 3.15 Communication flow between PT files, PACL TCAM and Flow Table 61

Figure 3.16 SGi and N6 protocol stack 62
Figure 3.17 Virtual Local Area Networks (VLANs) configured within the communication network 63
Figure 3.18 Communication flow of network elements to the PCRF 64
Figure 3.19 Architecture of data flow 65
Figure 3.20 RADIUS Integration point 67
Figure 3.21 Diameter Integration point 68
Figure 4.1 4G Network with Billing system 77
Figure 4.2 5G Network with Billing system 77
Figure 4.3 Integration to Circuit Switching (CS) platform from 2G/3G/4G/5G 78
Figure 4.4 2-day Usage for Online Banking 79
Figure 4.5 2-day Usage for Transportation of Uber 79
Figure 4.6 2-day Usage for Online Gaming 80
Figure 4.7 2-day Usage for Online Retail 80
Figure 4.8 2-day Usage for Food on Mr Delivery 81
Figure 4.9 2-day Usage for Online Schooling 81
Figure 4.10 Architecture of network including Breakout point 82
Figure 4.11 SDN Controller overlooking Metro domain, Core domain and Centre domain 83
Figure 4.12 Service Protocols, Transport Protocols, and Data Plane 85
Figure 4.13 Multi-Protocol Label Switching Routing based on Push, Swop and Pop with Segment Routing 85
Figure 4.14 Layered Software-Defined Network on virtual platform for service provisioning 86
Figure 4.15 Detailed layered Software-Defined Network setup 87
Figure 4.16 7-day user stats on average of a sample of (x) 88
Figure 4.17 7-day user stats showing network abuse 89
Figure 4.18 Charging nodes communication 94
Figure 4.19 Billing flow - High level 96
Figure 4.20 Traffic Classification Framework 97
Figure 4.21 ATLAS showing activities within 24 hours 107
Figure 4.22 Matched Objects 108
Figure 4.23 Host Detection 109
Figure 4.24 Summary report of High Alert 110
Figure 5.1 WireShark tool showing the pcap file 111
Figure 5.2 Software Tools 112
Figure 5.3 Gigamon engine for traffic intelligence 112
Figure 5.4 Pass and Drop rules for Applications 113
Figure 5.5 Splunk Enterprise showing IMEI subscriber ID 114
Figure 5.6 Process Flow Test Case 1 115
Figure 5.7 Network Traffic by protocol over time – ICMP traffic 117
Figure 5.8 Network Traffic by protocol over time – Other traffic 117
Figure 5.9 Network Traffic by protocol over time – TCP traffic 118
Figure 5.10 Network Traffic by protocol over time – UDP traffic 118
Figure 5.11 IP addresses - Source and Destination 118
Figure 5.12 IP 10.32.134.43 connections to Servers 118
Figure 5.13 Connectivity between IP addresses 119
Figure 5.14 Active Visibility_Policies_Conditions_Actions 120
Figure 5.15 Utilization over 30-days 123
Figure 5.16 Fraud detection High Utilization – Condition 124
Figure 5.17 Fraud detection high utilization – Actions 124
Figure 5.18 Enabled and Success policy 125
Figure 5.19 Cloud Framework on Fraud 126
Figure 5.20 Machine Learning techniques 128
Figure 5.21 Interconnect Type 1 and 2 130
Figure 5.22 NNI between Mobile Network and Cloud Provider 131
Figure 5.23 Fraud Detection & Presentation – Cloud 131
Figure 5.24 Cloud Fraud Detection process 132
Figure 5.25 Cloud Fraud detection sample process 132
Figure 5.26 Online Fraud Cloud prediction 134
Figure 5.27 Traffic cleaning and data filtering 137
Figure 5.28 DDOS Detection 139
Figure 5.29 DDOS mitigation 140

Figure 5.30 Alert levels 141

LIST OF TABLES

Table 2.1 Fraud Detection methods 13
Table 2.2 Fraud type and revenue loss 20
Table 2.3 Detection Methods on Advantages and Disadvantages 22
Table 2.4 Use cases of fraud environments 25
Table 2.5 Required feature for SDN controller 34
Table 2.6 Comparison of routing technology assessment based on jitter and packet loss on five devices 36
Table 2.7 Detection Methods and Description 38
Table 3.1 Spanning versus Tapping 53
Table 3.2 UDA Configuration for filtering traffic 56
Table 3.3 Map rule terms 56
Table 3.4 Compliance for the design and the study 68
Table 4.1 4G and 5G Network elements 76
Table 4.2 Fraud types in Tariff plans 95
Table 4.3 Overview of Cloud Key Services 100
Table 5.1 Mapped Record of Subscriber_Rated_Billed 115
Table 5.2 Subscriber Fraud 116
Table 5.3 Conditions for Traffic Intelligence 122
Table 5.4 Actions for Traffic Intelligence 123
Table 5.5 Traditional Approach and Hadoop Machine Learning Approach 129
Table 5.6 Training Data 132
Table 5.7 Phases in the detection of DDOS attack 135
Table 5.8 DDoS Attack Types 135
Table 6.1 Traffic classification, detection, and blocking 145

APPENDIX 188

GLOSSARY

ACCESS	Adaptive Connected Component Embedding Simplification Scheme
ACL	Access Control List
AI	Artificial Intelligence
AMF	Access and Mobility Management Function
API	Application Programmable Interface
ASIC	Application Specific Integrated Circuit
AWS	Amazon Web Services
BGP	Border Gateway Protocol
BLS	Broad Learning System
BOM	Bill Of Material
CCS	Converged Charging System
CDN	Core Data Network
CDR	Call Detail Record
CNF	Containerized Network Function
CPU	Central Processing Unit
CSO-DCNN	Competitive Swarm Optimization Deep Convolutional Neural Network
DDOS	Distributed Denial-of-Service
DNS	Domain Name System
DPI	Deep Packet Inspection
EIGRP	Enhanced Interior Gateway Routing Protocol
EPC	Evolved Packet Core
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
FIFO	First In First Out
FL	Federated Learning
FMW	Fusion Middleware
GAN	Generative Adversarial Networks
GGSN	Gateway GPRS Support Node
GNN	Graph Neural Networks
GPRS	General Packet Radio Service
GTP	General Tunnelling Protocol
HPMN	Home Public Mobile Network
HSS	Home Subscriber Server
HTML	Hyper Text Markup Language
IaaS	Infrastructure as a Service
IDS	Intrusion Detection System
IMS	IP Multimedia Core Subsystem
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
IPFS	Inter Planetary File System
IRSF	International Revenue Share Fraud
KPI	Key Performance Indicator
L2GRE	Layer 2 General Routing Encapsulation
LDP	Label Distribution Protocol
LPA	Label Propagation Algorithm
MFA	Multi-Factor Authentication
ML	Machine Learning
MLP	Multilayer Perceptron
MME	Mobility Management Entity
MO	Managed Object

MPLS	Multiprotocol Label Switching
NFA	Neural Factorization Autoencoder
NFV	Network Function Virtualisation
OOD	Out-Of-Distribution
OSI	Open Systems Interconnect
OSPF	Open Shortest Path First
PCAP	Packet Capture
PCEF	Policy and Charging Enforcement Function
PCF	Policy Charging Function
PCFG	Probabilistic Context Free Grammar
PCRF	Policy Charging Rule Function
P-CSCF	Proxy Call Session Control Function
PDN	Packet Data Networks
PGW	Packet Gateway
PLMN	Public Land Mobile Network
QoS	Quality of Service
RMCP	Recovering, Mining, Clustering, and Predicting
RNN	Recurrent Neural Network
RSVP	Resource Reservation Protocol
SDN	Software-Defined Networking
SDP	Software-Defined Parameter
SG	Smart Grid
SID	Segment ID
SMF	Session Management Function
SMOTE	Synthetic Minority Oversampling Technique
SOAP	Simple Object Access Protocol
SRGB	SR Global Block
SVM	Support Vector Machine
TCG	Test Call Generation
TCP	Transmission Control Protocol
TCP	Transmission Control Protocol
TEID	Tunnel Endpoint ID
TLS	Transport Layer Security
TMS	Threat Mitigation System
TPF	Traffic Plane Function
UDA	User Defined Attribute
UE	User Equipment
UPF	User Plane Function
URL	Uniform Resource Locator
VLAN	Virtual Local Area Network
VPMN	Visitor Public Mobile Network
VPN	Virtual Private Network

CHAPTER ONE
INTRODUCTION

The rise in electronic payment services has led to an increase in fraud cases. Traditional fraud detection methods, which are often manual or semi-automated, are insufficient. These methods involve:

- Change Management Process: Requires approval for network changes.
- Business Assurance: Ensures compliance with business regulations.
- Network Detection Process: Uses scripts to detect fraud based on mobile device connections.

Fraud detection is crucial due to the significant financial losses caused by skilled fraudsters. Traditional methods require manual intervention, which adds complexity and reduces the network's performance. The proposed research aims to develop an intelligent and automated solution that does not affect live networks. Various methods have been developed to protect networks from fraud. These methods are evaluated based on their effectiveness and relevance to modern networks, including 5G covered in Table 2.1 on Detection Methods:

- Random Sampling: Equal probability sampling technique.
- Graph Neural Networks (GNN): Operates on graph-structured data.
- Broad Learning System (BLS): Detects network anomalies.
- Label Propagation Algorithm (LPA): Detects communities in network structures.
- Rule-Based Method: Applies predefined rules to data.
- Call Detail Record (CDR) Analysis: Uses statistical analysis and machine learning.
- Vector Machine: Supervised learning model.
- Random Forest: Combines decision trees for classification and regression.

From section 2.2.1 on Random Sampling and Graph Neural Network, the growth of networks increases the amount of fraud cases. Key considerations include fraudsters staying updated with new services and real-time detection is necessary. The key challenges are managing user plane traffic in different locations and centralized systems for data aggregation.

Discussed in section 2.2.2 on Broad Learning System, Zhong et al. (2019) propose a Broad Learning System (BLS) for real-time fraud detection in phone calls, showing better results than traditional methods. However, it requires a constant updates and manual intervention.

From section 2.2.3 on Label Propagation Community Detection Algorithm and Rule-Based Method, Niu et al. (2016) and Peng and Lin (2018) propose methods for detecting fraud communities and profiling fraudulent SIM cards. These methods require a deep understanding of network connectivity. Also, covered in section 2.2.4 on Call Detail Record (CDR) Analysis, CDR analysis uses both content and occurrences of CDRs to detect fraud. It includes statistical analysis, machine learning, and data visualization. Collusive fraud, where multiple fraudsters target a service, is particularly challenging to detect.

From section 2.2.5 on Vector Machine and Random Forest, various Machine Learning models, including Vector Machines and Random Forests, are used for fraud detection. These models analyse data for classification and prediction, showing high accuracy in detecting fraudulent activities.

1.1 The impact of fraud

The impact of fraud on the communication network based on revenue loss, fraud and identity theft lead to significant financial losses for telecommunication companies and erode customer trust. Including:

- Roaming Fraud, causing revenue loss due to delays between Visitor Public Mobile Network (VPMN) and Home Public Mobile Network (HPMN), often involving identity theft.
- Blockchain solutions propose to secure data storage and optimize roaming services using smart contracts and private blockchains.
- Revenue loss statistics, various types of fraud result in substantial global revenue losses, with International Revenue Share Fraud (IRSF) being the highest.

In Table 2.2 it provides the fraud type and revenue loss in ZAR billion.

• IRSF:	R 193.5
• Interconnect Bypass Fraud:	R 107.46
• Premium Rate Service Fraud:	R 67.32
• Subscription Fraud:	R 144.9
• PBX Hacking:	R 134.46
• Wangiri Fraud:	R 31.86
• Phishing:	R 28.26
• Abuse of Service Terms:	R 21.06
• SMS Faking/Spoofing:	R 14.22

Covered in section 2.3.2 provides impact of fraud on communication network based on detection features, various methods and features are used to detect fraud, with research gaps identified in each approach. Table 2.3 provides the detection features, methods, and research gaps,

- Fake Images Online: Uses user profile features and HTML text extraction with Naïve Bayes and SVM, limited by dataset scope.
- Dating Profile Descriptions: Uses SVM, limited by profile data.
- Celebrity Detection: Uses image scanner APIs, limited by dating website scope.

- Cybersexual Interactions: Uses URL extraction, with challenges in underage usage.

The challenges include identifying fake profiles, handling private and sensitive data, and improving spam detection and false identity recognition. Covered in section 2.3.3 on impact of fraud on communication network based on fraud technology solutions, technological solutions for fraud detection include are:

- Smart meters use to detect power loss using sensors and recurrent neural networks.
- Online trading fraud uses Balance Generative Adversarial Networks (GAN) for better precision and recall in detecting fraudulent transactions.
- Big Data analytics that play a crucial role in detecting fraud in online financial transactions and social network platforms.
- Anomaly detection to identify deviations from normal patterns, applicable in various sectors like finance, healthcare, manufacturing, and networking.

Section 2.3.4 on the Impact of Fraud on Smart Energy and Service Applications, provide the rise of digital services, including Smart Grid (SG) energy applications, has increased the risk of attacks. The key points include:

- User Behaviour Analysis: Detecting fraud by analysing user and social characteristics.
- Graph Topological Structure: Lv et al. (2019) propose using multi-encoders to identify fraud cases.
- Brazilian Energy Distributors: Araujo, Almeida, and Mello (2019) highlight significant financial losses due to fraud and the need for technical inspections.
- Energy Theft: Defined by Ganguly et al. (2022) as data corruption, device tampering, and bypassing meter readings.
- Machine Learning Solutions: Korba and Karabadji (2019) propose using ML to detect fraud based on user consumption profiles.
- SVR-Based Detection: Liu, Hu, and Ho (2014) use Support Vector Regression to detect pricing fraud in electricity billing.
- Classification Models: Jokar, Arianpoo, and Leung (2013) use ML methods to improve detection robustness.

Also, the research explores Energy Theft scenarios in:

- Cyber-Attack: Entering the network to create panic.
- Device Tampering: Disabling billing functionality.
- Bypass Meter: Unauthorized use of services.

Covered in section 2.3.5 on Hybrid Cloud network offer benefits like lower costs, higher availability, and increased security based on:

- Effective cloud management: Xu, Su, and Zhang (2018) propose combining cloud management with a rule engine for cost reduction and flexibility.
- Challenges: Include vendor lock, migration issues, and interoperability standardization.

- OpenStack: Explored by Sitaram et al. (2018) for building hybrid clouds.

The challenges faced with Cloud Networks include data backup and recovery due to dedicated pipelines, cloud bursting and auto-scaling, migration of services and resource monitoring and management.

1.2 Software-Defined Network in the communication network

Discussed in section 2.5 explores Software-Defined Network (SDN), SDN introduces a new framework for managing network components through a centralized control system, enabling flexible and standardized deployment of network services. Key features include:

- Centralized Control to manage links, nodes, application slices, and configurations.
- Open Standard Interfaces to accommodate third-party applications.
- Separation of control and data planes for enhances network programmability and automation.
- Consistent Architecture Standards that are across fixed, mobile, cloud, enterprise, and security domains.

Section 2.5.3 covers the integration of communication network and SDN controller providing key features for SDN controllers include Audit Capabilities to identifies packet drops, Implementation on VMs to support scalability and modularity, high availability to ensure network reliability and traffic rerouting, integration with standard interfaces to support orchestration and multi-vendor environments, and security to provide strong authentication and integrity validation.

1.3 Introduction to Fraud Case Detection

Network fraud detection is a critical area of the research, aiming to design methods to detect and mitigate fraud. Traditional methods often fall short in addressing the complexities of modern, evolving systems. The integration of machine learning and deep learning offers robust solutions across various fields, including healthcare, entertainment, and agriculture.

Detection Methods covered in Table 2.7 on Detection Methods on the following to mention:

- XGBoost: Uses a gradient-boosting framework to build decision trees. Limitation: Tends to overfit data set.
- ADABoost: Converts weak learners to strong learners for binary classification. Limitation: Requires a noise-free dataset.
- Naive Bayes: Based on Bayes’ theorem to predict outcomes by probability. Limitation: Zero-frequency issue.
- Decision Tree Classifier: Uses a tree-based structure for classification and regression. Limitation: High computational resources.

- Random Forest Classifier: Aggregates multiple decision trees to improve predictions. Limitation: High computational resources and time.
- KNN: Classifies based on distance to known data points. Limitation: Ineffective with high dimensionality.
- Logistic Regression: Used for dichotomous variables. Limitation: Overfitting with many features.
- K-CGAN: Uses Conditional GAN architecture for balanced class distribution.
- SMOTE and ADASYN: Techniques for handling imbalanced datasets in credit card fraud detection.
- Machine Learning Models: Used to detect false positive advertisements and malicious URLs.
- Web Scraping: Collects data for analysis and model training.
- Smart Meters: Detect energy theft using records and time series data.
- Random Forest and U-net: Used for classification and segmentation problems.
- ACCESS and RMCP: Methods to predict fraud in online lending by analysing knowledge graphs and clustering algorithms.
- CSO-DCNN: An unsupervised learning method showing high accuracy in detecting credit card, insurance, and mortgage fraud.
- Neural Factorization Autoencoder (NFA): Analyses customer calling patterns to detect fraudulent calls in real-time.

These summaries highlight the various methods and technologies used in fraud detection, emphasizing the importance of machine learning and deep learning in developing an effective solution.

1.4 Introduction of Architecture Connectivity for Fraud Detection

The key components for network architecture, are routing and switching to route network packets from source to destination; tapping, aggregation, and probing, the platform is to copy traffic, apply filters, and stream traffic; and the packet core for subscriber access, authorization, and service fulfilment.

Three architectures were proposed for the study, namely, Blocker First Architecture, it Includes a Blocker layer after the 'User Charging' element. The challenges results in managing dynamic network environments, verifying correct profile blocking, and handling risks associated with 'Blocker' issues or failures.

Interconnect-Server Architecture, the Interconnect layer filters network interface traffic. The challenge is the Addition of infrastructure and software costs.

SDN Controller Architecture is the proposed solution, the SDN controller handles fraud detection and blocking signals. The advantages, it provides centralized control for traffic classification and fraud blocking and allowing integration with policy charging elements.

Further the SDN Controller Architecture is built with the following components,

- Leaf and spine topology: Spine switch connects to layer 3 routing network; leaf switch connects to SDN controller and other switches.
- Hybrid cloud integration: AWS and Google Cloud connect via firewall access, sharing resources for specific services.
- Tapping and aggregation platform: Provides traffic copies to the probe platform for analysis.
- SDN controller: Centralized control for traffic classification and fraud detection/blocking.

Figure 3.4 in section 3.2 provides the physical connectivity and Figure 3.4 provides the traffic flows of sites A, B, and C, connecting to the service node via the SDN controller, which performs fraud classification. This architecture ensures efficient fraud detection and blocking by leveraging SDN technology for centralized control and traffic management.

The architecture is made of the Hybrid cloud on AWS and Google Cloud, the integration of a hybrid cloud network provides flexibility, scalability, and security. It allows migration of applications to the cloud, use of public, private, or hybrid cloud environments, to run critical applications on-premises for billing and compliance. Low latency for remote applications. Local data processing for large datasets, and data centre extension for cloud bursting, backup, and disaster recovery.

The purpose of the aggregators captures and filter traffic at each site. They are essential for ensuring traffic is encapsulated within Layer 2 General Routing Encapsulation (L2GRE) tunnels. And reducing the need for core network upgrades and managing latency issues.

In order to format the source/input data for the purpose of testing, key features for traffic analysis were used,

- De-duplication by removing duplicate packets to ensure accurate traffic analysis.
- User Defined Attribute (UDA) used to create pass-and-drop map rules to filter specific sequences of bits in network packets.
- Correlation performing the separation of user-plane and control-plane traffic, reconstructing network packets for accurate subscriber session analysis.

The purpose of Software-Defined Network (SDN) controller manages traffic classification and fraud detection/blocking. It integrates with Policy Charging elements (PCRF/PCF) to enforce network policies. Its centralized control within the network, ensuring efficient traffic management and fraud detection.

The types of fraud based on applications

- Subscription Fraud: Unauthorized use of services.
- Roaming Fraud: Exploiting delays between networks.
- Premium Rate Service Fraud: Abusing premium numbers for revenue.

The architecture for fraud detection and blocking in communication networks leverages hybrid cloud integration, aggregator deployment, and SDN controllers to ensure efficient traffic management and fraud prevention. Key features like de-duplication, UDA, and correlation enhance traffic analysis, while addressing the challenges posed by 5G networks.

1.5 SDN Placement and Fraud Detection

The Software-Defined Network (SDN) controller is central to automating and programming the network. It provides flexibility, allowing service providers to work efficiently with different vendors and develop their hardware and feature requirements. The SDN controller manages WAN links and physical transmission/switching nodes, addressing critical demands such as bandwidth, performance, reliability, and automation.

Traditionally, the SDN controller is placed within the leaf switches on a physical level and above the routing layer on a logical level. For this study, the SDN controller is placed after the probing platform, enhancing its operational functionality. It sits on a virtual platform and scales according to traffic growth.

The research creates a model of the Flow Table Management using CPU and ASIC in section 3.3.3, the CPU learns network packets/flows and creates Flow Table entries and the ASIC exports flow records to the CPU. The type of Flow for the study are:

- New Flow: Permissible flow
- Allowed Flow: Uncertain flow
- Fraud Flow: Duplicate flow

Also, the research explores Service Application Fraud that involves critical services within a network domain, labelled as interfaces like S1-U and SGi (4G) or N3 and N6 (5G). The study focuses on SGi and N6 interfaces, capturing traffic between the firewall and switch. And Energy Abuse Fraud that involves unauthorized use or manipulation of energy services. From the protocol stack, SGi Interface is the connectivity between the Evolved Packet Core (EPC) and the data network, serving as a service gateway for deep packet inspection and policy-based service selection and the N6 Interface is the connectivity between the User Plane Function (UPF) and the data network or cloud.

The SDN controller's placement and functionality are crucial for effective fraud detection and network management. By centralizing control and leveraging virtual platforms, the SDN controller enhances network flexibility and programmability, addressing modern network challenges and enabling efficient fraud detection.

A summary on the types of Energy Fraud and Service Application Fraud discussed in section 3.4 are

- Direct Theft: Unauthorized consumption of power without proper meter measurement.
- Meter Tampering: Manipulating meter readings to show false consumption data.
- Billing Irregularities: Cyber-attacks altering account information during manual meter reading and billing processes.

Some proposed solution requirements are meter tampering alerts to detect and alert on tampering activities, complete automation to automate detection and response processes, remote control and blocking to enable remote intervention to block fraudulent activities, consumption analysis to analyse usage within the low-voltage grid and usage statistics to provide detailed statistics on energy consumption.

1.6 Communication Flow for Fraud Detection

The steps in summary of the flow process for fraud detection on the network session is subscriber session activation and data mapping; session information parsing and saving; probe receives and inspects IP packets; session information retrieval and caching; IP flow data writing and confirmation; monitoring for fraud and usage tracking and action messages for blocking fraudulent traffic.

The steps in summary of the flow process of the Microservice Architecture Components discussed in section 3.5 are:

- UI Service: Web application for management and analytics.
- Visualization Service: API access to reporting data.
- Monitoring Service: Tracks traffic usage and triggers blocking.
- Projector Service: Aggregates and augments data for storage.
- Blocking Service: Executes blocking actions based on monitoring triggers.
- Message Bus: Scalable messaging platform for component communication.
- OLAP DB: High-performance database for data ingestion and query processing.
- IMDB: High-performance storage for session information.
- SQL Database: Stores service configurations and user profiles.
- Probe: Processes real-time traffic and performs DPI.
- Subscriber Information Extractor: Extracts user information from control plane nodes.

And exploring the Blocking integration Protocols:

- RADIUS Protocol: Used for session information mapping at multiple sites.

- Diameter Protocol: More reliable, used for extracting session information and provisioning rules to PCRF.

The proposed solution for energy abuse fraud detection and blocking leverages SDN technology, DPI, and machine learning to automate and enhance fraud prevention. By integrating with PCRF/PCF and using robust protocols like Diameter, the solution ensures efficient and scalable fraud management in Smart Grid environments.

1.7 Compliance Specifications for Fraud Detection and Blocking Solution

To ensure the solution meets compliance standards and international specifications, the following RFC and 3GPP standards were reviewed and aligned with the design:

Diameter Protocols

- RFC 6733, RFC 3588: Diameter Base Protocol for authentication, authorization, and accounting services in 3G and 4G networks. (Fajardo et al., 2012)
- RFC 8506, RFC 4006: Diameter Credit-Control Application for real-time credit control, including quota management, balance checks, and price inquiries. (Bertz, Dolson and Lifshitz, 2019)
- RFC 7155, RFC 4005: Diameter Network Access Server Application for network access and forwarding authentication requests. (Zorn, 2014)

RADIUS Protocols

- RFC 3162: RADIUS and IPv6 for authorizing and authenticating network access. (Aboba, Zorn and Mitton, 2001)
- RFC 2866: RADIUS Accounting for handling accounting requests and responses. (Rigney, 2000)
- RFC 5176: Dynamic Authorization Extensions to RADIUS for supporting unsolicited messages like Disconnect and Change-of-Authorization (CoA) packets. (Chiba et al., 2008)
- RFC 2865: Remote Authentication Dial-In User Service for carrying authentication and authorization information. (Rigney et al., 2000)

HTTP Protocols

- RFC 1945: HTTP/1.0 for lightweight, stateless, object-oriented protocol. (Berners-Lee, Fielding and Frystyk, 1996)
- RFC 2616: HTTP/1.1 for more stringent requirements and MIME-like message formatting. (Fielding et al., 1999)
- RFC 7540: HTTP/2 for efficient network resource use and lower latency. (Belshe, Peon and Thomson, 2015)

Domain Name System (DNS)

- RFC 1034, RFC 1035: Concepts, facilities, and implementation of domain names. (Mockapetris, 1987a; Mockapetris, 1987b)
- RFC 8499, RFC 2308: DNS Terminology for naming schemes and distributed database representation. (Hoffman, Sullivan and Fujiwara, 2019; Andrews, 1998)

Transport Layer Security (TLS)

- RFC 2246: TLS Protocol Version 1.0 for communication protection. (Dierks and Allen, 1999)
- RFC 4346: TLS Protocol Version 1.1 for privacy and data integrity. (Dierks and Rescorla, 2006)
- RFC 5246: TLS Protocol Version 1.2 for improved cryptographic algorithm negotiation. (Dierks and Rescorla, 2008)
- RFC 8446: TLS Protocol Version 1.3 for enhanced security features and forward secrecy. (Rescorla, 2018)

Transmission Control Protocol (TCP)

- RFC 793: TCP for reliable host-to-host communication. (Postel, 1981)
- RFC 4960, RFC 2960, RFC 3309: Stream Control Transmission Protocol (SCTP) for reliable transport with multi-homing support. (Stewart, 2007; Stewart et al., 2000; Stone, Stewart and Otis, 2002)

3GPP Standards

- 3GPP 23.002: Network architecture for UTRAN and GERAN radio access technologies. (3GPP, 2025a)
- 3GPP 23.203: Policy and charging control architecture for flow-based charging and policy control. (3GPP, 2025b)
- 3GPP 23.401: GPRS enhancements for E-UTRAN access, covering roaming, mobility, policy control, and charging. (3GPP, 2025c)
- 3GPP 23.402: Architecture enhancements for non-3GPP accesses. (3GPP, 2025c)
- 3GPP 29.060: GPRS Tunneling Protocol (GTP) across Gn and Gp interfaces. (3GPP, 2025d)
- 3GPP 29.061: Interworking between PLMN and PDN. (3GPP, 2025e)
- 3GPP 29.210: Charging rule provisioning over the Gx interface. (3GPP, 2025f)
- 3GPP 29.212: Policy and Charging Control (PCC) reference points. (3GPP, 2025g)
- 3GPP 29.213: PCC signaling flows and QoS parameter mapping. (3GPP, 2025h)
- 3GPP 32.299: Charging management and Diameter charging applications. (3GPP, 2025i)

These standards ensure that the solution adheres to industry best practices and regulatory requirements, providing a robust framework for fraud detection and blocking in communication networks. The study emphasizes the need for intelligent, automated solutions to detect and block fraud in communication networks. These solutions should ensure network stability and compliance while minimizing manual intervention. SDN technology offers significant advantages in network management, flexibility, and efficiency. It simplifies the deployment of new services, reduces operational complexity, and enhances network performance. The integration of SDN with communication networks provides a robust framework for future network innovations.

Chapter 4 provides a detail study of the characteristics and solution scoping, Chapter 5 provides the design and Test Cases, and Chapter 6 provides the discussion on the Test Cases and results in comparison to other researchers.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction to Fraud Detection within Communication Network

Online services within the financial service market are increasing at a rapid rate, together with the rise in fraud cases, the traditional fraud detection process mentioned from Agomuo *et al.,* (2025) is not successful enough to handle the losses. The traditional fraud detection process is either a human driven task or partly automated system that requires an additional human intervention to perform detection from Afrin, Roksana, and Akram, (2025) and discussed by Indarjit, Balyan, and Adonis, (2025a). Traditional processes are governed by the following:

- Change management process
  To make a change in the communication network, an approval process needs to be followed by incorporating business units.

- Business Assurance
  To abide by the business compliance act when dealing with a possible fraudster.

- Network Detection process
  To build and run scripts to detect fraudsters based on mobile devices connected to the network at random intervals.

From research and current channels, financial payments are becoming more of a major need by making the process between the customer and the service easier, at the same time fraudsters are becoming more skilled in network abuse, causing millions in revenue loss for Service/Communication/Mobile Providers. Fraud detection is becoming a major high-ranked subject, however, the element of blocking the fraudster from the communication network is less studied, since the engineering of performing live blocking on a network becomes a point of discussion. This also drives the business case for compliance agreement and regulation standards.

The traditional fraud detection methods use many principle-based detection and human intervention features to predict fraud cases. However, the traditional approach has been around for many years and has many shortcomings that require human intervention and adds unnecessary complexity. Automation is an important study in networks and reduces the number of processes on the network device. Every network device has a capacity processing limit, the more capacity used will directly decrease the performance of the device. If the capacity processing limit is 50Gbps the network device becomes vulnerable once it exceeds 50Gbps. This proposed study will not affect the live network and focus on an smart solution that will extract user plane traffic and decipher traffic based on policies.

2.2.1 Fraud discussion on methods

With the rapid increase of fraud cases and network revenue loss, different methods are developed and designed to act on protecting the network, the methods listed in Table 2.1 are of previous researchers and their insight to minimize the impact of fraud. For this purpose of the literature review, the methods discussed are based on the key relation to the study depicting from their methodology and their performance to achieve detection.

The study correlates the methods to meet the requirement of fraud detection, also the proposed solution uses a exisitng technology of Software-Defined Network to produce better results and simplify the architecture by Indarjit, Balyan, and Adonis, (2025a). The question arises Can the listed methods solve evolved networks? Can these methods maintain their success in the next 10 years consisting of 5G networks?

Table 2.1: Fraud Detection methods

Method	Description	Section
Random sampling	It defined as a sampling technique in which each sample takes an equal probability of being admitted.	2.2.1
Graph Neural Networks (GNN)	To operate on graph-structured data at which a collection of nodes represents the relationship with the other.	2.2.1
Broad Learning System (BLS)	An extension to collate pattern recognition and detect network anomalies and intrusions.	2.2.2
Label Propagation community detection Algorithm (LPA)	The assignment of labels to unlabelled data points, to detect groups using network structure alone as its guide.	2.2.3
Rule-Based method	A rule-based system that applies human-designed rules to store, sort and manipulate the data.	2.2.3
Call Detail Record (CDR) analysis-based approach	The approach includes statistical analysis, Machine Learning, and data visualization to identify patterns and trends in the data of network interface traffic.	2.2.4
Vector Machine	Vector Machine provides a supervised learning model with associated learning algorithms that analyze data on classification.	2.2.5
Random Forest	A well-known Machine Learning algorithm which combines the output of multiple decision trees to reach a single result. It handles both classification and regression problems.	2.2.5

2.2.1 Random Sampling and Graph Neural Network

The growth of the network, which includes new customers, and new services plays an equivalent role in the increase in fraud cases. The introduction of a new service that could belong to a traditional communication network or cloud provider requires to be tested on many scenarios/test cases to ensure the subscriber pays correctly for their bundle of service or application usage. On the 3 keys, basic considerations are:

- Fraudster lives up to date with new services on the network,
- All subscribers have access to the network, including fraud users, and
- Detection requires to be in real-time.

Authors, Xuan *et al.*, (2018) use random sampling on a suspect of fraud on patterns to limit fraud cases in user interaction and social link. Gori, Monfardini and Scarselli, (2005) use Graph Neural Networks (GNN) to build a table of the topology of a network and to aggregate the subscribers characteristics. Park *et al.*, (2019) use a supervised neural network model to model the network nodes by Indarjit, Balyan, and Adonis, (2025a).

The issue identified on the user traffic at different locations which Tarmazakov and Silnov, (2018) provide a centralized main system to manage all sites and to enable up-to-date aggregation of data. Authors, Tarmazakov and Silnov, (2018) use the basic methods of detection, analysis of subscriber events, control of the best data routes, and analysis of subscriber behaviour profiles.

Figure 2.1 shows a simplified diagram of different service nodes that are used for network fraud detection.

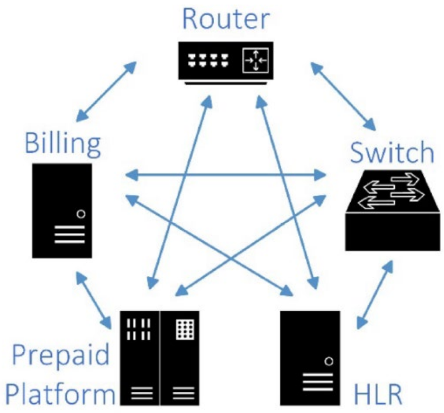


Figure 2.1: Equipment for mapping data in a communication network to detect phantom subscribers (Tarmazakov and Silnov, 2018, p.2)

The key components are the Billing platform and Home Location Register (HLR). For the purpose of this study, we will look deeply into these platforms and the role each plays in detecting fraud. One of the downfalls of the study from Figure 2.1 is the lack of knowledge of the communication physical connectivity in which more practical connectivity can be outlined.

2.2.2 Broad Learning System

While authors in Zhong *et al.*, (2019) explore a more precise and timely method of evolving fraud, proposing an identification method on phone calls based on broad learning than a less efficient and time-sensitive approach. The approach takes use of an in-depth algorithm for real-time prediction, the method shows increased results than traditional methods. Zhong *et al.*, (2019) proposed the Broad Learning System (BLS), the method of detection requires 15 seconds to monitor a call, in respect of additional features and enhancement nodes.

The executed code shows the process flow on data processing and text classifier model, both models are tidiest. In terms of text classifiers, a database of words/lines/tags needs to be constantly updated to ensure the efficiency of the model. The process mentioned requires human intervention.

2.2.3 Label propagation community detection algorithm and Rule-based method

Fraud upliftment schemes are evolving more by the night, Niu *et al.*, (2016) proposed a United Intelligent Scoring (UIS) algorithm for fraud detection pertaining to 3 merits of rules. Gonzalez (2005) explores fraud detection on 3rd Generation telecommunication standards. While, Peng and Lin, (2018) use a Label Propagation community detection Algorithm (LPA) to be able to generate a fraud gathering in the detection of fraud. Communication network companies from different regions have handover sites within their architecture to cross from one network to another or even one region to another.

Another method is the Rule-Based method in the fraud performance system consisting of establishing basic rules for the subscriber on profiling fraudulent sim cards, involving the analysis and monitoring of call behaviour patterns and their flows. This process required a full end-to-end understanding of the connectivity and based on the redundancy of the network for meaningful results.

2.2.4 Call Detail Record

Passive methods are partly man and system methods, the classification allows for Call Detail Record (CDR) analysis-based approach, audio analysis-based approach, and signaling data analysis-based approach. CDR analysis-based approach uses both content and occurrences of CDRs, unlike the rule-based method. CDR gathers voice, text messages, or data usage of a network. Shearer (2000) applies a CDR fraud methodology as a data preparation step followed by model building and evaluation; the data preparation includes data understanding and feature selection. Chapter 5 uses CDRs to perform the test case analysis.

Another line of fraud is collusive fraud, it's when a dedicated group of fraudsters target a critical service and threaten the communication service. Zhou *et al*., (2023) focus on the healthcare system, this type of fraud becomes difficult to identify since the fraudster's pattern/behaviour is alike a normal visit, which will require a technical expert's ability, and propose a 3-step model called Fraud Auditor.

In relation to work by Authors, Bindu, Mishra, and Thilagam, (2018) and Molloy *et al*., (2016) research a collusive fraud detection model that is divided into a statistics-based model and a Machine Learning (ML) model. The statistics-based model identifies abnormalities/positives by graphs and nodes while the Machine Learning model uses Graph Neutral Network (GNN) to identify fraud as its impactful for inspecting the traffic. Other related work by author Cao *et al*., (2015) shows a visual analytics approach that uses perspectives of user portraits, dramatic changes, and interpersonal events to detect fraud. The analogy of financial fraud can be used between a buyer and seller composed of nodes and edges.

From Zhou *et al*., (2023) the 3-step model by:

- Co-visit Network Understanding
    - Selecting patients of interest,
    - Initializing co-visit network
- Suspicious group identification
    - Filtering detection groups
    - Selecting suspicious groups
    - Refining suspicious groups
- Suspicious patients' examination
    - Redefining patients within a group
    - Investigating site behaviour
    - Reasoning and Annotating

2.2.5 Vector Machine and Random Forest

To bridge the study from author Smruthi (2019) and shows language processing to identify bogus information, the approach uses the process to analyse data using a variety of classification approaches from Probabilistic Context Free Grammar to detect bi-grams

(PCFG), and the model found at its best was able to detect 77.2% of non-trustworthy resources. From Facebook's dataset, it showed a precision of 74%.

More, Rao, Gyani, and Narsimha (2018) use a utilized ML algorithm and deploy the algorithm on application of Facebook, it detected at a rate of 81.7% accuracy in false detection based on imaging. Considering the layer 7 of the OSI model to show application inspection. The requirement for a suitable framework explores content material features/capabilities and consumer services from Twitter API, Sahoo *et al.*, (2022) use classification and analysis, fact mining algorithms, and upside-down photo searching and checking of news sources. Rather than exploring one means of detection to make a prediction, the system from Gurajala *et al.*, (2016) explores three separate methods:

- The Vector Machine
- Also, Random Forest, and
- Neural Networks

All mentioned finding positive results and prediction at a higher level. Other key approaches are phishing detection methods, blacklist and whitelist approaches, feature extraction, and sentiment analysis techniques. The researched method of feature-based classification paves a more promising way for detecting bots(machine learnt technology) and unauthorized users, also used on early identification within Software-Defined Network (SDN) and also found in big data networks. Figure 2.2 shows data from existing node relationships, based on the data distribution changes using the cross-validation technique, the raw data from 95% of common users and 1% of bogus users, changed to 74% of normal users and 28% of fraud identified.

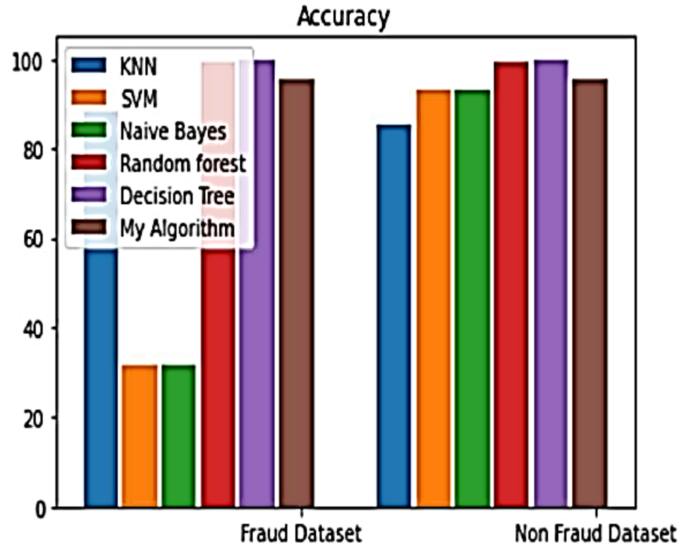


Figure 2.2: Accuracy analysis from existing methods

The impact of fraudster in energy fraud deeply affects the profits of a company, by the following examples:

- Fraudsters using a false identity and accessing the electrical energy network at zero payment or minimum of the tariff plan/pricing guide,
- Fraudster paid for a voucher for a lower energy service and the voucher exists for a higher purchase.
- Physical human theft tampering of meter, and calculation done incorrect readings leading to incorrect billing.

The requirement for a solution of application intelligence to provide assurance and return on investment to safeguard a network and its stability in an automated approach.

Summary of section

Random Sampling

Random sampling is used to select a representative subset of data from a larger dataset, ensuring that every element has an equal chance of being chosen. This helps in reducing bias and improving the accuracy of fraud detection models.
Example: In telecommunications, random sampling can be used to analyze call records to detect unusual patterns indicative of fraud, such as repeated calls to premium-rate numbers.

Graph Neural Networks (GNN)

GNNs are used to model and analyze complex relationships within graph-structured data, such as social networks or communication networks. They can identify patterns and anomalies that may indicate fraudulent activity.
Example: GNNs can be applied to detect fraud by analyzing the network of calls and messages between users, identifying suspicious clusters or connections that deviate from normal behavior.

Broad Learning System (BLS)

BLS is designed for efficient learning and classification, leveraging a flat network structure and incremental learning. It is particularly useful for handling large-scale data and detecting anomalies.
Example: BLS can be used to classify network intrusions and anomalies in telecommunications, such as detecting unusual data usage patterns that may indicate fraud.

Label Propagation Community Detection Algorithm (LPA)

LPA is used for community detection within networks, identifying groups of nodes that are densely connected. This helps in understanding the structure of the network and detecting fraudulent communities.
Example: In telecommunications, LPA can be used to detect fraud by identifying communities of users who frequently communicate with each other in suspicious ways, such as a group of users involved in SIM card cloning.

Rule-Based Method

Rule-based systems use human driven predefined rules to detect anomalies and fraud. These rules are based on expert knowledge and historical data.
Example: A rule-based system in telecommunications might flag accounts that make an unusually high number of international calls within a short period, indicating potential fraud.

Call Detail Record (CDR) for Analysis-Based Approach

CDR analysis involves examining detailed records of calls, messages, and data sessions to identify patterns and anomalies that may indicate fraud.
Example: Telecommunications companies use CDR analysis to detect fraud by identifying unusual call patterns, such as frequent calls to premium-rate numbers or calls made at odd hours.

Support Vector Machine (SVM)

SVMs are used for user traffic classification and regression tasks, effectively separating data into separate categories based on their features.
Example: In telecommunications, SVMs can be used to predict customer churn by analyzing usage patterns and identifying customers who are likely to switch providers due to fraudulent activities.

Random Forest

Random Forest is an well-known ensemble learning method that uses multiple decision trees to improve the accuracy and robustness of fraud detection models.
Example: Random Forest can be applied to detect fraud in telecommunications by analyzing various features of customer behavior, such as call duration, frequency, and location, to identify suspicious activities.

2.3 Impact of Fraud Case

2.3.1 Impact of Fraud on Communication Network – Revenue loss

The 2 main Service Provider revenue loss scenarios are accompanied by fraud and identity theft, implicating a huge financial loss to telecommunication companies, and losing the trust factor between the customer and Service Provider. Y. S *et al*., (2023) present a blockchain-based network that uses secure data storage, by using Inter Planetary File System (IPFS) on customer service reports. While telco providers introduce new cutting edge features and functions, whiles roaming fraud has an enormous impact on the network causing a loss of revenue discussed in Indarjit, Balyan, and Adonis, (2025a). The cause of roaming fraud occurs by the delay between the Visitor Public Mobile Network (VPMN) and Home Public Mobile Network (HPMN) where the subscriber is not charged correctly for extra service, also the subscriber may have forged his subscription by using another subscriber's access information. Blockchain is defined by author Ochôa *et al*., (2021) as a data structure where the data is stored in blocks in the form of transaction. A block contains 3 elements which are the data, the address of the current block, and the address of the next block which connects to the next to form a chain. The stages in roaming fraud detection are discussed in Maciá-fernández (2008) on the preventive stage and reactive stage which include the following:
Data collection stage,
Detection stage,
Supervision stage, and
Response stage.

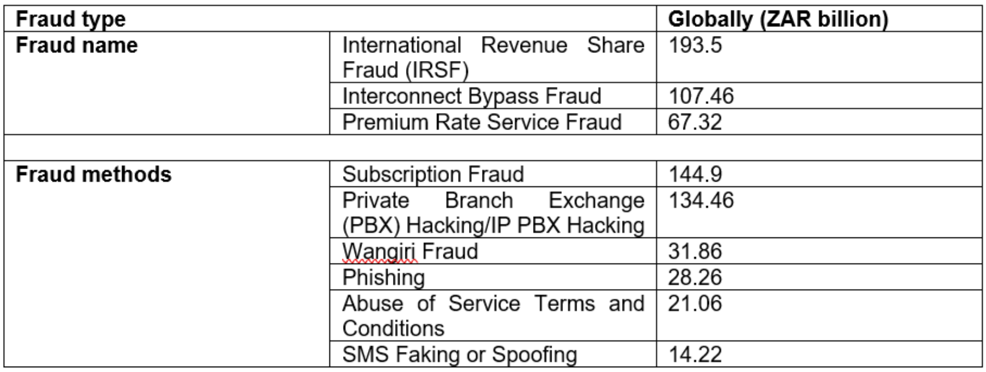
The different stages are applied to service restrictions in roaming, organizing detailed roaming tests at each step.

Roaming services can be improved with the use of features using smart contracts and private blockchains covered by Mark, Renier and Bailon, (2019). The mentioned approach used is a smart contract to solve the issue of the below interactions:

- To send a (carrier request) to the blockchain network, that adds higher processing, digital hash, and validation,
- To incorporate the user profile by using the existing contracts with the HPMN.

From Tarmazakov and Silnov, (2018) and reference to Table 2.2 showing the revenue loss in South African rand (ZAR) per year and an average of 7% of revenue over the year.

Table 2.2: Fraud type and revenue loss



Fraud type		Globally (ZAR billion)
Fraud name	International Revenue Share Fraud (IRSF)	193.5
	Interconnect Bypass Fraud	107.46
	Premium Rate Service Fraud	67.32
Fraud methods	Subscription Fraud	144.9
	Private Branch Exchange (PBX) Hacking/IP PBX Hacking	134.46
	Wangiri Fraud	31.86
	Phishing	28.26
	Abuse of Service Terms and Conditions	21.06
	SMS Faking or Spoofing	14.22

The amount of fraud results in a high loss of revenue, the type of fraud is explained below also discussed in Indarjit, Balyan, and Adonis, (2025a):

- International Revenue Share Fraud is levelled on fraudsters making unauthorized calls from one country in the globe to the next country.
- Interconnect Bypass Fraud, is when a fraudster is able to access the network and avoid the billing process.
- Premium Rate Service Fraud is when a fraudster abuses a premium number so that they can increase their profits they made from that particular tariff.

Another service that the mobile network provides is the online social networking platform, one to address is online engaging to find a potential partner. Application on layer 7 of the OSI model provides social platforms a risk and make it more at ease for fraudsters to take advantage of the platform. The impact of social platforms builds a false relationship and after engagement with the build of emotional feelings demands money from the victim. Bharne and Bhaladhare, (2022) provide a statical view of the impact of fraud and provide the challenges.

Currently, there are over 4 billion internet subscribers across the world and around 3.6 billion subscribers are on social media recorded in 2020, the forecast from Jain, Sahoo, and Kaubiyal, (2021) presents the increase to rise to 4.41 billion by 2025. From the paper, the different social applications explored are Tinder, Baddo, Bumble, and Facebook. The fraudster uses and exploits trust to gain the victim's emotions in building a false relationship. Figure 2.3 shows the number of reported cases from online dating platforms, in 2020, the top six dangerous scams were romance scams, and the victims lost a total over 304 million dollars. In 2018, the loss of 33 million dollars, presenting 9 times the increase over 5-year period from new FTC Data Show Consumers Reported Losing More Than $200 Million to Romance Scams in 2019 | Federal Trade Commission. (2019).

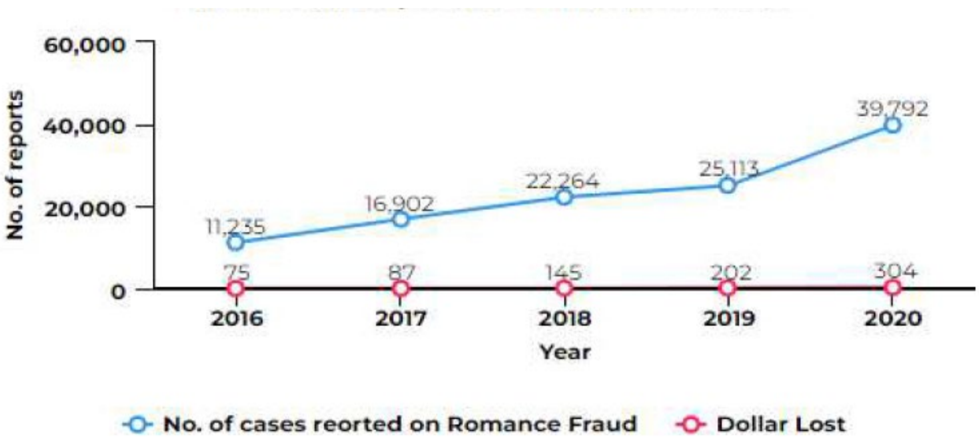


Figure 2.3: Number of reported fraud cases from Romance

2.3.2 Impact of Fraud on Communication Network – Detection Features

Table 2.3 shows the detection methods used, and disadvantage explored for the elaboration of the research also discussed from Indarjit, Balyan, and Adonis, (2025a).

Table 2.3: Detection Methods on Advantages and Disadvantages

Advantage	Methods Used	Disadvantage	Reference
To detect fake images online.	• Naïve bias, • Support Vector Machine (SVM).	The dataset is limited to a single dating website.	(De Jong, 2019); (Banerjee and Chua, 2023); (Kondamudi *et al*, 2023)
The detection is based on the description of the dating profile.	Support Vector Machine (SVM).	The limitation of dating website profiles.	(Suarez-Tangil *et al*., 2019); (Ellaky, Benabbou, and Ouahabi, 2023); (Liao *et al*., 2023)
To detect specific groups of celebrities.	Image scanner API	The limitation on what pertains to a dating website.	(Kahveci, (2019); (Meng *et al*., 2021); (Zhu, Hsu and Zhou, 2023)
Classify user-profiles and scan on celebrities.	Deep learning based on image recognition service.	The scammer is also on other websites.	(Al-Rousan *et al*., 2020); (Kumar *et al*., 2023); (Savchenko, Demochkin and Grechikhin, 2022)
Based on cybersexual interactions to encompass the victim on financial gains.	URL extraction	The platform can be used for underage minors.	(Pastrana *et al*., 2019); (Fuss, and Bőthe, 2022); (Ranney, 2021)

To highlight the challenges faced the authors Bharne and Bhaladhare (2022) provide the mentioned challenges,

- The network has millions of subscribers and also a percentage of criminal profiles behind the subscription. The identity of the fraudster is a main challenge to frame if it's real or a fake identity; it brings the challenge of identifying the fraudster at an early stage by looking into patterns. A good and effective system is a requirement for traffic identification.
- The key requirement for successful research and solution-driven implementation is on real-time raw data sets, unlike online dating platforms the information is confidential and sensitive, for example, a user's sexuality is posted on the dating site, but by law regulation, it's the platform's owner to keep the user's details within file and within process consent. There is also no pending procedure for how to conduct research on online dating sites.
- Another challenge among the many Machine Learning-based solutions is spam detection and false identities, it is within the minority and focuses on the reported dating fraud scammers, with the technicality of the websites and challenges to detect the fraudster.

Another paper from Anupriya *et al*., (2022) performs research on modern social platforms where users can showcase their hobbies, lifestyle, and image to the public, the user aims to gain many followers, not knowing which profile is real and which is not. Instagram is one of the popular applications that provides businesses and people to express themselves and their products. Which is at their fingertips for instance communication and forming partnerships. More above hackers are able to exploit users by making use of false accounts, and the impact of the damage becomes more problematic than any cybercrime.

2.3.3 Impact of Fraud on Communication Network – Fraud technology solutions

The impact of smart meters created power loss detection methods that show positive means, showing increase usage of the energy consumption time, however, study rarely investigates event logs using Machine Learning (ML) solutions. Acevedo *et al*., (2023) propose a method that considers the sequential nature of alarm logs using a recurrent neural network.

Smart meters are equipped with electrical mechanical sensors for detecting differential currents, it could be caused within the customer's home by current leakage, or by a bypass. The type of alarm will vary according to specifications. A differential current event begins when the current is higher than the set threshold. The threshold will define a current module to exceed a value and total consumption, if the restoration occurrence varies from start and end then there is a high possibility that the smart meter was manipulated.

The future of online trading is known on every platform, investors and potential buyers are communicated via Facebook, Instagram, and YouTube, these are public platforms that allow people to share information, the power of information also has its downfall, creating a platform for fraudulent attacks known as fraudulent trading. Online fraud accounts from a small number of transactions that shows imbalanced data. Teng *et al*., (2023) propose a Balance Generative Adversarial Network (GAN) that works by pretraining generated data and fine-tuning using transfer learning. In comparison to other techniques, Balance GAN performs 10% more in precision and recall. The fraudster acts on money transfers and withdrawals for money laundering, while the loss of the user's money, but the concern is on the identification of the fraudster mentioned by authors Zhao *et al*., (2022) and Cao *et al*., (2021). The challenges for the technique are:

- Oversampling – The method used to solve the problem of data imbalance by introducing kernel functions.
- Undersampling – Authors Sun *et al*., (2019) and Liu, Wu, and Zhou, (2008) use an ensemble learning mechanism known EasyEnsemble and BalanceCasade takes the negative samples and divides them into sets for different learners to use.
- Weighting – The method solves the data imbalance by weighting and redefining the loss function by parameters.
- GAN-Based – The GAN framework constructs different loss functions to make the data generation to attain better results by facilitating the training of models by Machine Learning.

The three-stage process consists of:

- Data generation
- Base detection model training
- Detection model enhancement

The increase in online financial transactions poses a challenge to big data analytics, to ensure safety during the online payment process. More, social network platforms are integrating with online platforms, with the existing trends of mobile evolution and rapid growth in new users, the velocity of big data can play a huge role in fraud detection of incidents. The lack of international standards to combat financial fraud provides a gap point for Singh, Alawami, and Kim, (2023) propose that domestic legislation to comply with international standards using Machine Learning in combination with banking legislation, to solve security and privacy governance difficulties over financial fraud. Figure 2.4 shows the different platforms ranked by the highest usage taken from (Singh, Alawami and Kim 2023, p.1).

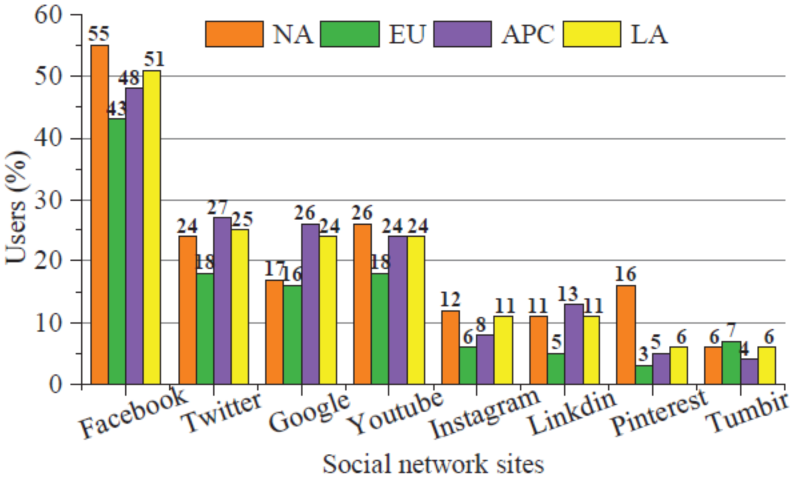


Figure 2.4: Social media platforms ranked by usage

(Singh, Alawami and Kim 2023, p.1)

Anomaly detection is to identify a deviation from normal patterns of a dataset and to help to find rare patterns, the approach can be applied to credit card faults and intrusion network detection. The idea of Isolation Forests is that anomalies are data points to isolate outliers in the data, the paper references the Isolation Forest for the detection of credit card fraud. Some of the use cases are listed in Table 2.4:

Table 2.4: Use cases of fraud environments

Use Case	**Description**	**Reference**
Finance	To identify any deviations from the trend of making a purchase based on traffic profiles, weekly usage and movement of cash.	(Ahmed, Mahmood, and Islam., 2016); (Shaohui *et al*., 2021); (Chang, Wang and Wang, 2022)
Healthcare	To identify false insurance payments and claims based on fake payments or payments bounce.	(Carvalho *et al*., 2015); (Georgakopoulos, Gallos and Plagianakos, 2020); (Ismail and Zeadally, 2021).
Manufacturing	To identify a prior failure on input and output parameters of machinery that behave abnormally by early detection.	(Susto *et al*., 2017); (Ding and Ming, 2019); (Boutaher *et al*., 2020)
Networking	The capability to detect network intrusion by monitoring unusual behaviour.	(Ahmed, Mahmood, and Hu., 2016); (Tekkali and Natarajan, 2023); (Zhao and Song, 2022)
Fraud	The fraud involves credit card and bank statements with the use of behavioural biometrics to identify irregularities in online purchases.	(Hilal *et al*., 2022); (Hussein, 2022); (Kataria and Nafis, 2019)
Medical Anomaly detection	The use of a density-based clustering method allows whether an aberrant care flow trace exists for a certain patient. The precise diagnosis is used to spot potentially fatal faults. A more detailed process is to create a standard list of	(Esmaeili *et al*., 2023); (Mohamed, Makhlouf and Fakhfakh, 2018); (Zachos *et al*., 2022)

	checks and sign off, by protecting the patient and also protecting the hospital.	
Image Processing	Used in banking sector and insurance industries to identify fraud by and deep anomaly detection system.	(Fontugne *et al*., 2008); (Mitra and Rao, 2021); (Vats and Tadepalli, 2022)
Isolation Forest	The approach uses Machine Learning called Isolation Forest, the use of outliers to search for abnormalities. The approach is built on a Decision Tree algorithm that operates with speed in detection. The Isolation forest is a ranking based on the path lengths of each data instance that reflects the level of an oddity.	(Xu *et al*., 2017); (Ghevariya *et al*., 2021); (Reddy and Kumar, 2022)

Authors Kouam, Viana, and Tchana, (2021) provide a comprehensive study of Simbox fraud strategies, fraud evolution, and fraud detection methods. A Simbox platform is made of three main components, the gateway, the simbank, and the control server. The control server is responsible for:

- Coordinates the system to establish connectivity to the network,
- Provides centralized destination administration of the Simbox architecture, and,
- The control server manages the VoIP traffic incoming into the architecture for traffic routing.

A Simbox fraudster invests a lot of money to gain feature sets that allow them to ease access of accessing networks, the main strategies mentioned by Okumbor, Anthony, and Olokunde (2019):

- Sim rotation – method for selecting the next SIM card,
- Sim activity limitations – a type of parameter limitation,
- Sim migration – method for selecting the next communication network,
- Base station switching/locking – method for selecting a base station,
- Changeable International Mobile Equipment Identity (IMEI) – method for changes in communication network module,
- Call forwarding – forwarding conditions.

Two methods exist for fraud prevention on countermeasures, active methods, and passive methods. Active methods require a permanent action by entities to mention:
Test Call Generation (TCG) – consists of setting up test phone numbers on a communication network to test numbers in different locations and countries through many different interconnect voice routes, TCG method worked more successfully in 2012 and 2013 from CSGi it became less significant since the fraudster figured out ways to avoid detection testing and the fraudster can allocate pools of Sim cards to be sacrificed, therefore allowing the detection by the mobile operating to be successful on the results.

2.3.4 Impact of Fraud Case on Smart Energy and Service Application

The internet has influenced consumers and organizations to implement digital services for ease of access to the application, from digital banking to international communication, and Smart Grid (SG) energy applications. Due to the high increase in usage of the communication network, manual intervention is less required. Introducing a higher risk of attacks on various applications. The study is applied to Smart Energy abuse and Service application abuse that will provide solutions for the detection and blocking of network abuse. The analysis of the user's behavioural characteristics and social characteristics plays a role in detection.
Lv *et al*., (2019) analyse graph topological structure to identify fraud cases and propose a method to learn expressive latent features for vertices, the method consists of multi-encoders to formulate the identification.

Araujo, Almeida, and Mello (2019) evaluated Brazilian Energy Distributors which result in a loss of billions of dollars from fraud and robbery, the first approach was to raise awareness among the public, but the effect was not fruitful and called for a more intelligent plan, Araujo, Almeida, and Mello (2019) took a more proactive approach to conduct technical inspections at each user to provide identity evidence.

Detecting fraud within the electrical system domain is a significant problem and is an active area of research, the key to understanding patterns of power usage is fundamental to the security goal of automation within the energy sector. Ganguly *et al*., (2022) define energy theft as including data corruption from cyber-attacks, device tampering, and bypassing meter readings. Further to add, energy fraud plays a role in accessing the network via an application, without correct billing and the use of false identification.

Also, Korba and Karabadji (2019) review new vectors of cyber risks and energy fraud, proposing a Machine Learning-based solution to detect fraud and abuse of a network. The solution takes advantage of the predictability of the user consumption profile, the performance evaluation on the historical consumption dataset showed a high detection rate. Liu, Hu, and Ho, (2014) explores an energy fraud scenario, where a fraudster attempts to misuse the billing system by using a lower cost of his electricity usage to other customers. The solution used a detection approach based on SVR to detect pricing electricity fraud of altered pricing curves.

Further, Jokar, Arianpoo, and Leung (2013) use a classification model that makes use of transformer meters with Machine Learning methods to improve the robustness of non-malicious detection.

To understand and tracking the power usage of a grid is a good lead for automation within the energy sector, energy theft is the following scenarios:

- Cyber-attack – the effect to enter the network and create a state of network panic,
- Device tampering – the ability to configure and disable billing functionality,
- Bypass meter – the act of utilizing a service by authorized and illegal system setup or alterations.

Authors in Park *et al*., (2019) perform a study on power usage using supervised and unsupervised techniques. This becomes an active form of research labelled and proposed a clustering algorithm that presents greater than 70% accuracy levels. The study uses a datasheet from a power company to analyse residential smart meters, performing the counts, in which the data showed 1,268 incorrect readings in Figure 2.5 of the distribution of incorrect counts.

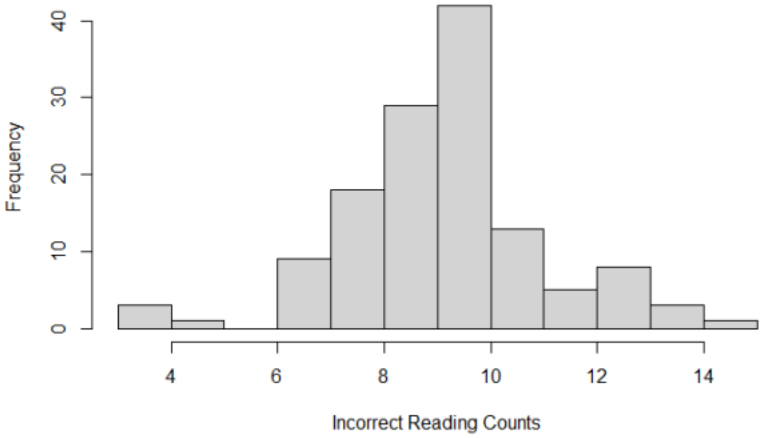


Figure 2.5: Distribution of incorrect reading counts

(Park *et al*., 2019)

The main task is to apply a cluster approach to energy readings to show power usage, the dimension assists with the interpretation and provides a hierarchical structure that indicates the source of the anomaly. When evaluating a Service Provider's network on fraud, the impact affects a cluster or many clusters of meters when dealing with threats and taking appropriate corrective measures. The clustering approach is created by building mathematical models to capture a pattern in the data and to account for hacking schemes.

2.3.5 Hybrid Cloud Network

The market of cloud computing and integration becomes more competitive and promotes the evolution of the Internet Protocol (IP) network to adapt in terms of flexibility and connectivity. Current trends make use of hybrid clouds to achieve lower cost, higher availability and increase security. Cloud providers will gain benefits in elasticity, efficiency, and flexibility. Previous researchers focus on single clouds or clouds in isolation which lack efficient management. The study will provide a comprehensive view of the hybrid cloud for fraud detection solutions.

Authors, Xu, Su, and Zhang (2018) provide a study on the effective way for cloud management, cost reduction, and improved flexibility by combining cloud management and a rule engine. Still, challenges arise, to state the functionality becomes limited, Duplyakin, Haney, and Tufo (2015) showcase a persistent and reliable configuration management system that shows limited functionality. Another challenge is the difficulty to migrate from one cloud to another, Xu, Su, and Zhang (2018) paper explores overcoming these challenges by providing a rule model defining rules and triggers, using Java reflection to interpret and translate the rule files. The model firstly is genetic with greater ease of use and the model supports more Infrastructure as a Service (IaaS) due to efficient architecture planning.

The cloud landscape provides the ability for traffic bursting, disaster recovery, migration, and monitoring, however, the challenge is the setup of a hybrid cloud by Sitaram *et al.*, (2018). The main limitations behind the scene are vendor lock, migration issues, and interoperability standardization. Authors in Sitaram *et al.*, (2018) explore the concept of hybrid clouds built on OpenStack.

2.4 Challenges faced with Cloud network

Some of the challenges faced are mentioned:

- Data backup and recovery
  Both feature service of backing data and system recovery serve of importance when handling customer data, the action required is dedicated pipelines to allow the service, which becomes very expensive. Hybrid cloud share resources and provide rapid scaling on a flexible platform by OpenStack Hybrid Cloud for Interoperability (2018).

- Cloud bursting and auto-scaling
  The action of cloud bursting becomes effective to utilize more cloud resources, due to higher usage of the platform and return resources on lower usage. The setup is complex and vendor lock creates many security challenges. Currently, there is no solution for hybrid scaling by Real Use cases: Why 50% of the enterprise are choosing hybrid cloud.
- Migration of services
  With new development in technology, a provider needs to consider new solutions and gain the best out of their investment, the challenge that arises is the difficulty of migrating a service from one platform to another. In some domains migration of data cannot be done since the element of vendor-specific.

- Resource monitoring and management

  Most cloud providers have their web-based propriety application, with a single sign-on. However, there is none for hybrid clouds and also no standardization between cloud providers. Which raises the challenge of redefining the framework between cloud and mobile providers.

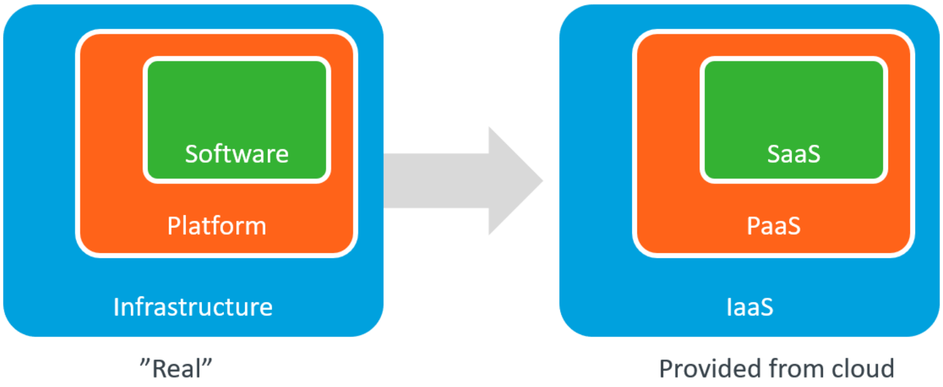
References on Cloud network:
Google Cloud, 2025. Backup and Disaster Recovery. Available at: https://cloud.google.com/backup-disaster-recovery/docs/concepts/backup-dr

DigitalOcean, 2025. Cloud Bursting: A Strategy for Handling Spikes. Available at: https://www.digitalocean.com/resources/articles/cloud-bursting

Oracle, 2023. What Is Cloud Migration? Importance, Benefits, and Strategy. Available at: https://www.oracle.com/cloud/cloud-migration

IBM, 2025. What is Cloud Monitoring? Available at: https://www.ibm.com/think/topics/cloud-monitoring

Figure 2.6 shows the different services offered by cloud platforms, the research is to compare different cloud providers and how to bring two cloud providers to collaboration and segmentation.



Fiaure 2.6: x as a Service on Infrastructure. Platform. Software

2.4.1 Software as a Service

Software as a Service is a distribution model in a cloud provider’s network to host applications and make them easy use to for the user over the internet, providing flexibility to the application user,

- Software is used remotely from the cloud.
- The customer does not have to focus on software maintenance.
- The service provider takes care of everything,
    - Setup
    - Updates
    - Accessibility
    - Capacity and scalability
- Examples
    - Online document converters
    - Online databases (Google, Bing, etc.)
    - Web-based email (Gmail, Yahoo, Outlook.com, etc.)
    - Web-based social media (Facebook, Twitter, etc.)
    - Enterprise support applications (accounting, invoicing, etc.)

2.4.2 Platform as a Service

A complete cloud environment to provide a build, run, and management of the service application.

- Aimed for customers to develop and offer their applications.
- Popular in providing documents and media over the web.
- May include a comprehensive set of easy-to-use building blocks,
    - Programming languages
    - Management tools
    - Quick time-to-market
- Examples
    - Microsoft Azure
    - RedHat Openshift
    - Amazon AWS+
    - Cloud Foundry

2.4.3 Infrastructure as a Service

A cloud service that offers full computing, storage, and networking resources when required and on pay-as-you-go service.

- The most comprehensive cloud service.
- Consists of a complete service-providing network with all components,
    - Network interfaces and links
    - Virtualized CPUs

    - Storage
    - Benefits from usual cloud features
    - Capacity management and scalability
    - Low initial cost
    - ”Pay as you grow”
    - Independence from a physical location
- Examples
    - Amazon EC2
    - Microsoft Azure

2.5 Software-Defined Network

Software-Defined Network (SDN) defines a new sphere in technology and highlights a new mindset. The design brings a new network framework, from authors Ruaro, Caimi and Moraes, (2020) and from Indarjit, Adonis, and Brandt, (2022), the research defines a way of managing network platforms by a centralized control system and creating an open, standardized way of deploying network communication services. Network Monitoring require a visible network to provide network services and troubleshoot to avoid new project costs and timeous rollouts. Without visibility, how does one plan the network. SDN is a cutting-edge technology that caters to self-managing network capacities through policy-driven rules, self-managing capacity availability, and thus avoiding unnecessary expansion investments.

The primary concept to deploy SDN is for communication network automation and to cater to network programmability discussed from thesis by authors in Indarjit, Adonis, and Brandt, (2022). The research is to use SDN on a horizontal level with the network infrastructure layer. Looking into centralized control plays a vital role in servicing updated tables of links, interfaces, nodes, application slices, and configurations. Other key features are mentioned below:

- The topology consists of open standard interfaces to accommodate third party Smart Grid/Service applications.
- The scope is on similar architecture standards across fixed, mobile, cloud, enterprise, and security domains.
- The controller defines centralized control of automation and programmability.
- The feature caters to the separation of the control plane from the data plane infrastructure or virtual infrastructure.

SDN is a technology that creates a new deployment and an implementation which drives a new approach. Authors in Ruaro *et al*., (2018), and Berestizshevsky *et al*., (2017) provide accredited research on centralized control. Authors Velloso *et al*., (2019) and Cong, Wen & Zhiying (2014) proposed the use of SDN that improve the scalability of networks in a straightforward method. Authors Berestizshevsky *et al*., (2017) and Sandoval-Arechiga *et al*., (2015) provide an off the shelve SDN model without the specification of a standardized approach. The focus is placed on the advantages and disadvantages of the SDN proposed paradigm.

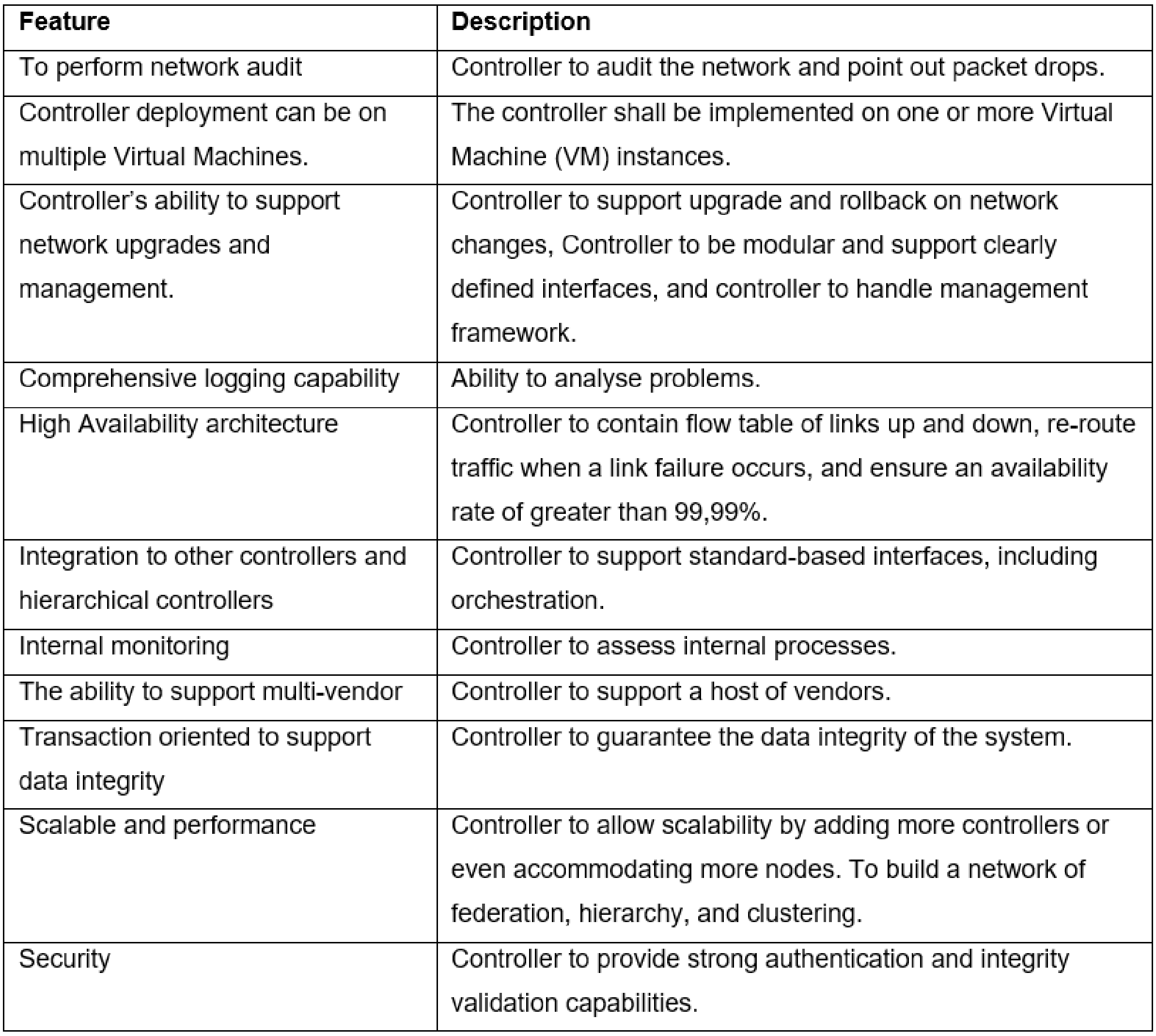
On the other front, Scionti, Mazumdar & Portero (2018) explain power-saving techniques by shutting down links, not in use. The paper addresses concern on the focus placed on the SDN management of the network from monitoring the status of each element, routing tables, and connections. This study will use SDN on the same layer for Performance Network Visibility using a vertical approach.

On SDN study propagates to present a new architecture and prove the ability of smart intelligence from the SDN controller and the mediation layer used between two different environments namely the communication network and Smart/Energy Grid. The research will expand on Ellinidou *et al*., (2019) using chiplet engineered design, performing software compilation of execution using fewer exchange messages between the SDN controller and switching environments.

2.5.1 Integration of Communication network and SDN controller

Table 2.5 shows the features of the SDN controller from Indarjit, Adonis, and Brandt, (2022).

Table 2.5: Required feature for SDN controller (Dinh & Park, 2021)



Feature	Description
To perform network audit	Controller to audit the network and point out packet drops.
Controller deployment can be on multiple Virtual Machines.	The controller shall be implemented on one or more Virtual Machine (VM) instances.
Controller's ability to support network upgrades and management.	Controller to support upgrade and rollback on network changes, Controller to be modular and support clearly defined interfaces, and controller to handle management framework.
Comprehensive logging capability	Ability to analyse problems.
High Availability architecture	Controller to contain flow table of links up and down, re-route traffic when a link failure occurs, and ensure an availability rate of greater than 99,99%.
Integration to other controllers and hierarchical controllers	Controller to support standard-based interfaces, including orchestration.
Internal monitoring	Controller to assess internal processes.
The ability to support multi-vendor	Controller to support a host of vendors.
Transaction oriented to support data integrity	Controller to guarantee the data integrity of the system.
Scalable and performance	Controller to allow scalability by adding more controllers or even accommodating more nodes. To build a network of federation, hierarchy, and clustering.
Security	Controller to provide strong authentication and integrity validation capabilities.

Authors Nugroho, Dian, and Setyawan (2017) take a practical view to compare the performance of the Open Shortest Path First (OSPF) protocol and to correlate with SDN technology. The analysis used virtualized GNS3 tool to measure the SDN performances. From the analysis of SDN, the results generated present a delay range of 0,3 ms to 6 ms and 0 % packet loss resulting in SDN performance is greater than traditional networks.

The layer 3 Internet Protocol (IP) network infrastructure uses a low-level configuration and unique syntax for each vendor. A network engineer will require a mix of skills to manage network layer 2 and 3 devices, adding a level of management complexity to the network. The centralization and use of a SDN controller or multiple SDN Controllers within a communication network become the Lego blocks for automation and network programmability on an SDN network.

SDN is a technology to cater on increased network expansion when building and catering for network growth, giving the Service Provider more of a upper hand with different vendors and allowing the company to develop its hardware and feature requirements.

The topology of *Software-Defined Network*ing is shown in Figure 2.7, some of the aspects of SDN from www.opennetworking.org (2020) are listed below:

- To separate the user plane and control plane.
- The practice of standardized interfaces to be able to program network devices.
- Auto-discovery feature allowing devices on the network to be polled for management purposes.
- To build a virtual platform.

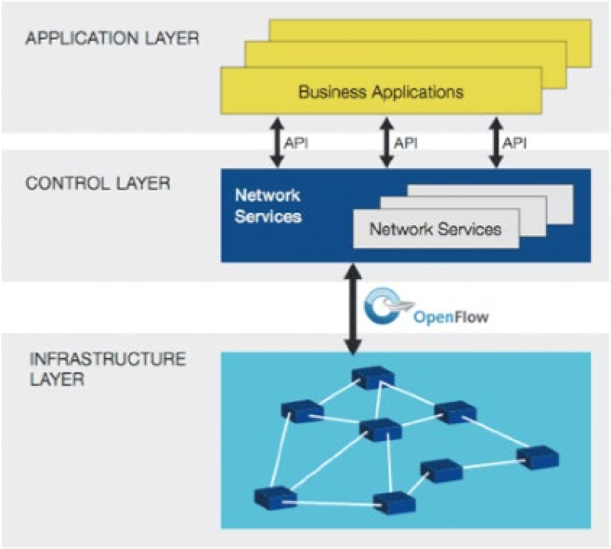
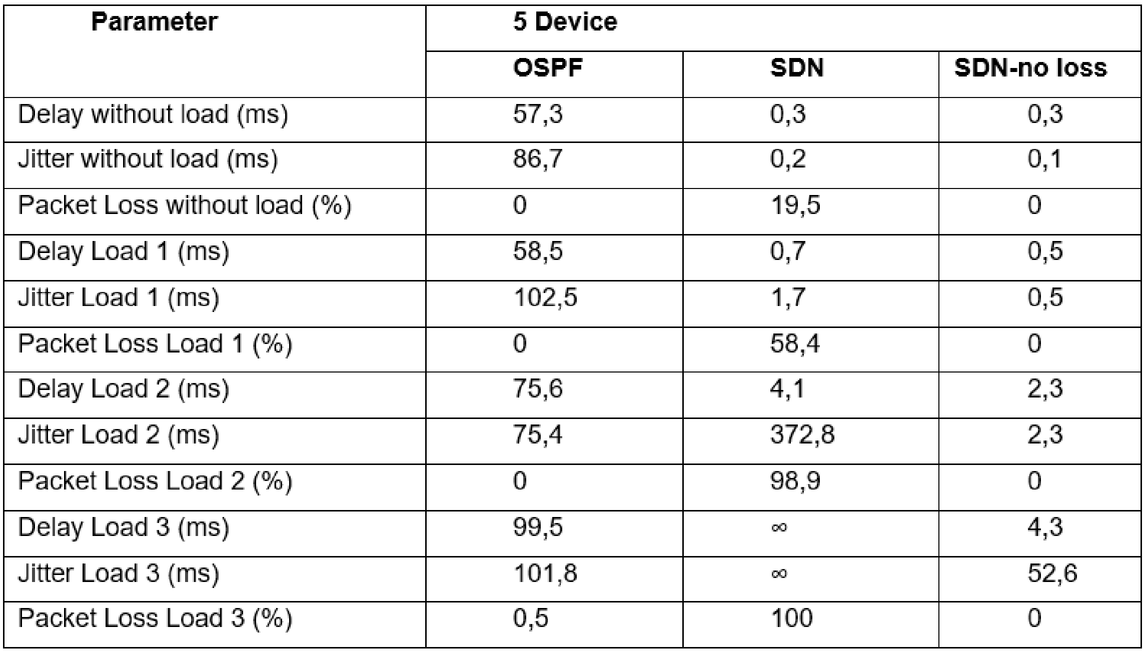


Figure 2.7 SDN Architecture (Open Networking Foundation, 2020)

Table 2.6 is the comparison of routing technology assessment on five devices taken from Nugroho, Dian & Setyawan (2017) shows the test results,

- The first column is Open Shortest Path First (OSPF), with OSPF cost metric configured on each interface within the network,
- The second, using SDN on condition of loss within each link, and the third,
- Using SDN on condition of no loss calculated.

Table 2.6: Comparison of routing technology assessment based on jitter and packet loss on five devices (Nugroho, Dian & Setyawan, 2017)



Parameter	5 Device		
	OSPF	SDN	SDN-no loss
Delay without load (ms)	57,3	0,3	0,3
Jitter without load (ms)	86,7	0,2	0,1
Packet Loss without load (%)	0	19,5	0
Delay Load 1 (ms)	58,5	0,7	0,5
Jitter Load 1 (ms)	102,5	1,7	0,5
Packet Loss Load 1 (%)	0	58,4	0
Delay Load 2 (ms)	75,6	4,1	2,3
Jitter Load 2 (ms)	75,4	372,8	2,3
Packet Loss Load 2 (%)	0	98,9	0
Delay Load 3 (ms)	99,5	∞	4,3
Jitter Load 3 (ms)	101,8	∞	52,6
Packet Loss Load 3 (%)	0,5	100	0

The results from Table 2.6 indicate a simulated relation for each delay parameter, jitter, and packet loss are in relation to each other. For all architectures, it indicates that SDN does not have any loss. The results show delay and jitter values, which are in the acceptable category according to ETSI standards. The results have proven that SDN condition without loss has comparable results and is suitable for packet delivery.

The key advantage of SDN technology is its practicality in building a network topology from Indarjit, Adonis, and Brandt, (2022). Unlike OSPF, which uses link-state principles and can rearrange traffic paths in terms of link failure, SDN looks at a topology that is not connected entirely between platforms. Further to the state, SDN is well known in Quality of Service (QoS) parameter results when compared to OSPF network (Nugroho, Dian & Setyawan, 2017).

Srikanth *et al*., (2018) research Software-Defined Networking, enabling the user to program the network nodes. With every hour increasing network traffic and vital congestion problems, the need for SDN becomes more pressing. The solution uses the least distance or lowest latency path algorithm from the SDN framework. Congestion on the network amounts to over-subscription of packets within a specific location due to the lack of capacity resources. The effect of congestion results in the subscriber having a negative experience on the network. A network node consists of a user plane and control plane. The user plane serves to route user traffic, while the control plane takes care of routing functionality.

To assist with the capacity processing challenges, the SDN controller requires to be re-designed to support the external environment and the outsourcing of packet processing. The re-design phase of the SDN controller is to allocate and divide into many sets of services instead of a centralized base. And allowing developers to select an arbitrary programming language instead of a forced programming language.

2.5.2 Summary of Software-Defined Network Challenges

The model of SDN serves to address many issues listed by authors in Yan *et al*., (2016) below:

- Layer 2/3/4 configuration is prone to human error, for a network architect to plan the network based on protocols by Open Shortest Path First (OSPF), consisting of a semi-automated traffic balancing scheme, which becomes exhausting to plan on scalability.
- New communication network services take weeks to deploy and many departments from a Service Provider consisting of Radio access networks, Core data networks, Transmission, and other departments that are required to implement a specific service, SDN serves to decrease the amount of time taken to arhitect and provide a single performance view.
- A wide variety of management platforms within the network management group leads to segmented skills and an uncentralized platform.
- Subscriber traffic recovery during a failure of a network, will require configuration change management process which takes hours and keep the network in low recovery states.
- Designing a product and vendor differences create a hybrid network design, and it becomes challenging to plan and manage.
- A wide variety of skills are required for each platform for each independent node. SDN offers further network improvements and becomes the platform to design and implement a scalable and efficient communication network.

2.5.3 Summary of Software-Defined Network Benefits

- The efficient use of resources. The SDN network contains an SDN controller that contains the characteristics of the network’s interfaces, links and nodes. The SDN network can detect high traffic and load balance subscriber traffic.
- The concept moves from costly proprietary equipment to cheaper and high-speed off-the-shelve hardware.
- Providing better network performance visibility on network (states) so that capacity may be used more efficiently and distributed.
- Reduce network complexity and operational overhead.
- Allows Service Providers to add new services for a new revenue source. To accommodate growing providers and their requirements.

- SDN allows feature implementation and increased deployment time.
- Partitioning of resources for safe experimentation.
- Implementation of well-known ease platforms/systems.
- Simplified operations, programming, etc., with centralized control.
- New possibilities
- Data plane/control plane that is decoupled and allows new decision algorithms creation.
- Enable application-level development of the network and systems.
- Vendor flexibility.
- Selection of requirements from different vendors serves to add innovation by giving the Service Provider more control of their technology strategy/roadmap.

2.6 Fraud case detection

Network Fraud detection is broadly studied across the world, to design detection methods and mitigate fraud occurrence, the challenge arises with old methods that cannot solve developing and growing systems. The crucial requirement for a solution that serves technology automation and flexibility robustness by Machine Learning and deep learning in fields such as healthcare, entertainment, and agriculture from authors Singh *et al*., (2021).

Table 2.7 shows the methods of the classification process used to detect fraud by Ghosh *et al*., (2023).

Table 2.7: Detection Methods and Description

Classification Method	**Description**	**Limitation**	**References**
XGBoost	The method uses a gradient-boosting framework to build full-scale decision trees and implement parallel decision trees.	Trends to overfit the data.	(Sheng and Yu, 2022); (Bao, 2020)
ADABoost	The technique handles binary classification problems and improves predictability by the conversion of a larger number of weak learners to strong learners.	The requirement of a dataset devoid of most of the noise.	(Yulita *et al*., 2021); (Chang and Fan, 2019)
Naive Bayes	The method is based on Bayes' theorem to predict the outcome by probability of occurrence.	The downfall it faces by zero-frequency issue, which is the missing variable as zero.	(Vijay and Verma, 2023); (Hairani *et al*., 2021)
Decision Tree Classifier	The algorithm uses classification and regression problems, it works	High in computation of resources and	(Zulfikar, Gerhana and Rahmania,

	on a tree-based structure which as the classifier of the dataset.	changes in data affect the outcome.	2018); (Indumathi, Ramalakshmi, and Ajith, 2021)
Random-Forest classifier	The technique allows aggregation of several Decision Tree classifiers, to improve the predictive capability of the algorithms.	The issue on the massive amount of computation of resources and the time required between periods.	(Mishra, Mallick and Gadanayak, 2020); (Lu *et al.*, 2019)
KNN	The method is used for classification on a distance-based approach to locate all unknown data points.	The downfall does not work on high dimensionality or large records.	(Lu, Tong, and Chen, 2015); (Altay, 2022)
Logistic Regression	The techniques is used for dichotomous and dependent variables.	The issue is on overfitting the count of features and recorded observations.	(Doss and Gunasekaran, 2023); (Bheemesh and Deepa, 2023)

Authors, Strelcenia and Prakoonwit (2023) explore a balanced class distribution making use of the undersampling and oversampling approach, by removing samples from the majority class to create a uniform dataset. The proposed method K-CGAN which is based on Conditional GAN architecture with the custom loss function Kilberg.

Figure 2.8 shows the imbalance data set of credit card transactions into training test data sets, which include the Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic (ADASYN), and novel K-C Generative Adversarial Networks (GAN) with XGBoost, Random Forest, Multilayer Perceptron (MLP), and Logistic Regression (LR). Figure 2.8 presents undersampling techniques followed by oversampling and balanced datasets, the classifier is used in training and testing datasets to identify fraudulent transactions.

One of the issues with credit card fraud, is the banking operator cannot disclose user information and is bound by security, sensitivity, and privacy. So the data attain is of public knowledge.

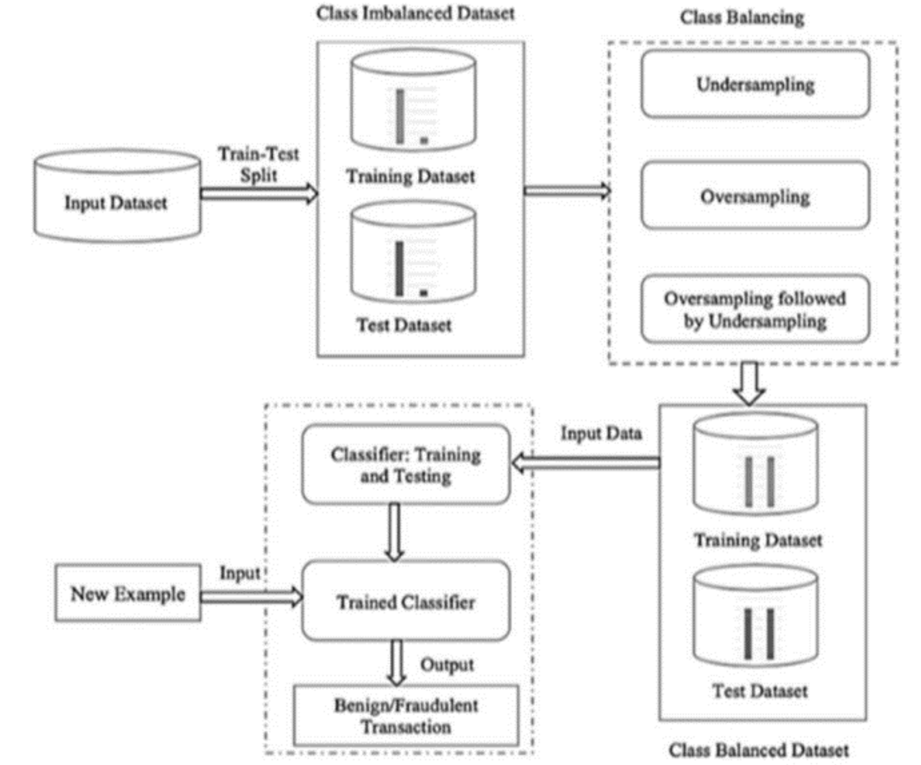


Figure 2.8: Imbalance issue
(Singh, Ranjan and Tiwari, 2021)

Another area of fraud detection on cybersecurity and defense technology aimed towards unwanted advertisements on the internet, the proposed study by Nguyen and Bein (2023) serves to build Machine Learning models on classification and prediction of data for the end user. The study builds trend detection on false positive advertisements. Authors Zeng, Kohno, and Roesner (2021) perform a study of bad advertisements by the University of Washington to show unfamiliarity with cybersecurity best practices and trust toward brands. One of the questions that arose is to what extent have cyberattacks evolved concerning our ability to detect them?

The combination of three methodologies where used to mitigate potentially malicious advent URLs, the requirements to meet a process of data collection to be automated, the feature of importance to trend the data to be observed, and the Machine Learning models have to be accurate.

- Data Science Aspect

  The requirement for a web scraper to scrape potentially malicious advertisements on the Google engine, using BeautifulSoul, the extraction of URL, company, title, and product description to a CSV file for data analysis.

- Machine Learning-Focused Aspect
  To employ an end-to-end Scikit-learn workflow using,
  - Attaining the data for cross-validation based on 70-30% train/test/ split,
  - Use Machine Learning algorithms,
  - Handling of NaN and categorical data,
  - Integrating the model to make predictions,
  - Evaluate the model, and
  - Display the results using graphs.
- Web Development Aspect
  To form a web application that contains keywords and the number of times to scrape into the Machine Learning model to avoid unbiased samples.

Energy theft is rising and going undetected for years causing huge loss of revenue to the electricity provider and to the government, the use of Smart meters contributes to assisting detection by records and time series.

Authors Ishkov, Terekhov, and Myshenkov (2023) explore the following techniques to harvest results,

- Exploring Decision paths, Random Forest on Attention maps is used under classification Machine Learning algorithms, based on decision trees to obtain forecast, the randomness is by selecting features and bootstrapping of training examples.
- U-net architecture is used to solve the segmentation problem, which contains an encoder and decoder. The task uses a binary mask with values ranging from 0 to 1.
- Evaluation Metrics are used to evaluate the quality of classifier models and include metrics such as accuracy, precision, recall, specificity, fi-measure, and Matthew's correlation coefficient. The classification metrics are divided into discrete forecast or binary classification.

A service is defined by the specification that is deployed to serve the system by well-defined Key Performance Indicators (KPIs), when evaluating the online platforms that serve investment and lending services, the key to predicting increasing fraud is groups among evolving platforms and developing systems. Wang *et al.*, (2023) propose Adaptive Connected Component Embedding Simplification Scheme (ACCESS) to predict fraud in groups, also proposing Recovering, Mining, Clustering, and Predicting (RMCP). In general, the loan application is a long process on whether the applicant can repay on time. The risk is built on whether the applicant or borrower is a possible gang fraud case. The contribution from authors Wang *et al.*, (2023) explored the following:

- The enhancement of the utility of the association of knowledge graphs to tackle online lending gang fraud,
- A Chinese address disambiguation to recover novel association representation methods together with graph clustering algorithms,
- The framework RMCP on a current online lending dataset to view the visual clustering performance and dynamic statistics.

Karthikeyan, Govindarajan, and Vijayakumar (2023) propose an unsupervised learning method called Competitive Swarm Optimization Deep Convolutional Neural Network (CSO-DCNN), by comparing inputs to outputs, working on real-time datasets. The results show an accuracy of 98,2% for credit cards, 99,77% for insurance, and 95,23% for mortgage data sets. The success of CSO resolves the optimization model by two parts in each iteration and bilateral competitiveness. Figure 2.9 shows the flow process of CSO-DCNN, the advantage of the method offers a higher level of accuracy on the detection of fraudulent use of credit cards and lower the risk of false internet payments. On the DCNN the learning rate is used to classify the best features by feature extraction and classification phases. In addition, the convolution applies multiple filters to pass over the data input by component multiplication approach.

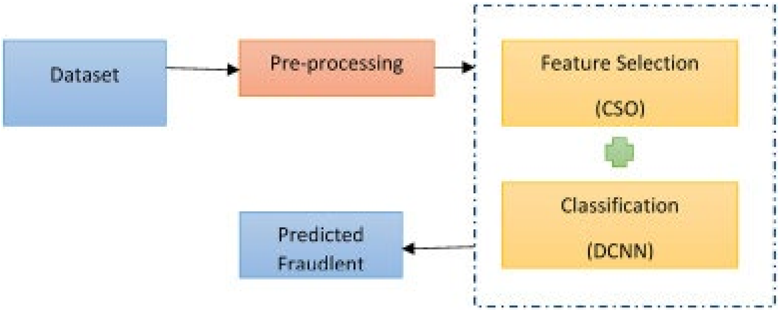


Figure 2.9: Flow process of Competitive Swarm Optimization Deep Convolutional Neural Network

Mobile network fraud consists of identifying a small number of fraud calls from a vast pool of traffic. Developing an effective and seamless approach to mitigate fraud has become challenging. Existing methods work on batch processing and real-time data for detection, to solve the pressing issue Wahid *et al.*, (2023) propose a model using Neural Factorization Autoencoder (NFA) is used to analyse customer calling trends and patterns in detecting fraudulent calls. The key contributions are the following:

- Real-time telephony fraud detection, to develop a system based on subscriber call patterns,
- Fraud pattern, to combine Neural Factorization Machine (NFM) and the Autoencoder (AE) to model,
- Memory module, to track fraud trends and patterns and update using First In First Out (FIFO) policy,
- Effectiveness, part of the validation exercise, to perform extensive experiments on raw CDR datasets.

To identify call patterns, by profile-based containing a database of attacks and anomaly detection uses each call's pattern as the baseline in detection and comparison by authors Burge and Shawe-Taylor (1997) and Fawcett and Provost (1997).

CHAPTER THREE
ARCHITECTURE AND CONNECTIVITY

3.1 Architecture and Connectivity

The research looks at three possible architectures that can be used to deploy the solution, within each architecture, pros and cons are looked at and the most suitable architecture is used in Figure 3.3 for the research and in publication from authors Indarjit, Balyan, and Adonis, (2025b). The selected architecture was based on number of equipment, cost, sending traffic vias the core network and real-world deployment. The first architecture was the Blocker component which were the first component of the flow, the disadvantage of the Blocker-First is on the need for pre-defined policies and no automation. The second architecture consisted of Interconnect servers which adds a bulk of cost to process the user plane traffic, exploring this long term is not feasible to perform expansions.

The key components to build the network:

- On layer 3 Routing and layer 2 switching – the infrastructure to route network traffic from source to destination,
- Network Performance Visibility of Tapping, Aggregation, and Probing – the platform to take a mirror of the traffic, apply filters, and stream the map,
- Evolved Packet Core – the infrastructure to handle subscriber access, authorization, and service fulfillment.

Figure 3.1 shows the Blocker First Architecture as covered in Indarjit, Balyan, and Adonis, (2025b), with the main component elements and the addition of the 'Blocker layer' is placed after the 'User Charging' element, the 'Blocker' contains a pre-defined set of rules that assess the user's profile. What are the issues with the Block First architecture?

The network environment changes on a fast rate; Can the Blocker handle the change management? What verification is used to ensure the correct subscriber or session is blocked? The environment of the 'User Charging' element is highly sensitive; Can the company handle the risk of losses from 'Blocker' issues or non-resilient failure?

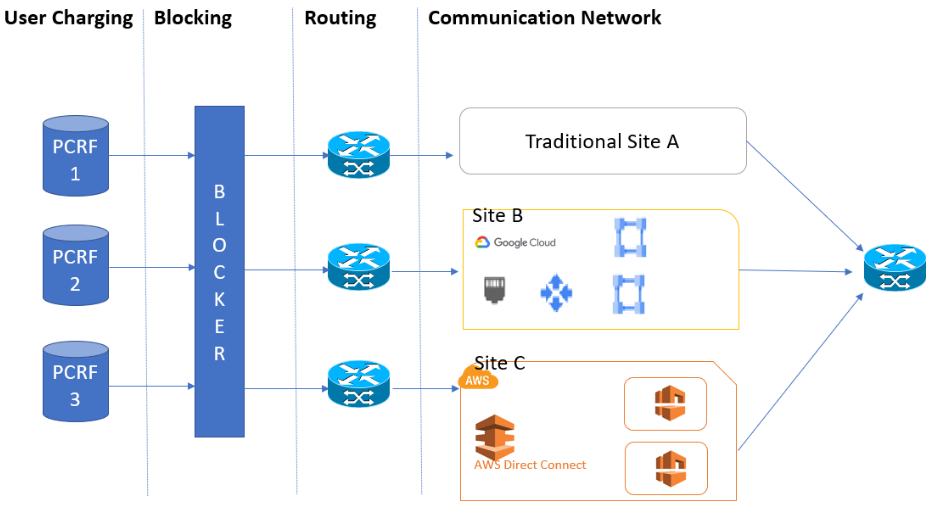


Figure 3.1: Blocker-First Architecture

Figure 3.2 shows the Interconnect-Server Architecture, the Interconnect layer serves to extract the network interface traffic and filter only required protocol data stack.

The Interconnect layer adds a higher level of investment to infrastructure, license and software.

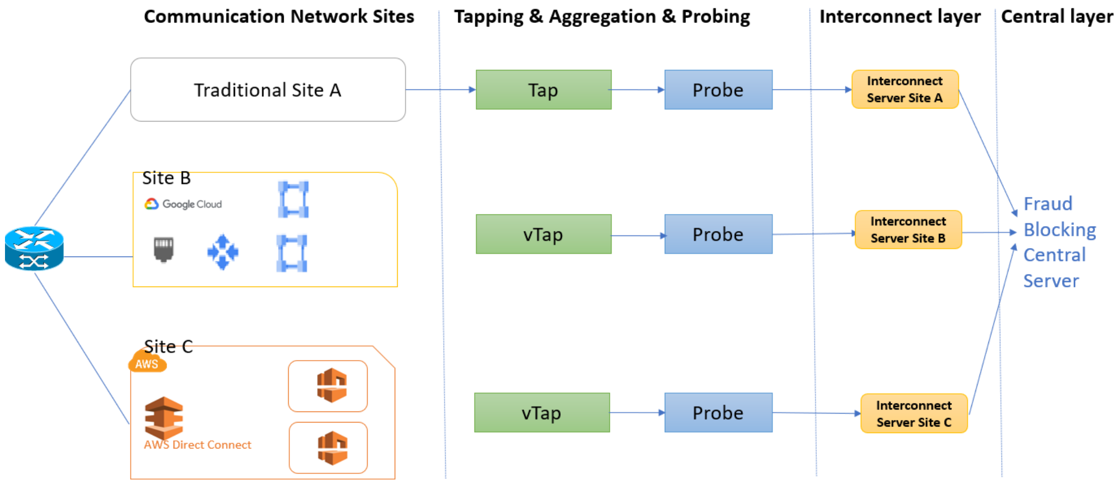

Communication Network Sites
Tapping & Aggregation & Probing
Interconnect layer
Central layer
Traditional Site A
Tap
Probe
Interconnect Server Site A
Site B
Google Cloud
vTap
Probe
Interconnect Server Site B
Fraud Blocking Central Server
Site C
AWS
AWS Direct Connect
vTap
Probe
Interconnect Server Site C

Figure 3.2: Interconnect-Server Architecture

Figure 3.3 shows the SDN controller Architecture from Indarjit, Balyan, and Adonis, (2025b), the SDN controller will handle fraud detection and blocking signals, and it becomes the proposed architecture to perform the study.

The SDN layer will encompass two parts of Fraud detection in terms of intelligent traffic classification and Fraud blocking in terms of integration to the Policy Charging 4G/5G element.

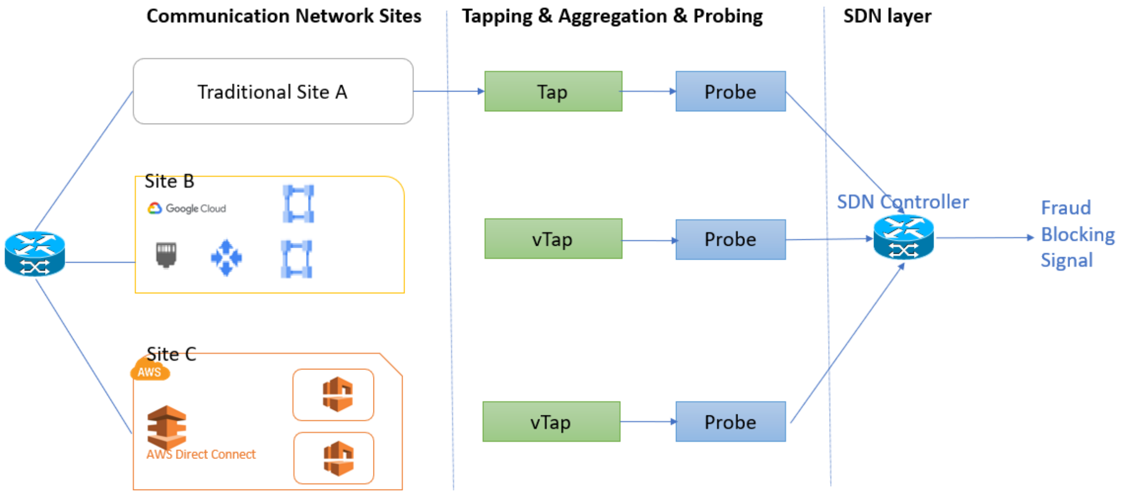


Figure 3.3: Software Defined Network Controller Architecture

3.2 Fraud Detection Selected Architecture

The traditional network is defined in a leaf and spine topology at which the spine switch connects to the layer 3 routing network and the leaf switch connects to the Software-Defined Network (SDN) controller and other switches to serve different services internally and externally. The spine and leaf network resides at a site, while the cloud providers, such as Amazon Web Services (AWS) and Google Cloud reside in another location in the study.

Both cloud providers connect to the router via a firewall access. The importance of hybrid cloud assists in running specific services on different platforms in terms of sharing resources. Each site, traditional or the cloud environment hosts the control-plane and user-plane traffic for its service.

In parallel to the spine and leaf network is positioned the tapping and aggregation platform. To provide a copy of the traffic to the probe. The requirement of physical tapping is used on the spine and leaf network and within the cloud is virtual tapping. From the probe platform, which deciphers the network packet/session sends specific traffic via the router to the SDN network to classify a fraud session. Figure 3.4 shows the physical connectivity for the study,

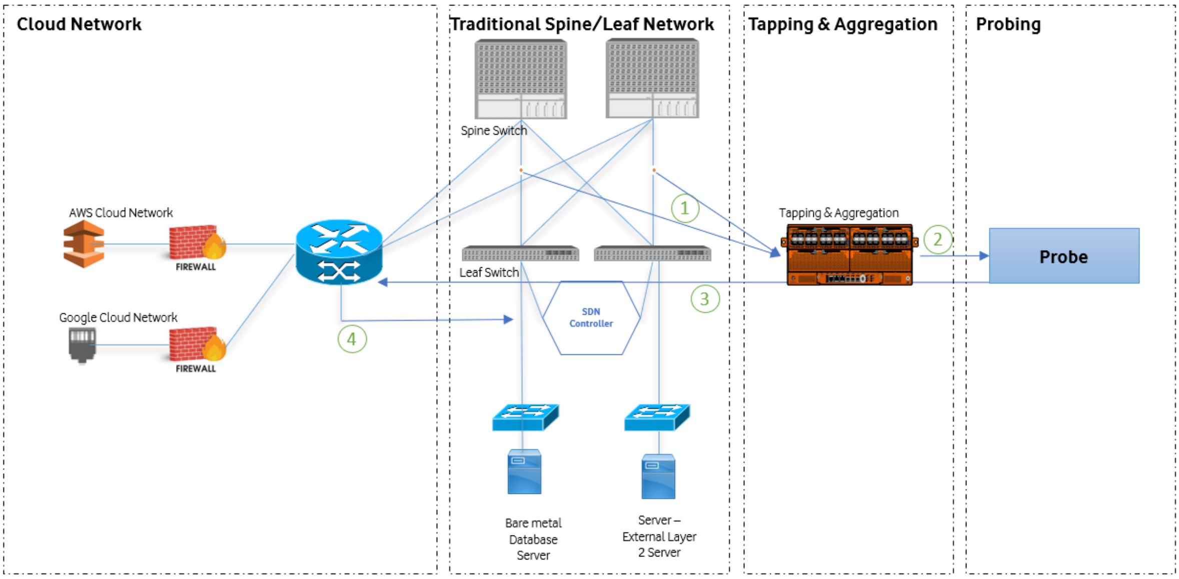


Figure 3.4: Hybrid cloud network and traditional Spine/Leaf network based on physical connectivity

Circle 1 takes a copy of the traffic that is captured between the spine and leaf links, the network traffic flows between the spine and leaf switches, these play a role in network resiliency. The type of traffic captured is interface traffic from the 4G and 5G networks. The traffic is sent to a packet broker that consists of an aggregator or a cluster of aggregators. These are used to filter the interface traffic, for the study, the interest is to send on the 4G traffic the Gi interface and on the 5G network to send the N3, N4, and N6 interface and part of the control plane from the Containerized Network Function (CNF). The 5G network becomes more complex when looking at Service Based Interface (SBI) tapping points. The SBI is encrypted and creates a requirement for virtual tapping.

Circle 2 shows the physical connections from the packet broker to the probe platform, the purpose is to filter the interface traffic by selecting specific traffic and sending the traffic to the probe platform. The aggregator contains intelligence filtering metrics to decode and strip the network packet. The importance of the study will involve network packet de-encapsulation and de-encryption.

Circle 3 shows the physical connections from the probe platform back to the routing network. The purpose of the probe platform is to decipher the network packet and stream the traffic.

Circle 4 shows the connection to the SDN controller, the concept of the SDN controller is based on centralized control, and the purpose of the study will redefine the SDN controller operation and base it on network traffic classification for fraud detection and blocking. The SDN controller will define a new level of feature management to apply filters on the network traffic. The architecture is not complete since blocking will occur on the 4G Policy and Charging Rule Function (PCRF) and the 5G Policy Control Function (PCF). The study will measure different protocols for communication between each network function.

Figure 3.4 shows the traffic flows between Sites A, B, and C to the service node, The SDN controller serves the different sites and plays a major role in performing fraud classification.

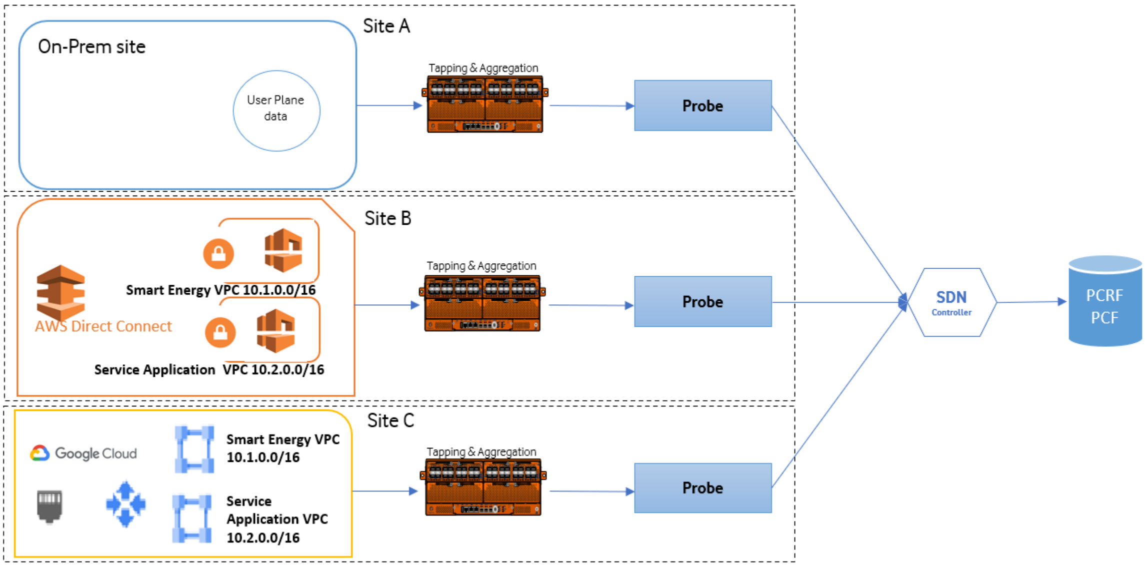


Figure 3.4: Logical network packet flow on tapping/aggregating and probing

3.3 Key Factors on the Architecture

The following key questions arise on the architecture,

The purpose of why to integrate a hybrid cloud network. How do we connect the cloud site to an aggregator?
Why should an aggregator or cluster of aggregators be placed at each site? What other traffic engineering techniques can be considered? What filters are required on the aggregator? What type of protocols are required and deployed?
The aim of the Software-Defined Network Controller and its placement? Define the types of fraud based on applications.
The role of the PCRF/PCF? What network protocols can be explored for the study? What are the challenges of the 5G network compared to the 4G network?

3.3.1 Hybrid Cloud

The hybrid cloud provides levels of flexibility to take advantage of scaling and security on the public cloud while keeping the data on-premises to oblige by governance policy standards. The hybrid cloud gains advantage based on the following:

- The ability to migrate applications to the cloud,
- The use of the public or private cloud or a combination of both,
- Flexibility to run critical applications on-prem in terms of billing,
- Running applications at remote locations for low latency on-demand.
- Local data processing, some data sets need to be processed within the same location for the ease of migration, due to size, cost, capacity, and timing constraints.
- Data center extension, used for cloud bursting, backup, and disaster recovery to the next cloud.

The connectivity from the cloud network to the traditional network is via the communication network, for any network to connect to the next network, the requirement of a firewall is deployed. The firewall uses an Access Control List (ACL) to permit or deny certain types of traffic. In practice, cloud infrastructure and applications can be deployed in locations where they never existed before, such as oil rigs, 5G cellular networks, or colocation facilities. After the firewall platform, the traffic enters the communication network by the router on layer 3 and then routes to the spine and leaf on layer 2. The tapping remains between the spine and the leaf network. The Firewall in Figure 3.4 will be one Firewall for the Cloud provider and another Firewall for the Communication network, the rules will require to be configured on both Firewalls.

Reverting Cloud Instances

An adversary may rollback changes configured on a cloud instance after performing malicious activities to evade detection and erase evidence of their presence. In highly virtualized environments of enclosure, such as cloud-based infrastructure, this can be easily achieved using restoration from VM or data storage snapshots via the cloud management tool. Another variation of this technique involves utilizing temporary storage attached to the compute VM instance. Many cloud providers offer various types of storage, including persistent, local, and/or ephemeral, with the latter types often reset upon stopping/restarting the VM.

Mitigations

This attack technique is challenging to mitigate with preventive controls since it exploits inherent system features.

Detection

To detect such events, to establish centralized logging of instance activity. This logging can be used to monitor and review system events even after reverting to a snapshot, reverting back to the original changes, or altering the persistence/type of storage. Specifically, monitor for events related to snapshots, rollbacks, and VM configuration changes that occur outside of the usage pattern. To reduce false positives, valid change management procedures could introduce a known identifier that is logged with the change (of a tag or header) if supported by the cloud provider, helping to distinguish legitimate actions from malicious ones.

Redundant Access

Fraudster may employ multiple remote access applications with varying configurations and control protocols or use subscriber credentialed access to remote services to maintain access if one mechanism is detected or mitigated. Fraudster will try every way to gain access to valid accounts to use external remote services, such as Virtual Private Networks, to maintain access despite interruptions to remote access tools configured within a hands-on network. Additionally, Fraudster may retain access by hiding behind an authorized account through cloud-based infrastructure and applications. Using a web shell is one method to maintain access to a network via an externally accessible web server.

Mitigation

Network intrusion detection and prevention systems is becoming a more known subject that use network signatures to identify traffic for specific adversary malware can mitigate activity at the network level. Signatures are often special indicators within protocols and vary across different malware families and versions. Fraudster may change tool signatures over time or construct protocols to avoid detection by practised defensive tools.

Detection

Detecting tools based on beacon traffic, command and control protocols, or adversary infrastructure requires prior threat intelligence on tools, Internet Protocol addresses, and domains the adversary may use, along with the ability to detect usage at the network boundary. Prior knowledge of indicators of compromise can also help detect adversary tools at the endpoint if tools are available to scan for those indicators. If an intrusion is entering and sufficient endpoint data or decoded command and control traffic is collected, defenders can likely detect additional application dropped as the adversary conducts their scam.

Valid Accounts

Fraudster can hijack the credentials of specific account or service accounts using credential access techniques or capture credentials earlier in their reconnaissance process through social engineering to gain initial access. The accounts an adversary may target can fall into three categories: default, local, and domain accounts.

- Default Accounts: These are built into an Operating System, such as the guest or administrator accounts on Windows systems, or default factory/provider set accounts on other platforms, systems, running software, or equipment.
- Local Accounts: Configured by an IT team for use by users, remote support, services, or administration on a single system or service.
- Domain Accounts: It is managed by Active Directory Domain Services, where roles, access and responsibility are configured across systems and services within that domain. These can include user, administrator, and service accounts.

Compromised credentials can be used to avoid access controls on various resources within the network and may provide easy access to remote systems and remote available services, such as VPNs, Outlook Web Access, and remote desktop. They may also grant an adversary increased role privileges to specific systems or access to restricted network areas. Fraudster may opt not to use malware or tools in conjunction with the legitimate access these credentials provide, making their presence smarter to detect.

Standardized accounts are not limited to client machines; they also include accounts to update logins for equipment such as network devices and computer applications, whether internal, open source, or commercial off-the-shelf (COTS). Appliances with preset usernames and passwords pose a serious threat if not changed post-installation, as they are easy targets. Similarly, Fraudster may use publicly disclosed or stolen private keys to connect to remote environments via remote services.

The overlap of account access, credentials, and permissions across a network is concerning because Fraudster may pivot across accounts and systems to reach high-level access (e.g., domain or enterprise administrator), bypassing access controls within the enterprise.

Mitigation

- Application Developer Guidance: Ensure applications do not store sensitive data or credentials insecurely.
- Audit: Routinely audit source code, application configuration files, open repositories, and public cloud storage for insecure use and storage of credentials.
- Filter Network Traffic: Implement network-based filtering restrictions to prohibit data transfers to untrusted VPCs.
- Multi-Factor Authentication (MFA): Integrate MFA as part of organizational policy to reduce the risk of Fraudster gaining control of valid credentials. MFA can also restrict access to cloud resources and APIs.
- Password Policies: Rotate access keys within a certain number of days to reduce the effectiveness of stolen credentials.

3.3.2 Aggregator deployment

The requirement for an aggregator or cluster of aggregators at each site serves to capture full portion of the traffic and filter the traffic. The placement of the taps is important for the capture of traffic. The question of why the solution requires an aggregator at each site. A feasibility study was performed to measure the cost of the Bill of Material (BOM) of per site base and use the Core Network (CDN) to route traffic. The BOM cannot be shared on the document since it contains confidential figures.

The disadvantage of routing user-plane traffic is listed below:

- When network packets are routed over the network, since it’s a copy of traffic, it holds no layer 3 routing functionality. Therefore, the traffic requires to be encapsulated within a Layer 2 General Routing Encapsulation(L2GRE) tunnel, Figure 3.6 shows the setup required.

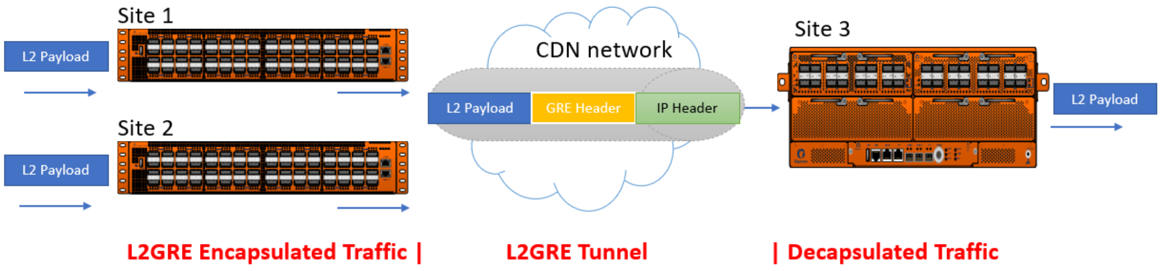
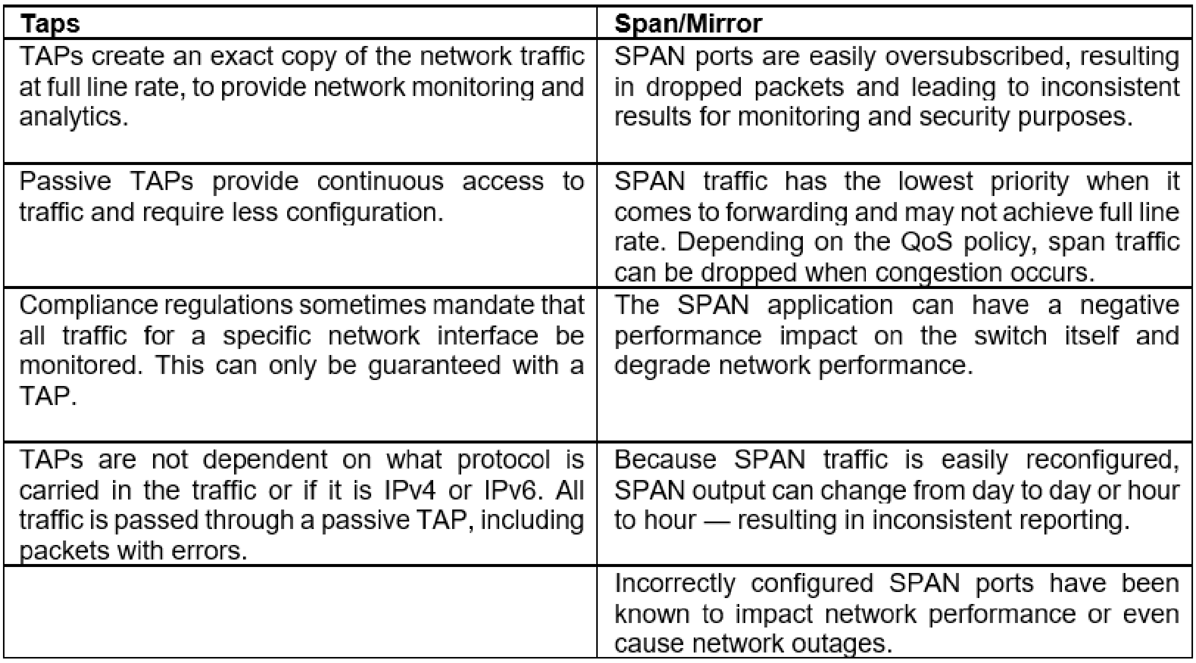


Figure 3.6: Layer 2 General Routing Encapsulation tunnel

- To enable the L2GRE tunnel session, a software license is required to perform the encapsulation and decapsulation. Site 1 and Site 2 add a GRE Header to the Layer 2 traffic, a virtual tunnel is created to route the traffic, and the decapsulation is a feature set to remove the GRE header off and provide the pure Layer 2 traffic to its destination.
- Another disadvantage to using the core network is the amount of capacity required to send the traffic from sites to the aggregator site. The core network will require to be upgraded and prioritized. From the feasibility study, it shows the cost of upgrading the core network is 3 X the price of placement of the aggregator at each site.
- The other disadvantage is when network traffic is sent through the core network, it is constantly affected by latency changes that add a level of complexity for re-constructing the network traffic at its destination.

Tapping of traffic is recommended as compared to mirror or span session that is required to be configured on the switch interface, the tapping of a network takes a copy of the traffic and does not degrade the performance of the network. Table 3.1 shows the advantages and disadvantages of tapping versus spanning.

Table 3.1: Spanning versus Tapping



Taps	Span/Mirror
TAPs create an exact copy of the network traffic at full line rate, to provide network monitoring and analytics.	SPAN ports are easily oversubscribed, resulting in dropped packets and leading to inconsistent results for monitoring and security purposes.
Passive TAPs provide continuous access to traffic and require less configuration.	SPAN traffic has the lowest priority when it comes to forwarding and may not achieve full line rate. Depending on the QoS policy, span traffic can be dropped when congestion occurs.
Compliance regulations sometimes mandate that all traffic for a specific network interface be monitored. This can only be guaranteed with a TAP.	The SPAN application can have a negative performance impact on the switch itself and degrade network performance.
TAPs are not dependent on what protocol is carried in the traffic or if it is IPv4 or IPv6. All traffic is passed through a passive TAP, including packets with errors.	Because SPAN traffic is easily reconfigured, SPAN output can change from day to day or hour to hour — resulting in inconsistent reporting.
	Incorrectly configured SPAN ports have been known to impact network performance or even cause network outages.

From Table 3.1 it becomes clear the preferred way of capturing the traffic and preserving the network performance for the study.

An anonymous pcap file was analyzed on WireShark tool, the selected tool for packet capture analysis is Wireshark, which stands out among alternatives like tcpdump, tshark, and NetworkMiner. Wireshark offers a user-friendly experience with an intuitive graphical interface, making network analysis more accessible. Additionally, it provides powerful filtering capabilities for deep packet inspection, allowing precise identification of source and destination traffic.

A network packet contains various protocols, and Wireshark excels in protocol recognition, supporting a broad range of them for detailed analysis. During my testing phase, I found that Wireshark is compatible across multiple environments, including Windows, macOS, and Linux, ensuring flexibility in software implementation.

And three key features were needed to be applied for consistent formatting of the data:

- De-duplication
- User Defined Attribute (UDA)
- Correlation

De-duplication feature

The collection of traffic from multiple points throughout the network can result in the same packet being sent more than once for monitoring. Traffic is monitored at TAP ports. Every time a packet traverses a TAP, it goes to the analysis tool. If a packet travels across more than one monitored link, the tool will receive multiple copies of the same packet.

Assume a simple 3-tier application in Figure 3.7. The switch sends a copy of all traffic for three servers over to a monitoring tool.

- On Port 1 the User query is seen inbound. Then the Web Server sends an outbound query to the Application Server.
- On Port 2 the same Web Server query is seen again inbound. The Application Server sends an outbound query to the Database Server.
- On Port 3 the same query is seen inbound.

In this five-packet example, there are two duplicates.

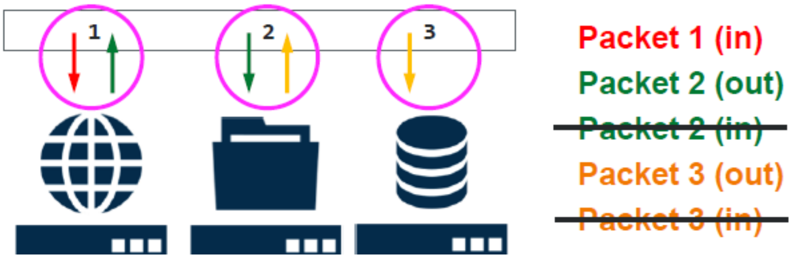
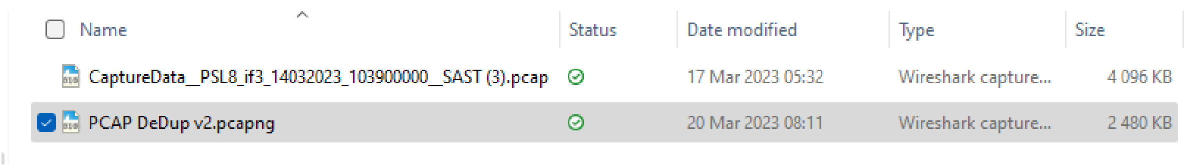


Figure 3.7: Application Server query

A test case was performed to analyze the source of the problem, the test case looked at the capture of a pcap file to identify the duplication and applied the De-Dup feature to isolate the problem.The testing used a pcap file and was performed in a live environment and not by simulation. The results showed favorable in terms of traffic duplication reduction.

Original Pcap with Duplicates = CaptureData__PSL8_if3_14032023_103900000__SAST (3).pcap were 4 MB. This pcap was then applied to Gigamon H-Series with GigaSMART DeDuplication enabled resultant “dedup’ed” file = PCAP DeDup v2.pcapng were 2,4 MB. Figure 3.8 shows the sizes.



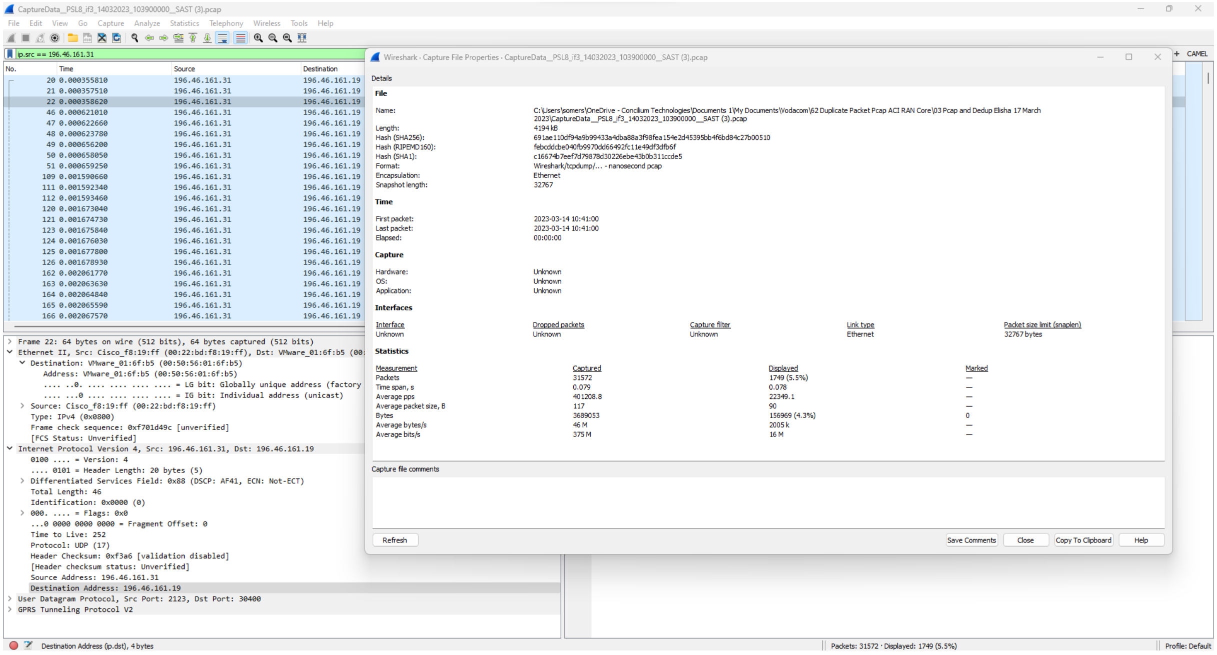
☐ Name	Status	Date modified	Type	Size
CaptureData__PSL8_if3_14032023_103900000__SAST (3).pcap	✓	17 Mar 2023 05:32	Wireshark capture...	4 096 KB
☑ PCAP DeDup v2.pcapng	✓	20 Mar 2023 08:11	Wireshark capture...	2 480 KB

Figure 3.8: Pcap files

Furthermore, to gain a better understanding of the benefit of De-Duplication, an examination of the IP towards the source and payload was monitored. Figure 3.9 shows the pcap file taken with an emphasis on IP 196.46.161.19, the number of packets under the IP showed 31572 and the average bits/s at 375 M. Figure 3.10 shows the IP after applying the De-Duplication feature, the results show the number of packets at 16654 and at average bits/s at 30 M. Showing a rate of Duplication reduction of 47, 2%

And

Showing the average bits/s reduction of 92%


CaptureData__PSL8_if3_14032023_103900000__SAST (3).pcap
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
ip.src == 196.46.161.31
CAMEL
Wireshark · Capture File Properties · CaptureData__PSL8_if3_14032023_103900000__SAST (3).pcap
Refresh
Save Comments
Close
Copy To Clipboard
Help

```
No.    Time          Source          Destination
   20 0.000355810   196.46.161.31   196.46.161.19
   21 0.000357510   196.46.161.31   196.46.161.19
   22 0.000358620   196.46.161.31   196.46.161.19
   46 0.000621010   196.46.161.31   196.46.161.19
   47 0.000622660   196.46.161.31   196.46.161.19
   48 0.000623780   196.46.161.31   196.46.161.19
   49 0.000656200   196.46.161.31   196.46.161.19
   50 0.000658050   196.46.161.31   196.46.161.19
   51 0.000659250   196.46.161.31   196.46.161.19
  109 0.001590660   196.46.161.31   196.46.161.19
  111 0.001592340   196.46.161.31   196.46.161.19
  112 0.001593460   196.46.161.31   196.46.161.19
  120 0.001673040   196.46.161.31   196.46.161.19
  121 0.001674730   196.46.161.31   196.46.161.19
  123 0.001675840   196.46.161.31   196.46.161.19
  124 0.001676030   196.46.161.31   196.46.161.19
  125 0.001677800   196.46.161.31   196.46.161.19
  126 0.001678930   196.46.161.31   196.46.161.19
  162 0.002061770   196.46.161.31   196.46.161.19
  163 0.002063630   196.46.161.31   196.46.161.19
  164 0.002064840   196.46.161.31   196.46.161.19
  165 0.002065590   196.46.161.31   196.46.161.19
  166 0.002067570   196.46.161.31   196.46.161.19
```

```
> Frame 22: 64 bytes on wire (512 bits), 64 bytes captured (512 bits)
v Ethernet II, Src: Cisco_f8:19:ff (00:22:bd:f8:19:ff), Dst: VMware_01:6f:b5 (00:
  v Destination: VMware_01:6f:b5 (00:50:56:01:6f:b5)
      Address: VMware_01:6f:b5 (00:50:56:01:6f:b5)
      .... ..0. .... .... .... .... = LG bit: Globally unique address (factory
      .... ...0 .... .... .... .... = IG bit: Individual address (unicast)
  > Source: Cisco_f8:19:ff (00:22:bd:f8:19:ff)
    Type: IPv4 (0x0800)
    Frame check sequence: 0xf701d49c [unverified]
    [FCS Status: Unverified]
v Internet Protocol Version 4, Src: 196.46.161.31, Dst: 196.46.161.19
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
  > Differentiated Services Field: 0x88 (DSCP: AF41, ECN: Not-ECT)
    Total Length: 46
    Identification: 0x0000 (0)
  > 000. .... = Flags: 0x0
    ...0 0000 0000 0000 = Fragment Offset: 0
    Time to Live: 252
    Protocol: UDP (17)
    Header Checksum: 0xf3a6 [validation disabled]
    [Header checksum status: Unverified]
    Source Address: 196.46.161.31
    Destination Address: 196.46.161.19
> User Datagram Protocol, Src Port: 2123, Dst Port: 30400
> GPRS Tunneling Protocol V2
```

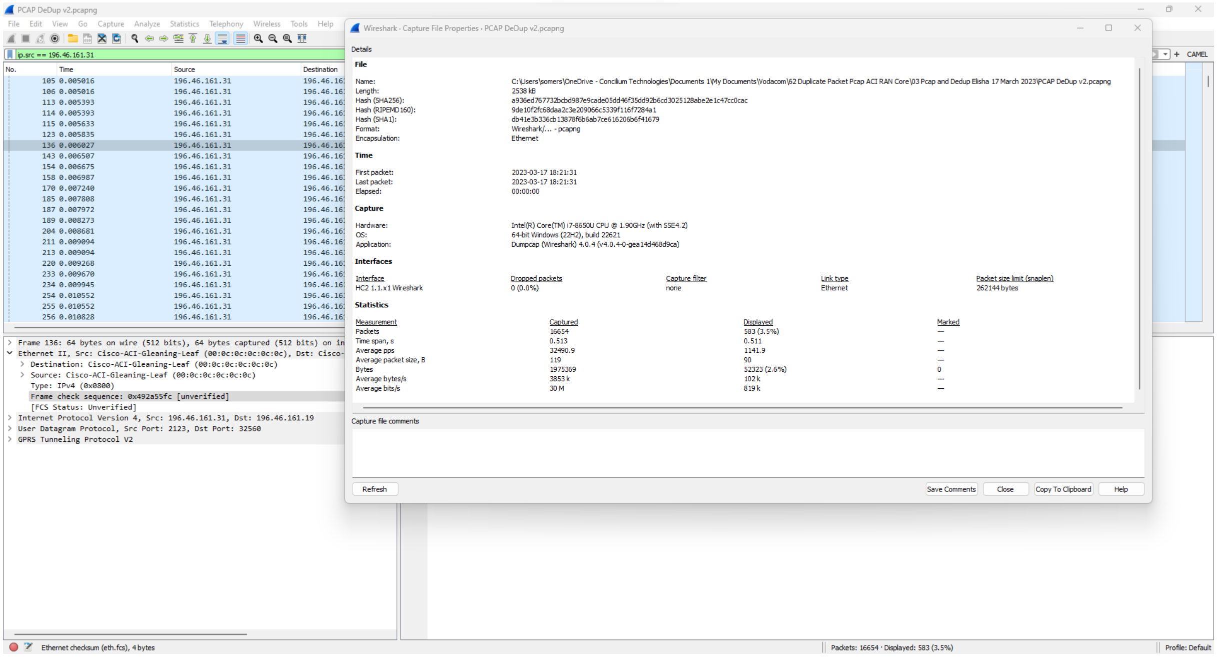
Destination Address (ip.dst), 4 bytes

Packets: 31572 · Displayed: 1749 (5.5%)

Profile: Default

```
Details
File
Name:                C:\Users\somers\OneDrive - Concilium Technologies\Documents 1\My Documents\Vodacom\62 Duplicate Packet Pcap ACI RAN Core\03 Pcap and Dedup Elisha 17 March
                     2023\CaptureData__PSL8_if3_14032023_103900000__SAST (3).pcap
Length:              4194 kB
Hash (SHA256):       691ae110df94a9b99433a4dba88a3f98fea154e2d45395bb4f6bd84c27b00510
Hash (RIPEMD160):    febcddcbe040fb9970dd66492fc11e49df3dfb6f
Hash (SHA1):         c16674b7eef7d79878d30226ebe43b0b311ccde5
Format:              Wireshark/tcpdump/... - nanosecond pcap
Encapsulation:       Ethernet
Snapshot length:     32767

Time
First packet:        2023-03-14 10:41:00
Last packet:         2023-03-14 10:41:00
Elapsed:             00:00:00

Capture
Hardware:            Unknown
OS:                  Unknown
Application:         Unknown

Interfaces
Interface            Dropped packets      Capture filter       Link type            Packet size limit (snaplen)
Unknown              Unknown              Unknown              Ethernet             32767 bytes

Statistics
Measurement              Captured        Displayed          Marked
Packets                  31572           1749 (5.5%)        —
Time span, s             0.079           0.078              —
Average pps              401208.8        22349.1            —
Average packet size, B   117             90                 —
Bytes                    3689053         156969 (4.3%)      0
Average bytes/s          46 M            2005 k             —
Average bits/s           375 M           16 M               —

Capture file comments
```

Figure 3.9: Pcap file – Before De-Duplication


PCAP DeDup v2.pcapng
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
ip.src == 196.46.161.31
CAMEL
Wireshark · Capture File Properties · PCAP DeDup v2.pcapng
Refresh
Save Comments
Close
Copy To Clipboard
Help

```
No.    Time       Source          Destination
  105 0.005016   196.46.161.31   196.46.16
  106 0.005016   196.46.161.31   196.46.16
  113 0.005393   196.46.161.31   196.46.16
  114 0.005393   196.46.161.31   196.46.16
  115 0.005633   196.46.161.31   196.46.16
  123 0.005835   196.46.161.31   196.46.16
  136 0.006027   196.46.161.31   196.46.16
  143 0.006507   196.46.161.31   196.46.16
  154 0.006675   196.46.161.31   196.46.16
  158 0.006987   196.46.161.31   196.46.16
  170 0.007240   196.46.161.31   196.46.16
  185 0.007808   196.46.161.31   196.46.16
  187 0.007972   196.46.161.31   196.46.16
  189 0.008273   196.46.161.31   196.46.16
  204 0.008681   196.46.161.31   196.46.16
  211 0.009094   196.46.161.31   196.46.16
  213 0.009094   196.46.161.31   196.46.16
  220 0.009268   196.46.161.31   196.46.16
  233 0.009670   196.46.161.31   196.46.16
  234 0.009945   196.46.161.31   196.46.16
  254 0.010552   196.46.161.31   196.46.16
  255 0.010552   196.46.161.31   196.46.16
  256 0.010828   196.46.161.31   196.46.16
```

```
> Frame 136: 64 bytes on wire (512 bits), 64 bytes captured (512 bits) on in
v Ethernet II, Src: Cisco-ACI-Gleaning-Leaf (00:0c:0c:0c:0c:0c), Dst: Cisco-
  > Destination: Cisco-ACI-Gleaning-Leaf (00:0c:0c:0c:0c:0c)
  > Source: Cisco-ACI-Gleaning-Leaf (00:0c:0c:0c:0c:0c)
    Type: IPv4 (0x0800)
    Frame check sequence: 0x492a55fc [unverified]
    [FCS Status: Unverified]
> Internet Protocol Version 4, Src: 196.46.161.31, Dst: 196.46.161.19
> User Datagram Protocol, Src Port: 2123, Dst Port: 32560
> GPRS Tunneling Protocol V2
```

Ethernet checksum (eth.fcs), 4 bytes

Packets: 16654 · Displayed: 583 (3.5%)

Profile: Default

```
Details
File
Name:                C:\Users\somers\OneDrive - Concilium Technologies\Documents 1\My Documents\Vodacom\62 Duplicate Packet Pcap ACI RAN Core\03 Pcap and Dedup Elisha 17 March 2023\PCAP DeDup v2.pcapng
Length:              2538 kB
Hash (SHA256):       a936ed767732bcbd987e9cade05dd46f35dd92b6cd3025128abe2e1c47cc0cac
Hash (RIPEMD160):    9de10f2fc68daa2c3e209066c5339f116f7284a1
Hash (SHA1):         db41e3b336cb13878f6b6ab7ce616206b6f41679
Format:              Wireshark/... - pcapng
Encapsulation:       Ethernet

Time
First packet:        2023-03-17 18:21:31
Last packet:         2023-03-17 18:21:31
Elapsed:             00:00:00

Capture
Hardware:            Intel(R) Core(TM) i7-8650U CPU @ 1.90GHz (with SSE4.2)
OS:                  64-bit Windows (22H2), build 22621
Application:         Dumpcap (Wireshark) 4.0.4 (v4.0.4-0-gea14d468d9ca)

Interfaces
Interface            Dropped packets      Capture filter       Link type            Packet size limit (snaplen)
HC2 1.1.x1 Wireshark 0 (0.0%)             none                 Ethernet             262144 bytes

Statistics
Measurement              Captured        Displayed          Marked
Packets                  16654           583 (3.5%)         —
Time span, s             0.513           0.511              —
Average pps              32490.9         1141.9             —
Average packet size, B   119             90                 —
Bytes                    1975369         52323 (2.6%)       0
Average bytes/s          3853 k          102 k              —
Average bits/s           30 M            819 k              —

Capture file comments
```

Figure 3.10: Pcap file – After De-Duplication

User Defined Attribute(UDA) Feature

UDA allows the creation of pass-and-drop map rules with pattern matches that perform a search of a specific sequence of bits at a specific offset in a network packet. The configuration of a 16-byte pattern matches a map rule, the pattern is a specific sequence of bits at a specific location in a frame.

The UDA filter is a conversion from digits to hex decimals, the so-called map is configured from IP addresses that use the inner or outer IP addresses, Table 3.2 shows the SIP and RTP stream, the SIP is the signaling portion and RTP can be seen as the data portion. The following conversions were applied.

Table 3.2: UDA Configuration for filtering traffic

SIP		10.128.3.69	rule add pass uda1-data 0a800345-00000000-00000000-00000000 uda1-mask ffffffff-00000000-00000000-00000000 uda1-offset 66
			rule add pass uda1-data 00000000-0a800345-00000000-00000000 uda1-mask 00000000-ffffffff-00000000-00000000 uda1-offset 66
RTP		10.116.184.132	rule add pass uda1-data 0a74b884-00000000-00000000-00000000 uda1-mask ffffffff-00000000-00000000-00000000 uda1-offset 66
			rule add pass uda1-data 00000000-0a74b884-00000000-00000000 uda1-mask 00000000-ffffffff-00000000-00000000 uda1-offset 66
		10.116.184.133	rule add pass uda1-data 0a74b885-00000000-00000000-00000000 uda1-mask ffffffff-00000000-00000000-00000000 uda1-offset 66
			rule add pass uda1-data 00000000-0a74b885-00000000-00000000 uda1-mask 00000000-ffffffff-00000000-00000000 uda1-offset 66
		10.116.184.134	rule add pass uda1-data 0a74b886-00000000-00000000-00000000 uda1-mask ffffffff-00000000-00000000-00000000 uda1-offset 66
			rule add pass uda1-data 00000000-0a74b886-00000000-00000000 uda1-mask 00000000-ffffffff-00000000-00000000 uda1-offset 66
		10.116.184.135	rule add pass uda1-data 0a74b887-00000000-00000000-00000000 uda1-mask ffffffff-00000000-00000000-00000000 uda1-offset 66
			rule add pass uda1-data 00000000-0a74b887-00000000-00000000 uda1-mask 00000000-ffffffff-00000000-00000000 uda1-offset 66
		10.116.184.136	rule add pass uda1-data 0a74b888-00000000-00000000-00000000 uda1-mask ffffffff-00000000-00000000-00000000 uda1-offset 66
			rule add pass uda1-data 00000000-0a74b888-00000000-00000000 uda1-mask 00000000-ffffffff-00000000-00000000 uda1-offset 66

Table 3.3 shows the definition of the terms used for the configuration.

Table 3.3: Map rule terms

Term	Description
Pattern	The map rule commands to set up the actual bit patterns in question. uda1-data and uda2-data
Mask	The map rule commands to specify which bits in the pattern will match the map rule. uda1-mask and uda2-mask
Offset	The map rule commands to specify where in the network packet bits are required to be matched. Uda1-offset and uda2-offset

The UDA filter allows the study to look into the network packet, The purpose of the research explore user-plane and control-plane traffic, giving the ability to open the protocol stack and provide the probe platform with the only required traffic. The configuration is pasted below to show the steps and the interface were linked:

```
(configuration) #                   map alias UDA01
(configuration map alias UDA01) #   type regular byRule
(configuration map alias UDA01) #   from 1/1/c1..c20
(configuration map alias UDA01) #   to 1/1/x1,1/1/x2,
(configuration map alias UDA01) #   rule add pass uda 1-data 12345678-12345678-
                                12345678-12345678 uda1-mask 0000ffff-0000ffff-0000ffff-
                                0000ffff uda 1-offset 10
(configuration map alias UDA01) #   exit
```

Correlation Feature

The feature is defined as the separation of the User-plane and Control-plane traffic, the purpose is to reconstruct the network packet since the control portion resides in a virtual or cloud environment. This means the User-plane Function can reside in one site and the control service function can reside in another site. The research requires looking into per subscriber session and the need for correlation feature gathers to perform filtering of the subscriber ID or tunnel ID. The protocol used is the General Tunnelling Protocol (GTP) for control-plane GTP-c and user-plane GTP-u to carry subscriber data from the user device to the internet.

The feature allows correlating the subscriber-specific attributes, including the subscriber ID which is the International Mobile Subscriber Identity (IMSI) and International Mobile Equipment Identity (IMEI). The GTP-c attributes are carried with the user-plane traffic which are identified as GTP tunnels also known as Tunnel Endpoint ID (TEID). Once the GTP-u TEIDs are identified and correlated to the subscriber attributes, then the subscriber traffic can be processed in a subscriber-readable manner.

Figure 3.11 shows a 5G network in which the control-plane network service functions are separated from the user-plane.

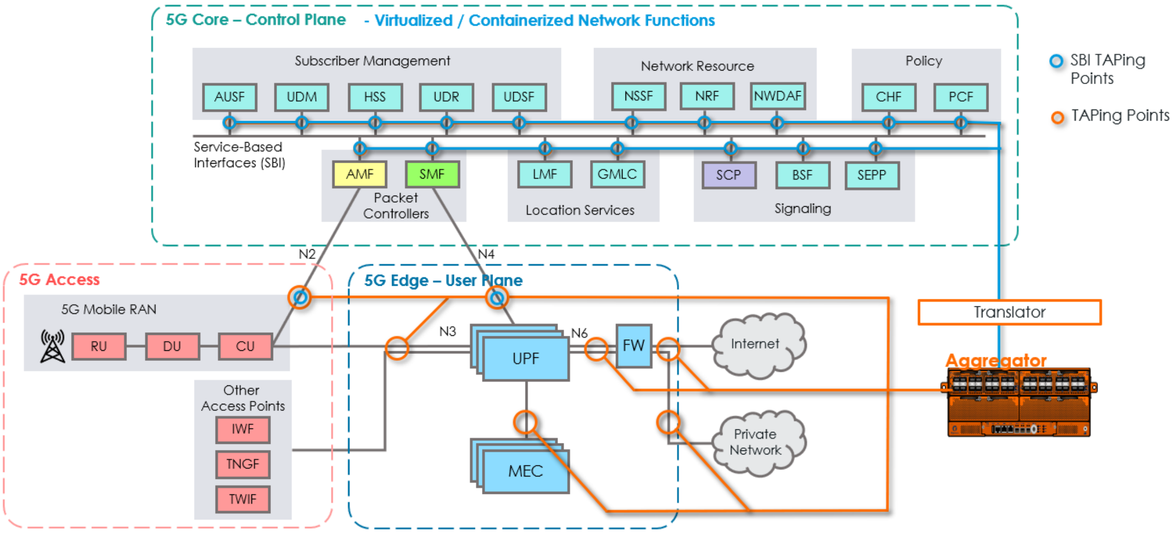


Figure 3.11: 5G Core control and userplane Network

The control-plane can be separated within domains and at each serve a function, these functions are either virtualized or containerized. The User-Plane Function (UPF) is connected to the Session Management Function (SMF) and Access and Mobility Management Function (AMF). The correlation feature is applied to the aggregator platform.

The 5G network becomes complex, each function serves the next function, the standard for a 5G Core is to have the control plane encrypted, which brings a higher level of complexity, since

the aggregation vendor needs to de-encrypt the traffic, the purpose of the Translator is shown to serve the de-encryption feature. The Translator will be required to de-encrypt the traffic and format the traffic, the research will expand on how to perform both features and the difficult steps when a 5G Core vendor cannot share development code with another vendor.
The correlation feature was applied to measure the number of sessions in use and the number of tunnels in use, the data showed the number that were processed.

```
GTP Resource Summary
==============================
Num Sessions In Use            8947852
Num Tunnels In Use             18738008
Tunnels Available              21139173
UPN CTunnels Available         0
Tunnels Pending Free           0
Tunnels Marked Free            13919808501
Tunnels Returned               13947980791
```

3.3.3 SDN Placement

The section provides the aim of the Software-Defined Network Controller and its placement in the network by Indarjit, Adonis, and Brandt, (2022). To define the types of fraud based on applications. To scope on SDN been the central component and use of an SDN controller or multiple SDN controllers within a communication network become the building blocks for automation and network programmability on an Software Defined Network communication network.

SDN technology caters to network automation, helping the Service Provider to work more closely with different vendors and allowing providers to develop their hardware and feature requirements. Software-defined transport networking consists of the network controller managing the Wide Area Network links and the physical interconnecting nodes. A survey conducted by Forrester Consulting on behalf of Juniper Networks, in January 2014, identified the key demands from a cloud infrastructure classified as capacity, performance, reliability, and automation/programmability. The SDN controller has challenges on modularity and does not support non-disruptive updates. Another challenge of the SDN controller is the reuse of SDN modules; if a module is used to collect topology link, interface, node data and requires to be used in another SDN environment, it will require to be re-programmed. Furthermore, external management needs to support external applications when an SDN change is encountered. The external system requires to have a higher reserve of processing to supporting any state.

To assist with the mentioned processing challenges, the SDN controller needs to be re-designed to support the external environment and the outsourcing of capacity packet processing. The re-design phase of the SDN controller is to allocate and divide into many sets of services instead of a centralized base. Software-defined network is favoured for their flexible operations and programmability layers; it provides room for virtual platforms that allow the selection of hypervisors and the fundamentals of Network Function Virtualisation (NFV) and Containerized Network Function (CNF). The careful approach to SDN requires measurements calculated for the applications and programs running on the computer. Blenk et al., (2018) review the research study of virtualization on SDN and hypervisor placement. There are three limitations of the research to mention:

- physcial location problem,
- SDN Controller placement in the topology, and
- Virtual network embedding.

The traditional location of the SDN controller in the network resides within the leaf switches on a physical connectivity level, but on a logical level, the SDN controller is placed above the routing layer. For the purpose of the research, the SDN controller is carefully placed after the probing platform. This places a higher operational functionality and feature required on the SDN controller. The SDN controller sits on a virtual platform and will require to be scaled out according to traffic growth. Figure 3.12 shows the network packet embedded into the ACI fabric consisting of the Outer and Inner IP, VXLAN header, and payload.

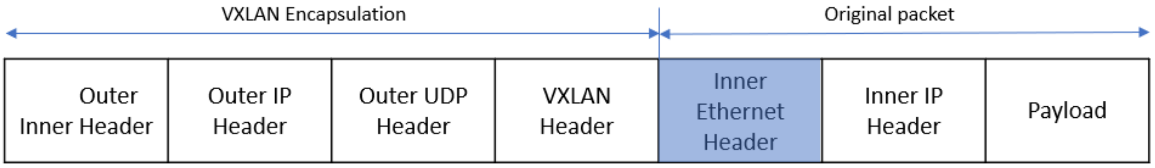


Figure 3.12: VXLAN Encapsulated network packet

Figure 3.13 shows the network packet encapsulated into a routing tunnel consisting of the tunnel IP, GRE Flags, Inner IP, and payload.

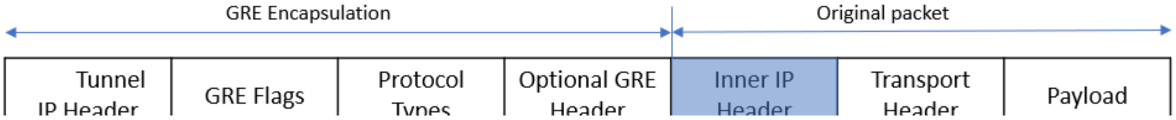


Figure 3.13: GRE Encapsulated network packet

From both network packets shown, the important focus is on the Inner IP Header which provides subscriber details. For the study on the ACI Fabric, the payload will require de-encapsulation. The next paragraphs build on network tables that contain routing information of the subscriber's traffic, the recommended tables measure a New flow, an Allowed flow, and a Fraud flow.

Figure 3.14 shows the Central Processing Unit (CPU) and Application Specific Integrated Circuit (ASIC) within the SDN controller, the CPU contains network packets/flows that are learned at first from its interface or Virtual Local Area Network (VLAN). The CPU creates two Flow Table tile entries. From the ASIC, flow records are exported to the CPU in bulk.

- A high-priority entry routes into FT files with 5-Tuple referring to the source IP address, source port, destination IP address, destination port, and transport protocol which uniquely identifies the TCP session flow with its interface and VLAN label with Permit action.

- A low-priority deny rule is installed into another FT file with 5-tuple and deny action.

- Both the permit and deny actions are derived from ft_collect_disable_action, a learned flow cannot be recorded in the flow table again.

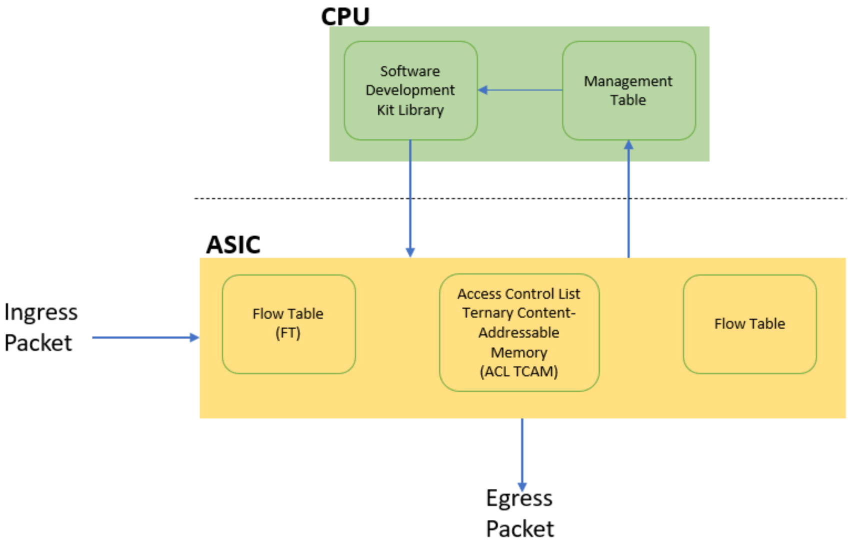

CPU
Software Development Kit Library
Management Table
ASIC
Ingress Packet
Flow Table (FT)
Access Control List Ternary Content-Addressable Memory (ACL TCAM)
Flow Table
Egress Packet

Figure 3.14: Flow between Central Processing Unit(CPU) and Application Specific Integrated Circuit(ASIC) of the SDN controller

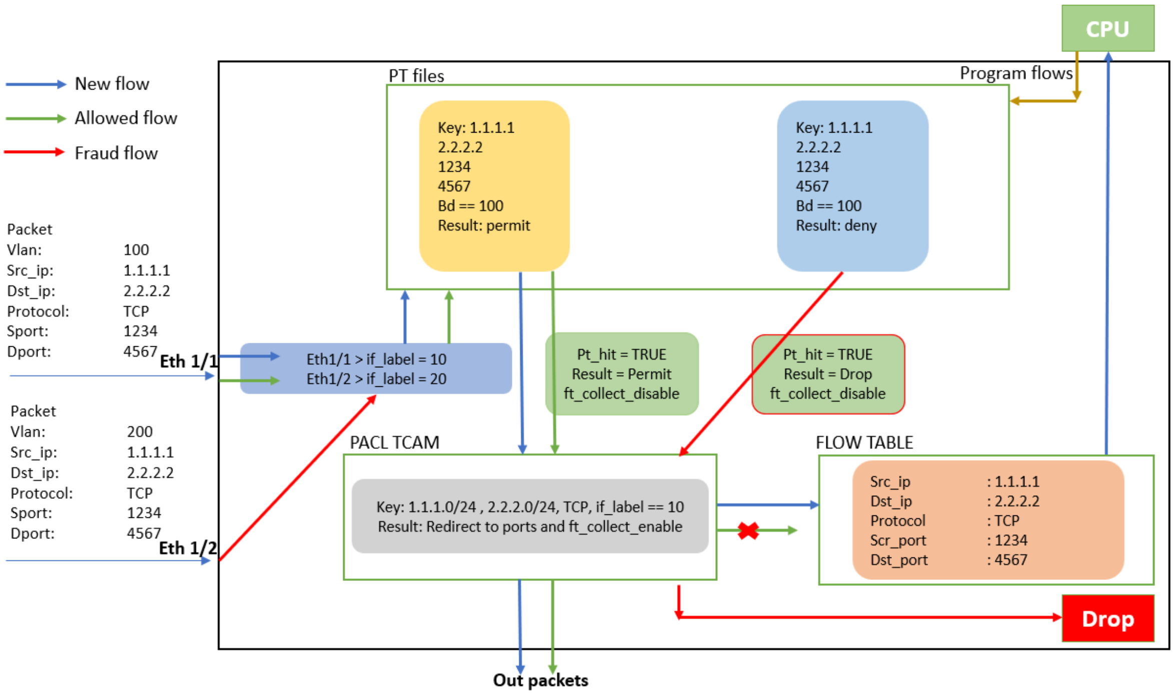


Figure 3.15: Communication flow between PT files, PACL TCAM and Flow Table

Figure 3.15 defines the communication flow between the PT files, PACL TCAM, and Flow Table.

The New flow is the permissible flow, the Allowed flow is defined as the uncertain flow, and the fraud flow is defined as the duplicate flow. The New flow and Allowed flow to enter vias Eth 1/1 to the PT file, the PT file will compare the network session details with the Fraud flow, if the same subscriber registers a duplicate, the flow will route to the PACL TCAM and be dropped. Further, the Allowed flow will be denied from the PACL TCAM and FLOW TABLE. In terms of Fraud flow a comparison is made to permit or deny the flow.

For the purpose of the study, two types of fraud are explored, Service Application fraud and Energy Abuse fraud. At which both play a major role in today's age, both present network challenges:

Service Application fraud is defined as critical services that belong to a network domain that serves as a network slice, within the network it is labelled as an interface such as S1-U and SGi interface from the 4G network, and N3 and N6 interface from the 5G network. The study will explore the SGi interface and N6 interface from both networks since the aggregation point will be captured between the firewall and the switch since the traffic requires to be pre-NAT. The firewall serves to allow certain traffic and deny certain traffic. The SGi interface is defined by 3GPP as the connectivity between the Evolved Packet Core (EPC) and the data network, the SGi interface can be defined as a service gateway for deep packet inspection and policy-based service selection. The N6 interface is the connectivity between the User Plane Function

(UPF) and the data network or public/private cloud. Figure 3.16 shows the protocol stack of the SGi interface and N6 interface, the PGW and UPF are service functions used for transporting the traffic and the data network is built from Layer 1 to Layer 7.

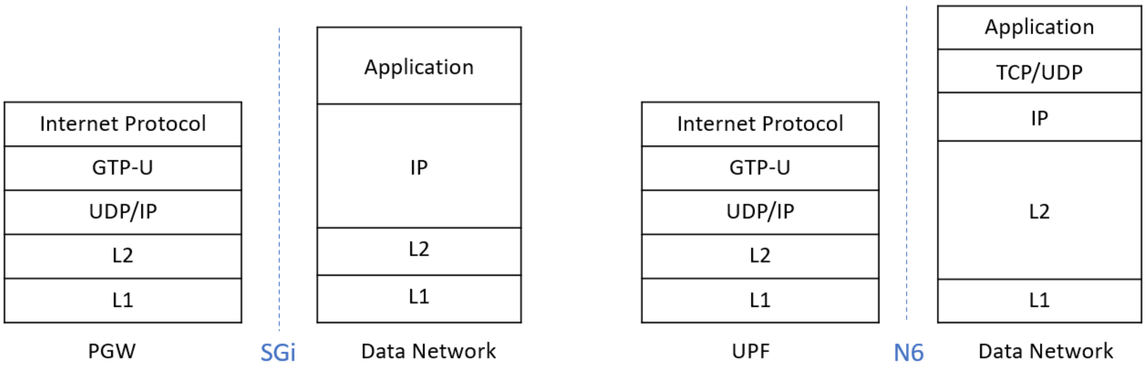


Figure 3.16: SGi and N6 protocol stack

3.4 Energy Abuse fraud

Energy fraud explores the Smart Grid environment with the attacks and to apply a better solution in not mitigation but blocking the fraudster. The types of attacks are mentioned below:

- Direct theft is defined as connectivity/consumption of power to premises without paying or meter measurement.
- Meter tampering is defined as false readings from the meter that are compromised.
- Billing irregularities, the process of manual meter reading and recording to billing provides an opportunity for cyber-attacks to change the account information.

The proposed solution needs to incorporate the following:

- Meter tampering and supply signal alerts
- Complete automation
- Remote control and blocking
- Analysis of consumption within low-voltage grid
- Usage statistics

To meet the above requirements, the question arises, Where can we attain the data flows from the communication network? What performance tool is required to be designed to meet analysis and usage? What deep inspection of the network packet is needed?

The communication network is built on Virtual Local Area Networks(VLANs) that are configured per service or application, the study will look at VLANs and perform Deep Packet Inspection using the aggregation and SDN platform. Figure 3.17 shows the different VLANs within the communication network.

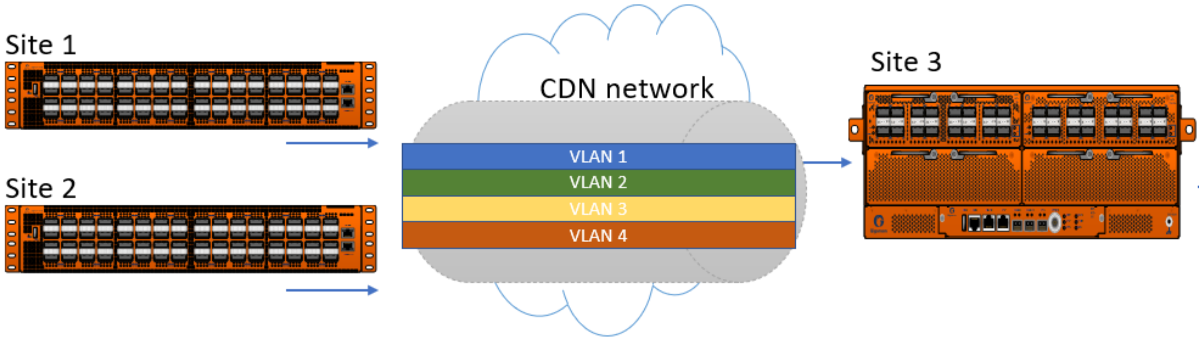


Figure 3.17: Virtual Local Area Networks (VLANs) configured within the communication network

Each VLAN has a unique number, and its traffic is separated from the next VLAN. This means VLAN 1 cannot see the traffic or subscriber information in VLAN 2/3/4.

3.5 PCRF/PCF

The section highlights the role of the PCRF/PCF, which network protocols can be explored for the study, and what are the challenges of the 5G network compared to the 4G network.

The Policy and Charging Rule Function (PCRF) within the 4G network and Policy Control Function (PCF) within the 5G network will be used to perform the blocking function, the study will explore both and their protocols to stop the fraudster. The PCRF relies on operator policies, subscription-specific data, or application data like the Proxy Call Session Control Function (P-CSCF) to provide the Software-Defined Parameter (SDP) for voice/video calls. While the PCF on the other hand, has several interfaces from network functions of AMF, SMF, AF, UDP, etc.

The next step is to look at the communication flow between the Network Functions in Figure 3.18. The solution will need to use Deep Packet Inspection (DPI) and Machine Learning (ML)/Artificial Intelligence (AI) to classify the application and user data.

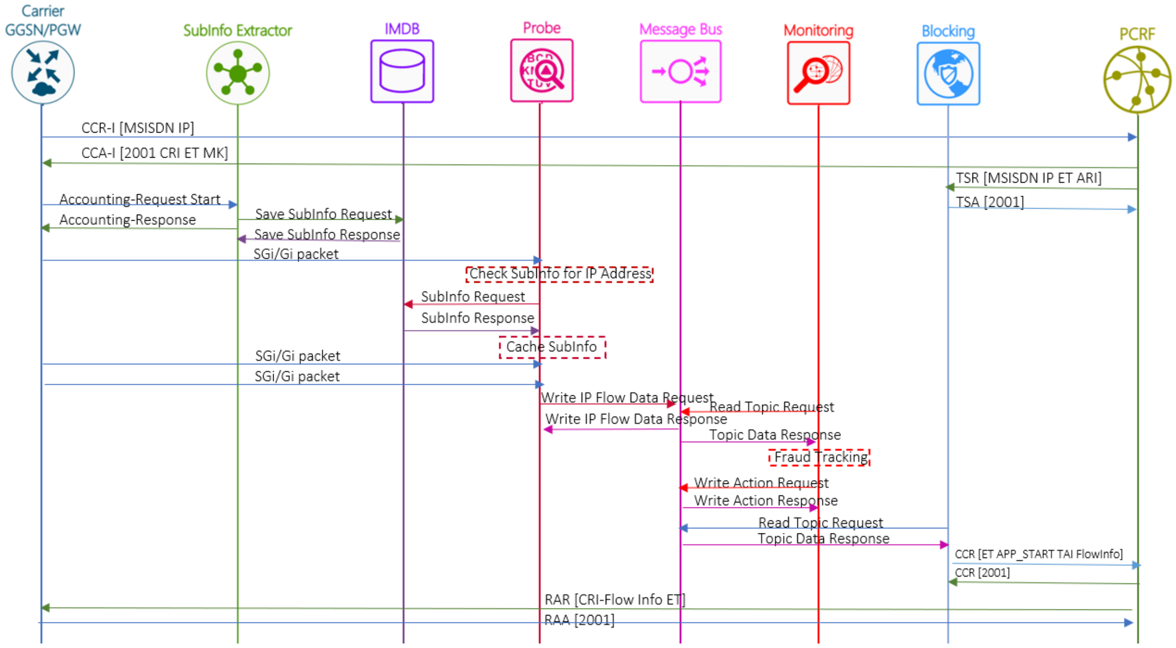


Figure 3.18: Communication flow of network elements to the PCRF

1. The subscriber activates the session, Gateway GPRS Support Node (GGSN)/ Packet Gateway (PGW) sends Accounting-Request Start to SubInfo Extractor with MSISDN: IP mapping and other session-related information.
2. SubInfo Extractor parses data and saves a new session object in the IMDB.
3. SubInfo Extractor confirms accounting and starts by sending Accounting-Response.
4. IMDB confirms data saving to SubInfo Extractor.
5. Probe receives a copy of the Gi/SGi IP packet from GGSN/PGW.
6. Probe checks session information for subscriber IP address.
7. Probe requests session information from IMDB.
8. IMDB provides session information to Probe.
9. Probe caches session information.
10. Probe receives subsequent copies of Gi/SGi packets from GGSN/PGW.
11. Probe writes IP flow data to Message Bus.
12. Message Bus confirms the writing of IP flow data.
13. Monitoring reads new data from Message Bus.
14. Monitoring performs fraud and usage tracking.
15. Monitoring detects usage tracking threshold or fraudulent traffic and writes action messages to Message Bus.
16. Message Bus confirms the writing of action data.
17. Blocking reads new data from Message Bus.

18. Blocking performs external API calls and provision traffic information to Carrier Network Equipment (NE).
19. Carrier NE performs necessary actions to block traffic on packet core/PCRF.

The architecture follows a microservice architecture wherein each service provides one or more functions and interacts with other services via the Message Bus or REST-based API. Figure 3.19 shows the components that make up the proposed solution of the compute nodes, data storage, data exchange, and services.

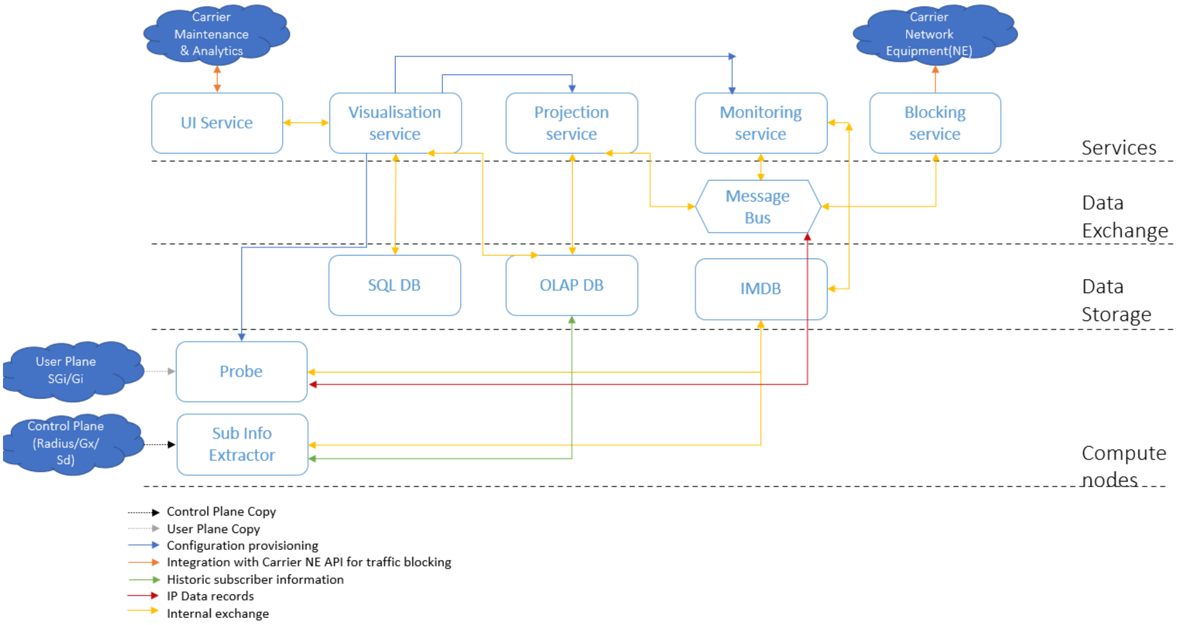


Figure 3.19: Architecture of data flow

UI Service

- The User Interface is represented by the WEB application and provides solution management and configuration capabilities for analytics purposes.

Visualisation Service

- The service provides unified API access to the reporting data stored inside OLAP BD as well as API’s for user and configuration management.

Monitoring Service

- The service is used to track traffic usage per rule and triggers the blocking service via Message Bus when a rule is breached.

Projector Service

- This service aggregates data from Probes, augments it with additional information, and inserts it into the OLAP DB.
- Data preaggregation is required to reduce the required bandwidth for site interconnection if central OLAP DB is used for data storage.

Blocking Service

- This service gets instructions to perform a blocking or notification action to Carrier NE via Message Bus. This could be from a Monitoring service or the probe.

Message Bus

- Message Bus is represented by a highly scalable, distributed messaging platform and is used between different components and services in order to communicate in a fault-tolerant way.
- Message Bus is deployed as a cluster for redundancy.
- Online Analytical Processing Database (OLAP DB)
- This database provides high-performance data ingestion and query processing for collected data by probes.
- The database also provides extremely efficient data compression to optimize disk usage and speed up querying for large data volumes.

In-memory Database (IMDB)

- This database is used as high-performance storage for subscriber session information lookups by Probes.

SQL Database

- This database is used to store and retrieve service configuration and user profiles.

Probe

- This component processes a feed of real-time User Plane traffic from the Carrier's subscribers that is represented by Gi/SGi interfaces of the mobile network and performs deep packet inspection to classify the traffic and find fraudulent traffic based on internal fraud detection methods.
- The component also correlates IP-flows with subscriber session information if this information was provisioned to SubInfo Extractor previously.
- Information about IP-flows with correlated subscriber session information is then inserted into the Message Bus for further processing and storing in the OLAP DB.

Subscriber Information Extractor (SubInfo Extractor)

- The component will process the user’s information from Carrier Control Plane nodes. The following is supported:
- GGSN/PGW or BRAS RADIUS flow
- GGSN/PGW Gx flow to/from PCRF
- TDF Sd interface to/from PCRF
- Depending on the configuration, the following can be extracted:
- MSISDN (RADIUS,Gx, Sd)
- IMSI (RADIUS, Gx, Sd)
- IMEI (RADIUS, Gx, Sd)
- Location including cell identifier (RADIUS, Gx, Sd)
- RAT (RADIUS, Gx, Sd)
- Charging-Rule-Base-Name
- User-Name (RADIUS)

To ensure the research meets the criteria, the study explores two blocking integration protocols of the GGSN/PGW RADIUS and PCRF Diameter. The first is RADIUS protocol used for the subscriber’s session information to map the Gi/SGi session mapping. If the GGSN/PGW nodes in Figure 3.20 are applied at many sites then RADIUS protocol will need to be implemented at each site independently.

Carrier
GGSN/PGW
SubInfo Extractor

Figure 3.20: RADIUS Integration point

When a packet data service fraud event occurs, the subscriber will have his data services disabled by the network node based on notifications sent by the Blocking Service. This will be done using the following process:

- Notification is sent to Fusion Middleware (FMW) via a Simple Object Access Protocol (SOAP) API to apply certain actions to subscriber or subscriber profiles on the network equipment.
- Notification is sent to SAMS via a file transfer aggregating records for the last 5 minutes on 5 minutes-based time intervals.
- Notification is sent to SAMS via a file transfer aggregating records including FMW result codes for the last 5 minutes on 5 minutes-based time intervals.

- File Names will be ()_DATA_FRAUD_<YYYYMMDDHHMM> and ()_DATA_FRAUD_RESULTS_<YYYYMMDDHHMM>

This integration point is used to notify and prevent fraud,
Data services charging bypass – once a charging bypass is detected, the Blocking Service will send notifications to FMS and SAMS.

Fair Usage Policy – once the subscriber reaches the preconfigured threshold for the preconfigured traffic matching profile on Blocking Service side, Blocking Service will send notifications to FMS and SAMS.

The second is the Diameter protocol used between the Blocking and PCRF nodes in Figure 3.21, the integration of Diameter is via the Sd interface to:

Extract subscriber session information – IP: MSISDN mapping, IMSI, IMEI, etc.

Provision 4tuple information about fraudulent calls to PCRF – based on this information PCRF dynamically creates a new rule on GGSN/PGW equipment acts as Policy and Charging Enforcement Function (PCEF) with 4tuple information and low available bandwidth.

The more reliable protocol to use is Diameter since its part of the Transmission Control Protocol (TCP) been a connection-oriented protocol and its ability to scale out.

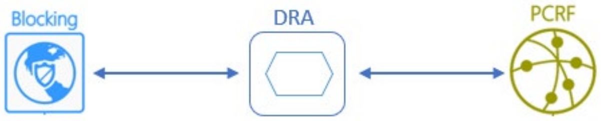


Figure 3.21: Diameter Integration point

3.6 Compliance Specifications

For the research to meet compliance standards and international specification standards the following RFC and 3GPP standards were reviewed and to align the solution to Table 3.4. The design takes into consideration the standards and specifications to meet compliance laws.

Table 3.4: Compliance for the design and the study

Specification	Name
RFC 6733, RFC 3588	Diameter Base Protocol Allows authentication, authorization, and account services on the 3G and 4G networks for network access and mobility. Diameter protocol will be used for the solution.

RFC 8506, RFC 4006	Diameter Credit-Control Application The networking protocol of the Diameter application is proposed to implement real-time credit control for the user of the service. The applications allow: • Quota management • Simplified debit and credit checks • Remaining balance checks • Price/tariff inquiries
RFC 7155, RFC 4005	Diameter Network Access Server Application To provide access to a network by receiving and forwarding authentication and authorization requests from the subscriber of the service.
RFC 3162	RADIUS and IPv6 The networking protocol authorizes and authenticates subscribers when accessing the network. It consists of rules that control the operation. RADIUS is implemented to make connections between devices and prevent unauthorized users from entering the network.
RFC 2866	RADIUS Accounting The protocol is responsible for receiving the accounting request and returning a response to the client and to notify if the request was successfully received. The RADIUS server can also act as a proxy client to other servers.
RFC 5176	Dynamic Authorization Extensions to Remote Authentication Dial-In User Service (RADIUS) From specification RFC2865, RADIUS protocol will not cater on support unsolicited messages sent from the purposed server to the Network Access Server(NAS), however with addition of RADIUS commands to enable unsolicited messages to be sent from the NAS, the extended commands provide high-level support for Disconnect and Change-of-Authorization (CoA) packets.
RFC 2865	Remote Authentication Dial-In User Service The protocol carries authentication, authorization, and configuration information between a NAS that desires to authenticate its links and a shared authentication server. The protocol suffers degraded performance and lost data when used in large-scale systems.
RFC 1945	Hypertext Transfer Protocol -- HTTP/1.0 The HTTP is an application-level protocol consisting of lightness and speed for distribution, collaborative, hypermedia information systems. The protocol is a stateless, object-oriented protocol that is used for server names and distributed object management systems. A feature of HTTP is the typing of data representation, allowing systems to be built independently.
RFC 2616	Hypertext Transfer Protocol -- HTTP/1.1 The specification defines more stringent requirements than HTTP/1.0 to ensure reliable deployment, the improved protocol allows messages to be formatted by MIME-like messages.
RFC 7540	Hypertext Transfer Protocol Version 2 (HTTP/2)

	HTTP/2 provides more efficient use of network resources and lower latency by introducing header field compression and allowing multiple concurrent exchanges on the same connection.
RFC 1034	Domain names - concepts and facilities The RFC introduces network domain style names, which are used for internet mail and host address support, and the servers used to implement domain name facilities as well as protocols.
RFC 1035	Domain names - implementation and specification The protocol serves for naming resources at which the names are usable in different hosts, networks, protocols, internet, and administrative organizations. A domain name is used to argue a local agent known as a resolver that is responsible for hiding the distribution of data among name servers from the user.
RFC 8499, RFC 2308	DNS Terminology The specification contains a collection of various DNS-related terms. It's a combination of used naming schemes for objects on the internet, and a distributed database representing the names and certain properties.
RFC 2246	The TLS Protocol Version 1.0 The specification of Transport Layer Security(TLS) protocol provides communications protection over the internet network, the protocol allows the client and server applications to communicate by preventing eavesdropping and tampering.
RFC 4346	The TLS Protocol Version 1.1 The protocol provides privacy and data integrity between two communicating applications, it consists of TLS Record Protocol and TLS Handshake. The connection is private, symmetric cryptography is used for data encryption, The connection is reliable, Keyed MAC provides integrity checks.
RFC 5246	The TLS Protocol Version 1.2 The protocol version 1.2 provides improved flexibility and negotiation of cryptographic algorithms, the main changes are: • Introduction of cipher-suite-specified PRFs use of P_SHA256, • Introduction of single hash that explicitly specifies the hash algorithm used, • The ability to specify the signature algorithms to accept, for cleanup to the client's and server's ability. • TLS Extensions and AES Cipher Suites integrated with TLSEXT and TLSAES, • Tighter checking of Encrypted Pre Master Secret version no.
RFC 8446	The TLS Protocol Version 1.3 The cipher suite concept has been changed to distance the authentication and key exchange mechanisms from the record protection algorithm and a hash to be used with both the key derivation function and handshake message authentication code (MAC).

	A zero round-trip time (0-RTT) mode was added to the connection setup. The public-key-based key exchange mechanisms are used to provide forward secrecy, static RSA and Diffie-Hellman cipher suites which is removed. Handshake messages are encrypted, the new Encrypted Extensions message allows various extensions. The key derivation functions have been redesigned to allow easier analysis by cryptographers due to improved key separation properties.
RFC 793	Transmission Control Protocol TCP provides a reliable host-to-host protocol between hosts in the data network and to interconnect systems.
RFC 4960, RFC 2960, RFC 3309	Stream Control Transmission Protocol SCTP is a well-known transport protocol operating on top of a connectionless network to provide: • To acknowledge error-free non duplicated transfer of user data, • To data fragmentation to conform to the discovered path MTU size, • Sequence delivery of user messages within streams, • Network-level fault tolerance with supporting multi-homing on both setup points.
RFC 3257	Stream Control Transmission Protocol Applicability Statement Provides the following: Data reliability Data sequence preservation and Flow and congestion control
3GPP 23.002	Network architecture The document shows the architectures of the 3GPP of UTRAN and GERAN radio access technologies.
3GPP 23.203	Policy and charging control architecture The architecture presents two main functions, Flow Based Charging, including charging control, online credit control, service data flows, application traffic, and Policy control, including gating control, QoS control, QoS signaling, etc.
3GPP 23.401	General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access The 3GPP standard provides stage 2 of the service description for the Evolved 3GPP Packet Switched domain, the specification covers roaming and non-roaming scenarios, including mobility between EUTRAN and pre-E-UTRAN radio access technologies, policy control, and charging. ITU-T describes a 3-stage step for the characterization of network standards.
3GPP 23.402	Architecture enhancements for non-3GPP access

	The (S5) and (S8) interfaces in the 4G architecture have GTP and PMIP variants, the GTP variant is documented in TS 23.401 and the PMIP variant is documented in this specification, The S2b reference point in the 4G architecture has also been defined to have both GTP and PMIP variants.
3GPP 29.060	General Packet Radio Service (GPRS); GPRS Tunneling Protocol (GTP) across the Gi/Gn and Gp interface The document presents the 2nd version of GTP used on: The Gi/Gn and Gp interfaces of General Packet Radio Service (GPRS), The Iu, Gi/Gn, and Gp interface of the Universal Mobile Telecom System.
3GPP 29.061	Interworking between the Public Land Mobile Network (PLMN) supporting packet-based services and Packet Data Networks (PDN) The standard shows the requirements for traffic environment interworking on: PLMN and PDN PLMN and PLMN
3GPP 29.210	Charging rule provisioning over the Gx interface The Gx reference point is for provisioning service data flow-based charging rules between the Traffic Plane Function (TPF) and Charging Rule Function (CFR), The protocol to be used between TPF and CRF over the Gx reference point, The information is to be exchanged between the TPF and CRF over the Gx interface.
3GPP 29.212	Policy and Charging Control (PCC); Reference points The standard presents the stage 3 specification of the Gx, Gy, and Sd reference points, the Gx reference point lies between the Policy and Charging Rule Function and the Policy and Charging Enforcement Function. The Gy interface is between the Policy and Charging Rule Function and the Bearer Binding and Event Reporting Function. The Sd interface is between the Policy and Charging Rule Function and the Traffic Detection Function.
3GPP 29.213	Policy and charging control signaling flows and Quality of Service (QoS) parameter proposal on mapping The specification provides details of the flows of Policy and Charging Control (PCC) over the Diameter-based Rx, Gx, Gy, Sd, Sy, S9, Nt, Diameter-based St and Np interfaces and the relations with signaling flows over Gn/Gp, S4, S5/S8, S2a and S2c interface. The specification describes the PCC reference architecture for roaming and non-roaming, The mapping of QoS parameters including SDP, UMTS parameters and authorization parameters, and The PCRF addressing using the DRA.
3GPP 32.299	Communication network management; Charging management; Diameter charging applications

	The standard is the Technical Specification of charging functionality and charging management in the GSM/UMTS network. To detail online charging on real-time charging messages, To detail online and offline charging for those domains, The interface that is used in the charging framework to transfer the charging information.

3.7 Introduction to 5G Standalone (5GSA)

The introduction of 5GSA presents new connectivity challenges, as a site supporting 5GSA may need to connect to three different core elements depending on whether the terminal is in NSA or SA mode: AMF and MME within the Single 5GC, and legacy MME within the legacy EPC. However, this triple core connectivity can be implemented by addressing the following aspects in different domains:

Core

- 5GC Blueprint: Defined to support NSA and SA coexistence.
- Control Plane Connectivity:
  - Legacy EPC (MME)
  - EPC embedded in the Single Core (New MME in SPC)
  - 5GC embedded in the Single Core (AMF)
- User Plane Connectivity:
  - Legacy EPC (S/P GW u)
  - 5GC (SPC performing roles for UPF and EPC S/P GW u)
- Traffic Routing: All traffic from the core towards gNB should be routed through the SecGW.

Transport

- Site Connectivity: Supported by the transport network to connect to three cores.
- Standard Routing: Transport needs to perform standard routing of gNB traffic/signaling to the core and vice versa.

Radio

- Base Stations: Current gNBs support dual mode NSA+SA.
- Software Dependency: Required to support NSA+SA.
- IP Endpoints: gNB needs to define IP endpoints for each core it connects to.

Enhancements and Benefits of 5GSA

5GSA and 5G Core enhance current 4G Core and Radio Technologies, significantly improving performance, introducing new services, and providing a richer experience. 5GSA is deployed as an overlay with the existing 4G network and will interwork with existing

IMS/Voice, User Management and Authentication (e.g., CSDB), Charging, and Lawful Interception.

New 5G Terminals

New terminals are needed to support 5GSA, including 5G Carrier Aggregation on SA (initially between 3.5GHz and a 5G-enabled FDD band). These terminals also support 2G/3G/4G/5GNSA. Devices with Carrier Aggregation capabilities are essential for a good 5GSA experience.

New RAN Software

To maximize coverage at 3.5GHz and low band 5G, Carrier Aggregation (CA) is essential. This requires upgrading RAN with new software releases and ensuring at least one band with Dynamic Spectrum Sharing (DSS) on 3.5GHz sites and a minimum of two suitable bands with DSS on all other sites. Depending on site readiness, RAN modernization may be required.

New 5G Core

The new 5G Core, delivered as "Single Core," includes core network functions for all mobile access data technologies, including 2G, 3G, 4G, NB-IoT, and 5G. It is based on a new Service-Based Architecture where core network functions are built as microservices and connect via standardized APIs.

5GSA Benefits and Drivers

The introduction of 5GSA is based on three pillars: 5G Core, Radio Carrier Aggregation, and Dynamic Spectrum Sharing features, and new terminals. These elements bring tangible benefits to our current 4G Core and Radio Technologies, significantly improving performance and introducing multiple new and richer services for subscribers.

5G Core Benefits and Use Cases

Single Core Capability:

- Single core network for mobile access technologies and fixed
- Optimizes 4G/5G Core interworking
- Smooth traffic migration and capacity growth on the 5G Core
- Cloud-native environment
- Legacy core with limited 3GPP/industry development

Enhanced Service Capabilities:

- 5G Network Slicing
- Advanced 5G MEC functions

- More granular QoS control
- Superior 5G SA security
- Service-Based Architecture
- API exposure for NaaP

CHAPTER FOUR
Characteristics and Solution Scoping

4. Characteristics and Solution Scoping

Fraud detection plays a major role in networks, subscribers, and revenue, the digital era is exploding with new technologies such as 5G/6G and requires automated solutions to suppress harmful elements that can arise. The research looks at Fraud detection and Fraud blocking; to gain a better perspective into the problem, the study builds on its critical characteristics:

• The large internet pool of users	4.1
• Correlation of the Billing system	4.2
• Characteristics of the user/traffic profiles	4.3

- Sim registration and swop 4.4
- Aggregation/interface points in the communication network 4.5
- SDN location and specification 4.6
- Energy fraud identification 4.7
- Packet de-encryption 4.8
- Blocking integration and rules 4.9
- Test case(s) proposals 4.10

4.1 The large internet pool of users

Considering a mobile network consisting of millions of users, each region belongs to an internet pool with network nodes on the 4G and 5G network in Table 4.1 providing the naming/new nodes. The traffic flow is from the User Equipment (UE) to the internet plays a role in the detection of fraud and capturing the correct user/interface traffic.

Table 4.1: 4G and 5G Network elements

4G Network elements	5G Network elements
eNodeB	gNB
Mobility Management Entity (MME)	Access and Mobility Management Function (AMF)
Serving Gateway (SWG)	
Home Subscriber Server (HSS)	Userplane Function (UPF)
PDN Gateway (PWG)	Session Management Function (SMF)
Policy Charging Rule Function (PCRF)	Policy Control Function (PCF)

The aim is to map from the IP address handed to each device, the subscriber ID, the study looks at the Radio Access Network to link both. An example of an internet pool is mentioned below:

```
vvvv  Pool Name                      Start Address Mask/End Address Used      Avail
----- ------------------------------- --------------  --------------          ---------------
RG00  internet_pvt_8                 10.152.0.0  255.255.0.0             27996    37538
RG00  internet_pvt_6                 10.145.0.0  255.255.0.0             27662    37872
RG00  internet_pvt_5                 10.144.0.0  255.255.0.0             28140    37394
```

4.2 Correlation of the Billing system

A part of the study requires linking up the subscriber's usage on the network to the billing system also presented in Test Case 2 in section 4.10, a user should not have access to the network when blacklisted or bypassing the billing system. The billing system is built on the Online Charging System (OCS)/Converged Charging System (CCS) in Figure 4.1 of the 4G network and Figure 4.2 of the 5G network.

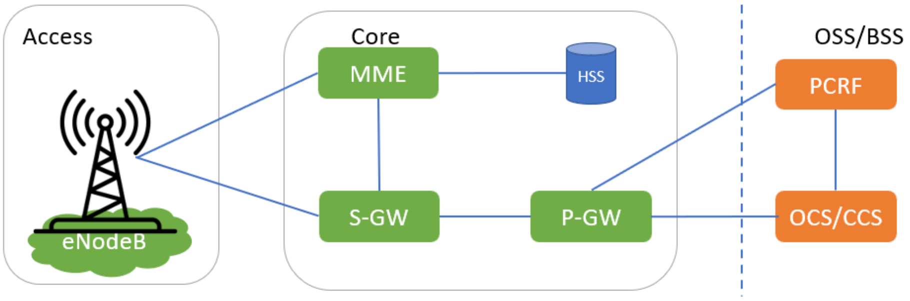


Figure 4.1: 4G Network with Billing system

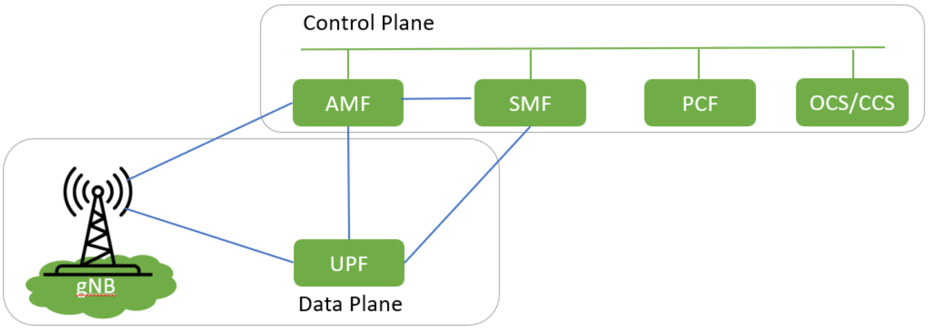


Figure 4.2: 5G Network with Billing system

The billing system plays a vital role in correlating the subscriber's account to the subscriber's usage on the network, the study requires looking at the billing and correct charges on the different tariff plans, the process is built with many components and provides challenges in the integration of billing and network use. A question arises if the OCS and CCS platforms is the same used for 4G and 5G networks. The research carefully looked at each technology to ensure protocols were aligned and compatible.

4.3 Characteristics of the User/Traffic Profiles

The architecture for a voice network and data network are different but share integration points. Figure 4.3 shows the platforms and the routing separation. The research deeply explores the characteristics of the user on social behaviour on data networks:

- The number of times in the week the phone is switched on and off,
- The number of sessions and time of each session,
- The applications or services subscribed to and used,
- The patterns of applications based on a week,
- Changes to the IP address when on the network,

The characteristics of the user on social behaviour for voice network:

- The number of times in the week the phone is switched on and off.
- The number of calls per day completed.
- The duration of a call.
- The pattern of calls and sessions with different contact numbers.

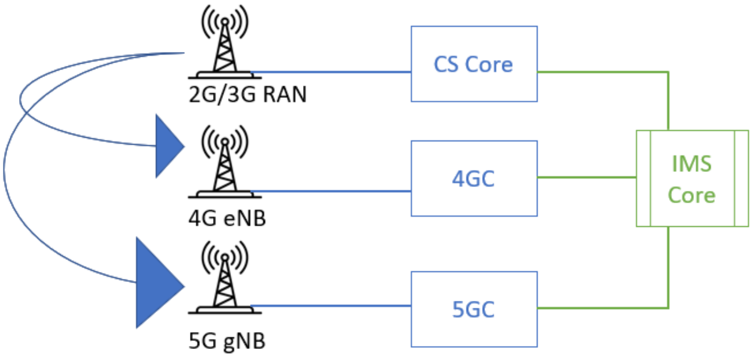


Figure 4.3: Integration to Circuit Switching (CS) platform from 2G/3G/4G/5G

The 2G/3G network is connected to the Circuit Switch platform for voice calls, and the 4G and 5G networks is a data network where the research is performed on, and the process of analysis is studied. The research required to understand the usage of different sectors on the network to profile and sample the patterns of usage in the following sectors:

- Online banking
- Transportation – Uber
- Online Gaming
- Online Retail
- Food – Mr Delivery
- Online schooling

Each sector shows a pattern of capacity trends from the network and is combined using Statista, from, https://www.statista.com/. In online Banking the platform serves for a 24-hour event, and reserves capacity for the banking users. Figure 4.4 show the usage of the banking platform over 2 days, the trend of banking is more random-stable state but within a threshold. The platform starts from 5:00 AM and takes an incline to 9:00 AM, from 9:00 AM the usage is more random at its peaks at 22:30 PM and takes a decline in using the platform at 01:30 AM.

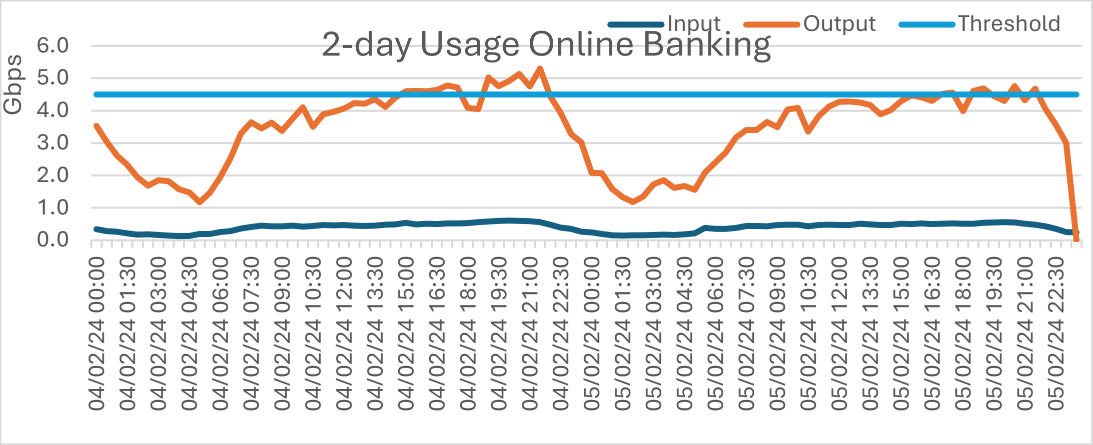


Figure 4.4: 2-day Usage for Online Banking

In Figure 4.5 it shows a 2-day usage for Transportation with Uber Service, the usage is defined by a smooth incline and the peak usage is between 19:30 PM to 21:30 PM.

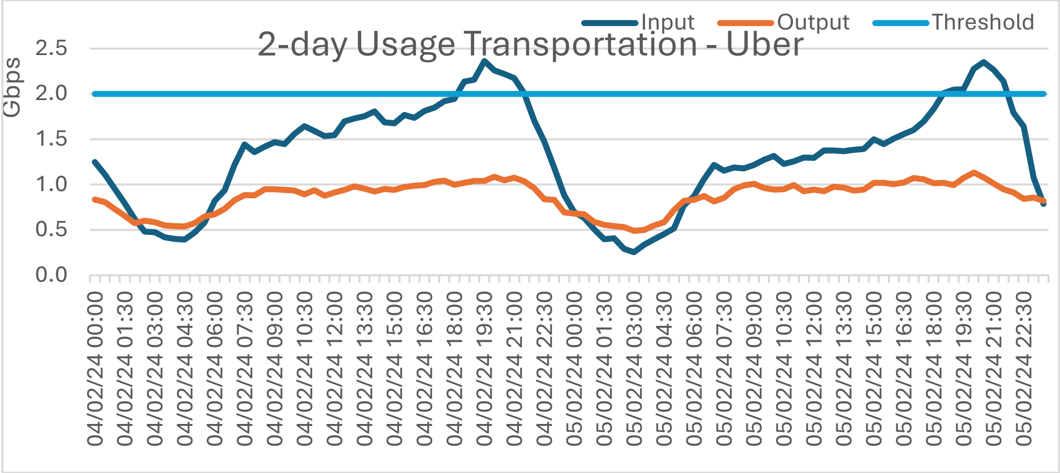


Figure 4.5: 2-day Usage for Transportation of Uber

Online Gaming using high throughputs of data, Figure 4.6 shows the usage at a steady incline from 8:00 AM to 22:00 PM. Gaming possesses the same trend each day.

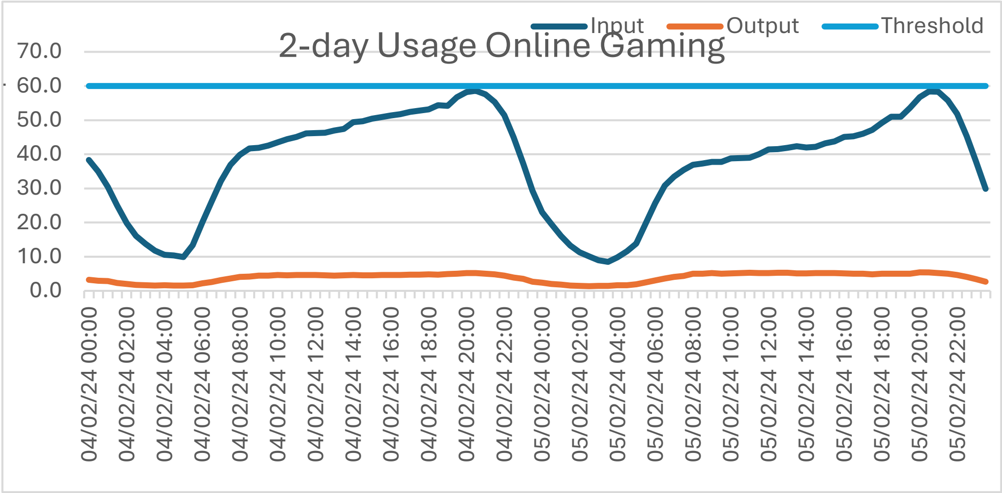


Figure 4.6: 2-day Usage for Online Gaming

The next usage is Online Retail, Figure 4.7 shows the growing trend of browsing or shopping from the internet. The capacity trend takes a steep incline from 5:00 AM to 9:00 AM, and then a moderate incline from 9:00 AM to 21:00 PM.

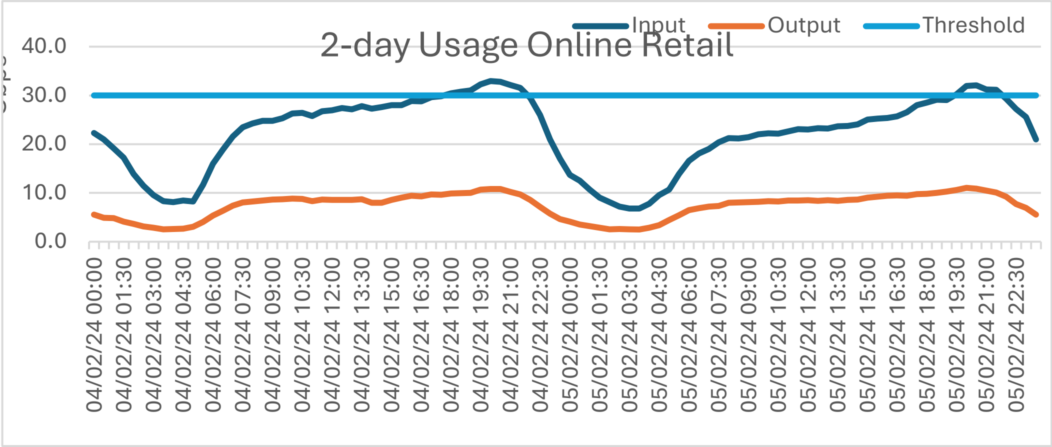


Figure 4.7: 2-day Usage for Online Retail

The next sector shows the usage of Mr Delivery application, Figure 4.8 shows the 2-day usage of the trend is slow-rise in the morning hours and slow-fall in the evening hours or described as parabola open down. The Mr Delivery is based on promotions that drive their business, the trend between 08:45 AM to 21:00 PM show the peak usage of the application.

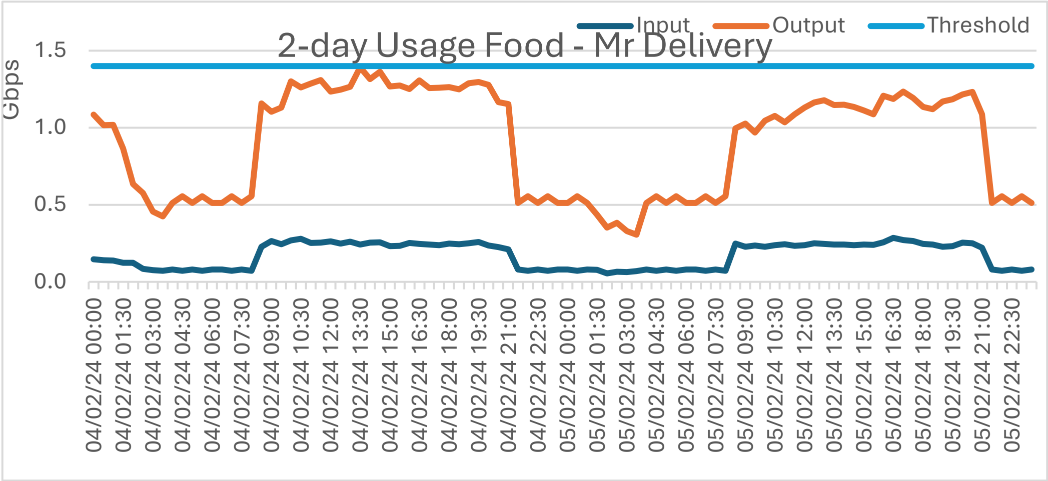


Figure 4.8: 2-day Usage for Food on Mr Delivery

The next sector is Online Schooling, which includes primary, secondary, and tertiary education. Figure 4.9 shows the peak usage between 9:00 AM and 21:00 PM and described by smooth incline.

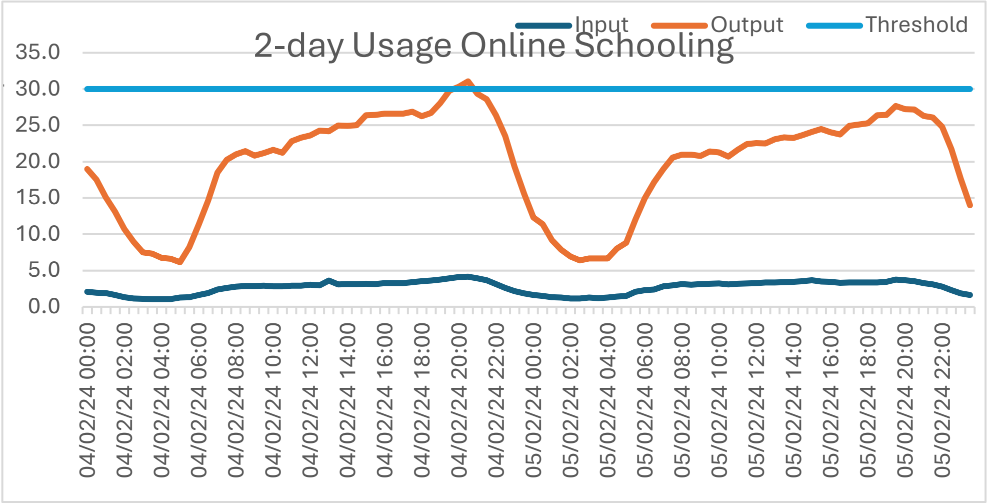


Figure 4.9: 2-day Usage for Online Schooling

4.4 Sim registration and swop

According to South Africa Telecommunications policy, the RICA Act was founded by Act 70 of 2002 to authorize a subscriber to have access to the network and to secure the system/network. The RICA process is easy and requires identity documents and proof of residence. However, the process can be flawed with a fake identity or using the identity of another person.

The study looks at entry points in the network:

- 4G and 5G contract sim
- 4G and 5G prepaid sim
- Application bundle
- Cloud provider bundle

Each plays a role in which the fraudster can access the network and utilize the other services not paid or accounted for.

4.5 Aggregation/interface points in the communication network

A Service Provider’s network consists of the Radio Access Network (RAN), Transmission (TX), and the Core, the three sectors are built of layers on the Open Systems Interconnect (OSI) model, for the purpose of the study and looking at Energy abuse or Service application abuse, the fundamental points is to identify the aggregation point in the network, a point of entry and exit. The Core consists of different platforms, namely IP Multimedia Core Subsystem (IMS), Billing, and Core Data Network. Within the Core Data Network, it has a routing and switching network that has breakout points to the internet. The research looks at the breakout points that route and control the traffic or services to the internet in Figure 4.10.
The aggregation points identified is between the Spine and Leaf switches where the user plane traffic or Gi traffic can be captured specifically for the research.

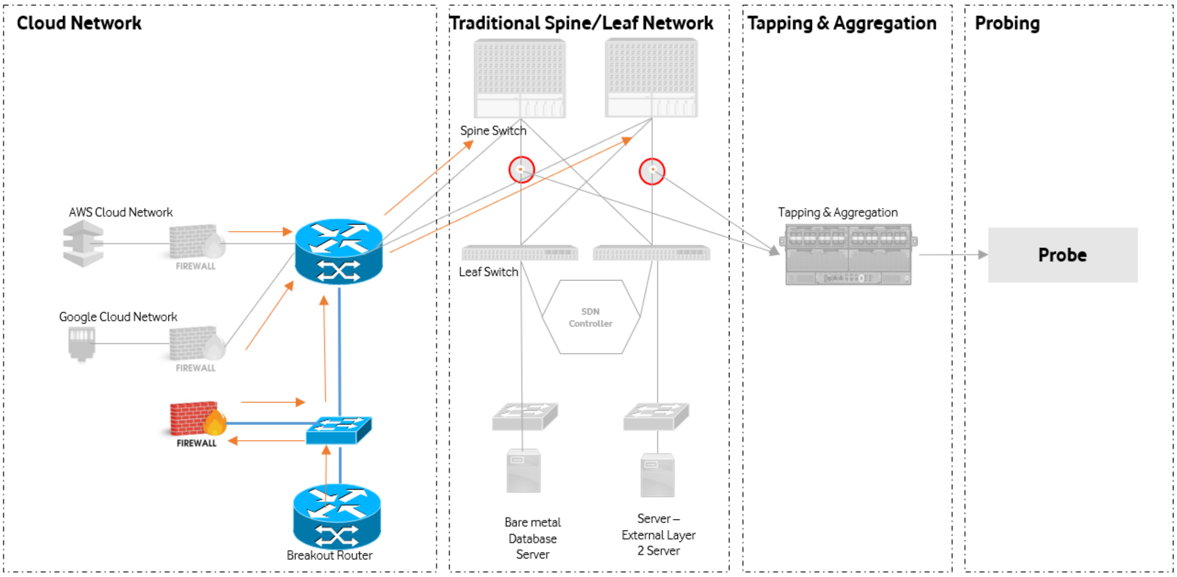


Figure 4.10: Architecture of network including Breakout point

The aggregation point is the location of where all traffic or services are routed, it becomes the point where the study taps the user plane traffic. The design of a spine and leaf topology it builds redundancy in the network from failure of links or failure of a service. The arrows indicate the flow of traffic from the cloud network or service provider network.

4.6 SDN location and specification

The study shows the SDN controller connected to the leaf switches, the physical connectivity remains the same, however the logical connectivity portraits changes and overlooks three domains in Figure 4.11 and discussed from Indarjit, Balyan, and Adonis, (2025d):

- Traffic is routed from the probes at different sites to the SDN controller, to note the SDN controller host routing tables of the nodes and links connected.
- The SDN controller will perform the traffic classification and fraud detection.

The SDN controller overlooks the Metro Network domain, Core Network domain, and Data Centre domain, for each a Virtual Private Network (VPN) and Segment Routing are created and applied, both protocol is of importance since during the process of traffic classification, traffic will be encrypted for the purpose of a safe/protected network, and also changes in routing play a role either via packet loss or packet duplication.

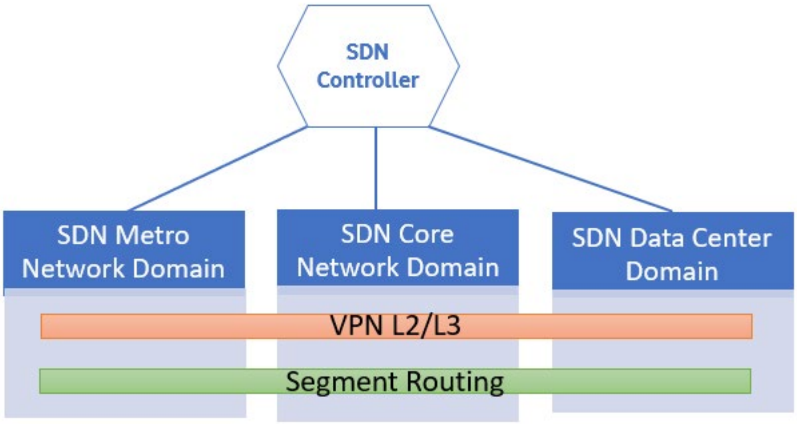


Figure 4.11: SDN Controller overlooking Metro domain, Core domain and Centre domain

The next layer involved is the protocols used to establish the network, the protocols fall either under Service or Transport Protocols. At which a brief summary is provided on each, the research needs to ensure that traffic is readable and within the correct format for traffic classification. Figure 4.12 shows the breakdown of Service Protocols, Transport Protocols, and Data-Plane, the 4G network is designed with the control and the data plane together, while 4G CUPS designs the both planes separated, and further 5G network encrypts the Service-Based Interfaces. Therefore, the research becomes complex, and each sector needs to be addressed or catered for.

Service Protocols are formulated with the following:

- Layer 2 Virtual Private Network (VPN) builds a topology of point-to-point connections that connect end users.

- Layer 3 Virtual Private Network (VPN), the Customer's Edge (CE) switch or router must be configured to exchange traffic with the Provider's Edge (PE) switch/router.

Transport Protocols are formulated with the following:

- Inter- Domain Traffic Protocols

These protocols manage the flow of data between different administrative domains or networks. Examples include:

- Border Gateway Protocol (BGP): The primary inter-domain routing protocol used to exchange routing information between different autonomous systems on the internet.
- Inter-Domain Routing Protocols: These handle routing policies and the distribution of routing information across multiple domains.

Intra-Domain Traffic Engineering Fast Re-Route

This involves optimizing the flow of traffic within a single domain or AS. It aims to use network resources efficiently and balance loads. Key components include:

- Resource Reservation Protocol (RSVP): Used in some networks for reserving resources for specific traffic flows.
- Traffic Engineering Protocols: Such as those using MPLS (Multiprotocol Label Switching) to control traffic paths.

Intra-Domain MPLS LSP

MPLS is used to create efficient data paths within a network. An LSP is a path through an MPLS network that data follows based on label switching. Key concepts include:

- Label Distribution Protocol (LDP): Used to establish and maintain LSPs.
- Constraint-Based Routing: MPLS can use constraints to make routing decisions based on resource availability and other factors.

IP Routing

This involves determining the path that IP packets take from the source to the destination across an IP network. Key elements include:

- Routing Protocols: Such as OSPF (Open Shortest Path First), EIGRP (Enhanced Interior Gateway Routing Protocol), and BGP for dynamic routing and path selection.
- Routing Tables: Tables that store network paths and routes used by routers to forward packets.

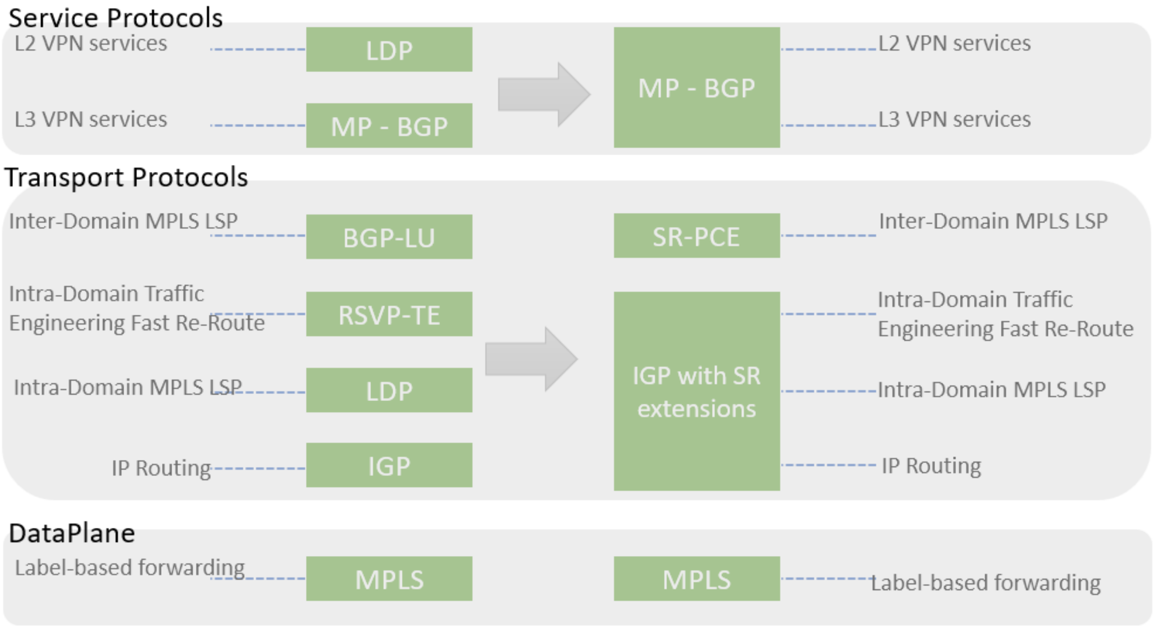


Figure 4.12: Service Protocols, Transport Protocols, and Data Plane

Segment Routing from Figure 4.13 is based on Source Routing, the control plane uses Segment Routing and the data plane uses Multi-Protocol Label Switching protocol in Figure 4.13 the process of Push, Swop, and Pop is showed, the Segment ID (SID) is created either by (Index + downstream router's SR Global Block (SRGB) outgoing label) else (Index + local SRGB for the local label) which is 16004. In Figure 4.10 the payload is routed from router to router, router 2 and router 3 advertise their loopback IP prefix with its SID=16004 which provides the best routing path.

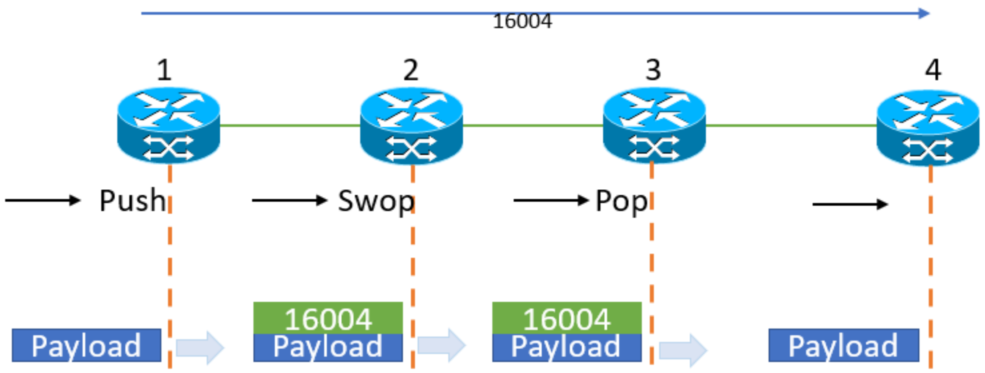


Figure 4.13: Multi-Protocol Label Switching Routing based on Push, Swop and Pop with Segment Routing

The SDN controller is built by a vendor or organization to provision services from the network since the design is based on virtual infrastructure, the layered architecture in Figure 4.14 provides service provision for end-to-end, implements automatic network deployment and unified O&M and meets other requirements.

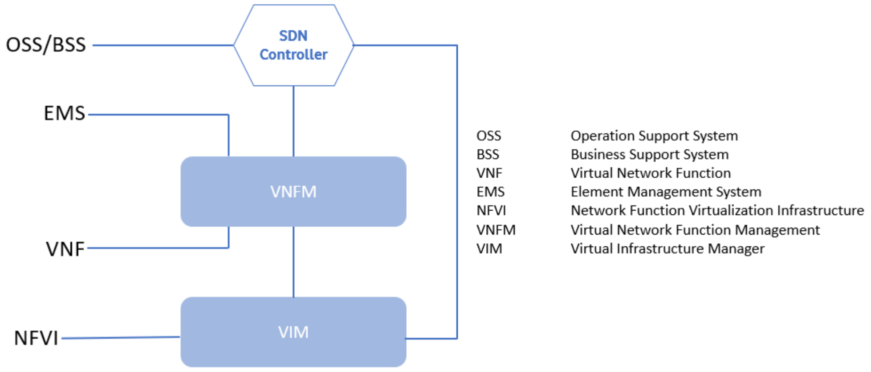

OSS/BSS
SDN Controller
EMS
VNFM
VNF
NFVI
VIM
OSS Operation Support System
BSS Business Support System
VNF Virtual Network Function
EMS Element Management System
NFVI Network Function Virtualization Infrastructure
VNFM Virtual Network Function Management
VIM Virtual Infrastructure Manager

Figure 4.14: Layered Software Defined Network on virtual platform for service provisioning

The SDN layered virtual network adopts an open architecture to help simplify and efficiently operate and maintain the NFV networks.

- For service provisioning,
- To provide network service life cycle management, and,
- Flexible applicability

Figure 4.15 shows a detailed layered Software-Defined Network architecture, showing the protocols used to communicate and its components:

- SDN – To manage the network service life cycle, including service deployment, scale in/out, and termination, and provide cross-DC resource scheduling capability.
- OSS/BSS – Serves as the support system of the live network and manages the VNFs and CNFs.
- EMS – Serve as the node management system, to monitor network outages at the Software as a Service and Infrastructure as a Service layer, to maintain the capabilities of the VNFs and CNFs.
- VIM – Serves as the Infrastructure layer management system of FusionSphere and FusionManager.
- VNFM – To manage the VNF life cycle for deployment, scale in/out, and termination.

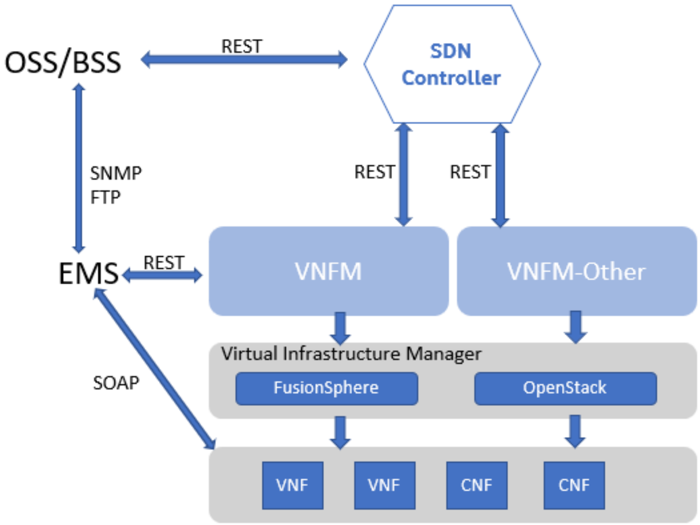


Figure 4.15: Detailed layered Software Defined Network setup

The research explores the 4G and 5G networks, the 5G network is built using Virtual Network Function (VNF) and Container Network Function (CNF), and on CNF deployment the following metrics are taken:

To prepare the deployment environment, the following steps occurred:

- Configuring the switch underlay network and installing and pre-configure.
- Deploy and install SDN software, PaaS, and Life Cycle Management.
- Manually create VPCs, host groups, and disk types on FusionSphere.
- Create tenants and projects and modify credentials on PaaS.

4.7 Energy Fraud identification

A fraudster seeks to bring harm to the Smart Grid and dismantle the operational aspects for their personal agenda, their objective is to utilize the Smart Grid without payment or cause harm to the infrastructure by Indarjit, Balyan, and Adonis, (2025d). For the purpose of the research, two main aspects are studied, Energy abuse and Application abuse. Both environments are of large scale and with many access points consisting of multiple systems, the study focuses on the communication network and identifies the aggregation point.

From the aggregation point, the traffic can be sampled to further identify traffic profiles for each environment. On Energy abuse and Service abuse, based on a sample of users the average trend on vlan(x) is shown in Figure 4.16, the peak utilization is important and needs to be understood.

a and b indicate normal utilization peaks occurring twice in a 7-day count, c indicates failover of traffic, a network is built on redundancy, failover occurs for a certain duration of time and the capacity integrated serves to accommodate user traffic.

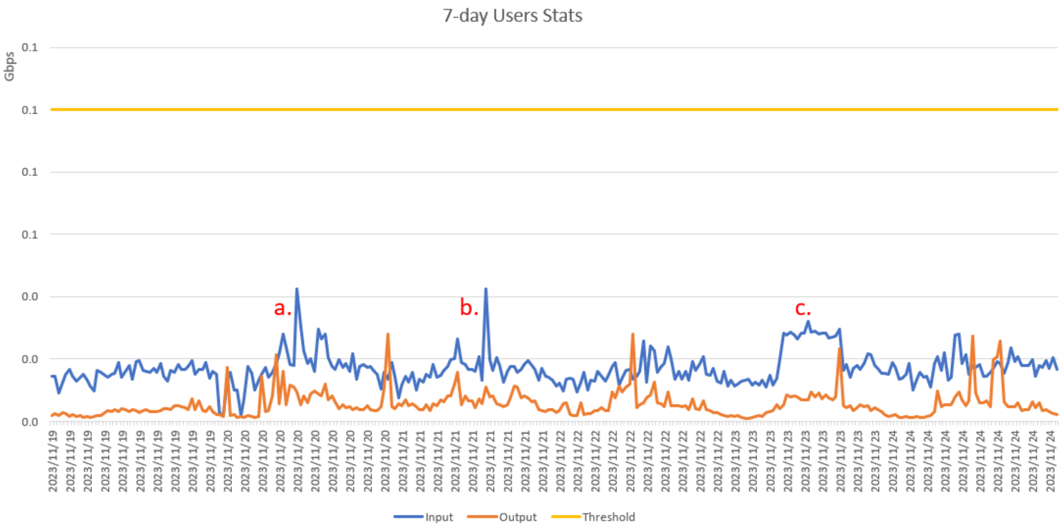


Figure 4.16: 7-day user stats on average of a sample of (x)

Figure 4.17 shows the footprint of abuse on the network over a 7-day average per period, d shows fraud on the network, the characteristics of fraud are:

The utilization exceeded the highest peaks over the 7-day period,

The high utilization over a longer period compared to other active sessions,

The changes within data usage considering hard-stops,

The policy of random peak usage.

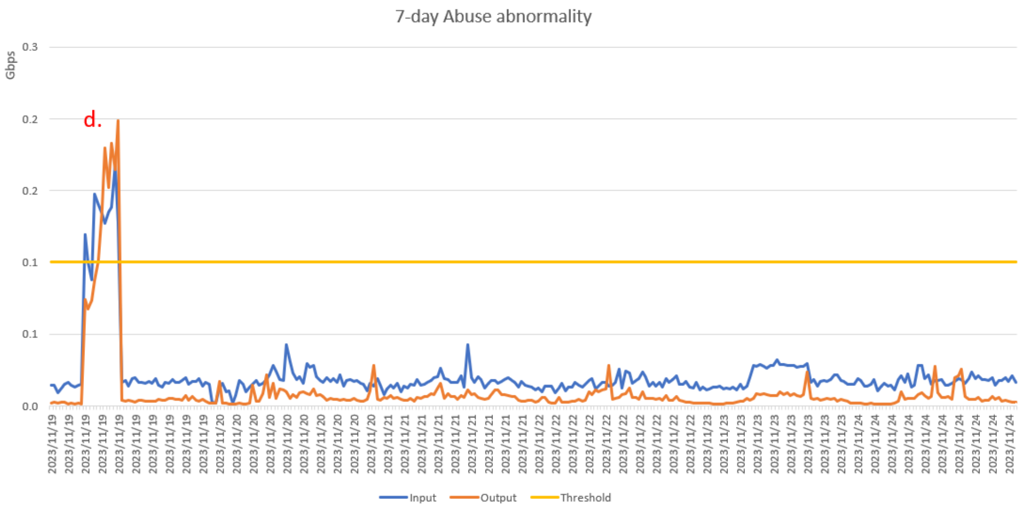


Figure 4.17: 7-day user stats showing network abuse

It was found when fraud is detected it can be two categories:
Ghost traffic, The Tx is indicated as 0. The sample is showed from the testing,

```
[local]GGPS03# show port util table
------ Average Port Utilization (in mbps) ------
Port Type Current 5min 15min
                          Rx    Tx  Rx     Tx   Rx      Tx
----- ---------------------- ------- ------- ------- ------- ------- -------
5/13 10G Ethernet      3002    0   2855   0    2817   0
```

On the Actual fraud traffic, the TX and RX both have a value.
Fraud requires to be looked at within a 2 to 7 days, the pattern of a fraudster can be categorized by random sampling.

4.8 Packet de-encryption

The expansive network of mobile devices poses significant security challenges. Encryption is a crucial tool for safeguarding data integrity and privacy, and each generation of mobile technology has introduced more advanced encryption methods to counter emerging threats and enhance user security.

2G (GSM) The second generation (2G) of mobile networks, known as GSM (Global System for Mobile Communications), marked the onset of mobile communication. In GSM, encryption was employed to protect user data during sending and receiving. The main algorithms used were A5/1 and A5/2:

- A5/1: This stream cipher was used in GSM networks to encrypt both voice and data. Although more secure than A5/2, A5/1 was still vulnerable to attacks due to the limitations of cryptographic techniques at the time.
- A5/2: A less secure variant intended for export markets, A5/2 was found to be susceptible to various attacks, making it less effective in protecting data.

3G (UMTS) The third generation (3G) of mobile networks introduced UMTS (Universal Mobile Telecommunications System), which improved on the security framework of GSM. Key advancements in 3G include:

- KASUMI: Is a block cipher used in 3G networks to encrypt both voice and user data. KASUMI was designed to be more resistant to cryptographic attacks than the A5 series.

- UEA1 and UIA1: These algorithms provide encryption and integrity protection for user data in 3G networks, enhancing security against tampering and interception.

While 3G marked a significant step forward in security, it set the stage for even more robust measures in subsequent generations.

4G (LTE) The fourth generation (4G) of mobile networks, known as LTE (Long-Term Evolution), brings major improvements in both speed and security. Key features of LTE include:

- Advanced Encryption Standard (AES): AES, a symmetric key encryption algorithm, is used in LTE networks to offer strong encryption for user data, with AES-128 and AES-256 providing robust protection against unauthorized access.
- Evolved Packet System (EPS) Encryption: LTE employs EPS for encryption, involving multiple layers of security, including encryption for both the user plane (data) and control plane (signaling).
- Integrity Protection: LTE includes integrity protection algorithms to ensure data is not altered during transmission.

These advancements in LTE represent a significant leap in securing mobile communications.

5G The fifth generation (5G) of mobile networks introduces enhanced security and performance. Critical features of 5G encryption include:

- Enhanced Encryption Algorithms: 5G utilizes AES-256 for encrypting user data, offering even stronger security than 2G/3G/4G generations.
- Network Slicing Security: 5G introduces network slicing, which creates multiple virtual networks within a single physical network, each with its own security measures. This adds a layer of customization and control.
- Improved Key Management and Authentication: 5G enhances key management and authentication protocols to ensure secure communication and data protection across different network environments.

5G represents the forefront of mobile network security, addressing the complexities and demands of modern digital communication.

Decryption Process for 4G LTE

In LTE networks, encryption is applied to both the user plane (data) and the control plane (signaling). AES (Advanced Encryption Standard) is used for encryption. The proposed decryption process follows:

a. Encryption Key and Algorithm:
    - The encryption key is derived during the authentication and key agreement process.
    - AES is used with a specific key length (typically AES-128 or AES-256).
b. Ciphering:
    - Data transmitted over the LTE network is encrypted using the AES algorithm.
    - The encryption key and initialization vector (IV) are used to encrypt the data.
c. Decryption Process:
    - The process to extract the Key and initialization vector: Obtain the encryption key and IV from the context (usually established during the authentication process).
    - To apply AES Decryption: Use the AES algorithm to decrypt the encrypted traffic. The same key and IV used for encryption are required for decryption.
    - And removal of Padding: If padding was added during encryption, it should be removed after decryption.

The below presents how to decrypt LTE traffic using AES in Python code:

```
python
from cryptography.hazmat.primitives.ciphers import Cipher, algorithms, modes
from cryptography.hazmat.primitives import padding
from cryptography.hazmat.backends import default_backend

def decrypt_4G_traffic(key, iv, ciphertext):
    """
    Decrypts 4G encrypted traffic using AES decryption.

    :param key: The AES key used for decryption. Must be 16 or 32 bytes long (AES-128 or AES-256).
    :param iv: The Initialization Vector used during encryption.
    :param ciphertext: The encrypted data to be decrypted.
    :return: The decrypted plaintext data.
    """
    # Ensure the key length is either 16 bytes (AES-128) or 32 bytes (AES-256)
    if len(key) not in [16, 32]:
        raise ValueError("Invalid key size. Key must be 16 or 32 bytes long.")

    # Create the AES cipher object
    cipher = Cipher(algorithms.AES(key), modes.CBC(iv), backend=default_backend())
    # Decrypt the ciphertext
    decryptor = cipher.decryptor()
    padded_plaintext = decryptor.update(ciphertext) + decryptor.finalize()

    # Unpad the plaintext
    unpadder = padding.PKCS7(algorithms.AES.block_size).unpadder()
    plaintext = unpadder.update(padded_plaintext) + unpadder.finalize()

    return plaintext.decode()

# Example usage
key = b'\x01' * 16  # Example AES-128 key
iv = b'\x02' * 16   # Example IV
ciphertext = b'\x03' * 32  # Example ciphertext (must include actual encrypted data)

# Decrypt the message
decrypted_message = decrypt_4G_traffic(key, iv, ciphertext)
print(f"Decrypted Message: {decrypted_message}")
```

Decryption Process for 5G

5G networks use advanced encryption and key management. The basic principles of decryption are similar but with enhanced security features:

a. Encryption Key and Algorithm:
    - 5G uses AES-256 for encryption, with a more complex key management and distribution process.

b. Ciphering:
    - Data is encrypted with AES-256, and the key is derived from a more secure key management framework.

c. Decryption Process:
    - To extract the Key and initialization vector: Obtain the AES key and IV used for encryption.
    - To apply AES Decryption: Use AES-256 to decrypt the data. Both key and IV are required.
    - And removal Padding: If padding was used, it should be removed after decryption.

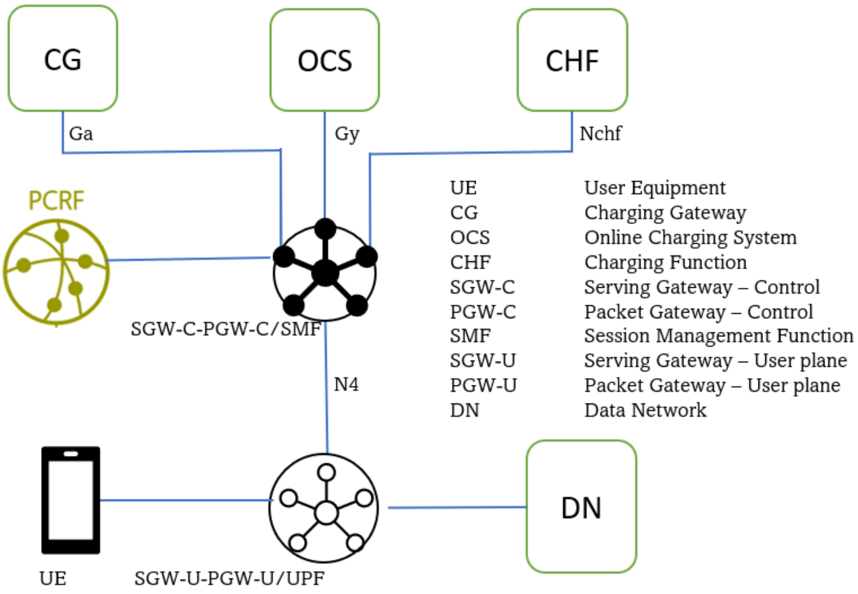
The below presents the python code similar to the LTE decryption but with AES-256:

python

```
def decrypt_5g_traffic(key, iv, ciphertext):
    """
    Decrypts 5G encrypted traffic using AES-256 decryption.

    :param key: The AES key used for decryption. Must be 32 bytes long (AES-256).
    :param iv: The Initialization Vector used during encryption.
    :param ciphertext: The encrypted data to be decrypted.
    :return: The decrypted plaintext data.
    """
    # Ensure the key length is 32 bytes (AES-256)
    if len(key) != 32:
        raise ValueError("Invalid key size. Key must be 32 bytes long for AES-256.")

    # Create the AES cipher object
    cipher = Cipher(algorithms.AES(key), modes.CBC(iv), backend=default_backend())

    # Decrypt the ciphertext
    decryptor = cipher.decryptor()
    padded_plaintext = decryptor.update(ciphertext) + decryptor.finalize()

    # Unpad the plaintext
    unpadder = padding.PKCS7(algorithms.AES.block_size).unpadder()
    plaintext = unpadder.update(padded_plaintext) + unpadder.finalize()

    return plaintext.decode()

# Example usage
key = b'\x04' * 32  # Example AES-256 key
iv = b'\x05' * 16   # Example IV
ciphertext = b'\x06' * 32  # Example ciphertext

# Decrypt the message
decrypted_message = decrypt_5g_traffic(key, iv, ciphertext)
print(f"Decrypted Message: {decrypted_message}")
```

4.9 Blocking Integration and Rules

The charging platform for subscribers allows the Service Providers to calculate the usage amount from the tariff structure, using online, offline, and converged charging. From the architecture in Figure 4.18 the user plane and control plane exchange service statistics with the OCS server.



UE	User Equipment
CG	Charging Gateway
OCS	Online Charging System
CHF	Charging Function
SGW-C	Serving Gateway – Control
PGW-C	Packet Gateway – Control
SMF	Session Management Function
SGW-U	Serving Gateway – User plane
PGW-U	Packet Gateway – User plane
DN	Data Network

Figure 4.18: Charging nodes communication

Online charging is between PGW-C with the OCS server over the Gy interface, Offline charging is between the SGW-C and PGW-C with the CG over the Ga interface, and Convergent charging is between the SMF and CHF over the N40 interface for the 5G network.

Charging is the process whereby:

- The subscriber users the data services from the network, and the network nodes SGW-U/PGW-U/UPF gather the flow characteristics such as traffic volume, online duration, and events to the SGW-C/PGW-C/SMF.
- The Control traffic of the SGW/PGW/SMF then interacts with the OCS server and generates charging data records.
- The subscriber initiates the request on the network. The network nodes determine to use online, offline, or converged charging to charge the user by PDU session on charging characteristics and the APN/DNN in a Create PDP Context Request/Create

Session Request/PDU Session establishment Request message, as well as PCRF/PCF.

- If online or convergent charging is used, the PGW-C/SMF creates a charging session on the OCS server and applies for quotas and charging events based on the Rating Group (RG), if offline charging is used, there is no need for quota.
- The SGW-C/PGW-C/SMF can map the RG to a Usage Reporting Rule (URR) ID. Followed by delivering the charging rules, quotas, and charging events to the user plane of the SGW/PGW/UPF of the N4 interface.

When the user accesses a service, the network functions SGW-U/PGW-U/UPF parses packets, to collect statistics on the traffic volume and online duration to the Control plane of the SGW/PGW/SMF. Followed by interworking with the OCS server/CG/CHF.

For offline charging, the SGW-C/PGW-C generates offline CDRs and sends them to the CG. For online charging, the PGW-C reports the charging data to the OCS service by applying new quotas. And for convergent charging, the SMF network function reports charging traffic to the CHF by applying new quotas.

The blocking rule is applied to the following Fraud types in Call fraud, Digital certificate fraud, Evasive protocol fraud, and Fair usage policy in Table 4.2.

Table 4.2: Fraud types in Tariff plans

Tariff plans				
Fraud type	**Data bundle**	**Zero rated**	**Reverse Billing**	**Uncategorized traffic**
Call Fraud	Degrade the session quality			
Digital certificate fraud	Block MSISIDN for data	Block MSISIDN for data	Block MSISIDN for data	log
Evasive protocol fraud	Block MSISIDN for data	Block MSISIDN for data	Block MSISIDN for data	log
Fair usage policy	Block MSISIDN for data	Block MSISIDN for data	Block MSISIDN for data	log

4.10 Test Case(s) proposal

4.10.1 Test Case 1

Fraud detection with an unpaid token on the Communication Network

Energy sector

Communication sector

Service Application sector

To use traffic classification and identify for each sector the patterns of a fraudster. To create an algorithm by Machine Learning on the detection.

Fraud detection with an unpaid token on the network requires bringing the network's subscriber platform and the billing platform together also discussed in Indarjit, Balyan, and Adonis, (2025d). The mapping of the subscriber to their billing becomes a fundamental point for the research and its implementation. The first step is to explore what components are active on the billing platform in Figure 4.19.

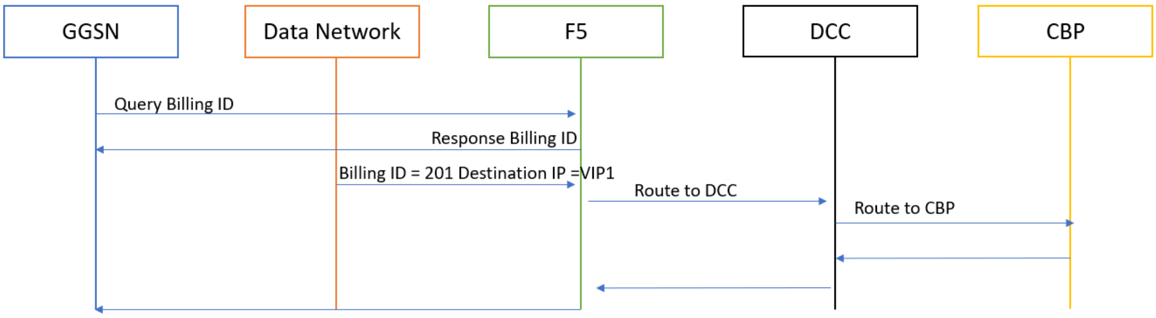


Figure 4.19: Billing flow - High level

The components of Billing flow are shown below:

CBP — Convergent Billing Point implements rating, charging, and accounting functions and supports both online charging and offline charging, also providing real-time QoS control, this is triggered based on the threshold configured in the tariff.

DCC Proxy — The function supports the specific demands of Diameter Charging – a dedicated online mediation instance, to take control of any internal routing of Diameter traffic based on subscriber.

GGSN — Manages data sessions and integrates with Online Charging System for real-time data rating and charging.

F5 — F5 Distributed Cloud Services Billing service enables a subscriber to understand usage reports, quotas, and pricing, obtain usage reports, and switch between subscription plans.

The test case needs to extract if the subscriber is billing correctly by performing checks from the Convergent Billing Point, where the Convergent Charging System (CCS) resides.

4.10.2 Test Case 2: Fraud detection utilizing the network

To evaluate the effects of fraud users for traffic classification and build an algorithm to automate the detection.

To build an accurate traffic classification model the following considerations need to be taken:

- Inaccurate traffic classification, from the complexity of different devices and firmware version upgrades, Service Providers make use of encryption to preserve customer traffic and personal data,

- Misclassification, considering traffic classification rules, which are static and do not appear to be automated, the flexibility to classify on port number, source-destination address, and domain names.
- Complex encryption algorithm, when traffic is encrypted, it becomes complex to perform classification, considering encryption by random numbers, hash functions, and other techniques.
- Traffic processing problem, the process to de-encrypt traffic can take much more processing power.
- Traffic congestion, subscribers that use multiple applications at the same time makes it much more challenging to classify and record the traffic for each application.

The test case requires the evaluation of traffic or utilization in a network, the process to analyse the traffic, cater on the requirements for positive fraud case, use previous work to formulate and compare a higher detection rate, and apply the algorithm on software tool. Test case 2 required to create a new framework to solve the deployment.

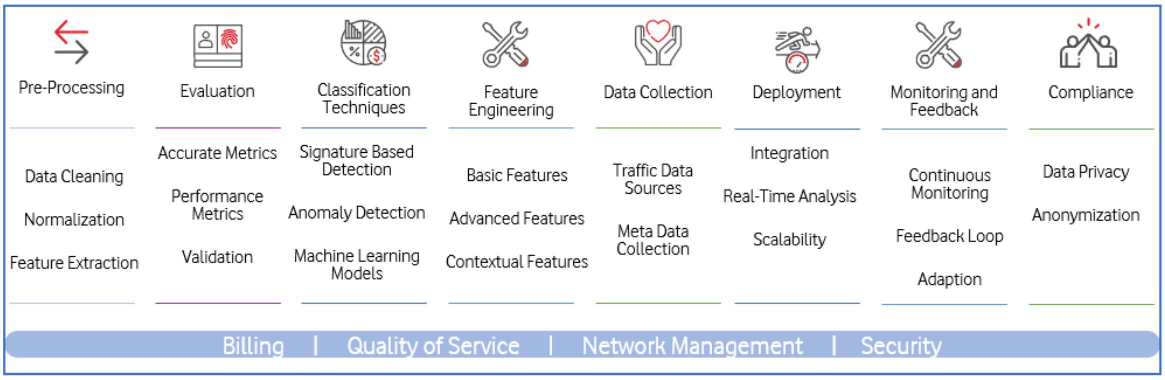


Figure 4.20: Traffic Classification Framework

The framework in Figure 4.20 for traffic classification involves categorizing the traffic to manage Billing, Quality of Service, Network Management, and Security. The following pillars are used to approach traffic classification:

The common grounds to host and manage the traffic are:

- Billing: Implement fair usage policies or tiered service plans.
- Quality of Service (QoS): allows traffic to be prioritized traffic for different user experiences (e.g., emergency service or government service)
- Network Management: To manage the network nodes and systems.

- Security: Caters on the detection and to mitigate malicious/unauthorized traffic.

The framework entails the following:

Preprocessing

- Data Cleaning: This is the process of removing noisy data.
- Normalization: To standardize data formats and scales.
- Feature Extraction: Extract relevant features from raw data, such as payload size, packet intervals, and protocol-specific attributes.

Evaluation

- Accuracy Metrics: To ensure measurements are precision and overall accuracy.
- Performance Metrics: Assess the model’s performance of the network in terms of latency, throughput, and resource utilization.
- Validation: The use of cross-validation techniques and holdout datasets to ensure generalizability.

Classification Techniques

- Signature-Based Detection: The use of predefined patterns or signatures to identify known traffic types.
- Anomaly Detection: To identify deviations from normal traffic patterns to detect unknown or emerging threats.
- Machine Learning Models:
    - Supervised Learning: To train models using labeled data to classify traffic (e.g., Random Forests, Support Vector Machines, Neural Networks).
    - Unsupervised Learning: The use of clustering techniques to group similar traffic types without prior labels (e.g., K-means, DBSCAN).
    - Deep Learning: To employ advanced models like Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) for more complex patterns (e.g., traffic flow prediction).

Feature Engineering

- Basic Features: The need to consider source/destination IP addresses, ports, protocols, packet lengths, and inter-arrival times.
- Advanced Features: Flow duration, average packet size, payload content (if inspection is performed), and entropy measures.

- Contextual Features: Application-specific patterns, user behavior patterns, and network topology.

Data Collection

- Traffic Data Sources: To gather data from various sources such as packet captures (PCAPs), flow records application logs.
- Metadata Collection: The collection of metadata like IP addresses, port numbers, protocol types, and timestamps.

Deployment

- Integration: The deployment of a classification system within the network infrastructure (e.g., in-line appliances, cloud services).
- Real-Time Analysis: To ensure the system can process traffic data in real-time or near-real-time as required.
- Scalability: Design the system to handle varying traffic loads and adapt to network changes.

Monitoring and Feedback

- Continuous Monitoring: To track the system's performance and accuracy over time.
- Feedback Loop: Use real-world data and user feedback to refine and retrain the classification models.
- Adaptation: The use of automation to update models and rules based on new traffic patterns and emerging threats.

Compliance and Privacy

- Data Privacy: To abide by the POPIA act and ensure compliance with regulations such as GDPR or CCPA when handling user data.
- Anonymization: Implementing techniques to anonymize sensitive information when possible.

4.10.3 Test Case 3: Fraudulent Access on Cloud Network

The means of using other forms of access to make use of the cloud environment also discussed in Indarjit, Balyan, and Adonis, (2025d) and its key functions on a virtual machine. The test case will allow the identification of a fraudster on the cloud network access.

To provide a summary of the services of the three main cloud providers is stated in Table 4.3.

Table 4.3: Overview of Cloud Key Services (Amazon Web Services, 2025; Microsoft Azure, 2025; Google Cloud, 2025)

Service	Description	AWS	Azure	GCP
Computing	Virtual machines and scalable computing resources	EC2	Virtual Machines	Compute Engine
Object Storage	Storage for unstructured data in objects	S3	Blob Storage	Cloud Storage
Block Storage	Storage for data in blocks, similar to traditional hard drives	EBS	Disk Storage	Persistent Disk
Database (relational)	Managed relational database services	RDS	SQL Database	Cloud SQL
Database (NoSQL)	Managed NoSQL database services	DynamoDB	Cosmos DB	Firestore / Datastore
Content Delivery Network (CDN)	Global distribution of content to reduce latency and improve performance	CloudFront	Azure CDN	Cloud CDN
Serverless Computing	Running code without managing server infrastructure	Lambda	Functions	Cloud Functions
Big Data Processing	Processing and analyzing large datasets	EMR	HDInsight	Dataproc
Machine Learning	Provision of services and tools for Machine Learning	SageMaker	Machine Learning Studio	AI Platform
Identity and Access Management	Management of users and permissions	IAM	Azure AD	Cloud IAM
Monitoring and Logging	Monitoring and logging of applications and infrastructure	CloudWatch	Azure Monitor	Stackdriver (Operations)
Networking	Management of networks and their security	VPC	Virtual Network	VPC
Container Orchestration	Management and orchestration of containers	EKS	AKS	GKE

Data Warehousing	Storage and analysis of large amounts of structured data	Redshift	Synapse Analytics	BigQuery
Backup and Disaster Recovery	Backup and recovery of data	Backup	Azure Backup	Backup

To define the difference between the following clouds and why it’s importance of a Hybrid cloud, in summary,

The Public Cloud requires no capital expenditure to scale up or expand, the applications running can be easily scaled out, and the cost is dependent on what resources are used, the Private Cloud, the hardware required to be purchased from the start of deployment, the owner has full control of the platform and it easily secures, also, the owner is responsible for maintenance and upgrades. And, Hybrid Cloud possesses the highest flexibility, on applications, the owner can deploy any applications without being bound by cloud provider limits, and the owner controls the security, legal aspects, and compliance.

Virtual Private Cloud (VPC) is a private network inside a public network (the internet), the VPC provides a high level of security and control over its network infrastructure, allowing to define of customized network configurations and control access resources and more. The VPC also provides a high level of isolation allowing to create secure and desolated environments for their applications and service. This makes the VPC a great solution for organizations that run sensitive workloads or comply with security and privacy regulations.

4.10.3.1 Test Case 3a

This test case aims to simulate fraudulent access attempts on a cloud network and identify such activities using various detection mechanisms. It will involve attempts to gain unauthorized access to a virtual machine and its functions. The preconditions can be seen as:

- A cloud environment with multiple VMs and users.
- Monitoring and logging tools configured for detecting access anomalies.
- A set of predefined access permissions and roles.
- Alerting mechanisms for unusual activities.

The test data is viewed as:

- Valid User Accounts: Pre-configured user accounts with legitimate access to the VM.
- Fraudulent User Accounts: Test accounts or simulated users with no legitimate access rights.

- VMs and Resources: Cloud-based VMs with key functions that are critical to test access controls.

The steps taken on detection:

a. Baseline Access Review:
    - To verify that legitimate subscriber has appropriate access to the VMs and resources.
    - Review current access logs to establish a baseline of normal activity based on a mean.

b. Simulate Fraudulent Access:
    - To attempt to access the VM using the fraudulent user accounts.
    - Test various methods such as:
        - Brute-force attacks on login credentials.
        - Exploiting known vulnerabilities or weak configurations.
        - Using stolen credentials.

c. Monitor and Log Activity:
    - Ensure that all access attempts, including failed and successful logins, are logged.
    - Observe how the monitoring tools respond to these fraudulent access attempts.
    - Check if alerts are generated for suspicious activities.

d. Analyse Detection Mechanisms:
    - Review the logs for entries related to fraudulent access attempts.
    - Verify if the monitoring tools and alerting systems correctly flagged the fraudulent access.
    - Assess the accuracy and timeliness of the detection alerts.

e. Response and Mitigation:
    - Check the response actions taken by the system or security team upon detection.

- Ensure that appropriate measures are taken, such as locking out the fraudulent accounts and notifying administrators.

f. Post-Test Cleanup:

- Remove any test accounts or simulated fraud activities.
- Restore the system to its normal operating state, configuration files should be saved before the testing for rollback purpose.
- Review and update security policies based on findings.

g. Expected Results:

- Fraudulent access attempts are detected by the monitoring and logging systems.
- Alerts are generated for unauthorized access.
- Correct response actions are initiated to address and mitigate the fraudulent access.
- No unauthorized changes or access are granted to the VM or its functions. Changes for a live system should be controlled and approvals on each level.

h. Post-Test Analysis:

- Evaluate the effectiveness of the detection and response mechanisms.
- Identify any gaps in the current access controls and monitoring systems.
- Recommend improvements to enhance security and prevent future fraudulent access.

i. Remarks:

- Ensure compliance with data protection regulations while conducting the test.
- Coordinate with relevant teams to avoid disruption to legitimate users.

4.10.3.2 Test Case 3b

This test case simulates different methods of unauthorized access to a cloud environment, focusing on access control bypass, privilege escalation, and suspicious activity detection. The goal is to assess the effectiveness of protective controls in identifying and responding to fraud. The preconditions are:

- Cloud environment with multiple virtual machines and many subscribers.
- Access controls and permissions configured on router or firewall.
- Security monitoring tools and Intrusion Detection Systems (IDS) in place.
- Well laid Incident response plan ready for action.

The test data is viewed as:

- Legitimate User Accounts: The accounts with various roles and permissions.
- Malicious User Accounts: Simulated or test accounts with no legitimate access rights.
- Sensitive Resources: VMs with critical applications and data.

The test plans are designed as follows:

a. Initial Setup:

    - To confirm that legitimate user roles and permissions are accurately assigned.
    - Ensure security monitoring and IDS are active and configured to detect anomalous behavior in the network.

b. Privilege Escalation Attempt:

    - Use a sample legitimate user account to attempt privilege escalation. This may include:
        - Exploiting misconfigurations or possible vulnerabilities.
        - The use of tools to gain elevated privileges.
    - Observe if the system detects and logs these attempts.

c. Access Control Bypass:

    - Attempt to bypass access controls using techniques such as:
        - Manipulating URLs or API requests.
        - Using SQL injection or similar attacks.
    - Check for fraud-access detection and logging of such attempts.

d. Suspicious Activity Simulation:

   - Replicate activities that are unusual for the role, such as:
     - Accessing files or data not normally required.
     - Performing actions that are not typical for a user’s profile.
   - Verify if the monitoring tools flag these activities as suspicious.

e. Monitor for Anomalies:

   - Use real-time monitoring tools to track all activities during the test.
   - Ensure that any unauthorized or suspicious access attempts generate alerts.
   - Review how the alerts are categorized and if they prompt the appropriate response.

f. Response Action Evaluation:

   - Simulate the activation of incident response protocols based on alerts.
   - Verify that response actions, such as isolating affected VMs or disabling fraudulent accounts by compliance, are executed promptly.

g. Review Access Logs:

   - Analyze access logs for signs of unauthorized access or anomalies by either logging into nodes or th use of WinSCP tool.
   - Check for any missed or false-negative detections.

h. System and Security Policy Review:

   - Evaluate the effectiveness of existing security policies and controls based on test findings.
   - Suggest improvements or updates to enhance fraud detection and response mechanisms.

i. Expected Results:

   - Detection of privilege escalation and access control bypass attempts.
   - Generation of alerts for suspicious activities by category.
   - Effective and timely response to fraudulent access scenarios.
   - Comprehensive logs that accurately reflect the test cases.

j. Post-Test Actions:

- Document findings and any gaps in detection or response.
- Update security controls and policies as necessary, like firewall access and permit/deny rules.
- Provide training or updates to the security team based on the test results.

k. Remarks:

- Ensure that the test does not disrupt normal operations or affect legitimate users. By best practise to perform testing on off-peak hours.
- Coordinate with the security and IT teams to ensure comprehensive coverage.

4.10.4 Test Case 4: DDoS attack

To identify when fraudsters attempt to limit the network's availability by identifying their pattern and scoping a proposed deep-learning technique on detection by Indarjit, Balyan, and Adonis, (2025d).

An attack by more than one person and/or machine in an attempt to make the network's resource temporarily or indefinitely unavailable to its intended users. This is achieved by several methods, but the basic principle is to overload the target server or something in the network path to the server so that it becomes unavailable to legitimate users.

The target is on:

- Network Link Capacity (1-100 Gbps link),
- Firewall session table or connection per second capacity,
- Load Balancer or Web Server session table or connection per second capacity,
- DDoS the DNS server that serves as the authoritative answer for the domain in question.

A typical attack is performed using botnets, which are either under the control of the attacker or hired for the duration of the attack. Botnets consist of 100s, 1000s, or 10,000s of zombie computers. These are internet connected devices that have been compromised and can be remotely issued with commands by their command-and-control servers.

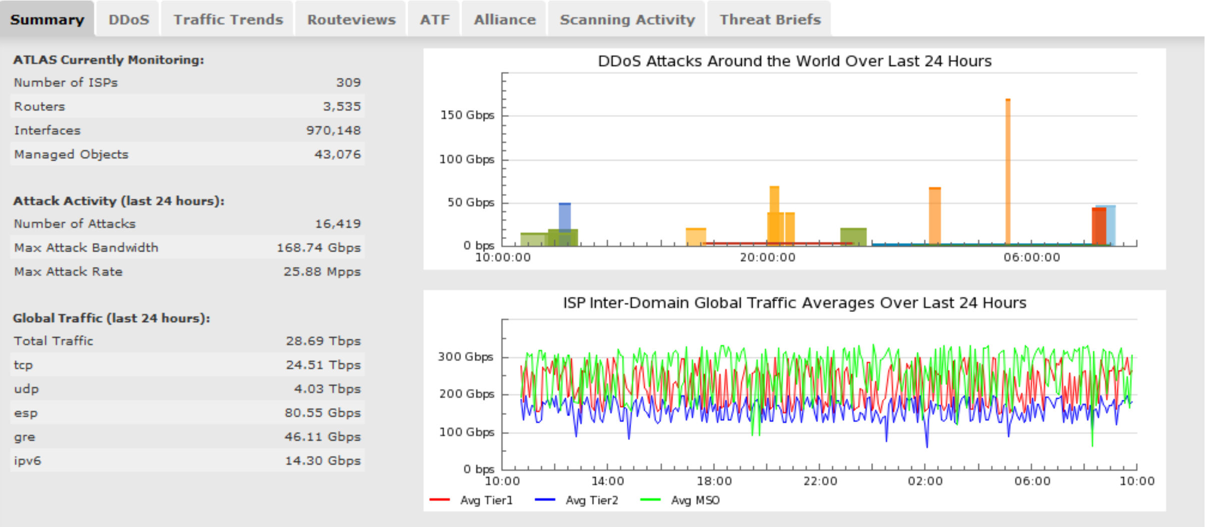
The ultimate goal of these attacks is varied and could be any of the following:

- Blackmail – A demonstration of the ability to take down a site, followed by a ransom email. (gaming, banking, and e-trading sites all employ robust DDoS mitigation services)
- State-sponsored/terrorist – During conflicts it is advantageous to take down opponents' websites.
- Personal – online computer gamers use DDoS to win against their opponents.
- 'Just because I can' – DDoS attacks can be purchased easily from darknet sites.

It can be identified the types of attack:

- Bandwidth versus packets per second to occupy the specified capacity.
- Reflection/Amplification attacks.
- Slow attacks – Slow loris: some attacks appear to be real connections, but they gradually fill the server's connection table.

Figure 4.21 is an example ATLAS summary report of all the Attack activity as seen by users of Arbor Peakflow equipment over a period of 24 hours. The largest attack was 168.74 Gbps in bandwidth and the max attack rate was 25.88Mpps/25.88 million packets per second.



ATLAS Currently Monitoring:	
Number of ISPs	309
Routers	3,535
Interfaces	970,148
Managed Objects	43,076
Attack Activity (last 24 hours):	
Number of Attacks	16,419
Max Attack Bandwidth	168.74 Gbps
Max Attack Rate	25.88 Mpps
Global Traffic (last 24 hours):	
Total Traffic	28.69 Tbps
tcp	24.51 Tbps
udp	4.03 Tbps
esp	80.55 Gbps
gre	46.11 Gbps
ipv6	14.30 Gbps

Figure 4.21: ATLAS showing activities within 24 hours

The highest attack rate in bps was almost 168Gbps (South America) and in pps 25.88Mpps (Asia). If an attack of this size targets customers in Data Centres without adequate DDoS mitigation service, the attack would impact all services.

DDoS mitigation is important not just for those customers who pay explicitly for it but also to prevent collateral damage within Data Centres/Shared environments/Reseller environments.

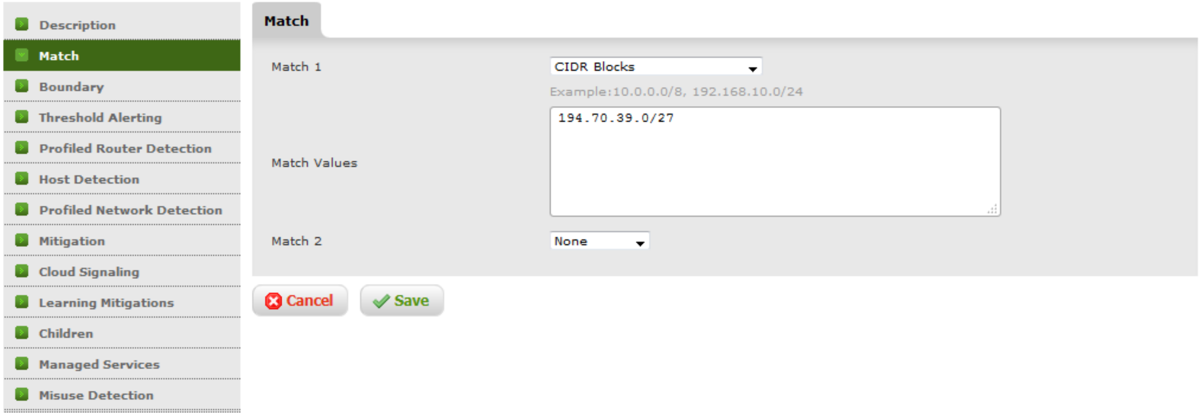
How do we know when there is an attack? How do we know when to mitigate an attack?

It all starts with Managed Objects (MOs) – (formerly called zones – the terms are interchangeable). A Managed Object is essentially the definition of the network object(s) that is to be protected. The Managed Object is a primary building block of the DDoS mitigation service.

Figure 4.22 shows the different elements that make up the managed objects. In the implementation, we use Description/Match/Boundary/Host Detection/Mitigation and Misuse

Detection. The Match tab is where to define the IP Prefixes/networks we want to protect. Generally, the use of CIDR blocks but there are other choices such as:

- CIDR groups
- Local ASN
- Peer ASN
- AS Path reg expression



Match 1: CIDR Blocks

Example:10.0.0.0/8, 192.168.10.0/24

Match Values: 194.70.39.0/27

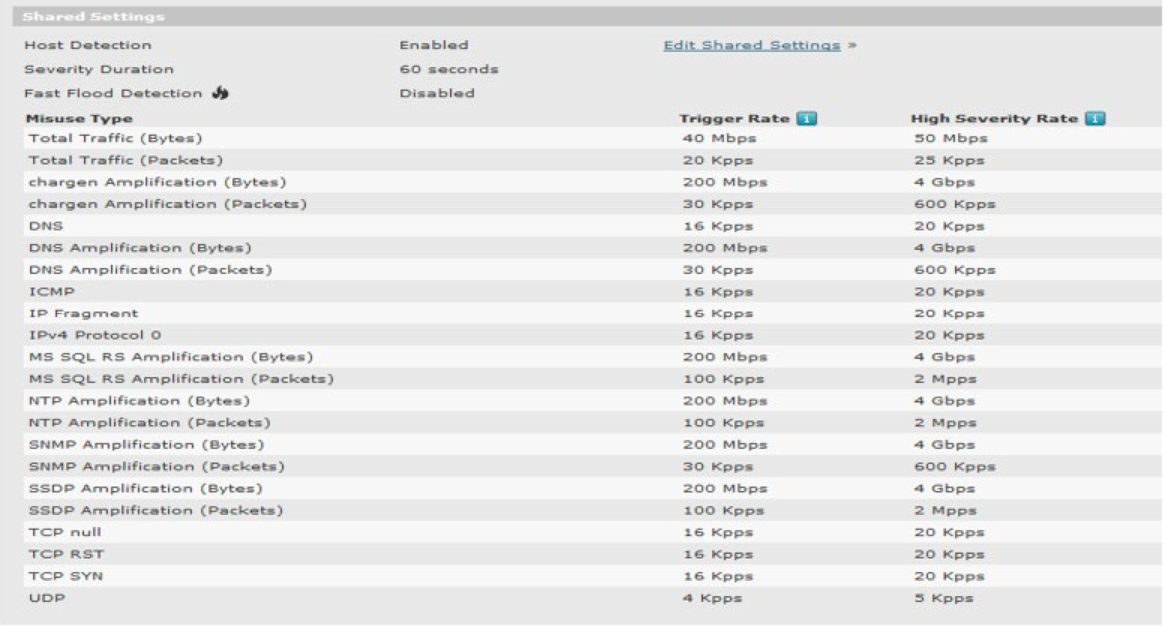
Match 2: None

Figure 4.22: Matched Objects

Three levels of alerts are built within the system:

- Low = triggered by the "Trigger Rate"
- High = triggered by the "High Severity Rate"
- Medium = triggered by the median of the Trigger and High Severity rate.

E.g.: Total traffic: Trigger = 10 and High = 20 then Medium = 15 (halfway between 10 and 20). The system will react to High alerts in Figure 4.23. The algorithm uses average peak traffic would be by using various reports the system can run, and then calculates the thresholds for the different categories of traffic.



Shared Settings

Host Detection	Enabled	Edit Shared Settings »
Severity Duration	60 seconds	
Fast Flood Detection	Disabled	

Misuse Type	Trigger Rate	High Severity Rate
Total Traffic (Bytes)	40 Mbps	50 Mbps
Total Traffic (Packets)	20 Kpps	25 Kpps
chargen Amplification (Bytes)	200 Mbps	4 Gbps
chargen Amplification (Packets)	30 Kpps	600 Kpps
DNS	16 Kpps	20 Kpps
DNS Amplification (Bytes)	200 Mbps	4 Gbps
DNS Amplification (Packets)	30 Kpps	600 Kpps
ICMP	16 Kpps	20 Kpps
IP Fragment	16 Kpps	20 Kpps
IPv4 Protocol 0	16 Kpps	20 Kpps
MS SQL RS Amplification (Bytes)	200 Mbps	4 Gbps
MS SQL RS Amplification (Packets)	100 Kpps	2 Mpps
NTP Amplification (Bytes)	200 Mbps	4 Gbps
NTP Amplification (Packets)	100 Kpps	2 Mpps
SNMP Amplification (Bytes)	200 Mbps	4 Gbps
SNMP Amplification (Packets)	30 Kpps	600 Kpps
SSDP Amplification (Bytes)	200 Mbps	4 Gbps
SSDP Amplification (Packets)	100 Kpps	2 Mpps
TCP null	16 Kpps	20 Kpps
TCP RST	16 Kpps	20 Kpps
TCP SYN	16 Kpps	20 Kpps
UDP	4 Kpps	5 Kpps

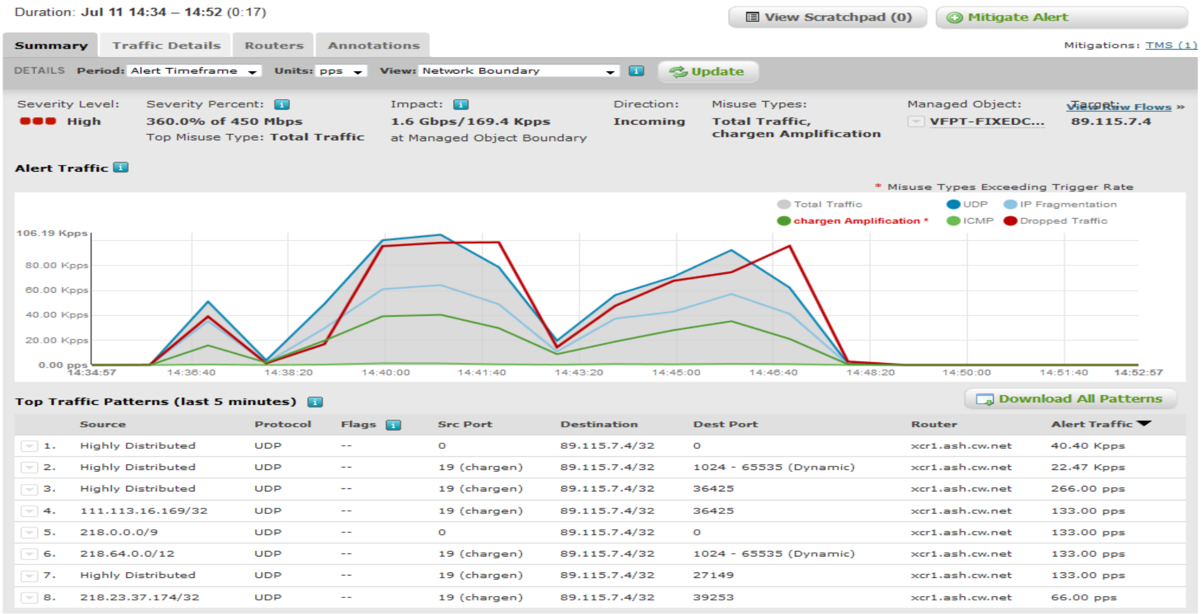
Figure 4.23: Host Detection

The mitigation tab is used to control actions should an IP prefix need to be under mitigation. It can define whether a mitigation is automatic or not. The default mitigation template that used for the MO, which may be different for user initiated or Auto-Mitigation. The MO template defines how to normally handle a mitigation for that particular MO.

- Which Threat Mitigation System (TMS) devices to use,
- IPv4 Black/White Lists (FCAP filters – allow TCP but only port 80 and 443),
- IPv4 address whitelists,
- IPv4 address blacklists,
- IPv4 Location filter lists,
- The different countermeasures can enable or not, and
- Traffic shaping.

The Alert Summary in Figure 4.24 provides the destination/target address, the type of alert, and the protocols being used. In this case, there were multiple sources, denoted by 0.0.0.0/0, protocol UDP, and the source port was 123 NTP. This indicates that this is almost certainly an NTP reflection attack. (gives the attacker a 550-bandwidth multiplier). When a High Severity alert is generated, the system performs a number of configuration notification actions.
Typically, three actions are automatically performed when a High Severity alert is raised.

- An email is sent to the relevant support teams and customer,
- Logs the alert to Syslog,
- Send an SNMP-TRAP to the VCHS Netcool Servers, and,
- This is used to generate a Remedy ticket, that is auto-routed to the relevant support team.



Duration: **Jul 11 14:34 – 14:52** (0:17)

Mitigations: TMS (1)

DETAILS **Period:** Alert Timeframe **Units:** pps **View:** Network Boundary

Severity Level:	Severity Percent:	Impact:	Direction:	Misuse Types:	Managed Object:	Target:
●●● **High**	**360.0% of 450 Mbps** Top Misuse Type: **Total Traffic**	**1.6 Gbps/169.4 Kpps** at Managed Object Boundary	**Incoming**	**Total Traffic, chargen Amplification**	**VFPT-FIXEDC...**	**89.115.7.4**

View Raw Flows »

Alert Traffic

Top Traffic Patterns (last 5 minutes)

	Source	Protocol	Flags	Src Port	Destination	Dest Port	Router	Alert Traffic
1.	Highly Distributed	UDP	--	0	89.115.7.4/32	0	xcr1.ash.cw.net	40.40 Kpps
2.	Highly Distributed	UDP	--	19 (chargen)	89.115.7.4/32	1024 - 65535 (Dynamic)	xcr1.ash.cw.net	22.47 Kpps
3.	Highly Distributed	UDP	--	19 (chargen)	89.115.7.4/32	36425	xcr1.ash.cw.net	266.00 pps
4.	111.113.16.169/32	UDP	--	19 (chargen)	89.115.7.4/32	36425	xcr1.ash.cw.net	133.00 pps
5.	218.0.0.0/9	UDP	--	0	89.115.7.4/32	0	xcr1.ash.cw.net	133.00 pps
6.	218.64.0.0/12	UDP	--	19 (chargen)	89.115.7.4/32	1024 - 65535 (Dynamic)	xcr1.ash.cw.net	133.00 pps
7.	Highly Distributed	UDP	--	19 (chargen)	89.115.7.4/32	27149	xcr1.ash.cw.net	133.00 pps
8.	218.23.37.174/32	UDP	--	19 (chargen)	89.115.7.4/32	39253	xcr1.ash.cw.net	66.00 pps

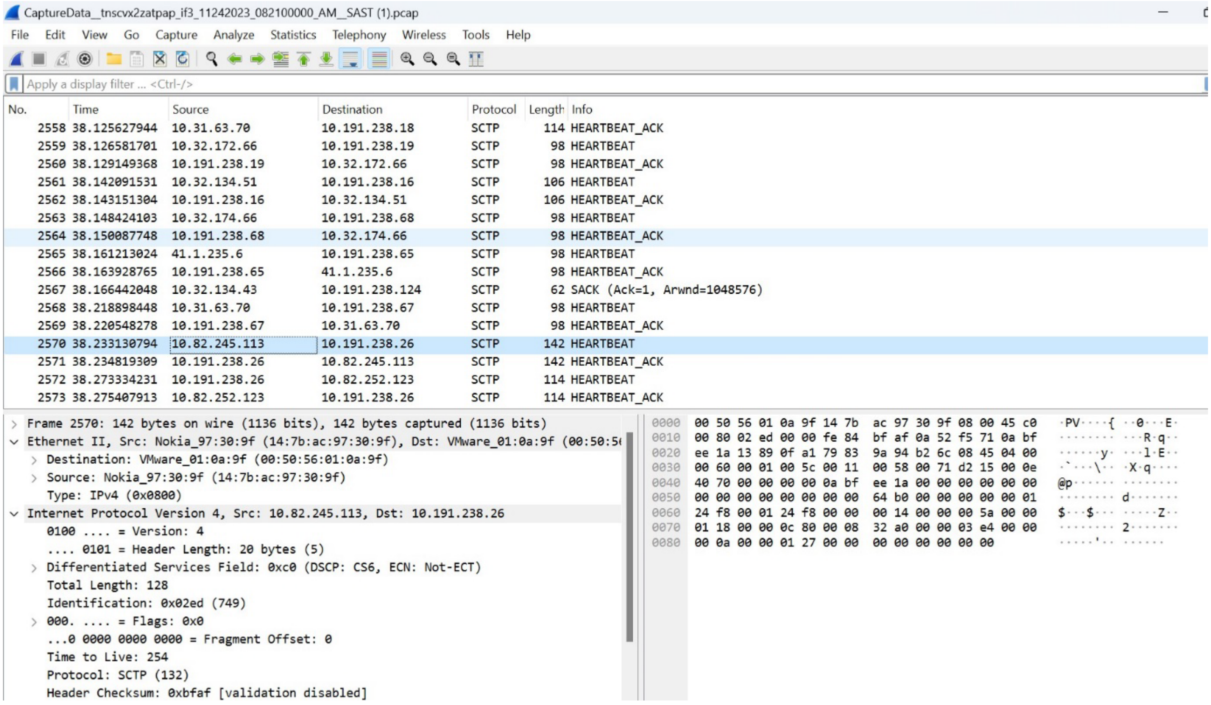
Figure 4.24: Summary report of High Alert

CHAPTER FIVE
DESIGN AND TEST CASES

5. Design and Test Cases 1,2,3,4

Chapter 5 presents the Test Case(s) of the concept to determine whether the design meets the thesis's specification. The design is built on Test Case(s) to replicate the concept carefully. The process of analysing, classifying, and blocking requires a consistent deep dive across each Test Case (TC) and in publication by authors Indarjit, Balyan, and Adonis, (2025d). An anonymous pcap file was extracted on User Plane traffic, the pcap file needed to be formatted discussed in Chapters 3 and 4.

WireShark tool was used to showcase and analyse the pcap file, In Figure 5.1 the source and destination IP addresses are shown, and not disclosing who the subscriber is. Also, the protocol and packet length are stated.


CaptureData__tnscvx2zatpap_if3_11242023_082100000_AM__SAST (1).pcap
File
Edit
View
Go
Capture
Analyze
Statistics
Telephony
Wireless
Tools
Help
Apply a display filter ... <Ctrl-/>

```
No.    Time          Source          Destination      Protocol Length Info
  2558 38.125627944  10.31.63.70     10.191.238.18    SCTP     114 HEARTBEAT_ACK
  2559 38.126581701  10.32.172.66    10.191.238.19    SCTP      98 HEARTBEAT
  2560 38.129149368  10.191.238.19   10.32.172.66     SCTP      98 HEARTBEAT_ACK
  2561 38.142091531  10.32.134.51    10.191.238.16    SCTP     106 HEARTBEAT
  2562 38.143151304  10.191.238.16   10.32.134.51     SCTP     106 HEARTBEAT_ACK
  2563 38.148424103  10.32.174.66    10.191.238.68    SCTP      98 HEARTBEAT
  2564 38.150087748  10.191.238.68   10.32.174.66     SCTP      98 HEARTBEAT_ACK
  2565 38.161213024  41.1.235.6      10.191.238.65    SCTP      98 HEARTBEAT
  2566 38.163928765  10.191.238.65   41.1.235.6       SCTP      98 HEARTBEAT_ACK
  2567 38.166442048  10.32.134.43    10.191.238.124   SCTP      62 SACK (Ack=1, Arwnd=1048576)
  2568 38.218898448  10.31.63.70     10.191.238.67    SCTP      98 HEARTBEAT
  2569 38.220548278  10.191.238.67   10.31.63.70      SCTP      98 HEARTBEAT_ACK
  2570 38.233130794  10.82.245.113   10.191.238.26    SCTP     142 HEARTBEAT
  2571 38.234819309  10.191.238.26   10.82.245.113    SCTP     142 HEARTBEAT_ACK
  2572 38.273334231  10.191.238.26   10.82.252.123    SCTP     114 HEARTBEAT
  2573 38.275407913  10.82.252.123   10.191.238.26    SCTP     114 HEARTBEAT_ACK
```

```
> Frame 2570: 142 bytes on wire (1136 bits), 142 bytes captured (1136 bits)
v Ethernet II, Src: Nokia_97:30:9f (14:7b:ac:97:30:9f), Dst: VMware_01:0a:9f (00:50:5
  > Destination: VMware_01:0a:9f (00:50:56:01:0a:9f)
  > Source: Nokia_97:30:9f (14:7b:ac:97:30:9f)
    Type: IPv4 (0x0800)
v Internet Protocol Version 4, Src: 10.82.245.113, Dst: 10.191.238.26
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
  > Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
    Total Length: 128
    Identification: 0x02ed (749)
  > 000. .... = Flags: 0x0
    ...0 0000 0000 0000 = Fragment Offset: 0
    Time to Live: 254
    Protocol: SCTP (132)
    Header Checksum: 0xbfaf [validation disabled]
```

```
0000  00 50 56 01 0a 9f 14 7b  ac 97 30 9f 08 00 45 c0   ·PV····{ ··0···E·
0010  00 80 02 ed 00 00 fe 84  bf af 0a 52 f5 71 0a bf   ········ ···R·q··
0020  ee 1a 13 89 0f a1 79 83  9a 94 b2 6c 08 45 04 00   ······y· ···l·E··
0030  00 60 00 01 00 5c 00 11  00 58 00 71 d2 15 00 0e   ·`···\·· ·X·q····
0040  40 70 00 00 00 00 0a bf  ee 1a 00 00 00 00 00 00   @p······ ········
0050  00 00 00 00 00 00 00 00  64 b0 00 00 00 00 00 01   ········ d·······
0060  24 f8 00 01 24 f8 00 00  00 14 00 00 00 5a 00 00   $···$··· ·····Z··
0070  01 18 00 00 0c 80 00 08  32 a0 00 00 03 e4 00 00   ········ 2·······
0080  00 0a 00 00 01 27 00 00  00 00 00 00 00 00         ·····'·· ······
```

Figure 5.1: WireShark tool showing the pcap file

The architecture explores software tools to assist the test case(s) within the SDN component shown in Figure 5.2. The software tools are built within the SDN controller as the central element for detection.

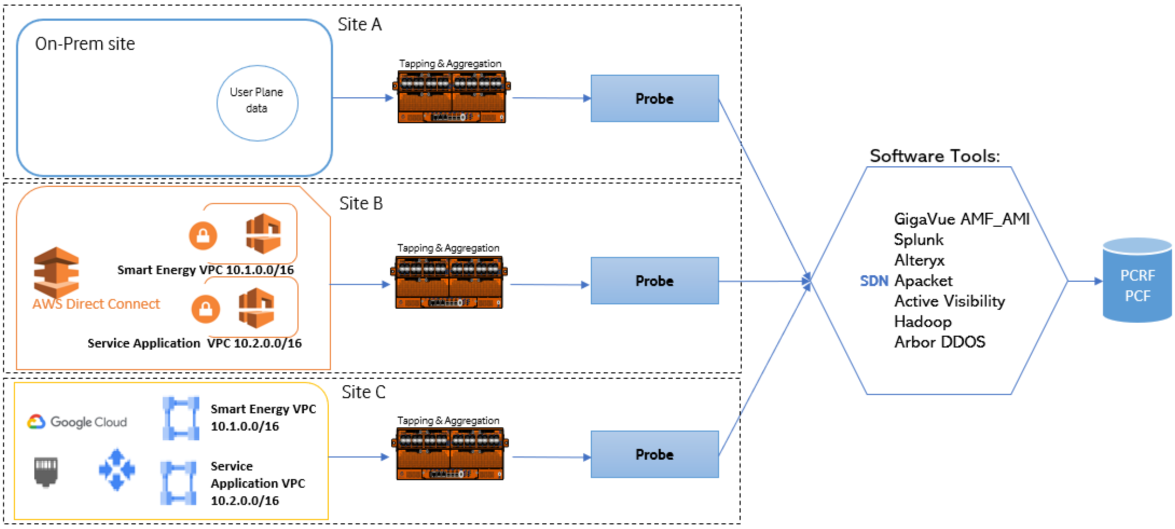
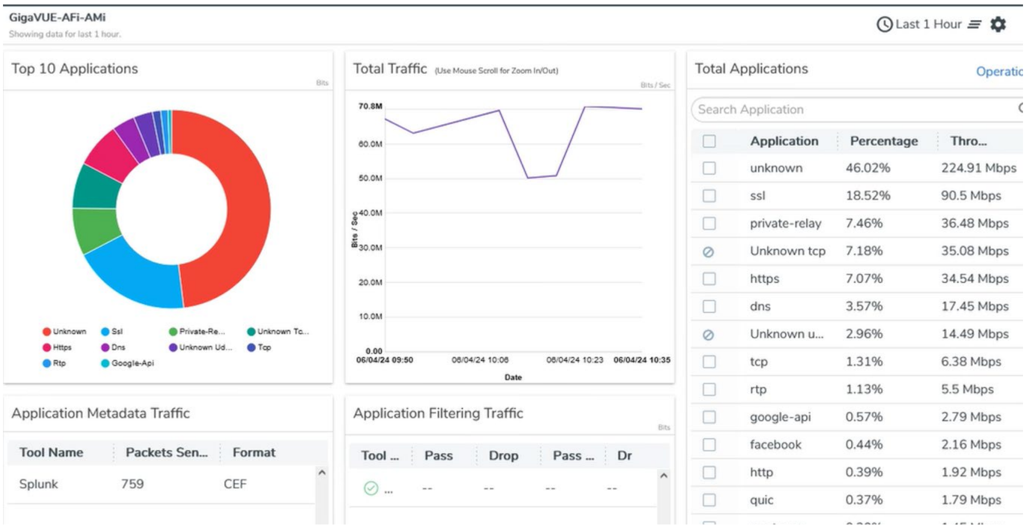


Figure 5.2: Software Tools

5.1 Background of design

To demonstrate each test case, different software tools were explored to identify the most suitable tool for the purpose of traffic classification and detection. During the search for a software tool, it became more challenging to find a tool that would serve as a blueprint for the research. The following traffic tools were explored, 'Zabbik', 'Icinga', 'Nagius', 'Cacti', and 'Checkmk'. Whilst these tools carry their value and assist companies in understanding their network resources but provide limits for traffic intelligence.

The category of traffic intelligence and subscriber intelligence work together to understand network fraud, using the Gigamon engine to analyse the pcap file in Figure 5.3 showing the applications used and the total traffic. An amount of traffic cannot be classified by the engine and other traffic belonging to ssl, tcp, https, dns, google-api, facebook, quic and ms-teams.



	Application	Percentage	Thro...
□	unknown	46.02%	224.91 Mbps
□	ssl	18.52%	90.5 Mbps
□	private-relay	7.46%	36.48 Mbps
⊘	Unknown tcp	7.18%	35.08 Mbps
□	https	7.07%	34.54 Mbps
□	dns	3.57%	17.45 Mbps
⊘	Unknown u...	2.96%	14.49 Mbps
□	tcp	1.31%	6.38 Mbps
□	rtp	1.13%	5.5 Mbps
□	google-api	0.57%	2.79 Mbps
□	facebook	0.44%	2.16 Mbps
□	http	0.39%	1.92 Mbps
□	quic	0.37%	1.79 Mbps

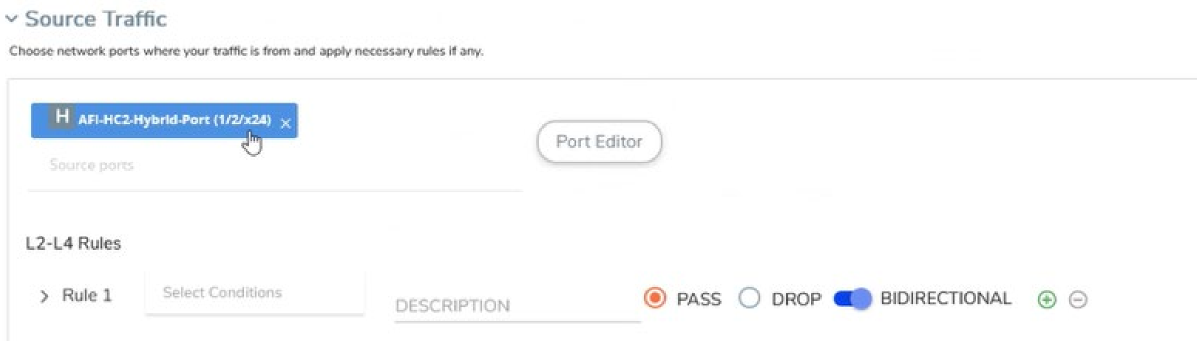
Tool Name	Packets Sen...	Format
Splunk	759	CEF

Tool ...	Pass	Drop	Pass ...	Dr
✓ ...	--	--	--	--

Figure 5.3: Gigamon engine for traffic intelligence

Certain traffic on the network from a malicious application can be labelled as fraud that abuses subscribers, either requesting money from the subscriber or utilizing network resources on capacity resulting in limited access for other subscribers plays a role in harm and network losses. The network is a shared platform, and the availability of the network is important, the test case arises to block an entire application due to the above, the Gigamon engine allows the 'Drop' of a certain application by rules in Figure 5.4.

The data needs to be understood via subscriber ID for the purpose of blocking a fraud-classified case, the data needed to be fed to another tool identified as Splunk Enterprise, the use of subscriber intelligence allows the research to identify each subscriber by IMEI number in Figure 5.5, which is the standard of identification used in mobile networks.



Source Traffic

Choose network ports where your traffic is from and apply necessary rules if any.

H AFI-HC2-Hybrid-Port (1/2/x24) ×

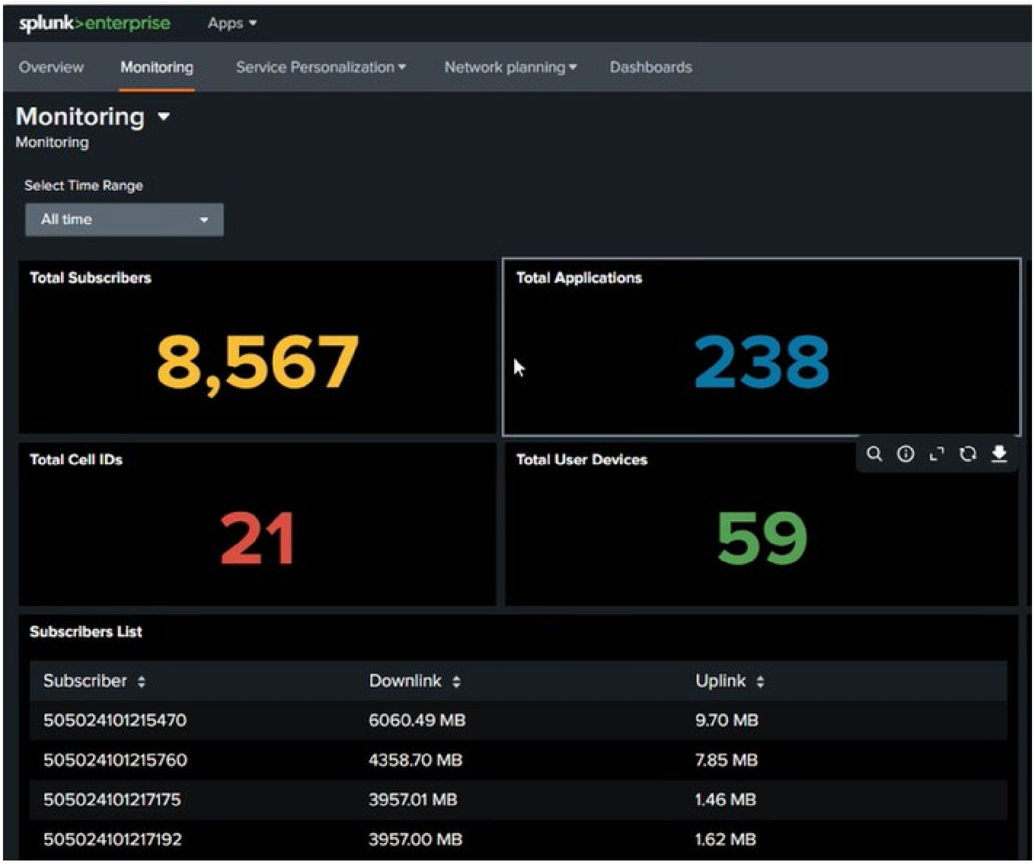
Source ports

Port Editor

L2-L4 Rules

Rule 1 | Select Conditions | DESCRIPTION | ● PASS ○ DROP | BIDIRECTIONAL

Figure 5.4: Pass and Drop rules for Applications



Monitoring ▾
Monitoring

Select Time Range: All time

Total Subscribers
8,567

Total Applications
238

Total Cell IDs
21

Total User Devices
59

Subscribers List

Subscriber	Downlink	Uplink
505024101215470	6060.49 MB	9.70 MB
505024101215760	4358.70 MB	7.85 MB
505024101217175	3957.01 MB	1.46 MB
505024101217192	3957.00 MB	1.62 MB

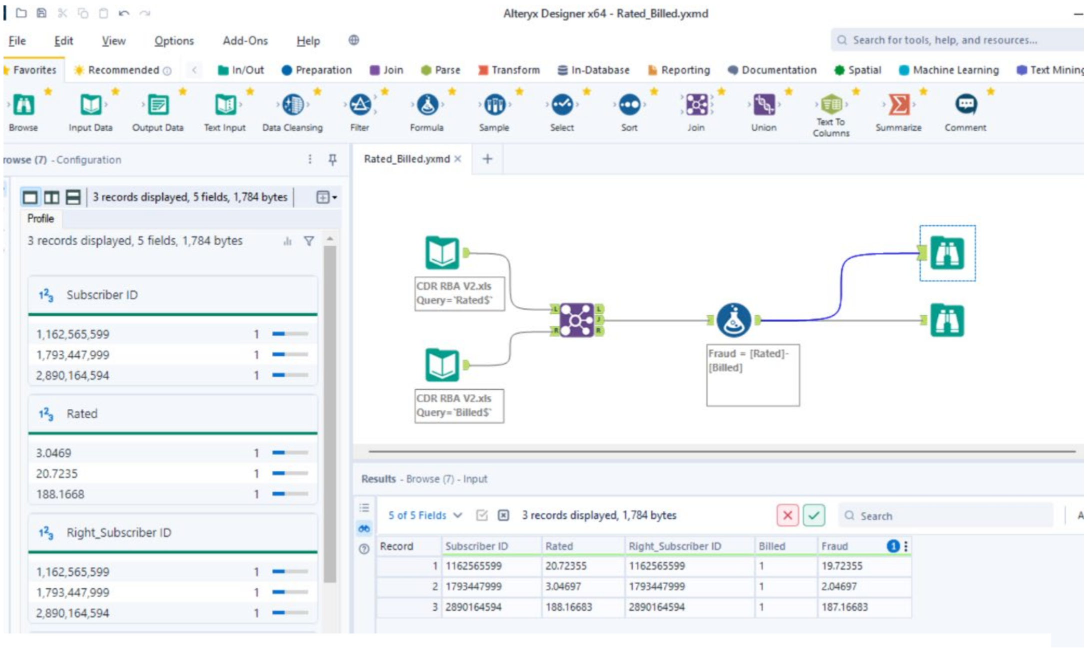
Figure 5.5: Splunk Enterprise showing IMEI subscriber ID

5.2 Test Case 1: Fraud detection with unpaid token on the Communication Network

- Energy sector
- Communication sector
- Service Application sector

To use traffic classification and identify for each sector the patterns of a fraudster. To create an algorithm by Machine Learning on the detection.

To showcase Test Case 1 and replicate the concept, many tools were explored to bring the billing platform and subscriber usage together, the software tool used to demonstrate the testing by the name of Alteryx Designer was incorporated and making use of its components. The data used is collected from Call Detail Records (CDR) which holds the subscriber ID, data usage, and details of the session of the Home Location Register (HLR). Further details on CDR creation and testing are within the Appendix. The process flow in Figure 5.6 uses Input data from the CDR file, which is converted into excel format, the two inputs used are the subscriber 'Rated' information and subscriber 'Billed' information. Since the data was formatted from previous steps, the process flow did not require removing packet duplicates. The 'Join' components performs the function to map the records based on the subscriber, each subscriber is unique and billing is to the subscriber account. When dealing with a large number of subscribers, the amount of processing to map the subscriber will take more time. The formula component performs the usage or rating from the tariff plan and does the subtraction from the billing platform. Finally, the ‘Browse’ allows to display of the data.



3 records displayed, 5 fields, 1,784 bytes

Subscriber ID	
1,162,565,599	1
1,793,447,999	1
2,890,164,594	1

Rated	
3.0469	1
20.7235	1
188.1668	1

Right_Subscriber ID	
1,162,565,599	1
1,793,447,999	1
2,890,164,594	1

3 records displayed, 1,784 bytes

Record	Subscriber ID	Rated	Right_Subscriber ID	Billed	Fraud
1	1162565599	20.72355	1162565599	1	19.72355
2	1793447999	3.04697	1793447999	1	2.04697
3	2890164594	188.16683	2890164594	1	187.16683

Figure 5.6: Process Flow Test Case 1

Table 5.1 shows the mapped subscriber data while Table 5.2 shows the Fraud case of subscribers that have incorrectly billed due to abuse of the platform of any sector.

Table 5.1: Mapped Record of Subscriber_Rated_Billed

Subscriber ID	Rated	Right_Subscriber ID	Billed
1132922998	7.05549	1132922998	7.05549
1142045999	1.92884	1142045999	1.92884
1158583999	9.62579	1158583999	9.62579
1162565599	20.72355	1162565599	1
1173668897	26.94287	1173668897	26.94287
1216092996	7.92081	1216092996	7.92081
1216092998	29.16317	1216092998	29.16317
1223595999	3.7413	1223595999	3.7413
1224381799	2.76811	1224381799	2.76811
1228894999	13.22671	1228894999	13.22671
1245659999	5.26962	1245659999	5.26962
1254896999	1.92171	1254896999	1.92171

1314799994	6.34951	1314799994	6.34951
1332566998	1.91142	1332566998	1.91142
1386820799	15.30599	1386820799	15.30599
1389813999	8.13821	1389813999	8.13821
xxxxxxxxxxx			
xxxxxxxxxxx			

Table 5.2: Subscriber Fraud

Subscriber ID	Rated	Right_Subscriber ID	Billed	Fraud
1162565599	20.72355	1162565599	1	19.72355
1793447999	3.04697	1793447999	1	2.04697
2890164594	188.16683	2890164594	1	187.16683

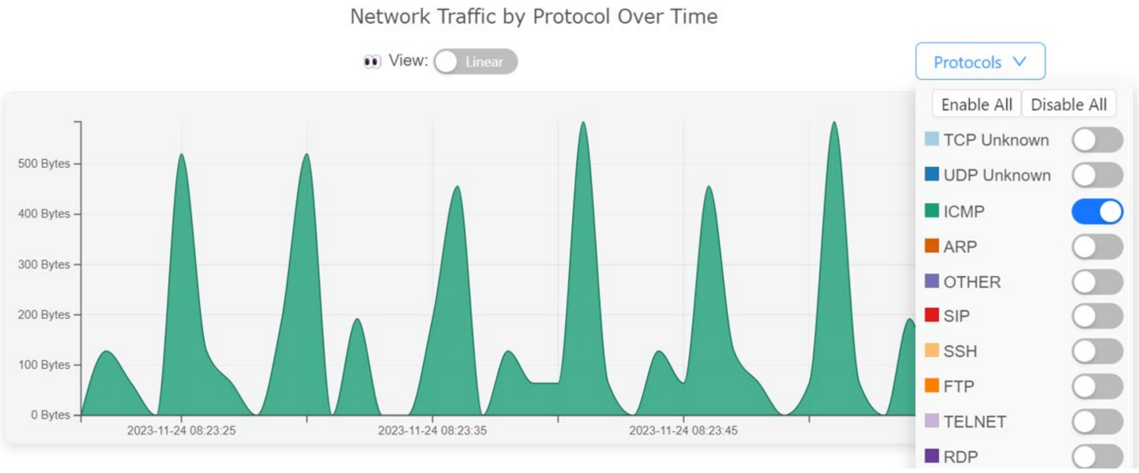
Telecom fraud, such as Wangiri fraud in Arafat, Qusef, and Sammour. (2019), poses significant challenges for operators, leading to revenue losses and difficulties in detection and prosecution. Wangiri fraud involves missed calls from premium numbers, enticing subscribers to call back and incur high charges. This paper suggests using ensemble classifiers, particularly the Extreme Gradient Boosting algorithm, to improve fraud detection accuracy and efficiency, especially during less monitored periods like holidays and weekends.
The study used a genuine Call Detail Record (CDR) dataset with 2,415,919 calls, of which 1.58% were fraudulent Wangiri calls. The dataset was split, with 70% used for training and 30% for evaluation. The Extreme Gradient Boosting (XGBoost) classifier showed the best performance, significantly reducing false negatives from 172 to 34 when the classification threshold was set to 0.2. This high calibration was further evidenced by XGBoost's lowest Brier score loss of 0.000201.

Test case 1 shows effective by correlation of the HLR and Billing platform; to note, the data were formatted in Chapter 3, creating a standard format for processing. Further consideration can be given to de-encryption process when focus is given on a 5G Standalone network.

5.3 Test Case 2: Fraud detection utilizing the network

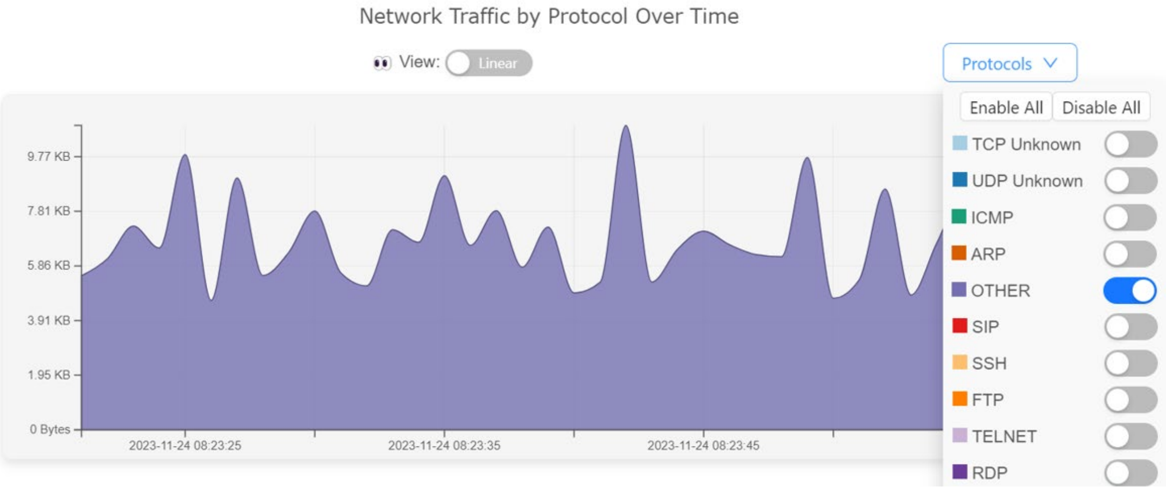
To evaluate the effects of fraud users for traffic classification and building an algorithm to automate the detection.

To explore the test case on Fraud detection utilizing the network requires to investigate the pcap file and identifying the protocol used, the traffic patterns, IP connections, and possible abuse. The combination of each assists the research to formulate an algorithm. The pcap file is a capture of subscriber traffic taken to filter the data and analyse the pattern of fraudster. To explore the pcap file, the APacket tool was used and sampled into a timeframe at which the classification can be read, for each protocol such as TCP/UDP/HTTP/HTTPS/DNS/SIP/SSH/TELNET/RDP/SMB/SSDP/ICMP/ARP/other, some protocols are secure and others are not but are used presently. From each protocol the utilization should be fairly the same on average or peak, when considering subscriber traffic, it’s best practice to look at the peak values in Figures 5.7, 5.8. 5.9. 5.10 of the traffic types and Figures 5.11, 5.12, 5.13 show the Internet Protocol mapping.



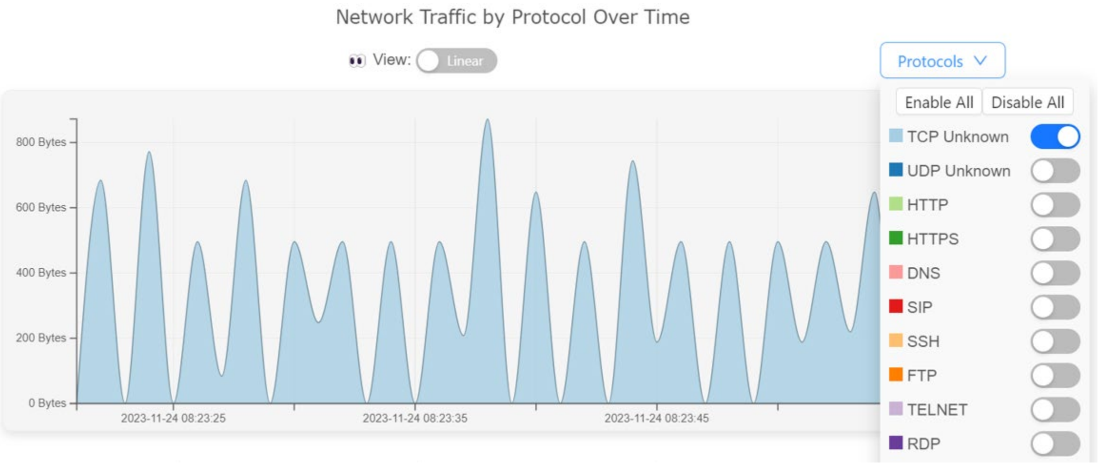
TCP Unknown ○
UDP Unknown ○
ICMP ●
ARP ○
OTHER ○
SIP ○
SSH ○
FTP ○
TELNET ○
RDP ○

Figure 5.7: Network Traffic by protocol over time – ICMP traffic



TCP Unknown ○
UDP Unknown ○
ICMP ○
ARP ○
OTHER ●
SIP ○
SSH ○
FTP ○
TELNET ○
RDP ○

Figure 5.8: Network Traffic by protocol over time – Other traffic



Network Traffic by Protocol Over Time

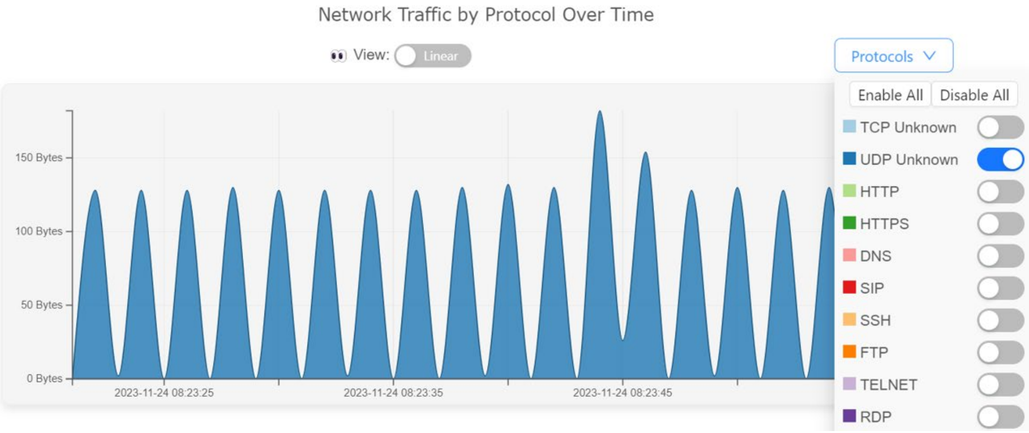
View: Linear

Protocols

Enable All | Disable All

- TCP Unknown ●
- UDP Unknown ○
- HTTP ○
- HTTPS ○
- DNS ○
- SIP ○
- SSH ○
- FTP ○
- TELNET ○
- RDP ○

Figure 5.9: Network Traffic by protocol over time – TCP traffic



Network Traffic by Protocol Over Time

View: Linear

Protocols

Enable All | Disable All

- TCP Unknown ○
- UDP Unknown ●
- HTTP ○
- HTTPS ○
- DNS ○
- SIP ○
- SSH ○
- FTP ○
- TELNET ○
- RDP ○

Figure 5.10: Network Traffic by protocol over time – UDP traffic

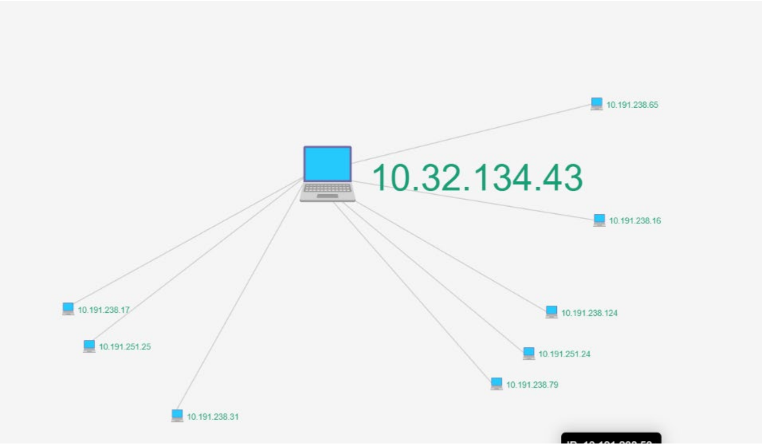


Figure 5.11: IP addresses - Source and Destination

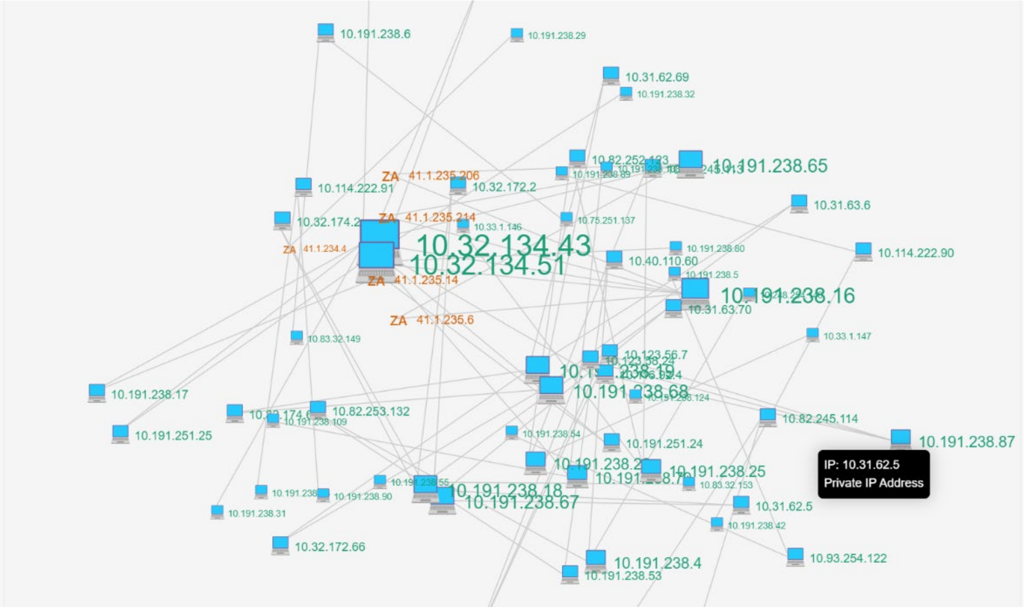


Figure 5.12: IP 10.32.134.43 connections to Servers

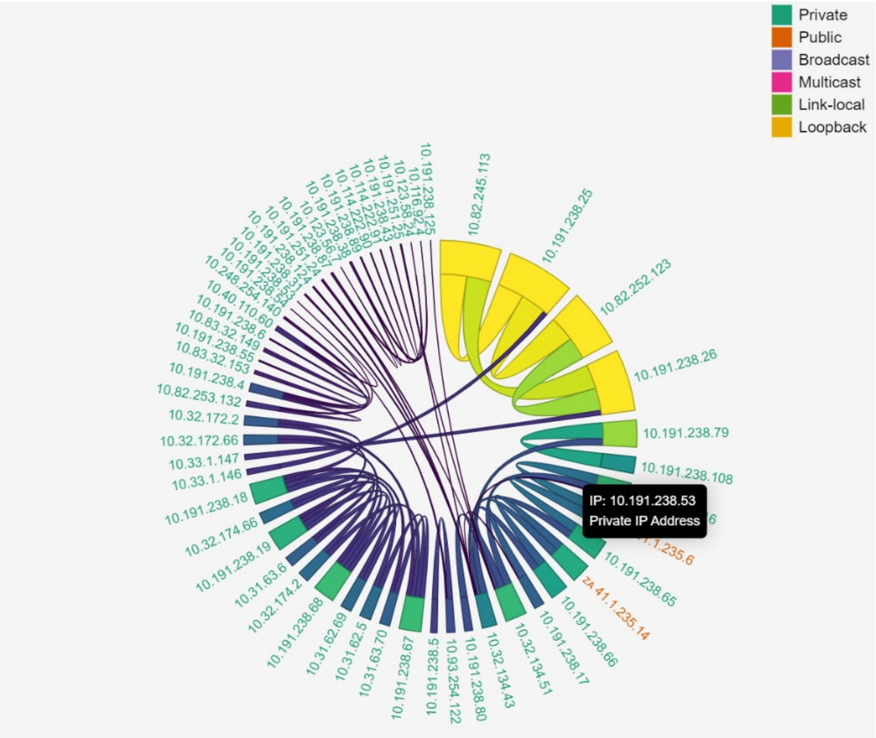
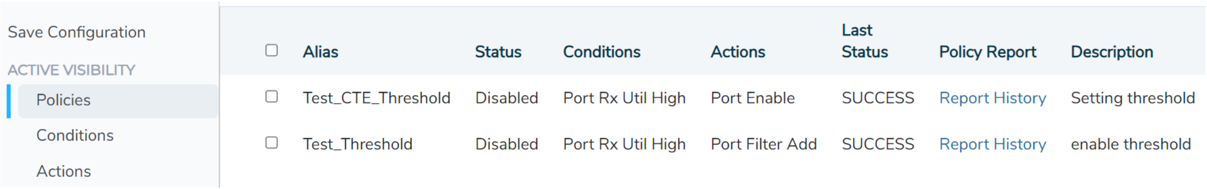


Figure 5.13: Connectivity between IP addresses

To perform the test case on traffic classification to identify a fraud case, the three areas required to be implemented are Policies, Conditions, and Actions. From research mentioned in Chapters 2 and 3 and the use of assessing traffic patterns and intelligence metrics must be applied to subscriber traffic. The aim is to filter and mark out fraud subscribers using automation and ensure the session is blocked. It can be noted that there are different techniques for blocking, to mention a few:

- Blocking the fraud subscriber from the network,
- Blocking the fraud subscriber from a certain application, and
- Blocking the fraud subscriber's session.

Each of the above will use a schema to enable the blocking, for the purpose of the test case, the session of a fraud subscriber is blocked.
Figure 5.14 shows the three areas.



□	Alias	Status	Conditions	Actions	Last Status	Policy Report	Description
□	Test_CTE_Threshold	Disabled	Port Rx Util High	Port Enable	SUCCESS	Report History	Setting threshold
□	Test_Threshold	Disabled	Port Rx Util High	Port Filter Add	SUCCESS	Report History	enable threshold

Figure 5.14: Active Visibility_Policies_Conditions_Actions

A **Policy** for fraud on a mobile network is a set of guidelines, rules, and procedures designed to prevent, detect, and respond to fraudulent activities within the mobile network environment. This policy includes:

- Fraud Prevention: Measures to prevent fraudulent activities from occurring. This involves implementing secure authentication processes and monitoring unusual patterns of subscriber usage.
- Fraud Detection: Mechanisms to identify fraudulent activities as they occur. This includes real-time monitoring systems that analyze usage patterns for anomalies, automated alerts for suspicious behavior, and periodic audits.
- Fraud Response: Is the procedure to address and mitigate the impact of fraud once it is detected. This includes steps for investigating incidents, mitigating damage, recovering losses, and taking corrective actions.

Overall, the policy aims to protect the network, its users, and its financial assets from fraudulent activities.

A **Condition** to identify a fraud case on the communication network refers to specific criteria or indicators used to detect suspicious or potentially fraudulent activity. These conditions help in flagging anomalies that may suggest fraud. Conditions for identifying fraud cases include:

- Unusual Usage Patterns: Is abnormal increases in call volume, data usage, or text messaging compared to a user's typical behaviour is a sign of fraud.
- Geographic Irregularities: Sudden or frequent changes in the geographic locations from which a user accesses the network, especially if they are geographically inconsistent with known patterns.
- High-Risk Transactions: Transactions or activities that involve high monetary value or frequent changes, such as large international calls or purchases, that deviate from normal behaviour.
- Multiple Accounts or Devices: Use of multiple accounts or devices from a single user or IP address, especially if they are linked in ways that are unusual or suspicious.
- Account Access Anomalies: Unusual login attempts, especially from unfamiliar locations or devices, or signs of unauthorized access to user accounts.
- Billing Discrepancies: Irregularities in billing, such as unexplained charges or unexpected changes in billing patterns, which indicate fraud.
- Sim Card Activity: Frequent SIM card swaps or activations in a short period of time, which could be an attempt to evade detection or gain unauthorized access.
- Use of Known Fraudulent Techniques: Detection of techniques commonly associated with fraud, such as spoofing, phishing, or social engineering attacks.
- Behavioural Anomalies: Patterns that deviate from established user behaviour, such as sudden changes in the frequency or type of services used.

An **Action** on a fraud subscriber on the mobile network refers to steps taken in response to detecting fraudulent activity associated with a subscriber. These Actions are aimed at mitigating the impact of fraud, recovering losses, and preventing further abuse. Actions include:

- Suspension or Termination of Service: Suspending or permanently terminating the subscriber's service to prevent further fraudulent activity and mitigate losses.
- Account Locking: Locking the subscriber's account to prevent access or usage until the fraud is investigated and resolved.
- Account Review and Reset: Reviewing and resetting account credentials, such as passwords or PINs, to secure the account against further unauthorized access.

- Collaboration with Law Enforcement: Working with law enforcement agencies to investigate the fraud, identify perpetrators, and take legal action if required.

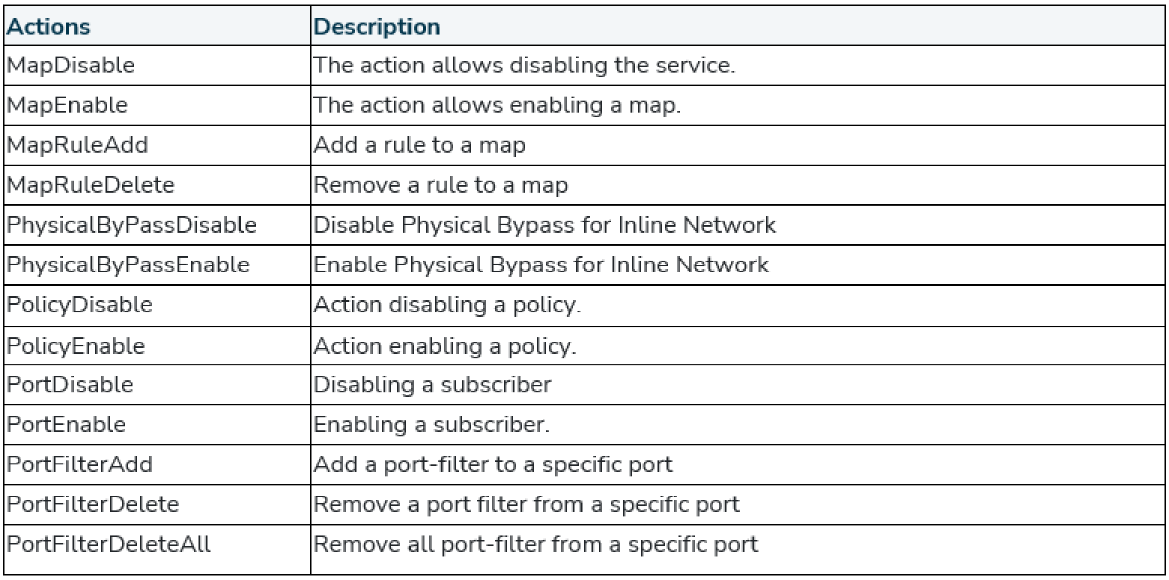
The actions are part of a broader fraud management strategy aimed at protecting the network, its users, and its financial assets from the impact of fraudulent activities.

The Condition(s) template is provided on Table 5.3 and defined Action(s) template is provided on Table 5.4.

Table 5.3: Conditions for Traffic Intelligence

Conditions	Description
GsCpuUtilHigh	To detect High gs group cpu utilization when a fraudster consumes high capacity.
GsCpuUtilLow	To detect Low gs group cpu utilization below normal threshold.
GsHbStatusDown	To detect Gs group heartbeat status down
GsHbStatusUp	To detect Gs group heartbeat status up
GsPktBufThHigh	To detect High gs group packet buffer utilization
GsPktBufThLow	To detect Low gs group packet buffer utilization
GsPktDropRateHigh	To detect High gs group packets drop rate
GsPktDropRateLow	To detect Low gs group packets drop rate
GsRxPktErrorHigh	To detect High gs group Rx error packets
GsRxPktErrorLow	To detect Low gs group Rx error packets
GsRxPktRateHigh	To detect High gs group Rx packet rate
GsRxPktRateLow	To detect Low gs group Rx packet rate
PortRxBufferHigh	To detect High port Rx buffer utilization
PortRxBufferLow	To detect Low port Rx buffer utilization
PortRxDiscardsHigh	To detect High port Rx discards
PortRxDiscardsLow	To detect Low port Rx discards
PortRxDropsHigh	To detect High port Rx drops
PortRxDropsLow	To detect Low port Rx drops
PortRxErrorsHigh	To detect High port Rx errors
PortRxErrorsLow	To detect Low port Rx errors
PortRxUtilHigh	To detect High port Rx utilization
PortRxUtilLow	To detect Low port Rx utilization
PortTxBufferHigh	To detect High port Tx buffer utilization
PortTxBufferLow	To detect Low port Tx buffer utilization
PortTxDiscardsHigh	To detect High port Tx discards
PortTxDiscardsLow	To detect Low port Tx discards
PortTxDropsHigh	To detect High port Tx drops
PortTxDropsLow	To detect Low port Tx drops
PortTxErrorsHigh	To detect High port Tx errors
PortTxErrorsLow	To detect Low port Tx errors
PortTxUtilHigh	To detect High port Tx utilization
PortTxUtilLow	To detect Low port Tx utilization

Table 5.4: Actions for Traffic Intelligence



Actions	Description
MapDisable	The action allows disabling the service.
MapEnable	The action allows enabling a map.
MapRuleAdd	Add a rule to a map
MapRuleDelete	Remove a rule to a map
PhysicalByPassDisable	Disable Physical Bypass for Inline Network
PhysicalByPassEnable	Enable Physical Bypass for Inline Network
PolicyDisable	Action disabling a policy.
PolicyEnable	Action enabling a policy.
PortDisable	Disabling a subscriber
PortEnable	Enabling a subscriber.
PortFilterAdd	Add a port-filter to a specific port
PortFilterDelete	Remove a port filter from a specific port
PortFilterDeleteAll	Remove all port-filter from a specific port

Test case 2 allows the creation of detecting fraudulent traffic from the network, using the tapped user plane traffic, to place Policies, Conditions, and Actions. A policy named Fraud_Detection_High_Utilization permits the identification of traffic above its mean threshold and high utilization over a period of time, once both Conditions meet the criteria, the configured Action will be applied. Figure 5.15 shows the utilization of traffic over a 30-day period, the data rate is in Gbps.

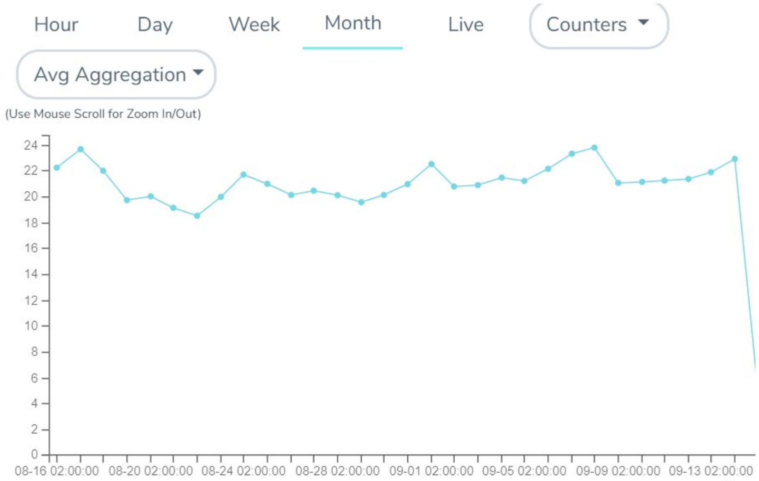
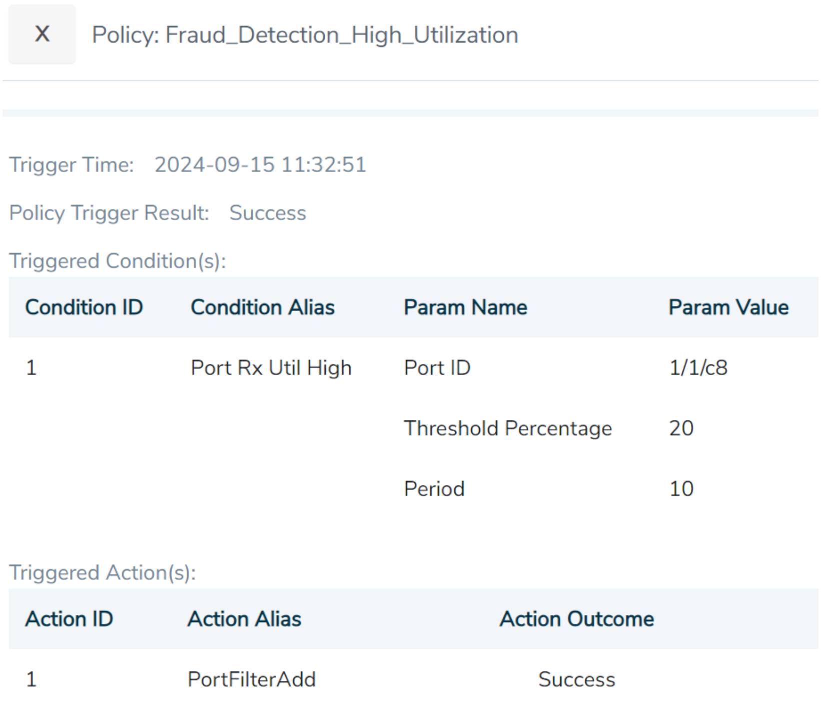


Figure 5.15: Utilization over 30-days

From the test case Fraud_Detection_High_Utilization, the Condition is established to monitor the traffic and identify when traffic is above the mean threshold in Figure 5.16.



X Policy: Fraud_Detection_High_Utilization

Trigger Time: 2024-09-15 11:32:51

Policy Trigger Result: Success

Triggered Condition(s):

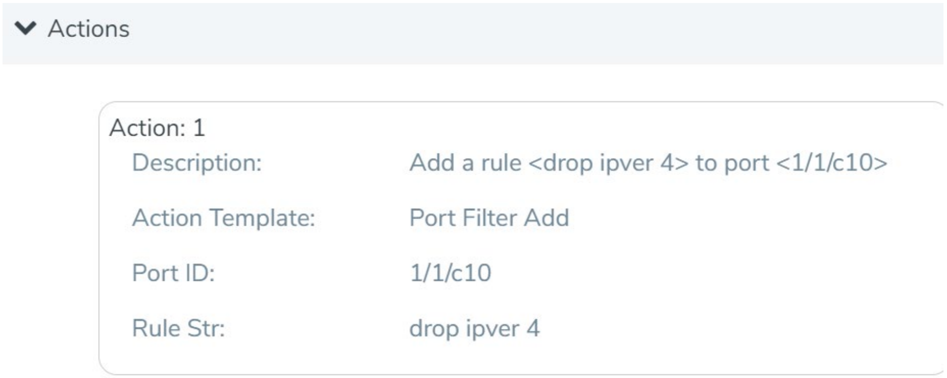
Condition ID	Condition Alias	Param Name	Param Value
1	Port Rx Util High	Port ID	1/1/c8
		Threshold Percentage	20
		Period	10

Triggered Action(s):

Action ID	Action Alias	Action Outcome
1	PortFilterAdd	Success

Figure 5.16: Fraud detection High Utilization - Condition

The Action on Fraud_Detection_High_Utilization is shown in Figure 5.17; to drop specific traffic from the network classified as fraudulent and Figure 5.18 shows the Success of the full policy.

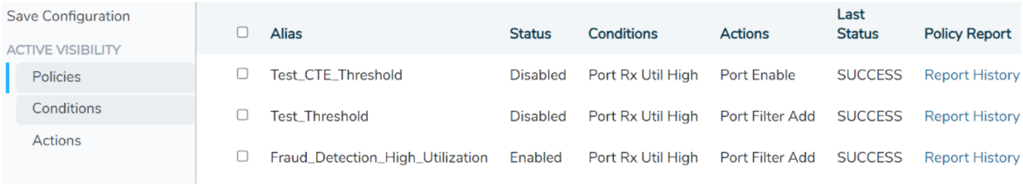


Actions

Action: 1

Description:	Add a rule <drop ipver 4> to port <1/1/c10>
Action Template:	Port Filter Add
Port ID:	1/1/c10
Rule Str:	drop ipver 4

Figure 5.17: Fraud detection high utilization - Actions



□	Alias	Status	Conditions	Actions	Last Status	Policy Report
□	Test_CTE_Threshold	Disabled	Port Rx Util High	Port Enable	SUCCESS	Report History
□	Test_Threshold	Disabled	Port Rx Util High	Port Filter Add	SUCCESS	Report History
□	Fraud_Detection_High_Utilization	Enabled	Port Rx Util High	Port Filter Add	SUCCESS	Report History

Figure 5.18: Enabled and Success policy

5.4 Test Case 3: Fraudulent Access on Cloud Network

The means of using other forms of access to make use of the cloud environment and its key functions on a virtual machine. The test case will allow the identification of a fraudster on the cloud network access.

Fraud is a growing threat to the Communication business, and its technology platforms and systems are becoming more vulnerable to these fraudulent attacks as the network moves to a digital, online strategy. The traditional tools intended to detect fraud are no longer adequate, as new types of fraud emerge, and as criminals find ways to bypass these systems. Such systems typically use rules and thresholds and are based on relatively limited datasets stored in stand-alone databases, which can be difficult to use to detect and prevent fraud in real-time. With more and more data to analyse and an increasing complexity in fraud, a new approach is needed.

The test case leverages Big Data and Machine Learning to deliver the next generation of real-time fraud detection and prevention solutions. And to provide,

- Suspect number
- Alert type
- Record
- Reason for alert
- Fraud event date
- Data protocol used

- Covering the selected fraud types (SIM Swap, Wangiri, PBX Hacking and APN Bypass) with a future option to scale up to detect unknown fraud types. Identifying revenue losses and associated fraud.
- Ability to adapt the cloud infrastructure to handle new use cases such as IoT, Account Takeover, Subscription and Dealer fraud
- Delivering rapid, substantial and measurable ROI.

The solution uses Machine Learning fraud and revenue threat analytics application that natively runs on Hadoop. Hadoop software allows distributed processing of large data sets across clusters of computers using simple programming models. If the underlying Big Data architecture is based on Hadoop the solution will leverage the Hadoop data lake and is fully certified with 2 of the largest Hadoop distributors, Cloudera and Hortonworks.

The solution can reside in the cloud or on-premises behind the firewalls. The framework shows the components in Figure 5.19.

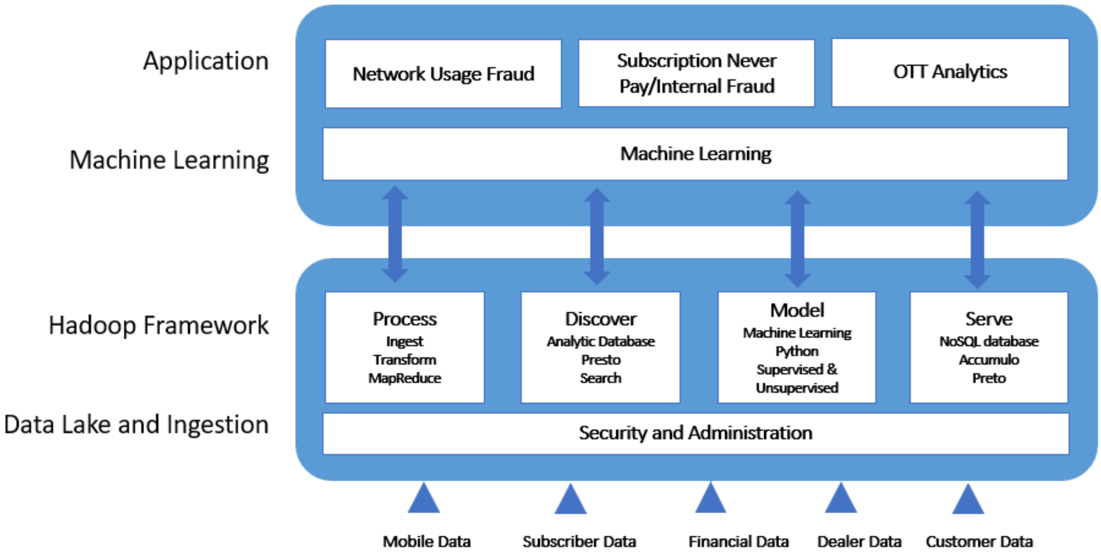


Figure 5.19: Cloud Framework on Fraud

Data Lake and Ingestion

Data such as Subscription applications, Internal Dealer data, online web account logs, CDRs, TD.35, NRTDRE files, SS7 ISUP, GTP, Gi, Gn, LTE, Diameter, NetFlow, Billing records and CRM data usually reside in different data sources across Operators IT platforms. Combining this data to a single view allows both a human and a Machine Learning algorithm to distinguish between an anomalous amount of roaming voice, SMS, or data used by a business user, who has paid their bill on time for the last five years, as opposed to a new post-paid user who has been a subscriber for 5 days. The key to data ingestion is closing the 24-hour window of fraud opportunity. This is where real-time data ingestion analytics is required. The CDRs and ISUP traffic are sent to the platform solution faster and can apply Machine Learning and alert on anomalous traffic behaviour.

In addition to CDRs, Deep Packet Inspection (DPI) technology is used to detect Over-The-Top (OTT) applications, which, although not fraudulent, take revenue away from Operators.

This includes the following data ingestion:

- Voice and Data CDRs
- International ISUP
- TD 35 / TD 57 data roaming CDRs
- Location based data
- MIS
- Customer Care Online logs

Big Data Framework

The middle tier of the architecture is where the data is stored and processed, this is called the Hadoop framework. Hadoop is designed to be scaled from a single server to thousands of machines, with a very high degree of fault tolerance. Rather than rely on high-end hardware, the flexibility of the clusters comes from the software's ability to detect and handle failures at the application layer.

Hadoop also provides superior flexibility for ingesting any type of data, no matter the structure, the source, or the multitude of sources. With the available new technology, Cloud providers prefer their Fraud Management software tools to ingest data in real-time, which Hadoop allows.

Key Value Database and distributed SQL query engine

To complete the Hadoop framework, the architecture includes two additional components to maintain the speed and scale required to be able to detect fraud in real-time. The Key Value Database (Accumulo) provides the ability to store data in massive tables for fast, random access. This helps sort and distribute the data through horizontal scaling across hundreds of machines, offering extremely high performance, particularly for the types of protocols. Accumulo maps rows and column keys, in a similar way to Google BigTable, but also timestamps values, making the database a three-dimensional mapping system. Accumulo has been scaled to process 100 million events per second.

Once the data has been processed, using a distributed SQL (Structured Query Language) query engine designed for ad-hoc analysis at interactive speed via the web-based User Interface, from gigabytes to petabytes using Presto DB. The platform can store data from multiple networks, the distributed SQL query engine will provide a response that can be as short as sub-seconds, rather than hours to provide the real-time functionality they require.

Machine Learning (ML)

In the middle to top layer of the architecture, this is where the Machine Learning algorithm to detect fraud resides. It is not feasible to have fraud analysts manually inspect an incredibly high number of individual calls and then create graphs to detect fraud. It simply doesn't scale; detecting known fraud types on small subsets of data is relatively easy. Detecting known fraud types when you have 24 hours to do it is relatively easy. Detecting known and new (unknown) types of fraud by looking at all the data all the time is very difficult and a reactive approach.

The most optimal way to detect anomalies on a large network is to apply Machine Learning at a massive scale in real-time. When there is enough accessible data available in real-time, fraud can be detected faster in real-time. The Machine Learning algorithm is applied to all new data ingested into the platform.

The solution analyses features as data are being stored and generates fraud probabilities by looking at the new entry and a historic pattern in real-time in Figure 5.20.

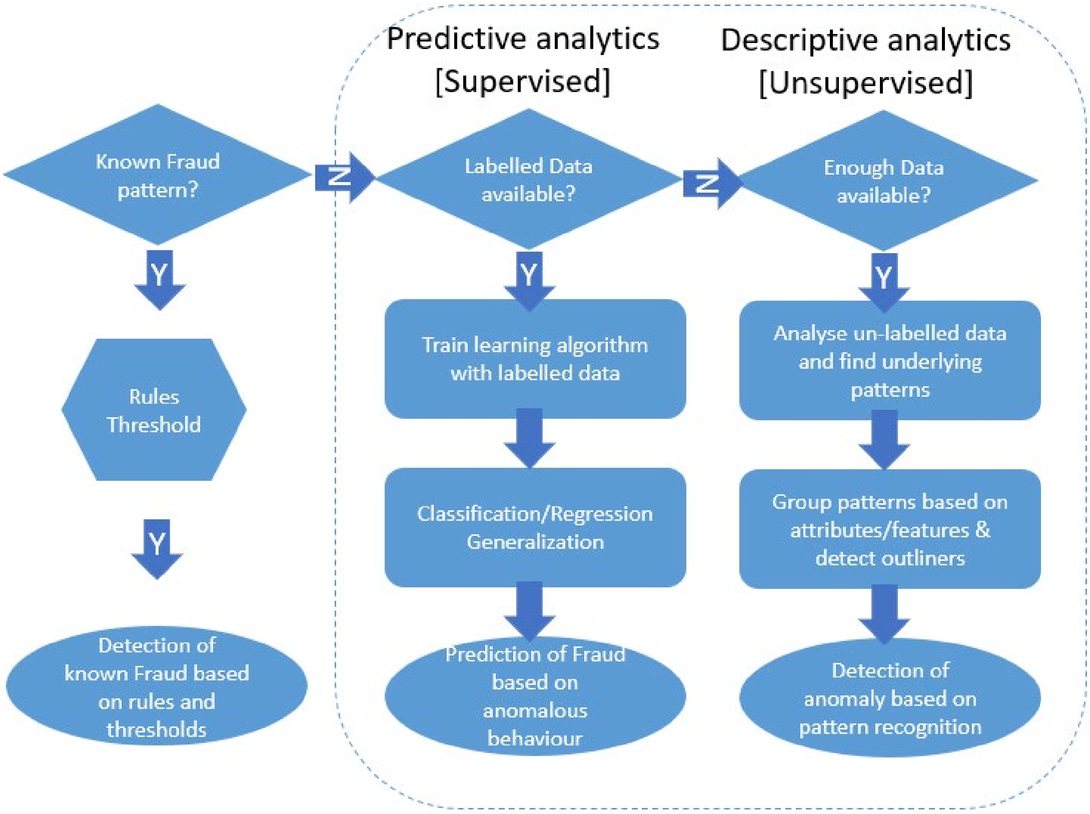

Predictive analytics
[Supervised]
Descriptive analytics
[Unsupervised]
Known Fraud pattern?
N
Labelled Data available?
N
Enough Data available?
Y
Y
Y
Rules Threshold
Train learning algorithm with labelled data
Analyse un-labelled data and find underlying patterns
Y
Classification/Regression Generalization
Group patterns based on attributes/features & detect outliners
Detection of known Fraud based on rules and thresholds
Prediction of Fraud based on anomalous behaviour
Detection of anomaly based on pattern recognition

Figure 5.20: Machine Learning techniques

Application Layer

Once the data has been collected and analysed, the output of the results helps to identify attack points and potential losses with a dashboard composed for the following fraud types (not limited to):

- Immediate use cases to address:
  - SIM Swap, Wangiri, PBX Hacking and APN abuse
- The following are additional use cases:
- Roaming Fraud

- Premium Rate Service Fraud
- Abuse or Arbitrage Fraud
- International Revenue Share Fraud
- Call and SMS Spamming
- Subscription Fraud
- Dealer Fraud
- Internal Fraud

The process of detecting these fraud types includes:

- A combination of many mobile and business data sources as explained in the data ingestion section.
- Applying Machine Learning on all data collected in real-time for anomaly detection against these data sources.
- Leveraging the Hadoop cluster for real-time graph analysis and powerful visualization

Table 5.5 shows the Traditional Approach and using Hadoop Machine Learning Approach. The Hadoop Approach provides better metrics on real-time processing and real-time Anomaly detection.

Table 5.5: Traditional Approach and Hadoop Machine Learning Approach

Traditional Approach	Hadoop Machine Learning Approach
Batch File Ingestion • Time-delayed • Discover of threats over lengthy time	Real-Time packet Ingestion • Real-time processing • Discover in seconds versus hours
Rules-based Known Fraud detection	Real-Time Anomaly detection
Data silos	Enterprise Data lake
Flat world forensics	Graph-based visual analytics
High-cost proprietary architecture	Native Hadoop architecture

- Network Address Translation (NAT) is required for access to Public Cloud Services.
- NAT Provides IP translation inbound and outbound (and both at the same time) between the customers addressing and the public addressing required by the CSP.

Further, Figure 5.21 shows the physical connections of Microsoft Azure on Interconnect 1, which hosts an SVLAN with customer traffic that is either private or public, while Interconnect

2 hosts public and private VLANs. Both are configured with the specification of a Network-to-Network Interface to bridge the communication network to the cloud network. Figure 5.22 shows the NNI connectivity between the communication network using the PE router to the Cloud Service Provider, the configuration is shown:

```
interface GigabitEthernet0/0/0/0
 description test_case_3
 bandwidth 1000000
 monitor-session SPAN_2 ethernet direction tx-only port-level
 !
 load-interval 30
!
interface GigabitEthernet0/0/0/0.14
 description Link to cloud through TX node CCT16567
 vrf voip_Cloud
 ipv4 address 41.1.102.129 255.255.255.252
 encapsulation 802.1q
 logging events link-status
```

Figure 5.23 provides a more insightful view of the Fraud detection framework for Cloud Service Provider consisting of Threat Domain Intelligence, AI/ML Controls, Data Ingestion, Rule Management, Visualization, Alarm & Case Management, and Signalling Security.

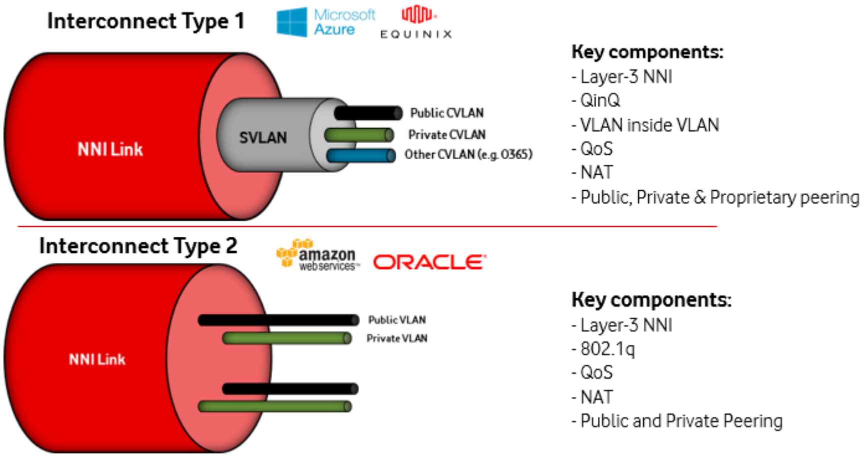


Figure 5.21: Interconnect Type 1 and 2

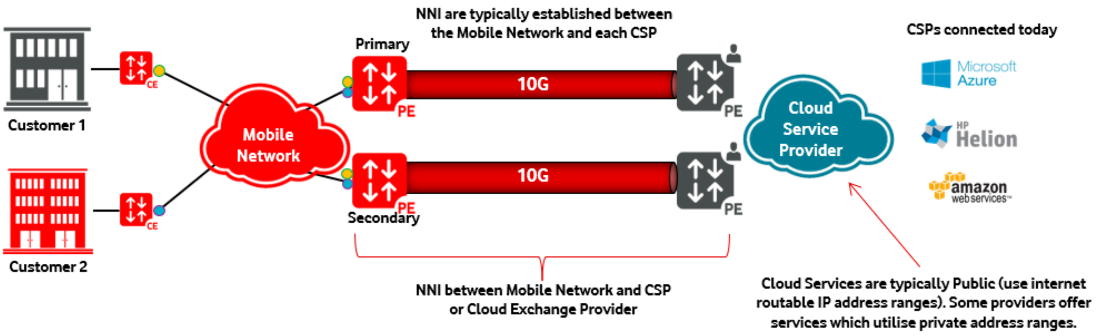

NNI are typically established between the Mobile Network and each CSP
Primary
CSPs connected today
Customer 1
CE
Mobile Network
PE
10G
PE
Cloud Service Provider
Microsoft Azure
HP Helion
amazon web services
Customer 2
CE
10G
PE
Secondary
PE
NNI between Mobile Network and CSP or Cloud Exchange Provider
Cloud Services are typically Public (use internet routable IP address ranges). Some providers offer services which utilise private address ranges.

Figure 5.22: NNI between Mobile Network and Cloud Provider

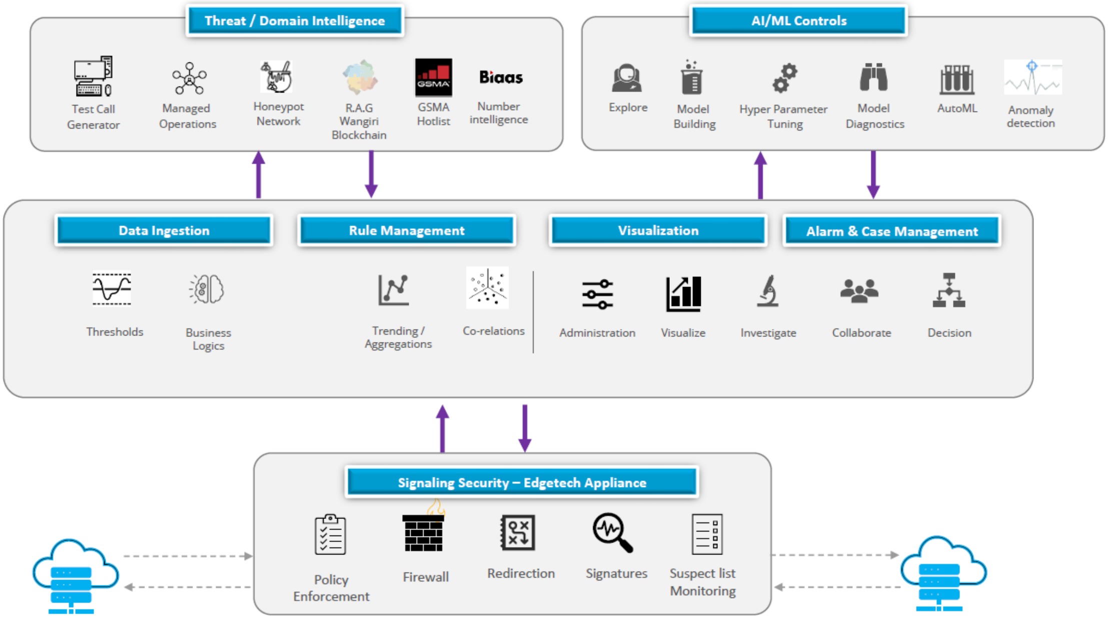


Figure 5.23: Fraud Detection & Presentation - Cloud

Furthermore, Figure 5.24 details The Cloud Fraud Detection process, with Detect and Investigate as its core functions. Whiles Figure 5.25 shows an example of the following:

Source	Subscriber data from 2G/3G/4G/5G
Ingest	Integration
Detect	Rule-based Detection and Deep Packet Inspection
Investigate	Analysis and decision
Protect	Analyst decision
Impact	Post checks

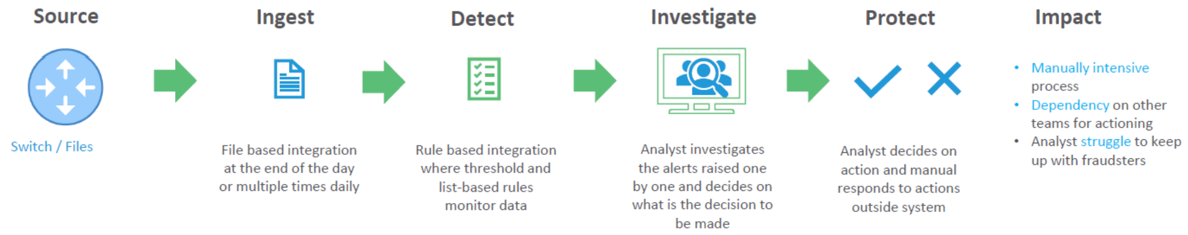

Source
Switch / Files
Ingest
File based integration at the end of the day or multiple times daily
Detect
Rule based integration where threshold and list-based rules monitor data
Investigate
Analyst investigates the alerts raised one by one and decides on what is the decision to be made
Protect
Analyst decides on action and manual responds to actions outside system
Impact
Manually intensive process
Dependency on other teams for actioning
Analyst struggle to keep up with fraudsters

Figure 5.24: Cloud Fraud Detection process

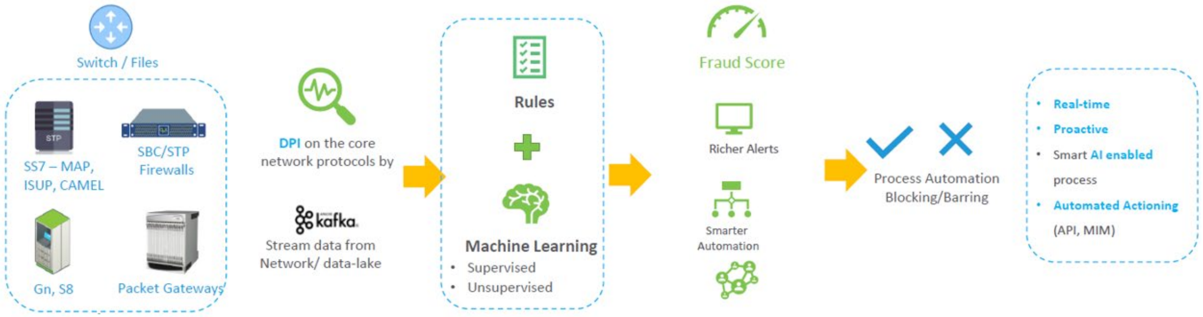


Figure 5.25: Cloud Fraud detection sample process

In terms of Google Cloud, their products are serverless and managed by Google, allowing more time to prepare the data, build the fraud detection model, host online predictions on streaming data, set up the fraud notification, and create operational dashboards. The first stage is to gather the historical data on credit card transactions as training data, containing credit card numbers, transaction amount, merchant information, category and subscriber demographics, the data will be encrypted to protect the record. The last column shows the fraud labelled in Table 5.6.

Table 5.6: Training Data

Subscriber ID	Time stamp	Rated	Billed	Merchant	Province	Fraud
1132922998	16:00	7,05549	7,05549	SB	WC	0
1142045999	16:30	1,92884	1,92884	ABSA	KZN	0
1158583999	17:00	9,62579	9,62579	FNB	NGA	0
1162565599	17:30	20,72355	1	Capetic	SGA	19,72355
1173668897	18:00	26,94287	26,94287	SB	NGA	0
1216092996	18:30	7,92081	7,92081	FNB	KZN	0
1216092998	19:00	29,16317	29,16317	ABSA	WC	0
1223595999	19:30	3,7413	3,7413	FNB	WC	0

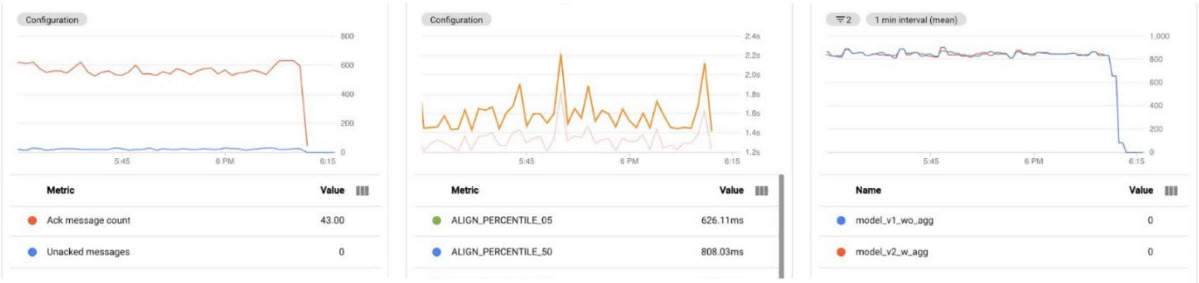
Using the transactional info and subscriber demographics, the possibility of modelling using SQL and BigQuery ML, and different classification algorithms, such as logistic regression, XGBoost, deep neural network, and ML tables for automation, can be considered. For this use case, the XGBoost model retains levels of explainability. Using the model and a few lines of code, the model can be trained. In the following Fraud Model, XGBoost is trained to predict is 1 if fraud else 0 based on various features like transaction category, amount, and demographics.

```
FRAUD MODEL
  `[PROJECT_ID].[DATASET].simplemodel`
OPTIONS(
  model_type='BOOSTED_TREE_CLASSIFIER',
  num_parallel_tree=8,
  max_iterations=50,
  input_label_cols=["is_fraud"]
) AS
SELECT
  *
FROM `Training data`
```

However, from the test case, it requires building a strong performing model, to incorporate a better approach, an additional feature is used on the subscriber's historical data/transactions.

```
FRAUD MODEL
  `[PROJECT_ID].[DATASET].model_w_aggregates`
OPTIONS(
  model_type='BOOSTED_TREE_CLASSIFIER',
  num_parallel_tree=8,
  max_iterations=50,
  input_label_cols=["is_fraud"]
) AS
  SELECT * FROM `Training Data`
SELECT
  "simplemodel" AS model_name,
  *
FROM
  ML.EVALUATE(
    MODEL `[PROJECT_ID].[DATASET].[MODEL_NAME_WITHOUT_AGG]`,
    (SELECT * FROM `Training data`))
UNION ALL
SELECT
  "model_w_aggregates" AS model_name,
  *
FROM
  ML.EVALUATE(
    MODEL `[PROJECT_ID].[DATASET].[MODEL_NAME_WITH_AGG]`,
    (SELECT * FROM  `Training Data`))
```

The model can be exported for online predictions in Figure 5.26.



	Metric	Value
●	Ack message count	43.00
●	Unacked messages	0

	Metric	Value
●	ALIGN_PERCENTILE_05	626.11ms
●	ALIGN_PERCENTILE_50	808.03ms

	Name	Value
●	model_v1_wo_agg	0
●	model_v2_w_agg	0

Figure 5.26: Online Fraud Cloud prediction

5.5 Test Case 4: DDoS attack

To identify when fraudsters attempt to limit the network’s availability by identifying their pattern and scoping a proposed deep-learning technique on detection.

A Distributed Denial-of-Service (DDoS) attack is an attempt to make a network resource unavailable to its intended users. A typical DDoS attack uses a Botnet, (in some cases, hundreds of thousands of compromised computers) to send a constant barrage of data packets to the intended target. The aim is to swamp the target's communication links or consume all its resources by serving bogus requests. The result is genuine users cannot access the target system or performance is so slow that the system is unusable.

DDoS attacks on the communication network are usually the stories that make the evening news but small and medium-sized businesses are often targets as well. These days, any company using its website as a primary method for business transactions is a target, especially during busy transactional periods or high-profile company events such as new product launches. Attackers can use these critical events as opportunities to extort vulnerable businesses that cannot afford to lose their credibility during these important times.

The impact of a successful DDoS attack can be far-ranging and severe:

- Loss of Revenue
- Loss of customers
- Reputational Damage
- Increased Cost of Business

DDoS Attacks also cause collateral damage, affecting network components that are in the transit path of targeted components. Such network components include:

- Business Critical Firewalls
- Business Critical L3/L4 load-balancers and other forms of IP devices

The solution is a network-wide infrastructure security platform that measures and monitor's traffic. It uses both flow and deep packet inspection (DPI) technologies and to provide macro- and micro-level visibility, allowing to identify threats and to improve the performance of the network.

In order to protect networks from DDoS attacks, the following steps are followed in Table 5.7.

Table 5.7: Phases in the detection of DDoS attack

Phases	Description
Preparation	Learn what "normal" traffic is on the network. Preparation offers a means to gain pervasive network visibility and recognize normal traffic patterns.
Identification	Once "normal" traffic is known, abnormalities are identified. Identification models network behaviour, creates a baseline, and alerts when network anomalies are identified.
Classification	Classification determines whether an anomaly is a threat. Classification identifies DDoS and zero-day threats and determines the type, severity, and size.
Trace Back	Trace Back allows real-time historical analysis of all network activity.
Reaction	Reaction enables the initiation of the appropriate mitigation process to stop a threat.
Post mortem	Post-mortem provides detailed mitigation reports that explain what happened and how an attack was alleviated. This knowledge can be used to mitigate future attacks.

DDoS Detectable Attacks

The solution is able to detect the following classes of DDoS attacks against in Table 5.8:

- All IP addresses that are advertised through the Internet Peering Edge
- IPv4-based attacks

Table 5.8: DDoS Attack Types

DDoS Attack Type	Direction	IP version	Description
UDP flood/misuse attack	From the Internet, via Internet Backbone	IPv4	The attacker attempts to overwhelm link bandwidth by sending huge amounts of UDP packets which due to the stateless nature of UDP are simpler to spoof than TCP. It is not possible to eliminate all UDP from a network since a number of core protocols use UDP such as DNS, SIP, real-time multimedia, NTP, etc.
TCP SYN flood/misuse attack	From the Internet, via Internet Backbone	IPv4	The attacker initiates a TCP connection with the victim, but only sends the initial SYN packet and never completes the TCP three-way handshake. This results in the victim reserving memory/connection buffers unnecessarily and eventually is unable to accept new TCP connections from legitimate users.
TCP ACK flood/misuse	From the Internet, via Internet Backbone	IPv4	This attack is also known as a SYN-ACK flood. The attacker tries to hide their presence by spoofing the source IP address of a "relay" node in SYN packets sent to the

			victim. This results in SYN-ACK packets being returned by the victim to the “relay” host. Both the “relay” and the victim are impacted and the attacker’s true IP address is hidden from the telemetry of both the victim and the “relay”.
TCP RST flood/misuse attack	From the Internet, via Internet Backbone	IPv4	In this attack, the attacker sends TCP RST (reset) packets to the victim, using spoofed source IP addresses in an attempt to reset TCP sessions between the victim and legitimate hosts.
Fragmentation flood/misuse attack	From the Internet, via Internet Backbone	IPv4	Re-assembling fragmented IP packets consumes CPU resources for both security devices such as IPS, firewalls as well as servers. The DDoS Platform is able to detect fragmentation attacks using Netflow since a flow of fragmented packets has TCP port field set to 0 in the Netflow telemetry data.
ICMP flood/misuse	From the Internet, via Internet Backbone	IPv4	Attackers can use simple ICMP messages such as Echo Request (“ping”) packets to consume link bandwidth since all nodes with IP stacks are able to generate ICMP traffic.

Countermeasures include the following:

Malformed IP packet filtering (malformed IP header, incomplete fragments, bad IP checksum, duplicate fragments, fragment too long, short packet, short TCP packet, short UDP packet, short ICMP packet, bad TCP checksum, bad UDP checksum)

- Check Layer 3/4 FCAP-based stateless packet filtering (source and destination IP).
- Check IP source address filtering (blacklisting/whitelisting) based on lists of attacking IP addresses.
- Check IP source address filtering (blacklisting/whitelisting) based on country of origin (as determined by the MaxMind GeoIP database built into the DDoS Platform).
- Check IP source address rate-limiting based on country of origin.
- Check Zombie detection (source IP addresses sending “high” traffic).
- Check DNS authentication (active and passive) to protect DNS servers.
- Check TCP idle connection teardown.
- Check Payload regular expression (“regex”) filtering.
- Check Malformed DNS packet filtering.
- Check DNS query rate-limiting per source IP.
- Check DNS packet regular expression (“regex”) filtering.
- Check Malformed HTTP packet filtering.
- Check HTTP object rate-limiting per source IP.
- Check HTTP request rate-limiting per source IP.
- Check HTTP header regular expression (“regex”) filtering.

- Check Malformed SIP packet filtering.
- Check SIP message rate-limiting per source IP.

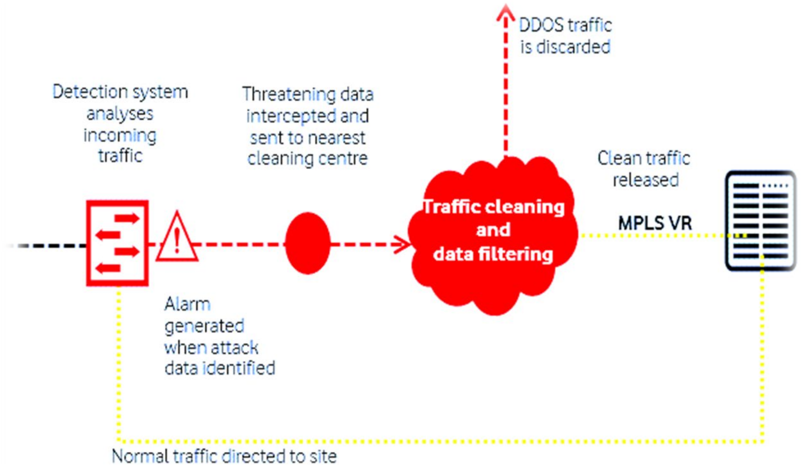

DDOS traffic is discarded
Detection system analyses incoming traffic
Threatening data intercepted and sent to nearest cleaning centre
Clean traffic released
Traffic cleaning and data filtering
MPLS VR
Alarm generated when attack data identified
Normal traffic directed to site

Figure 5.27: Traffic cleaning and data filtering

The solution employs Border Gateway Protocol route injection within AS1273 to advertise a more specific route to the host or network block under attack. The injection of a new route will result in all traffic destined for the victim to be attracted to Cleaning Centre. Clean traffic is then forwarded to the host/IP block under attack via a private MPLS VRF in Figure 5.27. The approach of utilizing a cleaning VRF rather than a GRE tunnel minimizes latency during an attack.

The features are designed to offer the following DDoS mitigation:

- Blocking – Source, Source suspend, per packet, and combinations of source, header, and rate based
- Attack Protection - Flood Attacks (TCP, UDP, ICMP, DNS Amplification), Fragmentation Attacks (Teardrop, Targa3, Jolt2, Nestea), TCP Stack Attacks (SYN, FIN, RST, SYN ACK, URG-PSH, TCP Flags), Layer 7 Application Attacks (HTTP GET floods, SIP Invite floods, DNS attacks, HTTPS protocol attacks), DNS Cache Poisoning, Vulnerability attacks, Resource exhaustion attacks (Slowloris, Pyloris, LOIC, etc.). Flash crowd protection.
- Mitigation - Blacklist/Whitelist, Geo Location reporting and blocking, Zombie blocking, packet content filtering, packet header filtering, Botnet removal (AIF feed), Malformed packet removal (TCP, UDP, DNS, DNSSEC, HTTP, HTTPS, SIP), multiple anti-spoofing countermeasures, blended attack protection, proxy aware countermeasures, rate limiting.

Severity 1: Threat Level 'High' as identified, this includes, but shall not be limited, to the following threats. Service down or major service impact, caused by:

- Flood attacks, such as TCP, UDP, ICMP, Spoofed SYN Flood, Non-Spoofed Syn Flood, UDP Flood, FIN, SYNACK Flood (Spoofed and Non-Spoofed), Ping Flood, Smurf Flood, Combined UDP/TCP/ICMP.
- Fragmentation attacks, such as: IP/UDP, IP/ICMP, IP/TCP
- HTTP attacks, such as Connection Flood (client Attack), HTTP errors 404, etc., HTTP half connections
- BGP Attacks
- DNS Attacks

Severity 2: Threat Level 'Medium' or 'Low', as identified, these do not result in alarms being raised into the SOC. This will include non-service impacting anomalies.

Severity 3: Threat Level 'Non-Urgent' but recorded.

When a subscriber is provisioned on the DDoS mitigation platform, a list of IP addresses can be configured to protect from DDoS attacks. The traffic levels and traffic types flowing towards the protected IP addresses will be baselined to determine the expected traffic usage, and then thresholds set, above which point a DDoS attack alert is generated.

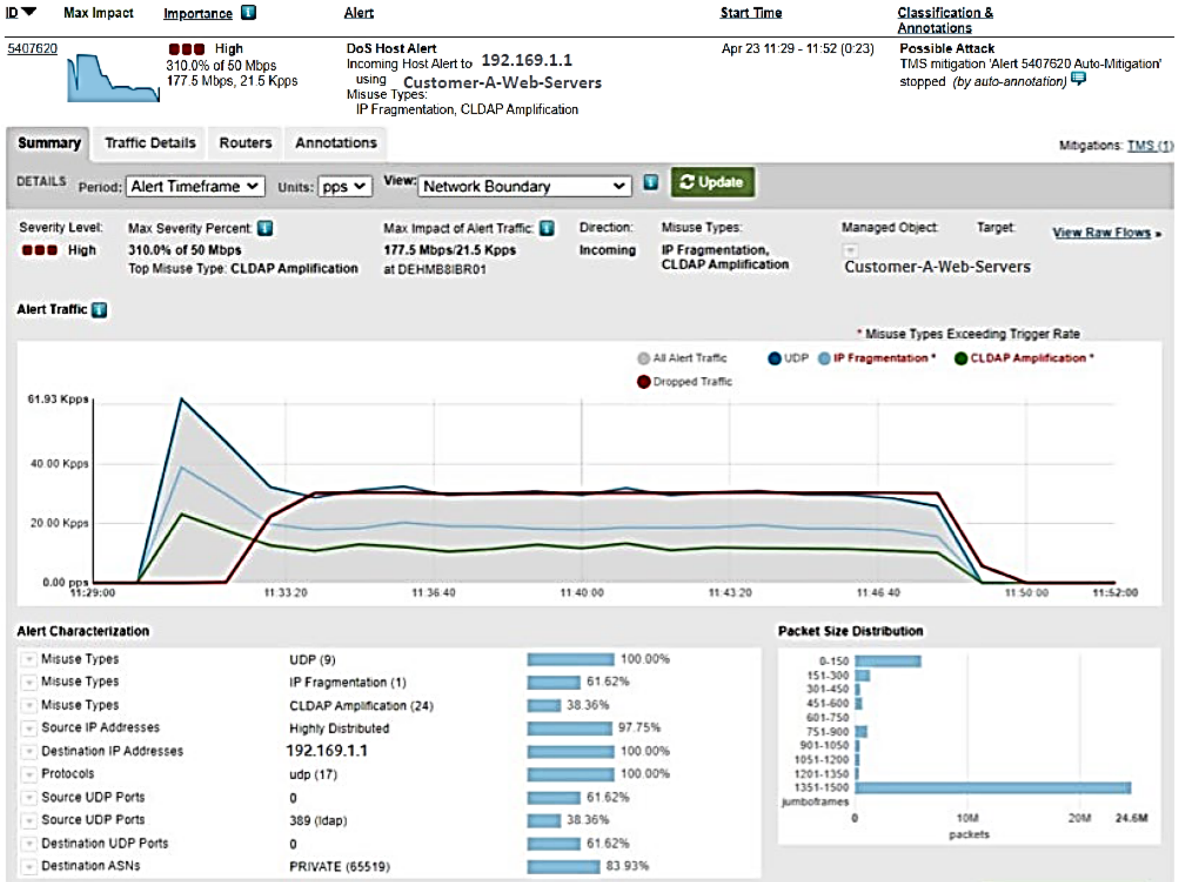
Countermeasures are crafted that are specific to the IP addresses and applications being protected to ensure only legitimate traffic is permitted during a DDoS attack mitigation.
The IP ranges being protected, alerting thresholds, and countermeasures are all linked using what is known as a 'Managed Object'. In Test Case 4, subscriber 'Acme' have a 'Managed Object' called 'Acme-ecommerce-Webservers'. This Managed Object would contain the IP addresses used for the public-facing Acme Web servers and have alerting thresholds and countermeasures associated specifically with those servers. Acme may then have further Managed Objects protecting their SNMP mail servers or VPN concentrators – each with its customized thresholds and countermeasures.

When the DDoS platform detects an attack (based on the pre-defined thresholds set in the Manage Object), mitigation will be initiated automatically or started manually.
The DDoS solution mitigates attacks by diverting the attack traffic to TMS cleaning devices located in strategic points of presence. The TMS devices inspect all inbound traffic to separate the attack traffic from normal traffic.
The separation of attack traffic from normal traffic is achieved by the application of DDoS-specific filters (countermeasures). These countermeasures are defined when the DDoS service is provisioned and can be customized specifically to the customer's needs.

Once the attack traffic has been removed, the cleaned (legitimate) traffic is placed back onto the network where it continues to its original destination.

When an attack occurs, the TMS platform attracts in all the traffic, removes the dirty traffic and passes only clean traffic. This solution removes attack traffic in the network before it hits the customer firewall, routers and servers. This means that the customer and their customers can enjoy uninterrupted service even when attacks happen.

Figure 5.28 shows the start of DDOS detection and the start of mitigation in Figure 5.29.



ID ▼	Max Impact	Importance	Alert	Start Time	Classification & Annotations
5407620		●●● High 310.0% of 50 Mbps 177.5 Mbps, 21.5 Kpps	DoS Host Alert Incoming Host Alert to 192.169.1.1 using Customer-A-Web-Servers Misuse Types: IP Fragmentation, CLDAP Amplification	Apr 23 11:29 - 11:52 (0:23)	Possible Attack TMS mitigation 'Alert 5407620 Auto-Mitigation' stopped *(by auto-annotation)*

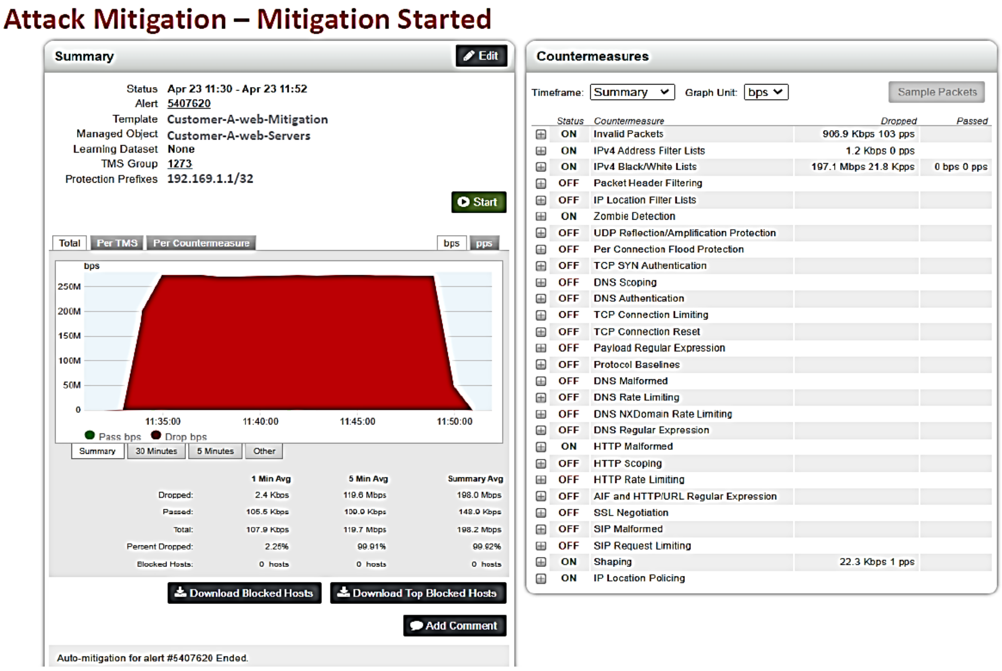
Period: Alert Timeframe
Units: pps
View: Network Boundary

Severity Level:	Max Severity Percent:	Max Impact of Alert Traffic:	Direction:	Misuse Types:	Managed Object	Target
●●● High	310.0% of 50 Mbps Top Misuse Type: CLDAP Amplification	177.5 Mbps/21.5 Kpps at DEHMB8IBR01	Incoming	IP Fragmentation, CLDAP Amplification	Customer-A-Web-Servers	

Alert Characterization

Misuse Types	UDP (9)	100.00%
Misuse Types	IP Fragmentation (1)	61.62%
Misuse Types	CLDAP Amplification (24)	38.36%
Source IP Addresses	Highly Distributed	97.75%
Destination IP Addresses	192.169.1.1	100.00%
Protocols	udp (17)	100.00%
Source UDP Ports	0	61.62%
Source UDP Ports	389 (ldap)	38.36%
Destination UDP Ports	0	61.62%
Destination ASNs	PRIVATE (65519)	83.93%

Figure 5.28: DDOS Detection



Attack Mitigation – Mitigation Started

Summary

Status Apr 23 11:30 - Apr 23 11:52
Alert 5407620
Template Customer-A-web-Mitigation
Managed Object Customer-A-web-Servers
Learning Dataset None
TMS Group 1273
Protection Prefixes 192.169.1.1/32

	1 Min Avg	5 Min Avg	Summary Avg
Dropped:	2.4 Kbps	119.6 Mbps	198.0 Mbps
Passed:	105.5 Kbps	100.0 Kbps	148.0 Kbps
Total:	107.9 Kbps	119.7 Mbps	198.2 Mbps
Percent Dropped:	2.25%	99.91%	99.92%
Blocked Hosts:	0 hosts	0 hosts	0 hosts

Auto-mitigation for alert #5407620 Ended.

Countermeasures

Timeframe: Summary Graph Unit: bps

Status	Countermeasure	Dropped	Passed
ON	Invalid Packets	906.9 Kbps 103 pps	
ON	IPv4 Address Filter Lists	1.2 Kbps 0 pps	
ON	IPv4 Black/White Lists	197.1 Mbps 21.8 Kpps	0 bps 0 pps
OFF	Packet Header Filtering		
OFF	IP Location Filter Lists		
ON	Zombie Detection		
OFF	UDP Reflection/Amplification Protection		
OFF	Per Connection Flood Protection		
OFF	TCP SYN Authentication		
OFF	DNS Scoping		
OFF	DNS Authentication		
OFF	TCP Connection Limiting		
OFF	TCP Connection Reset		
OFF	Payload Regular Expression		
OFF	Protocol Baselines		
OFF	DNS Malformed		
OFF	DNS Rate Limiting		
OFF	DNS NXDomain Rate Limiting		
OFF	DNS Regular Expression		
ON	HTTP Malformed		
OFF	HTTP Scoping		
OFF	HTTP Rate Limiting		
OFF	AIF and HTTP/URL Regular Expression		
OFF	SSL Negotiation		
OFF	SIP Malformed		
OFF	SIP Request Limiting		
ON	Shaping	22.3 Kbps 1 pps	
ON	IP Location Policing		

Figure 5.29: DDOS mitigation

The DDoS Mitigation service performs the following configured specification in response to a DDoS attack:

Auto Mitigation

The platform will automatically start mitigation within 120 seconds of a High Severity Alert being generated. The mitigation will end at a pre-defined time after the alert has ended.

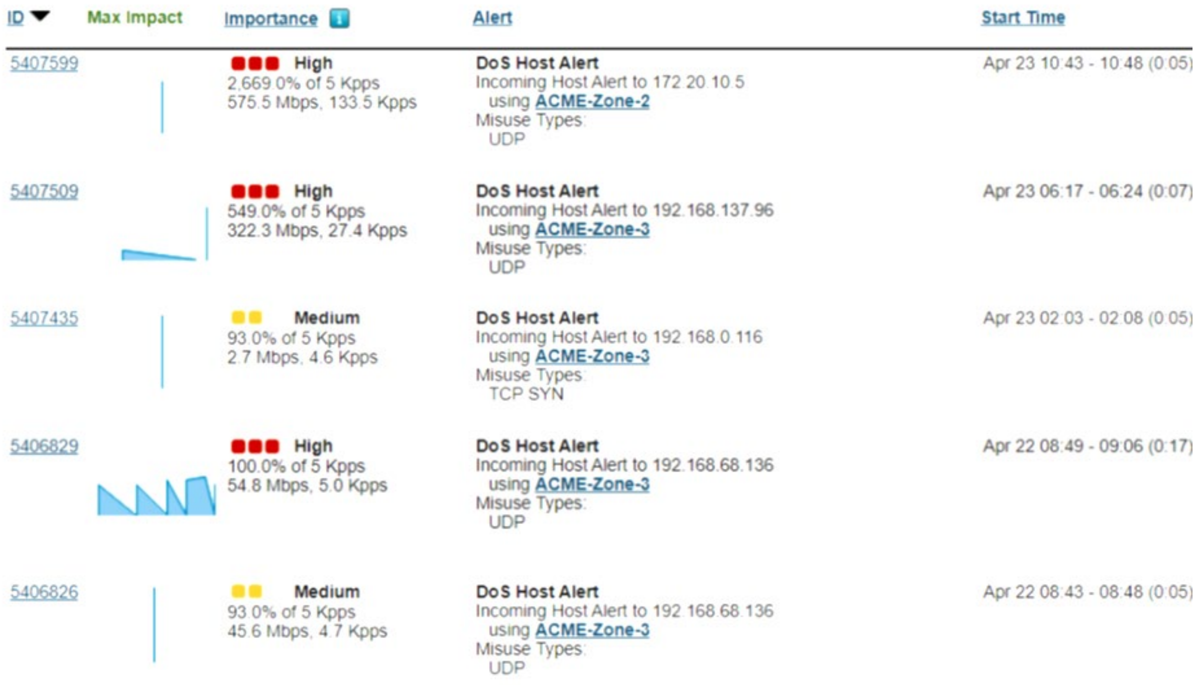
Manual Mitigation (Pre-Authorised)

The support group will need to carry out an initial triage of a High Severity Alert and will start mitigation within 15 minutes if they deem the alert to be genuine. Contact with the subscriber after mitigation to advise them of the detected alert and running mitigation.

Manual Mitigation – (Non-Pre-Authorised)

The support group will carry out an initial triage of High Severity Alert and contact the subscriber within 15 minutes if they deem the alert to be genuine. The customer must then confirm that a mitigation can be started – or that the alert is a false positive requiring no further action.

Figure 5.30 shows the alert levels of High/Medium and the sample.



ID ▼	Max Impact	Importance	Alert	Start Time
5407599		■■■ High 2,669.0% of 5 Kpps 575.5 Mbps, 133.5 Kpps	DoS Host Alert Incoming Host Alert to 172.20.10.5 using ACME-Zone-2 Misuse Types: UDP	Apr 23 10:43 - 10:48 (0:05)
5407509		■■■ High 549.0% of 5 Kpps 322.3 Mbps, 27.4 Kpps	DoS Host Alert Incoming Host Alert to 192.168.137.96 using ACME-Zone-3 Misuse Types: UDP	Apr 23 06:17 - 06:24 (0:07)
5407435		■■ Medium 93.0% of 5 Kpps 2.7 Mbps, 4.6 Kpps	DoS Host Alert Incoming Host Alert to 192.168.0.116 using ACME-Zone-3 Misuse Types: TCP SYN	Apr 23 02:03 - 02:08 (0:05)
5406829		■■■ High 100.0% of 5 Kpps 54.8 Mbps, 5.0 Kpps	DoS Host Alert Incoming Host Alert to 192.168.68.136 using ACME-Zone-3 Misuse Types: UDP	Apr 22 08:49 - 09:06 (0:17)
5406826		■■ Medium 93.0% of 5 Kpps 45.6 Mbps, 4.7 Kpps	DoS Host Alert Incoming Host Alert to 192.168.68.136 using ACME-Zone-3 Misuse Types: UDP	Apr 22 08:43 - 08:48 (0:05)

Figure 5.30: Alert levels

5.6 Summary of Test Case(s)

Test Case 1: Fraud Detection with Unpaid Token

Uses traffic classification to identify fraud patterns and create a machine learning algorithm for detection. Integrates subscriber and billing platforms, performs component analysis, and validates correct billing. Results: Identified incorrectly billed subscribers due to platform abuse.

Test Case 2: Fraud Detection Utilizing the Network

Evaluates the effects of fraud on traffic classification and builds an algorithm to automate detection. Addresses complexities from different devices, encryption, processing power, and traffic congestion. The framework: Includes preprocessing, evaluation, classification techniques, feature engineering, data collection, deployment, monitoring, feedback, compliance, and privacy. Tools Used: APacket tool, Fabric Manager, formatted pcap file. Policies, Conditions, and Actions:

- Policy: Guidelines to prevent, detect, and respond to fraudulent activities.
- Condition: Criteria to detect suspicious activity.
- Action: Steps taken in response to detected fraud.

Results: Successfully dropped fraudulent traffic over a 30-day period.

Test Case 3: Fraudulent Access on Cloud Network

Ensures high security and control over network infrastructure in a hybrid cloud environment.

Test Case 3a: Simulating Fraudulent Access Attempts Steps:

- Baseline access review
- Simulate fraudulent access attempts
- Monitor and log activity
- Analyze detection mechanisms
- Response and mitigation
- Post-test cleanup Expected Results: Detection of fraudulent access, generation of alerts, and appropriate response actions.

Test Case 3b: Unauthorized Access Simulation Steps:

- Initial setup
- Privilege escalation attempt
- Access control bypass
- Suspicious activity simulation
- Monitor for anomalies
- Response action evaluation
- Review access logs
- System and security policy review

Output: Identifies attack points and potential losses. Includes various fraud types like SIM Swap, Wangiri, PBX Hacking, etc. Detection Process: Combines multiple data sources, applies machine learning for real-time anomaly detection, and uses Hadoop for real-time graph analysis and visualization.

Test Case 4: DDoS Attack

Identifies and mitigates DDoS attacks using deep-learning techniques.

Attack Methods: Botnets, blackmail, state-sponsored, personal, or random attacks. Types of Attack: Bandwidth vs. packets per second, reflection/amplification attacks, slow attacks. Detection and Mitigation: Uses Managed Objects (MOs) to define network objects, set up alerts, and implement mitigation actions.

Mitigation Techniques: Blocking of source, header, and rate-based blocking, attack protection, and anti-spoofing measures.

CHAPTER SIX
Discussion

6. Discussion

Fraud detection algorithms have evolved significantly, adapting to various types of fraud and the increasing sophistication of fraudulent activities. The below are summary points applied to the Test case(s) on methods:

Rule-Based Systems: These systems use predefined rules to identify fraudulent activities. For example, a rule might flag transactions over a certain amount or from unusual locations. While straightforward, these systems can be easily circumvented by sophisticated fraudsters.
On Test Case 1: Fraud detection with unpaid token on the Communication Network uses Alteryx Designer software tool and making use of its components. The data used is collected from Call Detail Records (CDR) which holds the subscriber ID, data usage, and details of the session. The process flow uses input data from the CDR file, which is converted into excel format, the two inputs used are the subscriber 'Rated' information and subscriber 'Billed' information. The method is able to compare and filter out the fraud identified.

Machine Learning Algorithms: a more advanced methods, including supervised learning (like decision trees and support vector machines) and unsupervised learning (like k-means clustering), can identify complex patterns in large datasets. These algorithms learn from historical data and can improve over time but require substantial data and computational power.
Test Case 2: Fraud detection utilizing the network applies an algorithm based on Policy, Condition and Action, the algorithm identifies traffic patterns on thresholds configured over a monitoring time. The Condition values can be per-defined to enable accurate automation of fraud identification.

Anomaly Detection Algorithms: These algorithms identify outliers in data that deviate from established patterns, often using methods like isolation forests or autoencoders. They are effective in identifying new or evolving fraud patterns but may produce false positives.
Test Case 3: Fraudulent Access on Cloud Network presents its method to detect fraud on a cloud network using Ingest, Detect, Investigate, Protect and Impact.

Deep Learning: Neural networks, particularly those designed for sequential data (like LSTMs), can detect intricate patterns in transaction sequences. They are powerful but often seen as "black boxes," making interpretation and transparency a challenge.

Test Case 4: DDoS attack, it uses both flow and deep packet inspection (DPI) technologies and to provide macro-level and micro-level visibility, allowing to identify threats and to improve the performance of the network.

Graph-Based Algorithms: These analyse relationships between entities (like transactions) to uncover fraud rings and collusion. They provide valuable insights but require a thorough understanding of graph theory also applied to Test case 2 and Test case 4.

While current algorithms are still relevant, their effectiveness can be diminished by the below factors:

- Evolving Tactics: Fraudsters continuously adapt to their methods, making it crucial for algorithms to be flexible and updated on a 6-month base. Relying solely on traditional algorithms can lead to gaps in detection.
- Data Privacy Concerns: Increasing regulations around data privacy (like POPIA) can limit access to the data necessary for effective fraud detection, making it harder for algorithms to perform optimally.
- False Positives: Many algorithms struggle with balancing sensitivity and specificity, leading to high rates of false positives that can frustrate legitimate users and waste resources. One of the areas that showed during testing were zero rated websites that is requested by government.
- Integration Challenges: Implementing advanced algorithms often requires significant investment in technology and training, which can be a barrier for some organizations. From the test case(s) it was identified that the cost to expand the solution due to growing capacity plays a negative role to meet tocxf the initial investment, though, different techniques can be used to mitigate the cost factor. Such as:

    - Internet Protocol pooling

    The process is to configure on the Tapping platform a selection of Internet Pools, each site within a communication network had a number of Internet pools, let’s take an example, site a had 40 Internet pools, to ensure the capacity processing limits are met, 20 Internet pools are configured within a map rule for month 1 and the next 20 Internet pools are configure for month 2. Each Internet pool has an equal capacity. The process of alternating the pools allows a huge saving on capacity expansions.

    - Site Reduction

The next approach explores 10 sites that has equipment at each site, to meet growing capacity demands, equipment can be spread across 8 sites and not 10 by re-distribution of equipment at sites, allowing configuration for on the full complement of Internet pools. The challenge is moving the equipment from site to site or province to province.

Traffic classification in the communication networks is essential for several reasons, primarily to enhance security, improve quality of service, and manage network resources effectively. Here’s a summary into the requirement for traffic classification, detection, and blocking on Table 6.1.

Table 6.1: Traffic classification, detection, and blocking

Security Enhancement	• Threat Detection caters on classifying traffic to identify fraud activities and patterns, such as DDoS attacks, malware, or unauthorized access attempts. By detecting anomalies in traffic patterns, so the mobile networks can respond actively. • Data Protection of sensitive data transmitted over mobile networks can be targeted by cyber dark world. Traffic classification helps to identify and secure subscriber data.
Quality of Service (QoS)	• Prioritization of different applications require different levels of service. For example, ShowMax/Netflix need low latency and high bandwidth. Classifying traffic enables the network to prioritize these applications over less critical traffic, ensuring a good user experience. • Resource Allocation provides effective traffic classification allows mobile networks to allocate specific capacity resources dynamically based on current demands, reducing link congestion and improving overall network experience.
Network Management	• Network Monitoring and Analytics to understand traffic patterns helps operators monitor network performance and user behaviour to cater for more popular services. • On Billing and Policy Enforcement, Traffic classification aids in implementing data caps, zero-rated traffic, fair usage policies, and targeted marketing based on government, user behaviour and preferences.
Regulatory Compliance	• Many sites have regulations requiring mobile operators to ensure certain levels of service and security. Traffic classification can help demonstrate compliance and avoid penalties.
Blocking Malicious Traffic	• Prevention of attacks by classifying and blocking harmful traffic, networks can mitigate the risks of attacks, ensuring a safer environment for account holders.

	- Content Filtering for Operators can block access to inappropriate or harmful content, protecting subscribers and adhering to lawful intercept

Challenges

While traffic classification provides key benefits, it also presents issues, listed:

- Privacy concerns of subscribers may be wary of how their data is being classified and monitored.
- Evasion techniques of online criminals are increasingly using encryption and other methods to hide malicious traffic, making classification more challenging task. The process of decryption serves importance for the solution. Chapter 5 provided a deeper explanation and type of de-encryption.
- Resource Intensive by implementing effective traffic classification systems can require higher capacity processing resources and experts.

Traffic classification based on detection and blocking is crucial for maintaining secure, efficient, and user-friendly mobile communication networks. By effectively managing and analysing network traffic, the communication network can enhance security, optimize performance, and provide a better overall experience for subscribers on speed, network availability and data protection. In summary, while fraud detection algorithms remain crucial, their effectiveness depends on continuous adaptation, integration with real-time data, and consideration of evolving regulatory and subscriber base.

Sections 6.1, 6.2, 6.3 and 6.4 compare previous research for Test Case(s).

6.1 Test Case 1: Fraud detection with unpaid token on the Communication Network

Fraud detection using Machine Learning in various sectors involves analysing traffic patterns to identify fraudulent activities. Here are some detailed references and insights for each sector:

Energy Sector

In the energy sector, fraud detection focuses on identifying anomalies in consumption patterns, which can indicate meter tampering or unauthorized usage. According to the Communications Fraud Control Association 2019, the global telecom industry has reportedly suffered a loss of 28.3 billion a year on account of fraud, 1.74% of its total annual revenue. In percentage terms, the increase in fraud loss is estimated to be 37%, when compared to their 2017 report. (Machine Learning algorithms analyse large datasets from smart meters to detect unusual patterns that suggest fraud, 2021)

Communication Sector

The communication sector faces significant challenges with various types of fraud, such as subscription fraud, identity theft, and international revenue-sharing fraud. Machine Learning techniques, including anomaly detection and predictive analytics, are crucial for identifying and mitigating these threats. In a recent report by Gartner, it was estimated that in 2024, new implementations of ML within CSP fraud management would help reduce fraud losses by 10%. (AI-based systems can track fraudsters faster and reduce revenue loss by differentiating between genuine customers and fraudsters, 2022)

Service Application Sector

In service applications, fraud detection involves monitoring user behaviour and transaction patterns to identify suspicious activities. Revenues in 2023 are estimated at 498 billion dollars – and it is still growing, at an expected rate of 2.08% year-on-year. And while fraud attacks usually evolve quickly over time as companies squash them, telco fraud is unique as it's often taken as a given. Its costs are absorbed by operators, who would rather not embark on integrating complex risk management systems into their architectures. (Machine Learning models can classify traffic and detect deviations from normal usage patterns, helping to prevent fraud in real-time, 2024)

To create an effective fraud detection algorithm using Machine Learning, the following steps are gathered from the research and training model:

a. Data Collection: Gather historical data on normal and fraudulent activities. The test case used an anonymous CDR file and pcap file.

b. Feature Engineering: Identify relevant features that can help distinguish between legitimate and fraudulent behaviour. The data required to be formatted and cleaned for processing.

c. Model Training: Use Machine Learning techniques such as supervised learning (e.g., decision trees, neural networks) or unsupervised learning (e.g., clustering, anomaly detection) to train the model. The test case incorporated Alteryx Designer to design the components used on the training model.

d. Evaluation: Validate the model using a separate dataset to ensure its accuracy and reliability. The test case compared the CDR user report to the billing system.

e. Deployment: Implement the model in a real-time system to monitor and detect fraud continuously. The test case is built on real-time architecture.

Telecom fraud, such as Wangiri fraud from Arafat, Qusef, and Sammour. (2019), poses significant challenges for operators, leading to revenue losses and difficulties in detection and prosecution. Wangiri fraud involves missed calls from premium numbers, enticing subscribers to call back and incur high charges. This paper suggests using ensemble classifiers, particularly the Extreme Gradient Boosting algorithm, to improve fraud detection accuracy and efficiency, especially during less monitored periods like holidays and weekends.

The study analysed a Call Detail Record (CDR) dataset with 2,415,919 calls, including 1.58% fraudulent Wangiri calls. The data was split into 70% for training and 30% for evaluation. The Extreme Gradient Boosting (XGBoost) classifier performed best, reducing false negatives from 172 to 34 at a 0.2 classification threshold and achieving the lowest Brier score loss of 0.000201, indicating high calibration.

The paper in Guo, Sui, and Shi. (2011) discusses the rapid development of IP-based value-added and data services in mobile networks, which has led to an increase in chargeable services and associated billing attacks. Traditional fraud detection methods struggle with these attacks. The proposed solution in Guo, Sui, and Shi. (2011) combines identity authentication, key process monitoring, and anomaly service traffic identification to detect and prevent billing attacks. The system inspects fraudulent traffic and can alarm or block billing attempts.

The threat model shows attackers can exploit the air interface, core network, and service providers (SPs) by impersonating legitimate users or SPs. The solution includes a rule-based correction engine to analyse identity detection, attack inspection, and traffic investigation results. An Integrity Monitor module checks the integrity of service messages and data, while a Prevention Module blocks attack traffic and can intervene in the billing process to mitigate losses.

Whiles the authors provide successfully testing and its approaches to meet fraud detection, their research does not explore the communication network or a centralized base or an end-to-end solution, the purpose of this thesis were to explore a solution that caters overall.

6.2 Test Case 2: Fraud detection utilizing the network

Fraud detection utilizing communication networks involves evaluating the effects of fraudulent users on traffic classification and building algorithms to automate detection. Here are some detailed insights:

Fraud Detection in Communication Networks

Fraud detection in communication networks often involves identifying patterns of fraudulent behaviour, such as unusual call patterns, subscription fraud, and identity theft. Machine Learning and deep learning techniques are employed to analyse large volumes of data and detect anomalies that indicate fraud. The test case uses an anonymous pcap file to present the test case.

Evaluating the Effects of Fraud Users

To evaluate the effects of fraud users on traffic classification, it was essential to analyse historical data and identify patterns that differentiate legitimate users from fraudsters. This involved:

- Data Collection: Gathering data on normal and fraudulent activities.
- Feature Engineering: Identifying features that can help distinguish between legitimate and fraudulent behavior.
- Traffic Classification: Using Machine Learning models to classify traffic based on identified features.

To Create the algorithm to automate fraud detection involved the following steps:

- Model Training: Using supervised learning (e.g., decision trees, neural networks) and unsupervised learning (e.g., clustering, anomaly detection) to train the model. The test case uses the configured Fabric Manager to present the Policy, Condition and Action.
- Evaluation: Validating the model using a separate dataset to ensure accuracy and reliability.
- Deployment: Implementing the model in a real-time system to monitor and detect fraud continuously.

The article by Gandhar et al., (2024) provides a summary overview of the latest Machine Learning and deep learning methods for fraud detection, including their positive and negatives outcomes. An Optimized Deep Learning Approach for Detecting Fraudulent Transactions: The study by El, Tayebi and Sulimani (2024) proposes an intelligent system for detecting fraudulent transactions using various deep learning architectures, including Artificial Neural Networks (ANNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM). The imbalanced European credit card dataset was transformed into a balanced dataset. This happened after removing several normal transactions from the majority class to test the three deep learning architectures with hyperparameter optimization using the Bayesian optimization approach. The random undersampling method with a sampling strategy, From these results, the best accuracy score, 95.93%, is achieved by the RNN architecture, while the second-best score, 89.93%, is obtained by the ANN model. Moreover, the lowest score, 75.62%, is obtained using the LSTM model.

Traffic classification is cutting edge key in cybersecurity for preventing various threats. Traditional models struggle with rapid learning and recognizing ‘unknown’ traffic. The paper by Miao, et al., (2019) introduces the Siamese Prototypical Network (SPN), a few-shot traffic multi-classification method that supports out-of-distribution (OOD) detection. SPN integrates twin networks into a meta data learning framework and uses margin loss to enhance detection success.

The lab testing on three benchmark datasets (ISCX2012, CIC-IDS2017, and USTC-TFC2016) show that SPN excels in few-shot multi-classification, OOD detection, and intrusion detection. SPN outperforms traditional methods, demonstrating strong generalization capabilities and low resource over-head, making it effective for present traffic scenarios.

The research addresses the increasing security threats by Sah and V. (2024) to networked systems and the importance of detecting fraud in network behaviour to prevent fraud and unauthorized access. It highlights the limitations of traditional security measures and the need for advanced intrusion detection performance systems (IDS).

Using the CICIDS-2017 dataset, the study compares Decision Trees and Random Forests for anomaly-based intrusion detection. Both algorithms achieved over 99% accuracy and F1 scores. The Random Forest algorithm performs better on the Decision Tree in multiclass classification, with an accuracy of 0.99888 compared to 0.99867, despite a longer execution time. These results demonstrate the high effectiveness of both algorithms, with Random Forest being more efficient for multiclass classification.

Kilinc, H.H. (2022) conducted a detailed call traffic analysis using K-means clustering to segment multi-valued categorical variables. For anomaly detection, it employed several supervised models, including XGBoost, Extra Trees, and Random Forest, as well as an unsupervised model known as Isolation Forest. Additionally, it introduced a mixture of key model to compute average prediction probabilities. Anomaly scores were generated by summing the predictions from 5 models, and 1% of the calls were labelled as suspected fraud, consistent with industry reports. Our analysis revealed that the top 10 countries by the number of anomaly calls were the USA, the UK, the Netherlands, Uzbekistan, Libya Russia, Germany, and Switzerland. When considering the ratio of anomaly calls to total calls, the top suspect countries were Martinique, Ecuador, Belize Sri Lanka, Latvia, Mauritius, Chad, Marshall Islands, Portugal, and Ethiopia. The combined recall and F1-scores for the Random Forest and Extra Trees models were 83% and 89%, respectively, while the 3-model combination yielded scores of 79% and 88%. The research findings align with the CFCA 2019 Global Fraud Loss Survey, indicating that Indian Ocean Island countries, Caribbean and some African and Eastern European countries have similar fraud interests. This analysis provides valuable insights into suspicious call trends and model performance.

Learning from other authors and their research, it was proven the use of Policy, Condition and Action applied to traffic classification is able to detect fraud for any service and remains the central point of gathering the data and automation.

6.3 Test Case 3: Fraudulent Access on Cloud Network

Fraudulent access on cloud networks is a significant concern, especially as more organizations rely on cloud environments for their operations. Fraudsters often exploit vulnerabilities in cloud access points to gain unauthorized access to sensitive data and resources.

One common method used by attackers is token theft. This involves compromising and replaying a token issued to an identity that has already completed multifactor authentication (MFA). By doing so, the attacker can bypass MFA and gain access to organizational resources. This method is particularly concerning because it requires minimal expertise, becoming challenging to detect, and many organizations lack proper mitigations in their incident response plans. (Microsoft, 2022)

Another prevalent tactic is password spraying and brute forcing. Fraudsters use these techniques to gain access to machine driven system accounts and accounts not in use over 60 days, which often lack MFA and have weak passwords. These accounts are easy targets for hackers, allowing them to infiltrate cloud environments and orchestrate sophisticated attacks. (PYMNTS, 2024)

To mitigate these risks, companies should implement up-to-date security measures, such as enforcing weekly password change policies, enabling MFA for all user accounts, and regularly monitoring for suspicious activities.

Dawood M et al., (2023) presents the different security challenges and concerns associated with cloud computing in their comprehensive guideline. They highlight challenges such as data breaches, data confidentiality, and authentication problems. The researchers duplicate the importance of understanding these security threats to successfully deploy cloud computing solutions.

PYMNTS, (2024) experiment on how fraudsters exploit cloud access handover to orchestrate sophisticated attacks. The paper underscores the growing complexity of cloud security as more providers migrate data in the cloud, making it an easy target for online attacks.

StrongDM (2024) provides an overview of unlicensed access, including trends, patterns, types, examples, and prevention techniques. They discuss how online predators exploit vulnerabilities through phishing, social engineering, and criminal force attacks. The researchers also highlights the severe consequences of non-permitted accounts, such as profit losses and brand damage.

These discussions provide valuable insights into the various techniques used by attackers to gain unauthorized access to cloud providers and the fundamental of implementing robust security measures to safeguard against these threats.

Federated Learning (FL) has been applied to various test cases in corporate networks. Authors in Talluri, Zhang and Chen (2023) have proposed integrating federated learning into the 3GPP fifth Generation network data analytics. The test cases for federated learning in fifth Generation networking include time-series of future building for predictive analysis, classification for traffic management, and Quality of Experience modelling.

The paper proposes an in-depth cloud-native federated learning topology consisting of a federated learning server with serverless microservices on the cloud and clients with private IT/Machine Learning platforms. A unique feature of this architecture is that federated learning clients initiate local training by pulling text from a designated message queue created by the server, thus avoiding direct access by the FL server to the clients' IT/Machine Learning systems. Additionally, the federated learning server employs serverless microservices, where model aggregation computation is triggered passively upon receiving a defined set of local models, leading to profit preservation compared to an old method, persistently running FL server.

The experimental of cloud-native federated learning topology by training a global Machine Learning model across many providers for telecom fraud detection. The test bench resulted in showing that the global model trained via federated learning achieved up to a 23% improvement in F1-score compared to traditions models trained at individual providers.

Based on their analysis and research, authors in Abitova *et al*., (2023) concluded that the aim of using Artificial Intelligence for fraud detection shows positive with the adoption of advanced technologies. Other technologies such as Machine Learning algorithms, including deep learning and neural networks, and cloud computing platforms play a pivotal role. These technologies are particularly effective due to their ability to process long segments speedily and to identify mix trends/patterns indicative of fraudulent behaviour, which are crucial for real-time detection.

Australia's largest financial institutions from Abitova *et al*., (2023) transformed its simplistic, rules-based risk impact into more nuanced, Artificial Intelligence-powered fraud detection. They achieved a 400% increase in identified fraud cases, a 33% reduction in flagged transactions, and a 76% reduction in the original review financial year budget.

Current models and computing platforms typically lack the flexibility and scalability offered by Artificial Intelligence and cloud-based solutions. Current methods may not process data as in packet sequence or adapt as quickly to changing fraud patterns as Artificial Intelligence-driven systems do. With computational theory, a simple logistic regression formula is used in Artificial Intelligence for fraud detection. This formula represents how Artificial Intelligence models measure the probability of a transaction being fraudulent, based on various software inputs.

The research of cloud Artificial Intelligence in the context of real-time fraud detection reveals a dynamic and speedily environment adaptation. This research has investigated the multifaceted nature of cloud Artificial Intelligence, highlighting its robust capabilities and potential for innovation in combating financial fraud.

The experiment conducted by Roy *et al*., (2024) using Python, leveraging popular deep learning libraries such as TensorFlow and PyTorch. The infrastructure specifications were not explicitly detailed, as they depend on available resources and the computational demands of the chosen deep autoencoder topology. The dataset comprised anonymized bank transaction records, totalling 1 million entries. Key features included transaction amount, recipient, location, timestamp, and account type.

A Deep Autoencoder learns on this dataset, employing a 3-layer encoder and decoder structure with 128 neurons per layer. The model demonstrated promising results in anomaly detection, achieving a perfect recall of 1.0, meaning it identified all actual anomalies in the test data. However, the precision was 0.59, indicating some legitimate transactions were incorrectly flagged as anomalies.

The aim is the protect the cloud provider's network and its subscribers, the most optimal way to detect anomalies on a large network of the cloud is to apply Machine Learning at a massive scale in real-time. When there is enough accessible data available in real-time, fraud is detected faster in real-time. The Hadoop Approach provides better metrics on real-time processing and real-time Anomaly detection. Figure 5.24 details the Cloud Fraud Detection process, with Detect and Investigate as its core functions. The results show positive on the detection process.

6.4 Test Case 4: DDoS attack

Deep learning has emerged as a powerful tool for detecting DDoS attacks due to its ability to analyse large volumes of data and identify complex patterns.

Convolutional Neural Networks (CNNs)

CNNs are used to analyse traffic patterns and detect anomalies that indicate a DDoS attack. They can process large segment datasets and identify subtle variations in traffic that current methods might miss by Mittal (2023).

Recurrent Neural Networks (RNNs)

RNNs, particularly Long Short-Term Memory (LSTM) networks, are impactful in analysing sequential data, making them suitable for detecting patterns over time. The research from Alghazzawi et al., (2021) excel in identifying temporal patterns in network traffic, which is crucial for detecting slow and low-rate DDoS attacks.

Multi-Layer Perceptron (MLP)

MLPs are used to define traffic based on features such as packet size, Internet Protocol addresses, and port numbers. Studies have shown that MLPs can achieve high accuracy in detecting DDoS attacks, with some models reaching up to 98.99% detection accuracy by Ahmed et al., (2023).

Hybrid Models

Combining different deep learning models/frameworks can improve detection accuracy. For instance, a hybrid model might use CNNs for feature extraction and RNNs for sequence analysis. Hybrid models leverage the strengths of multiple approaches, providing robust detection capabilities by Alghazzawi et al., (2021).

Key Considerations

- Dataset Quality is the effectiveness of deep learning models/frameworks heavily depends on the quality and diversity of the training datasets. Commonly used datasets include CICDDoS2019 and CTU-13 provided by Ahmed et al., (2023).
- Feature Selection is important features such as packet size, flow duration, and protocol type is crucial for building model performance recommended by Mittal (2023).
- Real-time Detection for implementing selected models in real-time systems requires optimizing for speed and accuracy to minimize false positives and negatives by Ahmed et al., (2023).

Deep learning techniques provides promising impacts for detecting DDoS attacks by leveraging their ability to process and analyse large traffic segments and identify complex trends/patterns. By using models like CNNs, RNNs, and MLPs, and considering hybrid approaches, which can enhance the detection and mitigation of the negative customer effect, ensuring better network capacity availability and security.

The paper from Marleau, Rahman and Lung (2024) propose a DDoS detection and mitigation mechanism using BCP 38 and Software Defined Network. Experiments were conducted using Mininet, the Ryu SDN controller, and Scapy to create ICMP ping packets. The simulation topology included nine hosts (h one-h nine), divided into 3 subnets with different end addresses.

Initially, experiments were performed without IP spoofing prevention to observe traffic monitoring effects using ICMP ping tests. A mean reference point was established with normal traffic using an ICMP ping test without a DDoS attack.

To demonstrate the impact of a DDoS attack, a text command was added to the Mininet CLI to execute a python configuration on all hosts, specifying the destination IP and whether to spoof the source IP address. The results showed that over 66% of packets were dropped, and the delay increased to around 1270ms, significantly affecting h1's communication with the server.

When BCP 38 was implemented, the ICMP ping test at host h one showed 0% packet loss and a delay of 200ms, indicating successful mitigation of the DDoS attack.

The experiment from Aslam, Srivastava and Gore (2024) was conducted on a PARAM SHAVAK DLGPU system running 64-bit Ubuntu 18.04 LTS, equipped with 96 GB RAM and a 40-core 2 GHz Intel Skylake Processor. ONOS version 2.6.0 was used as the SDN controller, and Mininet was employed to create network topologies, providing a real-world network environment. The GEANT Zoo topology, consisting of 24 switches and 19 hosts, was in usage.

Timely detection of DDoS attacks is crucial, as early detection significantly enhances mitigation efforts. Validation is essential to verify the model's effectiveness. The efficacy of various ensemble Machine Learning algorithms (Random Forest, XGBoost, AdaBoost, Gradient Boosting, and Light_GBM) was evaluated using accuracy, F-score, recall, and precision metrics.

The analysis showed that XGBoost, AdaBoost, and Light_GBM achieved the highest accuracy on the self-generated dataset (99.9%, 99.7%, and 99.1%, respectively). This great performance is likely due to the SDN dataset's characteristics, which include flow-based features well-suited for these algorithms. Conversely, Random Forest and Gradient Boosting performed best on the CIC-DDoS 2019 dataset, with accuracies of 99.1% and 99.2%, respectively.

Effective DDoS mitigation relies on promptly identifying the attack source and implementing proactive steps. The approach integrates a mitigation module within the ONOS Flood Defender application. After identifying the attacker, the module dynamically enforces new flow rules on the attacker's connected switch, enabling it to discard or deny packets from the attacker aimed at the victim. Rapid detection and mitigation are crucial for the OFD application's effectiveness, as delays can result in the target being flooded with fake packets.

The solution employed Border Gateway Protocol route injection within AS1273 to advertise a more specific route to the host or network block under attack. The injection of a new route resulted in all traffic destined for the victim to be attracted to Cleaning Centre. Clean traffic is then forwarded to the host/IP block under attack via a private MPLS VRF in Figure 5.27. The separation of attack traffic from normal traffic is achieved by the application of DDoS-specific filters (countermeasures). When an attack occurs, the TMS platform attracts in all the traffic, removes the dirty traffic and passes only clean traffic. This solution removes attack traffic in the network before it hits the customer firewall, routers and servers. This provides uninterrupted service even when attacks happen. The solution proven with positive results and protected the network and its subscribers effectively.

6.5 Cross References

Fraud detection

(Xuan *et al.*, 2018); (Gori, Monfardini and Scarselli, 2005); (Park *et al.*, 2019); (Tarmazakov and Silnov, 2018); (Zhong *et al.*, 2019); (Niu *et al.*, 2016); (Peng and Lin, 2018); (Shearer 2000); (Bindu, Mishra, and Thilagam, 2018); (Molloy *et al.*, 2016); (Cao *et al.*, 2015); (Zhou *et al.*, 2023); (Smruthi 2019); (Rao, Gyani, and Narsimha 2018); (Sahoo *et al.*, 2022); (Gurajala *et al.*, 2016); (Y. S *et al.*, 2023); (Ochôa *et al.*, 2021); (Maciá-fernández 2008); (Tarmazakov and Silnov, 2018); (Bharne and Bhaladhare, 2022); (Kaubiyal, 2021); (De Jong, 2019); (Banerjee and Chua, 2023); (Kondamudi *et al*, 2023); (Suarez-Tangil *et al.*, 2019); (Ellaky, Benabbou, and Ouahabi, 2023); (Liao *et al.*, 2023); (Kahveci, (2019); (Meng *et al.*, 2021); (Zhu, Hsu and Zhou, 2023); (Al-Rousan *et al.*, 2020); (Kumar *et al.*, 2023); (Savchenko, Demochkin and Grechikhin, 2022); (Pastrana *et al.*, 2019); (Fuss, and Bőthe, 2022); (Ranney, 2021); (Anupriya *et al.*, 2022); (Acevedo *et al.*, 2023); (Teng *et al.*, 2023); (Zhao *et al.*, 2022); (Cao *et al.*, 2021); (Sun *et al.*, 2019); (Liu, Wu, and Zhou, 2008); (Singh, Alawami, and Kim, 2023); (Ahmed, Mahmood, and Islam., 2016); (Shaohui *et al.*, 2021); (Chang, Wang and Wang, 2022); (Carvalho *et al.*, 2015); (Georgakopoulos, Gallos and Plagianakos, 2020); (Ismail and Zeadally, 2021); (Susto *et al.*, 2017); (Ding and Ming, 2019); (Boutaher *et al.*, 2020); (Ahmed, Mahmood, and Hu., 2016); (Tekkali and Natarajan, 2023); (Zhao and Song, 2022); (Hilal *et al.*, 2022); (Hussein, 2022); (Kataria and Nafis, 2019); (Esmaeili *et al.*, 2023); (Mohamed, Makhlouf and Fakhfakh, 2018); (Zachos *et al.*, 2022); (Fontugne *et al.*, 2008); (Mitra and Rao, 2021); (Vats and Tadepalli, 2022); (Xu *et al.*, 2017); (Ghevariya *et al.*, 2021); (Reddy and Kumar, 2022); (Kouam, Viana, and Tchana, 2021); (Sheng and Yu, 2022); (Bao, 2020); (Yulita *et al.*, 2021); (Chang and Fan, 2019); (Vijay and Verma, 2023); (Hairani *et al.*, 2021); (Zulfikar, Gerhana and Rahmania, 2018); (Indumathi, Ramalakshmi, and Ajith, 2021); (Mishra, Mallick and Gadanayak, 2020); (Lu *et al*., 2019); (Lu, Tong, and Chen, 2015); (Altay, 2022); (Doss and Gunasekaran, 2023); (Bheemesh and Deepa, 2023); (Strelcenia and Prakoonwit 2023); Nguyen and Bein 2023); (Zeng, Kohno, and Roesner 2021); (Ishkov, Terekhov, and Myshenkov 2023); (Wang *et al*., 2023); Karthikeyan, Govindarajan, and Vijayakumar 2023); (Wahid *et al*., 2023).

Automation

(Arafat, Qusef, and Sammour. 2019); (Guo, Sui, and Shi. 2011); (Gandhar et al., 2024); (El, Tayebi and Sulimani 2024); (Sah and V. 2024); (Kilinc, H.H. 2022); (Dawood M et al., 2023); (Talluri, Zhang and Chen 2023); (Abitova *et al*., 2023); (Roy *et al*., 2024); (Alghazzawi et al., 2021); (Ahmed et al., 2023); (Alghazzawi et al., 2021); (Marleau, Rahman and Lung (2024).

Software-Defined Networking

(Yan *et al*., 2016); (Dinh & Park, 2021); (Nugroho, Dian, and Setyawan 2017); (Srikanth *et al*., 2018).

Hybrid cloud

(Xu, Su, and Zhang 2018); (Duplyakin, Haney, and Tufo 2015); (Xu, Su, and Zhang 2018); (Sitaram *et al*., 2018).

Smart Energy and Service Application

(Lv *et al*., 2019); (Araujo, Almeida, and Mello 2019); (Ganguly *et al*., 2022); (Korba and Karabadji 2019); (Jokar, Arianpoo, and Leung 2013); (Park *et al*., 2019); (Ruaro, Caimi & Moraes 2020); (Ruaro *et al*., 2018); Berestizshevsky *et al*., 2017); (Velloso *et al*., 2019); (Cong, Wen & Zhiying 2014); (Sandoval-Arechiga *et al*., 2015); (Scionti, Mazumdar & Portero 2018); (Ellinidou *et al*., 2019).

6.6 Future Recommendations

To enhance fraud detection and blocking in communication networks, the need to consider the following detailed future recommendations:

A system to perform advanced tariff linking on 6G intelligent network.
Due to competitive markets, the requirement is to correlate complex patterns for fraud prevention, the 6G network contains sophisticated components on the Billing platform. The ability of 6G network to link the contract or post-paid plans to the tariff charges, enabling the solution to fast track new changes.

Software intelligence networks to code policies, AI threat detection.
The solution is to detect on real-time monitoring and predictive assessments, by also increasing security of the platform. Using AI-powered to identify threat and cater on real-time, the solution should be allow polling on a 2-minute time scale for a well-updated system.

Software detector smart application intelligence, pattern detection on applications.
The use of detection on behavioural patterns are called Smart to preform sophistication techniques. It should assess the behaviour of each user and their habits in detecting anomalies. To provide an example, when a user has many failed access attempts.

Advanced Machine learning enforcement
The method to cater for authorized users to access sensitive data. To protect the user accounts by the use of Multi-Factor Authentication (MFA) providing a higher layer of security.

Rapid fraud detection in Automated Audits
By connecting Audit process to rapid detection, it has the potential to address vulnerabilities. To build a process of Security by design and conduct checks on security patches and vulnerable hardware.

Public Awareness and Education
By launching awareness campaigns to educate users about common fraud tactics and how to protect themselves. Provide resources and training to help users recognize and respond to potential threats.

By integrating these detailed strategies, mobile networks can significantly enhance their fraud detection and blocking capabilities, ensuring a more secure environment for subscribers. This comprehensive approach not only improves immediate security but also builds a resilient framework for future challenges.

CHAPTER SEVEN
Conclusion

The information presented proposes a new integrated platform that will serve the cloud and traditional networks to build a fraud detection and blocking solution. To provide a platform that improves the fraud detection scheme. Using a hybrid build of sites for connectivity that consist of virtualization for network flexibility. Lastly to map SDN technology to a re-design fraud detection framework.

The agreed approach looked at three architectures and performed a feasibility study of what's more suitable. The main focus was to reduce the rate of fraud in a network, the solution builds a new platform by using the existing systems. In this thesis, a comprehensive study was conducted to evaluate the novelty of SDN for fraud detection and blocking. The use case of Smart Energy abuse and Service application abuse were explored and to propose new mitigation scheme for the protection of the subscriber, the network and its resources.

7.1 Characteristics

Chapter 4 provided the characteristics and solution scoping for Fraud Detection and Blocking, covering the following characteristics:

The large internet pool of users, mobile networks consist of millions of users, with each region belonging to an internet pool with network nodes on the 4G/5G network. The network elements include:

- 4G Network elements: eNodeB, MME, SWG, PWG, HSS, PCRF.
- 5G Network elements: gNB, AMF, SMF, UPF, PCF.

Section 4.1 discussed mapping the IP address to each device and subscriber ID is crucial for fraud detection and Test case 1.

Correlation of the billing system, by linking the subscriber usage to the billing system is essential to prevent blacklisted users from accessing the network. The billing system, built on the Online Charging System (OCS) and Converged Charging System (CCS), plays a vital role in correlating subscriber accounts with network usage.

Characteristics of User/Traffic Profiles, by understanding user behaviour on data and voice networks helps in profiling and detecting anomalies. The characteristics include:

- Data network: Frequency of phone usage, session duration, subscribed applications, weekly usage patterns, IP address changes.
- Voice network: Frequency of phone usage, call frequency and duration, call patterns.

Different sectors like online banking, transportation, gaming, retail, food delivery, and online schooling show distinct usage patterns, which can be analysed for fraud detection in section

4.3. SIM registration and swap, The RICA Act in South Africa mandates SIM registration with identity documents and proof of residence. However, the process can be flawed with fake identities. Entry points for fraud include:

- 4G/5G contract SIMs
- 4G/5G prepaid SIMs
- Application bundles
- Cloud provider bundles

Aggregation/Interface points in the communication network, identifying aggregation points in the network is crucial for capturing user plane traffic. The core network consists of RAN, Transmission, and Core Data Network, with breakout points to the internet. The aggregation point, typically between spine and leaf switches identified in the research, is where user plane traffic is tapped for analysis.

SDN location and specification, The SDN controller is placed after the probing platform, enhancing its operational functionality. It manages traffic classification and fraud detection/blocking, integrating with Policy Charging elements (PCRF/PCF) covered in section 4.9.

Energy Fraud Identification, Energy fraud involves unauthorized consumption, meter tampering, and billing irregularities. The solution must detect and block such activities through automation, remote control, and detailed consumption analysis.

Packet De-encryption, Deep Packet Inspection (DPI) and de-encryption are necessary for analysing network packets and identifying fraudulent activities. This involves inspecting VLANs and using aggregation and SDN platforms.

The Blocking integration and rules, to block fraudulent activities involves integrating with PCRF/PCF and using protocols like RADIUS and Diameter. The solution must support dynamic rule creation and enforcement based on detected fraud.

Test case proposals, the test cases have been created to validate the solution's ability to detect and block fraud, ensuring compliance with billing systems and network protocols. This includes simulating various fraud scenarios and measuring the effectiveness of detection and blocking mechanisms.

The proposed solution for fraud detection and blocking leverages SDN technology, DPI, and machine learning to automate and enhance fraud prevention. By understanding the subscriber’s behaviour, integrating with billing systems, and identifying key aggregation points, the solution ensures efficient and scalable fraud management in communication networks.

7.2 SDN Controller Placement

The SDN controller is connected to the leaf switches, with logical connectivity changes to oversee three domains, the Metro network domain, the Core network domain and the Data centre domain. For each domain, a Virtual Private Network (VPN) and Segment Routing are created to ensure traffic is encrypted and routing changes are managed effectively covered in section 4.6.

The protocols for network establishment, the protocols used fall under service or transport protocols, ensuring traffic is readable and correctly formatted for classification discussed in with the following protocols and routing:

Service Protocols:

- Layer 2 VPN for Point-to-point connections for end users.
- Layer 3 VPN to configures Customer Edge (CE) and Provider Edge (PE) routers for traffic exchange.

Transport Protocols:

- Inter-Domain traffic protocols to manage data flow between different administrative domains (e.g., BGP).
- Intra-Domain traffic engineering to optimize traffic flow within a single domain (e.g., MPLS).
- IP routing that determines the path for IP packets (e.g., OSPF, BGP).

Segment routing:

- Based on source routing, using Segment IDs (SIDs) for efficient routing paths.

The SDN controller is built on a virtual infrastructure, providing service provisioning, automatic network deployment, and unified operations and maintenance (O&M). The research explored Energy Fraud Identification, Energy fraud involves unauthorized use or harm to the Smart Grid. The research focuses on the communication network aggregation points to sample traffic and identify fraud patterns on the following characteristics, exceeded utilization peaks, high utilization over extended periods, changes in data usage with hard stops and random peak usage policies discussed in section 4.7.

Packet De-encryption

Encryption is crucial for data integrity and privacy covered in section 4.8. Each mobile generation has introduced advanced encryption methods, and the research explores proposed decryption

Decryption Process for 4G LTE:

- Encryption Mechanism: Uses AES for both user plane and control plane encryption.
- Decryption Process: Involves extracting the key and initialization vector, applying AES decryption, and removing padding.

Decryption Process for 5G:

- Encryption Mechanism: Uses AES-256 with advanced key management.
- Decryption Process: Like 4G but with enhanced security features.

The SDN controller's strategic placement and layered architecture ensure efficient traffic management and fraud detection. By leveraging advanced encryption and decryption techniques, the solution enhances security and operational efficiency in modern communication networks.

7.3 Blocking Integration and Rules

The charging platform allows Service Providers to calculate usage based on tariff structures using online, offline, and converged charging. The architecture involves:

- Online Charging: Between PGW-C and OCS server over the Gy interface.
- Offline Charging: Between SGW-C, PGW-C, and CG over the Ga interface.
- Convergent Charging: Between SMF and CHF over the N40 interface for 5G networks.

The Charging process in section 4.9 of data collection, control traffic interaction, session initiation, quota management and statistics collection. Blocking rules are applied to various fraud types in different tariff plans gathered and provided on Table 4.2 is Fraud types in tariff plans.

Fraud Type	Data Bundle	Zero Rated	Reverse Billing	Uncategorized Traffic
Call Fraud	Degrade session quality	-	-	-
Digital Certificate Fraud	Block MSISDN for data	Block MSISDN for data	Block MSISDN for data	Log
Evasive Protocol Fraud	Block MSISDN for data	Block MSISDN for data	Block MSISDN for data	Log
Fair Usage Policy	Block MSISDN for data	Block MSISDN for data	Block MSISDN for data	Log

7.4 Test Case Proposals

To determine if the design meets the thesis specifications, various software tools were explored for traffic classification and detection. Tools like Zabbix, Icinga, Nagios, Cacti, and Checkmk were considered, but they had limitations in traffic intelligence. The Gigamon engine

was used to analyse the pcap files, identifying applications and total traffic. For subscriber intelligence, Splunk Enterprise was used to identify subscribers by IMEI number.
Chapter 5 explores and performs the testing.

Test Case 1: Fraud Detection with Unpaid Token
Uses traffic classification to identify fraud patterns in each sector and create a machine learning algorithm for detection. By integration of Mapping the subscriber platform to the billing platform, component analysis by identifying active components on the billing platform, validating that ensure correct billing by performing checks from the Convergent Billing Point.
Insights from Research

- Gandhar et al. (2024): Provided an overview of Machine Learning and deep learning techniques for fraud detection, highlighting strengths and weaknesses.
- El, Tayebi, and Sulimani (2024): Proposed intelligent system for detecting fraudulent transactions using deep learning architectures, achieving high accuracy with RNNs.
- Miao et al. (2019): Introduced the Siamese Prototypical Network (SPN) for few-shot traffic multi-classification and OOD detection, showing strong generalization capabilities.
- Sah and V. (2024): Compared Decision Trees and Random Forests for anomaly-based intrusion detection, with Random Forests showing higher efficiency in multiclass classification.
- Kilinc, H.H. (2022): provided a comprehensive call traffic analysis using K-means clustering and various supervised models, identifying suspicious call patterns and model performance.

The results mapped record of Subscriber_Rated_Billed and identified incorrectly billed subscribers due to platform abuse.

Example for Test Case 1:

- Subscriber ID: 1162565599
- Rated: 20.72355
- Billed: 1
- Fraud: 19.72355

Test Case 2: Fraud Detection Utilizing the Network
To evaluate the effects of fraud on traffic classification and build an algorithm to automate detection of inaccurate classification to address complexities from different devices and firmware versions, misclassification to ensure flexibility in classification rules, complex encryption handle encrypted traffic with advanced techniques, processing power to manage the processing requirements for decryption and traffic congestion to classify and record traffic from multiple applications. The framework covered in Figure 4.20 consisted of preprocessing,

evaluation, classification techniques, feature engineering, data collection, deployment, monitoring and feedback, compliance and privacy.

The proposed blocking integration and rules, along with the test case proposals, provide a comprehensive approach to fraud detection and blocking in communication networks. By leveraging advanced traffic classification techniques and ensuring compliance with privacy regulations, the solution aimed to enhance network security and operational efficiency.

To enable the Test Case the tools Used were APacket tool, Fabric Manager and using the formatted pcap file. The process allowed investigation of the pcap file to identify protocols, traffic patterns, IP connections, and possible abuse. Traffic classification to implement Policies, Conditions, and Actions to filter and mark fraud subscribers.

Policies, Conditions, and Actions:

- Policy: Guidelines to prevent, detect, and respond to fraudulent activities.
- Condition: Criteria to detect suspicious activity (e.g., unusual usage patterns, geographic irregularities).
- Action: Steps taken in response to detected fraud (e.g., suspension of service, account locking).

Example for Test Case 2:

- Policy: Fraud_Detection_High_Utilization
- Condition: Monitor traffic above mean threshold.
- Action: Drop specific traffic classified as fraudulent.

Results:

- Utilization of Traffic: Monitored over a 30-day period.
- Condition: Established to identify high utilization.
- Action: Successfully dropped fraudulent traffic.

Test Case 3: Fraudulent Access on Cloud Network

The importance of Hybrid Cloud provides the highest flexibility, deployment of any applications, control over security, legal aspects, and compliance. A private network within a public network, providing high security and control over network infrastructure, ideal for sensitive workloads and compliance with security regulations.

Insights from Research

- Dawood M et al. (2023): Discussed various security threats in cloud computing and the importance of understanding these threats for successful deployment.
- StrongDM (2024): provided an overview of unauthorized access, types, examples, and prevention methods.

- PYMNTS (2024): Reported on how fraudsters exploit cloud access points to orchestrate sophisticated attacks.
- Abitova et al. (2023): Highlighted the effectiveness of AI-powered fraud detection, achieving significant improvements in identified fraud cases and reducing manual review budgets.
- Roy et al. (2024): Performed an experiment with deep autoencoders for anomaly detection in bank transactions, achieving high recall but moderate precision.

Test Case 3a: Simulating fraudulent access attempts

The test case identifies fraudulent access attempts on a cloud network using detection mechanisms. The test data is used to valid user accounts of legitimate access to VMs, identify fraudulent user accounts of fraud legitimate access rights and VMs and resources of critical functions to test access controls.

The steps taken:

- Baseline access review to verify legitimate access and establish normal activity baseline.
- Simulate fraudulent access that attempts on access using fraudulent accounts (brute-force, exploiting vulnerabilities, stolen credentials).
- Monitor and log activity that ensure all access attempts are logged and monitored.
- Analyse detection mechanisms to review logs and verify detection accuracy.
- Response and mitigation to check response actions and ensure appropriate measures are taken.
- Post-test cleanup that removes test accounts and restore system state.
- Expected results for detection of fraudulent access, generation of alerts, and appropriate response actions.
- Post-test analysis to evaluate effectiveness and recommend improvements.
- Remarks that ensure compliance with data protection regulations and coordinate with relevant teams.

Test Case 3b: Unauthorized Access Simulation

The test case assesses effectiveness of controls in identifying and responding to unauthorized access. The Test data validates legitimate user accounts of various roles and permissions, malicious user accounts of fraud legitimate access rights and to protect sensitive resources of critical applications and data.

The Test plan involved the following:

- Initial Setup: Confirm legitimate roles and permissions, ensure monitoring tools are active.
- Privilege Escalation Attempt: Attempt escalation using legitimate accounts.
- Access Control Bypass: Attempt to bypass controls using various techniques.

- Suspicious Activity Simulation: Replicate unusual activities for the role.
- Monitor for Anomalies: Track activities and ensure alerts are generated.
- Response Action Evaluation: Simulate incident response protocols.
- Review Access Logs: Analyse logs for unauthorized access.
- System and Security Policy Review: Evaluate and suggest improvements.

The output of the analysis helps identify attack points and potential losses. The dashboard includes various fraud types:

- Immediate use cases: SIM Swap, Wangiri, PBX Hacking, APN abuse.
- Additional use cases: Roaming Fraud, Premium Rate Service Fraud, Abuse or Arbitrage Fraud, International Revenue Share Fraud, Call and SMS Spamming, Subscription Fraud, Dealer Fraud, Internal Fraud.

The detection process combines multiple data sources, applied Machine Learning for real-time anomaly detection and used Hadoop for real-time graph analysis and visualization. Test Case 3 demonstrates the use of Big Data and Machine Learning for real-time fraud detection in cloud networks. By leveraging Hadoop, distributed SQL query engines, and advanced ML algorithms, the solution provides robust and scalable fraud detection capabilities, ensuring enhanced security and operational efficiency.

Test Case 4: DDoS Attack

To identify and mitigate DDoS attacks using deep-learning techniques.

The attack methods identified are botnets for large-scale attacks, blackmail, state-sponsored, personal, or random attacks. The types of attack are bandwidth vs. packets per second, reflection/amplification attacks, and slow attacks. The detection and mitigation used Managed Objects (MOs) to define network objects to protect, set up alerts for different severity levels, and Implementation of mitigation actions.

Research Insights

- Marleau, Rahman, and Lung (2024): Proposed a DDoS detection and mitigation mechanism using BCP 38 and SDN. Experiments showed successful mitigation of DDoS attacks with 0% packet loss and reduced delay.
- Aslam, Srivastava, and Gore (2024): Evaluated ensemble ML algorithms for DDoS detection, with XGBoost, AdaBoost, and LightGBM achieving the highest accuracy on a self-generated dataset.

The results concluded:

- Detection of DDoS attacks.
- Generation of alerts and appropriate response actions.
- Effective mitigation to prevent service disruption.

Mitigation Techniques applied were blocking of source, header, and rate-based blocking, attack protection against various flood, fragmentation, and application layer attacks, and mitigation to blacklist/whitelist, geo-location blocking, packet filtering, and anti-spoofing measures.

The severity levels phased were:

- Severity 1: High threat level, major service impact.
- Severity 2: Medium/Low threat level, non-service impacting.
- Severity 3: Non-urgent, recorded anomalies.

This comprehensive approach ensures robust protection against DDoS attacks, maintaining network availability and performance. These test cases provide a comprehensive approach to identifying and mitigating fraudulent access and DDoS attacks in cloud networks, ensuring robust security and operational efficiency.

7.5 Summary discussion on Fraud Detection Algorithms

Fraud detection algorithms have evolved to address various types of fraud and the increasing sophistication of fraudulent activities. Here's a summary of the methods applied to the test cases:

Rule-Based Systems

These systems use predefined rules to identify fraudulent activities, such as flagging transactions over a certain amount or from unusual locations. They are straightforward but can be easily circumvented by sophisticated fraudsters. Demonstrated in Test Case 1: Fraud Detection with Unpaid Token on the Communication Network

- Uses Alteryx Designer software to process Call Detail Records (CDR).
- Compares subscriber 'Rated' and 'Billed' information to identify fraud.

Machine Learning Algorithms

Advanced methods like supervised learning (decision trees, support vector machines) and unsupervised learning (k-means clustering) identify complex patterns in large datasets. These algorithms learn from historical data and improve over time but require substantial data and computational power. Demonstrated in Test Case 2: Fraud Detection Utilizing the Network

- Applies an algorithm based on Policy, Condition, and Action.
- Identifies traffic patterns based on configured thresholds.

Anomaly Detection Algorithms

These algorithms identify outliers in data that deviate from established patterns, using methods like isolation forests or autoencoders. They are effective in identifying new or evolving fraud patterns but may produce false positives. Demonstrated in Test Case 3: Fraudulent Access on Cloud Network

- Uses a method involving Ingest, Detect, Investigate, Protect, and Impact to detect fraud on a cloud network.

Deep Learning

Neural networks, particularly those designed for sequential data (like LSTMs), can detect intricate patterns in transaction sequences. They are powerful but often seen as "black boxes," making interpretation and transparency a challenge. Demonstrated in Test Case 4: DDoS Attack

- Uses flow and deep packet inspection (DPI) technologies to provide macro-level and micro-level visibility, identifying threats and improving network performance.

Graph-Based Algorithms

These analyse relationships between entities (like transactions) to uncover fraud rings and collusion. They provide valuable insights but require a thorough understanding of graph theory. Applied to Test Case 2 and Test Case 4

7.6 Challenges and Considerations

Evolving Tactics, the fraudsters continuously adapt, making it crucial for algorithms to be flexible and updated regularly.

Data Privacy Concerns, the regulations like POPIA can limit access to necessary data, affecting algorithm performance.

False Positives that balancing sensitivity and specificity are challenging, leading to high rates of false positives.

Integration Challenges to implementing advanced algorithms requires significant investment in technology and training.

Deep learning techniques offer promising solutions for detecting and mitigating DDoS attacks by leveraging their ability to process and analyse large datasets and identify complex patterns. By using models like CNNs, RNNs, and MLPs, and considering hybrid approaches, networks can enhance their detection and mitigation capabilities, ensuring better availability and security.

While fraud detection algorithms remain crucial, their effectiveness depends on continuous adaptation, integration with real-time data, and consideration of evolving regulatory and subscriber bases. Implementing effective traffic classification and detection systems is essential for maintaining secure, efficient, and user-friendly communication networks.

7.7 Future considerations for researchers

By implementing these advanced techniques, communication networks can significantly enhance their fraud detection and prevention capabilities, ensuring robust security and operational efficiency for the future researcher.

Advanced Machine Learning Models

- The ability to deploy sophisticated algorithms: Utilizing advanced machine learning algorithms such as XGBoost and LightGBM. These ensemble methods combine multiple models to improve predictive performance and can handle large, complex datasets effectively.
- Data Integration: Integrate diverse data sources, including call detail records, user behavior logs, and network traffic data, to provide a comprehensive dataset for training the models.
- Continuous Learning: To implement continuous learning mechanisms to update the models with new data, ensuring they adapt to evolving fraud patterns.

Benefits:

- It will enhanced detection accuracy: These advanced models can identify complex and subtle patterns in the data that simpler models might miss, leading to more accurate fraud detection.
- Scalability: the capacity of handling large volumes of data, these models can scale with the network's growth, maintaining high performance.
- To reduced false positives: By accurately distinguishing between legitimate and fraudulent activities, these models reduce the number of false positives, minimizing unnecessary interventions.

Real-Time Monitoring and Analytics

- AI-Powered Systems: to deploy AI-powered threat detection systems that offer real-time monitoring and predictive analytics. These systems use machine learning algorithms to continuously analyze network traffic and user behavior.

- Automated Alerts: The power of real time will set up automated alert systems that notify security teams of suspicious activities in real-time, enabling immediate investigation and response.
- Predictive Analytics: It will use predictive analytics to forecast potential fraud incidents based on historical data and current trends, allowing for proactive measures.

Benefits:

- Immediate Threat Response: Real-time monitoring allows for the immediate detection and response to potential threats, preventing fraudulent activities before they cause significant damage.
- Proactive Security Measures: Predictive analytics enable the anticipation of fraud attempts, allowing for the implementation of preventive measures.
- Comprehensive Visibility: Continuous monitoring provides a comprehensive view of network activities, helping to identify and address vulnerabilities promptly.

Behavioural Biometrics

- User Behavior Analysis: It allows incorporation of behavioral biometrics into security protocols by analyzing user behavior patterns such as typing speed, mouse movements, and navigation habits.
- Anomaly Detection: The ability to develop algorithms to detect deviations from normal user behavior, flagging potential fraudulent activities.
- Multi-Factor Authentication: Can be combined with behavioral biometrics with other authentication methods, such as passwords and tokens, to enhance security.

Benefits:

- Additional Security Layer: Behavioral biometrics add an extra layer of security by identifying fraudulent activities based on behavioral anomalies, which are difficult for fraudsters to replicate.
- Sophisticated Fraud Detection: This method is particularly effective in detecting sophisticated fraud attempts that may bypass traditional security measures.

8. REFERENCES

3GPP, 2025. Architecture enhancements for non-3GPP accesses. 3GPP TS 23.402. Available at: https://www.3gpp.org/ftp/Specs/archive/23_series/23.402

3GPP, 2025. Charging management and Diameter charging applications. 3GPP TS 32.299. Available at: https://www.3gpp.org/ftp/Specs/archive/32_series/32.299

3GPP, 2025. Charging rule provisioning over the Gx interface. 3GPP TS 29.210. Available at: https://www.3gpp.org/ftp/Specs/archive/29_series/29.210

3GPP, 2025. GPRS enhancements for E-UTRAN access. 3GPP TS 23.401. Available at: https://www.3gpp.org/ftp/Specs/archive/23_series/23.401

3GPP, 2025. GPRS Tunneling Protocol (GTP) across Gn and Gp interfaces. 3GPP TS 29.060. Available at: https://www.3gpp.org/ftp/Specs/archive/29_series/29.060

3GPP, 2025. Interworking between PLMN and PDN. 3GPP TS 29.061. Available at: https://www.3gpp.org/ftp/Specs/archive/29_series/29.061

3GPP, 2025. Network architecture. 3GPP TS 23.002. Available at: https://www.3gpp.org/ftp/Specs/archive/23_series/23.002

3GPP, 2025. PCC signaling flows and QoS parameter mapping. 3GPP TS 29.213. Available at: https://www.3gpp.org/ftp/Specs/archive/29_series/29.213

3GPP, 2025. Policy and Charging Control (PCC) reference points. 3GPP TS 29.212. Available at: https://www.3gpp.org/ftp/Specs/archive/29_series/29.212

3GPP, 2025. Policy and charging control architecture. 3GPP TS 23.203. Available at: https://www.3gpp.org/ftp/Specs/archive/23_series/23.203

Abitova, G.A., Abalkanov, M., Abitova, G.A Shuteyeva, G., Aitmukhanbetova, E and Kulniyazova, K. (2024). Review of Cloud AI for Real-Time Fraud Detection," *2024 10th International Conference on Automation, Robotics and Applications (ICARA)*, Athens, Greece, pp. 454-460, doi: 10.1109/ICARA60736.2024.10553149.

Aboba, B., Zorn, G. and Mitton, D., 2001. RADIUS and IPv6. RFC 3162. Available at: https://www.rfc-editor.org/rfc/rfc3162.html

Acevedo, E., Massaferro, P., Fernández, A., Martins, A and Caudullo, G. (2023). Fraud Detection Using Event Logs with LSTM and Gradient Boosting. *2023 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT),* Washington, DC, USA, pp. 1-5.

Afrin, S., Roksana, S and Akram, R. (2025). AI-Enhanced Robotic Process Automation: A Review of Intelligent Automation Innovations. in *IEEE Access*, vol. 13, pp. 173-197, doi: 10.1109/ACCESS.2024.3513279.

Agomuo, O.C., Uzoma, A.K., Khan, K., Otuomasirichi, A.I and Muzamal, J.M. (2025). Transparent AI for Adaptive Fraud Detection. *2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, Bangkok, Thailand, pp. 1-6, doi: 10.1109/IMCOM64595.2025.10857433.

Ahmed, M., Mahmood, A. N., & Islam, M. R. (2016). A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems*, 55, 278-288.

Ahmed, S.; Khan, Z.A.; Mohsin, S.M.; Latif, S.; Aslam, S.; Mujlid, H.; Adil, M.; Najam, Z. (2023). Effective and Efficient DDoS Attack Detection Using Deep Learning Algorithm, Multi-Layer Perceptron. *Future Internet*, *15*, 76. https://doi.org/10.3390/fi15020076

Alghazzawi, D.; Bamasag, O.; Ullah, H.; Asghar, M.Z. (2021). Efficient Detection of DDoS Attacks Using a Hybrid Deep Learning Model with Improved Feature Selection. *Appl. Sci.11*, 11634. https://doi.org/10.3390/app112411634

Al-Rousan, S., Abuhussein, A., Alsubaei, F., Kahveci, O., Farra, H and Shiva, S. (2020). Social-Guard: Detecting Scammers in Online Dating. *IEEE Int. Conf. Electro Inf. Technol.*, vol. 2020-July, no. August, pp. 416–422.

Altay, O. (2022). Performance of different KNN models in prediction english language readability. *2022 2nd International Conference on Computing and Machine Intelligence (ICMI)*, Istanbul, Turkey, pp. 1-5.

Andrews, M., 1998. Negative Caching of DNS Queries (DNS NCACHE). RFC 2308. Available at: https://www.rfc-editor.org/rfc/rfc2308.html

Anupriya, E., Kumaresan, N., Suresh, S., Dhanasekaran, S., Ramprathap, K, and Chinnasamy, P. (2022). Fraud Account Detection on Social Network Using Machine Learning Techniques. *2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC),* Bhubaneswar, India, pp. 1-4.

Arafat, M., Qusef, A and Sammour, G. (2019). Detection of Wangiri Telecommunication Fraud Using Ensemble Learning. *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, Amman, Jordan, pp. 330-335, doi: 10.1109/JEEIT.2019.8717528.

Araujo, de, B. Almeida, de, H. and Mello, de, F. (2019). Computational Intelligence Methods Applied to the Fraud Detection of Electric Energy Consumers. *in IEEE Latin America Transactions*, vol. 17, no. 01, pp. 71-77.

Aslam, N., Srivastava, S and Gore, M.M. (2024). Evaluating DDoS Detection and Mitigation in SDN at Various Attack Rates," *2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT)*, Delhi, India, pp. 569-574, doi: 10.1109/IC2SDT62152.2024.10696232.

Banerjee, S and Chua, Y.K. (2023). Understanding online fake review production strategies. *Journal of Business Research*, Volume 156.

Bao, J. (2020). Multi-features Based Arrhythmia Diagnosis Algorithm Using Xgboost. *2020 International Conference on Computing and Data Science (CDS)*, Stanford, CA, USA, pp. 454-457.

Belshe, M., Peon, R. and Thomson, M., 2015. Hypertext Transfer Protocol Version 2 (HTTP/2). RFC 7540. Available at: https://www.rfc-editor.org/rfc/rfc7540.html

Berners-Lee, T., Fielding, R. and Frystyk, H., 1996. Hypertext Transfer Protocol -- HTTP/1.0. RFC 1945. Available at: https://www.rfc-editor.org/rfc/rfc1945.html

Bertz, L., Dolson, D. and Lifshitz, Y., 2019. Diameter Credit-Control Application. RFC 8506. Available at: https://www.rfc-editor.org/rfc/rfc8506.html

Bharne, S and Bhaladhare, P. (2022). Investigating Online Dating Fraud: An Extensive Review and Analysis. *2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC)*, Hyderabad, India, pp. 141-147.

Bheemesh. R.K, and Deepa. N. (2023). Accurate SMS Spam Detection Using Support Vector Machine in Comparison with Logistic Regression. *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, Chennai, India, pp. 1-5.

Bindu, P., Mishra, R., and Thilagam, P.S. (2018). Discovering spammer communities in twitter. *Journal of Intelligent Information Systems*, vol. 51, no. 3, pp. 503–527.

Blockchain and Smart Contracts. Int. *J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 3, p. 544.
Boutaher, N., Elomri, A., Abghour, N., Moussaid, K, and Rida, M. (2020). A Review of Credit Card Fraud Detection Using Machine Learning Techniques. *2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech)*, 2020, pp. 1-5.

Burge, P, and Shawe-Taylor, J. (1997). Detecting cellular fraud using adaptive prototypes, in: *Proceedings of the AAAI-97 Workshop on AI Approaches to Fraud Detection and Risk Management*, pp. 9–13.

Cao, N., Shi, C., Lin, S., Lu, J., Lin, Y.,-R., and Lin C,-Y. (2015). Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 280–289.

Cao, R., Liu, G., Xie, Y, and Jiang, C. (2021). Two-level attention model of representation learning for fraud detection. *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 6, pp. 1291–1301.

Carvalho, L. F., Carlos Teixeira, Ester C. Dias, Wagner Meira, and Osvaldo Carvalho. (2015). A simple and effective method for anomaly detection in healthcare. In *Proceedings of the SIAM International Conference on Data Mining Workshop*, vol. 2015, pp. 16-24.

Chang, K.-C, and Fan, C.-P. (2019). Cost-Efficient Adaboost-based Face Detection with FPGA Hardware Accelerator. *2019 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*, Yilan, Taiwan, pp. 1-2.

Chang, S., Wang, C, and Wang, C. (2022). Automated Feature Engineering for Fraud Prediction in Online Credit Loan Services. *2022 13th Asian Control Conference (ASCC)*, Jeju, Korea, Republic of, pp. 738-743.

Chiba, M., Dommety, G., Eklund, M., Mitton, D. and Aboba, B., 2008. Dynamic Authorization Extensions to RADIUS. RFC 5176. Available at: https://www.rfc-editor.org/rfc/rfc5176.html

Dawood, M.; Tu, S.; Xiao, C.; Alasmary, H.; Waqas, M.; Rehman, S.U. (2023). *Cyberattacks and Security of Cloud Computing: A Complete Guideline*. Symmetry, 15, 1981. https://doi.org/10.3390/sym15111981

DDoS Flood Detection and Mitigation using SDN and Network Ingress Filtering - an Experiment Report," *2024 IEEE 4th International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB)*, Taipei, Taiwan, pp. 67-72, doi: 10.1109/ICEIB61477.2024.10602663.

De Jong, K. (2019). Detecting the online romance scam: Recognising images used in fraudulent dating profiles.

Dierks, T. and Allen, C., 1999. The TLS Protocol Version 1.0. RFC 2246. Available at: https://www.rfc-editor.org/rfc/rfc2246.html

Dierks, T. and Rescorla, E., 2006. The Transport Layer Security (TLS) Protocol Version 1.1. RFC 4346. Available at: https://www.rfc-editor.org/rfc/rfc4346.html

Dierks, T. and Rescorla, E., 2008. The Transport Layer Security (TLS) Protocol Version 1.2. RFC 5246. Available at: https://www.rfc-editor.org/rfc/rfc5246.html

Ding, Z and Ming, M. (2019). Accelerometer-Based Mobile Device Identification System for the Realistic Environment," in *IEEE Access*, vol. 7, pp. 131435-131447.

Dinh, P.T. & Park, M. (2021). BDF-SDN: A Big Data Framework for DDoS Attack Detection in Large-Scale SDN-Based Cloud. *2021 IEEE Conference on Dependable and Secure Computing (DSC)*, pp. 1-8.

Doss, L.M, and Gunasekaran, M. (2023). Evasion and Poison attacks on Logistic Regression-based Machine Learning Classification Model. *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, Chennai, India, pp. 1-8.

El Kafhali, S.; Tayebi, M.; Sulimani, H. (2024). *An Optimized Deep Learning Approach for Detecting Fraudulent Transactions*. *Information* , *15*, 227. (https://doi.org/10.3390/info15040227)

Ellaky, Z., Benabbou, F, and Ouahabi, S. (2023). Systematic Literature Review of Social Media Bots Detection Systems*, Journal of King Saud University - Computer and Information Sciences*, Volume 35, Issue 5.

Esmaeili, M, *et al*. (2023). Generative Adversarial Networks for Anomaly Detection in Biomedical Imaging: A Study on Seven Medical Image Datasets," in *IEEE Access*, vol. 11, pp. 17906-17921.

Fajardo, V., Arkko, J., Loughney, J. and Zorn, G., 2012. Diameter Base Protocol. RFC 6733. Available at: https://www.rfc-editor.org/rfc/rfc6733.html

Fawcett, T., and Provost, F. (1997). Adaptive fraud detection, *Data mining and knowledge discovery* 1 (3), 291–316.

Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P. and Berners-Lee, T., 1999. Hypertext Transfer Protocol -- HTTP/1.1. RFC 2616. Available at: https://www.rfc-editor.org/rfc/rfc2616.html

Fontugne, Romain, Toshio Hirotsu, and Kensuke Fukuda. (2008). An image processing approach to traffic anomaly detection. In *Proceedings of the 4th Asian Conference on Internet Engineering*, pp. 17-26.

Fuss, J, and Bőthe, B. (2022). 15 - Cybersex (including sex robots), In Global Mental Health in Practice, *Mental Health in a Digital World*, Academic Press.

Gandhar, A., Gupta, K., Pandey, A.K. *et al.* (2024). *Fraud Detection Using Machine Learning and Deep Learning*. *SN COMPUT. SCI.* **5**, 453 (https://doi.org/10.1007/s42979-024-02772-x)

Ganguly, P. Dutta, S. Nasipuri, M and Tewari, S. (2022). Modeling Fraud in Residential Power Usage," *2022 IEEE 10th International Conference on Smart Energy Grid Engineering (SEGE)*, Oshawa, ON, Canada, 2022, pp. 125-130.

Georgakopoulos, S.V., Gallos, P, and Plagianakos, V.P. (2020). Using Big Data Analytics to Detect Fraud in Healthcare Provision. *2020 IEEE 5th Middle East and Africa Conference on Biomedical Engineering (MECBME)*, Amman, Jordan, pp. 1-3.

Ghevariya, R., Desai, R., Bohara, M.H, and Garg, D. (2021). Credit Card Fraud Detection Using Local Outlier Factor & Isolation Forest Algorithms: A Complete Analysis. *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, pp. 1679-1685.

Ghosh, S., Bilgaiyan, S., Gourisaria, M.K, and Sharma, A. (2023). Comparative Analysis of Applications of Machine Learning in Credit Card Fraud Detection. *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, pp. 1-7.

Gonzalez, P.A. (2005). Fraud detection method for mobile telecommunication networks, EP.
Gori, M., G. Monfardini, and F. Scarselli. (2005). A new model for learning in graph domains. in Proceedings. *2005 IEEE International Joint Conference on Neural Networks*. IEEE.

Guo, D., Sui, A.F, and Shi, L. (2011). Billing attack detection and prevention in mobile communication network. *2011 IEEE 13th International Conference on Communication Technology*, Jinan, China, pp. 687-691, doi: 10.1109/ICCT.2011.6157964.

Gurajala, S., White, J., Hudson, H, and Matthews, J. (2016). Profile characteristics of fake Twitter accounts. *Big Data Society*, pp. 1–13.

Hairani, H., Anggrawan, A., Wathan, A.I., Latif, K.A., Marzuki, K, and Zulfikri, M. (2021). The Abstract of Thesis Classifier by Using Naive Bayes Method. *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM)*, Pekan, Malaysia, pp. 312-315.

Hilal, Waleed, S. Andrew Gadsden, and John Yawney. (2022). *Financial Fraud:: A Review of Anomaly Detection Techniques and Recent Advances.*

Hoffman, P., Sullivan, A. and Fujiwara, K., 2019. DNS Terminology. RFC 8499. Available at: https://www.rfc-editor.org/rfc/rfc8499.html

Hussein, O. (2022). A Proposed Anti-Fraud Authentication Approach for Mobile Banking Apps," *2022 4th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, Giza, Egypt, pp. 56-61.

Indarjit, E., Balyan, V. and Adonis, M. (2025d). Establishing software-defined network for fraud detection on energy fraud and traffic classification test case(s*). International Journal on Smart Sensing and Intelligent Systems*, Sciendo, Vol. 18 (Issue 1). https://doi.org/10.2478/ijssis-2025-0043

Indarjit, E., Baylan, V. and Adonis, M. (2025b). 'Impact of fraud on revenue loss, detection features, and energy fraud identification', *Applications of Artificial Intelligence in 5G and Internet of Things* [Preprint]. doi:9781003532521.

Indarjit, E., Balyan, V., and Adonis, M. (2025c). 'Fraud detection and blocking solution within the communication network in reference to test and an analysis'

Indarjit, E., Balyan, V., and Adonis, M. (2025a). 'Review of the latest techniques that address network fraud detection and proposed architectures', *Applications of Artificial Intelligence in 5G and Internet of Things*, Taylor & Francis, 1, doi: 9781003532521.

Indarjit, E. (2022). Software Defined Networking based on centralized control applied to Smart Grid Applications. Cape Peninsula University of Technology. Dataset. https://doi.org/10.25381/cput.21602703.v1

Indarjit, E., Adonis, M., and Brandt, A. (2022). Software Defined Networking based on centralized control applied to Smart Grid Applications. Master's thesis. Cape Peninsula University of Technology.

Indumathi, N., Ramalakshmi, R, and Ajith, V. (2021). Analysis of risk factors in the Firework Industries: Using Decision Tree Classifier. *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, pp. 811-814.

Ishkov, D.O., Terekhov, V.I., and Myshenkov, K.S. (2023). Energy Theft Detection in Smart Grids via Explainable Attention Maps. *2023 5th International Youth Conference on Radio*

Electronics, Electrical and Power Engineering (REEPE), Moscow, Russian Federation, pp. 1-6.

Ismail, L and Zeadally, S. (2021). Healthcare Insurance Frauds: Taxonomy and Blockchain-Based Detection Framework (Block-HI). in *IT Professional*, vol. 23, no. 4, pp. 36-43.

Jain, A.K., Sahoo, S.R., and Kaubiyal, J. (2021). Online social networks security and privacy: comprehensive review and analysis. *Complex Intell. Syst.*, vol. 7, no. 5, pp. 2157–2177.

Jokar, P. Arianpoo, N and Leung, V.C. (2013). Intrusion detection in advanced metering infrastructure based on consumption pattern. *in IEEE International Conference on Communications (ICC)*, IEEE, pp. 4472–4476.

Kahveci, O. (2019). The Repository at St.Cloud State An Approach For Detecting Online Dating Scams.

Karthikeyan, T., Govindarajan, M, and Vijayakumar, V. (2023). An effective fraud detection using competitive swarm optimization based deep neural network, Measurement: Sensors.

Kataria, S. and Nafis, M.T. (2019). Internet Banking Fraud Detection Using Deep Learning Based on Decision Tree and Multilayer Perceptron. *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, pp. 1298-1302.

Kilinc, H.H. (2022). Anomaly Pattern Analysis Based on Machine Learning on Real Telecommunication Data," *2022 7th International Conference on Computer Science and Engineering (UBMK)*, Diyarbakir, Turkey, pp. 43-48, doi: 10.1109/UBMK55850.2022.9919564.

Kondamudi, M.R., Sahoo, S.R., Chouhan, L., and Yadav, N. (2023). A comprehensive survey of fake news in social networks: Attributes, features, and detection approaches. *Journal of King Saud University - Computer and Information Sciences*,Volume 35, Issue 6.

Korba, Amara, A. And Karabadji, El, N. (2019). Smart Grid Energy Fraud Detection Using SVM. *2019 International Conference on Networking and Advanced Systems (ICNAS)*, Annaba, Algeria, pp. 1-6.

Kouam, A.J., Viana, A.C., and Tchana, A. (2021). SIMBox Bypass Frauds in Cellular Networks: Strategies, Evolution, Detection, and Future Directions," *in IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2295-2323, Fourthquarter.

Kumar, C.R., Saranya, N., Harshini, P., Gilchrist, D., Rahman, M. (2023). Face recognition using CNN and siamese network, *Measurement: Sensors*, Volume 27.

Li, M. Sun, M. Liu, Q. and Zhang, Y. (2021). Fraud Detection Based on Graph Neural Networks with Self-attention. *2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT),* Shanghai, China. pp. 349-353.

Liao, C., Wang, J., Shan, B., Shang, J., Dong, T, and He, Y. (2023). *Near real-time detection and forecasting of within-field phenology of winter wheat and corn using Sentinel-2 time-series data*, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 196.

Liu, Y. Hu, S and Ho, -Y, T. (2014). Vulnerability assessment and defense technology for smart home cybersecurity considering pricing cyberattacks," in IEEE/ACM International Conference on Computer-Aided Design (ICCAD), IEEE, 2014, pp. 183–190.

Lu, S., Tong, W, and Chen, Z. (2015). Implementation of the KNN algorithm based on Hadoop. *2015 International Conference on Smart and Sustainable City and Big Data (ICSSC)*, Shanghai, pp. 123-126.

Lu, T., Huang, H., Zhao, W, and Zhang, J. (2019). The Metering Automation System based Intrusion Detection Using Random Forest Classifier with SMOTE+ENN. *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, Dalian, China, pp. 370-374.

Lv, L. Cheng, J. Peng, N. Fan, M. Zhao, D and Zhang, J. (2019). Auto-encoder based Graph Convolutional Networks for Online Financial Anti-fraud. *2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, Shenzhen, China, pp. 1-6.

Maciá-fernández, G. (2008). Roaming fraud: assault and defense strategies," vol. 241000, pp. 1–8.

Mark Renier, L.M. and Bailon, M. (2019). International Roaming Services Optimization Using Private

Meng, L., Duan, S., Zhao, Y., Lü, K, and Chen, S. (2021). The impact of online celebrity in livestreaming E-commerce on purchase intention from the perspective of emotional contagion, *Journal of Retailing and Consumer Services*, Volume 63.

Miao, G., Wu, g., Zhang, Z., Tong, Y, and Lu, B. (2023). SPN: A Method of Few-Shot Traffic Classification With Out-of-Distribution Detection Based on Siamese Prototypical Network. in *IEEE Access*, vol. 11, pp. 114403-114414, doi: 10.1109/ACCESS.2023.3325065.

Microsoft (2022). Available at: https://www.microsoft.com/en-us/security/blog/2022/11/16/token-tactics-how-to-prevent-detect-and-respond-to-cloud-token-theft/ (Accessed: 09 November 2024).

Mishra, S., Mallick, R.K, and Gadanayak, D.A. (2020). Islanding Detection of Microgrid using EMD and Random Forest Classifier. *2020 International Conference on Computational Intelligence for Smart Power System and Sustainable Energy (CISPSSE)*, Keonjhar, India, pp. 1-5.

Mitra, S and Rao, K. (2021). Experiments on Fraud Detection use case with QML and TDA Mapper. *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, Broomfield, CO, USA, pp. 471-472.

Mittal, M., Kumar, K. & Behal, S. (2023). *Deep learning approaches for detecting DDoS attacks: a systematic review*. *Soft Comput* **27**, 13039–13075. https://doi.org/10.1007/s00500-021-06608-1

Mockapetris, P., 1987. Domain names - concepts and facilities. RFC 1034. Available at: https://www.rfc-editor.org/rfc/rfc1034.html

Mockapetris, P., 1987. Domain names - implementation and specification. RFC 1035. Available at: https://www.rfc-editor.org/rfc/rfc1035.html

Mohamed, M.B., Makhlouf, A.M, and Fakhfakh, A. (2018). Correlation for efficient anomaly detection in medical environment. *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Limassol, Cyprus, pp. 548-553.

Molloy, I., Chari, S., Finkler, U., Wiggerman, M., Jonker, C., Habeck, T., Park, Y., Jordens, F and Schaik, R. (2016). Graph analytics for realtime scoring of cross-channel transactional fraud," in *Proceedings of International Conference on Financial Cryptography and Data Security*, pp. 22–40.

New FTC Data Show Consumers Reported Losing More Than $200 Million to Romance Scams in 2019 | Federal Trade Commission. (2019). https://www.ftc.gov/newsevents/news/pressreleases/2020/02/new-ftc-data-show-consumers-reported-losingmore-200-million-romance-scams-2019

Nguyen, S and Bein, D. (2023). Data Science Analysis of Malicious Advertisements and Threat Detection Automation for Cybersecurity Progress," *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA, pp. 0695-0704.

Niu, K., Jiao, H., Deng, N., and Gao, Z. (2016). A Real-Time Fraud Detection Algorithm Based on Intelligent Scoring for the Telecom Industry. *2016 International Conference on Networking and Network Application(NaNA),* Hakodate, pp. 303-306.

Nugroho, A.S., Dian Safitri, Y. & Setyawan, T.A. (2017). Comparison analysis of *Software-Defined Network* and OSPF protocol using virtual media. *2017 IEEE International Conference on Communication, Networks, and Satellite (Comnetsat)*, pp. 106-111.

Ochôa, I.S *et al*., (2021). Performance and security evaluation on a blockchain architecture for license plate recognition systems," *Appl. Sci.,* vol. 11, no. 3, pp. 1– 21.

Okumbor, F., Anthony, N., and Olokunde, A.A.J. (2019). Grappling with the challenges of interconnect bypass fraud. *IOSR J. Mobile Comput. Appl.*, vol. 6, pp. 35–41.

Park, N., *et al*. (2019). Estimating node importance in knowledge graphs using graph neural networks." in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Pastrana, S., Thomas, D., Hutchings, A., and Tapiador, J. (2019). Measuring ewhoring', *Proc. ACM SIGCOMM Internet Meas.Conf. IMC*, pp. 463–477.

Peng, L, and Lin, R. (2018). Fraud Phone Calls Analysis Based on Label Propagation Community Detection Algorithm. *2018 IEEE World Congress on Services, San Francisco*, CA, 2018, pp. 23-24.

Postel, J., 1981. Transmission Control Protocol. RFC 793. Available at: https://www.rfc-editor.org/rfc/rfc793.html

PYMNTS (2024). Available at: https://www.pymnts.com/cybersecurity/2024/fraudsters-exploit-cloud-access-points-to-orchestrate-sophisticated-attacks/ (Accessed: 09 November 2024).

Ranney, J.D. (2021). Chapter 3 - The process of exploitation and victimization of adolescents in digital environments: the contribution of authenticity and self-exploration, *Child and Adolescent Online Risk Exposure*, Academic Press.

Raviprakash, R. (2022) *AI-based systems can track fraudsters faster and reduce revenue loss by differentiating between genuine customers and fraudsters*. Available at: https://www.subex.com/article/how-ai-can-help-in-detecting-predicting-preventing-telecom-fraud/ (Accessed: 02 November 2024).

Reddy, P.R, and Kumar, A.S. (2022). Credit Card Fraudulent Transactions Prediction Using Novel Sequential Transactions by Comparing Light Gradient Booster Algorithm Over Isolation Forest Algorithm. *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, Gautam Buddha Nagar, India, pp. 563-567.
Rescorla, E., 2018. The Transport Layer Security (TLS) Protocol Version 1.3. RFC 8446. Available at: https://www.rfc-editor.org/rfc/rfc8446.html

Rigney, C., 2000. RADIUS Accounting. RFC 2866. Available at: https://www.rfc-editor.org/rfc/rfc2866.html

Rigney, C., Willens, S., Rubens, A. and Simpson, W., 2000. Remote Authentication Dial-In User Service (RADIUS). RFC 2865. Available at: https://www.rfc-editor.org/rfc/rfc2865.html

Roy, S., G S, N., Shankar, S.S., Adnan, M.M and Umaeswari, P. (2024). Leveraging Deep Autoencoders for Security in Big Data Framework: An Unsupervised Cloud Computing Approach," *2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT)*, Bengaluru, India, pp. 1-5, doi: 10.1109/ICDCOT61034.2024.10515899.

Ruskin, A. (2021) *Machine Learning algorithms analyze large datasets from smart meters to detect unusual patterns that suggest fraud.* Available at:

https://www.tcs.com/content/dam/global-tcs/en/pdfs/insights/whitepapers/digital-twin-approach-mitigate-telecom-fraud-risks.pdf (Accessed: 02 November 2024).

Sah, A.K and V. K. (2024). Anomaly-Based Intrusion Detection in Network Traffic using Machine Learning: A Comparative Study of Decision Trees and Random Forests. *2024 2nd International Conference on Networking and Communications (ICNWC)*, Chennai, India, pp. 1-7, doi: 10.1109/ICNWC60771.2024.10537451.

Sahoo, P. K., Mishra, S., Panigrahi, R., Bhoi, A. K., & Barsocchi, P. (2022). *An Improvised Deep-Learning-Based Mask R-CNN Model for Laryngeal Cancer Detection Using CT Images. Sensors*, 22(22), 8834.

Savchenko, A.V., Demochkin, K.V, and Grechikhin, I.S. (2022). Preference prediction based on a photo gallery analysis with scene recognition and object detection, *Pattern Recognition*, Volume 121.

SEON. (2024) *Machine Learning models can classify traffic and detect deviations from normal usage patterns, helping to prevent fraud in real-time*. Available at: https://seon.io/resources/telecommunications-fraud-detection-and-prevention/ (Accessed: 02 November 2024).

Shaohui, D., Qiu, G., Mai, H, and Yu, H. (2021). Customer Transaction Fraud Detection Using Random Forest," *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, Guangzhou, China, 2021, pp. 144-147.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining,"*J. Data Warehousing*, vol. 5, no. 4, pp. 13–22.

Sheng, C and Yu, H. (2022). An optimized prediction algorithm based on XGBoost. *2022 International Conference on Networking and Network Applications (NaNA)*, Urumqi, China, pp. 1-6.

Singh, A., Ranjan, R.K., and Tiwari, A. (2021). Credit card fraud detection under extreme imbalanced data: a comparative study of data-level algorithms. *J. Exp Theor. Artif. Intell*., pp. 1-28.

Singh, A.K., Kaur, R., Sahu, D., Bilgaiyan, S. (2021), “Real-Time Emotion Detection and Song Recommendation Using CNN Architecture”, In: Swain, D., Pattnaik, P.K., Athawale, T.

(eds) Machine Learning and Information Processing. *Advances in Intelligent Systems and Computing*, vol 1311. Springer, Singapore.

Singh, N., Alawami, M.A, and Kim, H. (2023). When social networks meet payment: a security perspective. *2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM),* Seoul, Korea, Republic of, pp. 1-6.

Smruthi, N. Harini. (2019). A Hybrid Scheme for Detecting Fake Accounts in Facebook. *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, 7:5S3.

Srikanth, A., Varalakshmi, P., Somasundaram, V., *et al*. (2018). Congestion Control Mechanism in Software-Defined Networking by Traffic Rerouting. *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 55-58.

Stewart, R., 2007. Stream Control Transmission Protocol. RFC 4960. Available at: https://www.rfc-editor.org/rfc/rfc4960.html

Stewart, R., Xie, Q., Morneault, K., Sharp, C., Schwarzbauer, H., Taylor, T., Rytina, I., Kalla, M., Zhang, L. and Paxson, V., 2000. Stream Control Transmission Protocol. RFC 2960. Available at: https://www.rfc-editor.org/rfc/rfc2960.html

Stone, J., Stewart, R. and Otis, D., 2002. Stream Control Transmission Protocol (SCTP) Checksum Change. RFC 3309. Available at: https://www.rfc-editor.org/rfc/rfc3309.html

Strelcenia, E and Prakoonwit, S. (2023). A New GAN-based data augmentation method for Handling Class Imbalance in Credit Card Fraud detection. *2023 10th International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, India, pp. 627-634.

StrongDM (2024). Available at: https://www.strongdm.com/blog/unauthorized-access (Accessed: 09 November 2024).

Suarez-Tangil, G *et al*. (2019). Automatically dismantling online dating fraud." *IEEE Transactions on Information Forensics and Security* 15: 1128-1137.

Sun, C., Cui, H., Zhou, H., Nie, W., Wang, X, and Yuan, Q. (2019). *Epileptic seizure detection with EEG textural features and imbalanced classification based on easyensemble learning*," Int. J. Neural Syst., vol. 29, no. 10.

Susto, Gian Antonio, Matteo Terzi, and Alessandro Beghi. (2017). Anomaly detection approaches for semiconductor manufacturing." *Procedia Manufacturing* 11: 2018-2024.

Talluri, S., Zhang, Q and Chen, R. (2023). A Cloud-Native Federated Learning Architecture for Telecom Fraud Detection. *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, Miami, FL, USA, pp. 1-3, doi: 10.1109/NOMS56928.2023.10154302.

Tarmazakov, E, I. and Silnov, D, S. (2018). Modern approaches to prevent fraud in mobile communications networks*. 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, Moscow and St. Petersburg, Russia. pp. 379-381.

Tekkali, C.G, and Natarajan, K. (2023). Smart Payment Fraud Detection using QML – A Major Challenge. *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India, pp. 523-526.

Teng, H., Wang, C., Yang, Q., Chen X, and Li, R. (2023). Leveraging Adversarial Augmentation on Imbalance Data for Online Trading Fraud Detection. *in IEEE Transactions on Computational Social Systems*.

Vats, P and Tadepalli, S.K. (2022). Palm Vein Image Processing and Enhancement in Vein Pattern Recognition System on FPGA. *2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET)*, Patna, India, pp. 1-5.

Vijay, V, and Verma, P. (2023). Variants of Naïve Bayes Algorithm for Hate Speech Detection in Text Documents. *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, Greater Noida, India, pp. 18-21.

Wang, C., Zhu, H., Hu, R., Li, R, and Jiang, C. (2023). LongArms: Fraud Prediction in Online Lending Services Using Sparse Knowledge Graph. in *IEEE Transactions on Big Data*, vol. 9, no. 2, pp. 758-772.

Xu, Dong, Yanjun Wang, Yulong Meng, and Ziying Zhang. (2017). An improved data anomaly detection method based on isolation forest. In *2017 10th international symposium on computational intelligence and design (ISCID)*, vol. 2,pp. 287-291. IEEE.

Xuan, S., *et al*. (2018). Random forest for credit card fraud detection. in *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*. IEEE.

Y. S, S. S. U, S. P. G and S. KS. (2023). Blockchain based Roaming fraud prevention using LSTM model in 4G LTE Network. *2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence),* Noida, India, 2023, pp. 222-229.

Yulita, I.N., Paulus, E., Sholahuddin, A, and Novita, D. (2021). AdaBoost Support Vector Machine Method for Human Activity Recognition. *2021 International Conference on Artificial Intelligence and Big Data Analytics*, Bandung, Indonesia, pp. 1-4.

Zachos, G., Mantas, G., Essop, I., Porfyrakis, K., Ribeiro, J.C, and Rodriguez, J. (2022). Prototyping an Anomaly-Based Intrusion Detection System for Internet of Medical Things Networks," *2022 IEEE 27th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Paris, France, pp. 179-183.

Zeng, E., Kohno, T., and Roesner, F. (2021). What Makes a "Bad" Ad? User Perceptions of Problematic Online Advertising. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21),* Association for Computing Machinery, New York, NY, USA, Article 361, 1-24.

Zhao, M, and Song, H. (2022). Semantic Analysis based on Artificial Intelligence – Prediction of Telecom Fraud. *2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC)*, Qingdao, China, pp. 924-927.

Zhong, R., Dong, X., Lin, R. and Zou, H. (2019). An Incremental Identification Method for Fraud Phone Calls Based on Broad Learning System. *2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an*, China, pp. 1306-1310.

Zhou, J *et al*. (2023). FraudAuditor: A Visual Analytics Approach for Collusive Fraud in Health Insurance. in *IEEE Transactions on Visualization and Computer Graphics*.

Zhu, H., Hsu, C, and Zhou, Z. (2023). Bystander pro-celebrity cyberbullying: An integrated perspective of susceptibility to retaliation and social capital gains, *Information & Management*, Volume 60, Issue 5.

Zorn, G., 2014. Diameter Network Access Server Application. RFC 7155. Available at: https://www.rfc-editor.org/rfc/rfc7155.html

Zulfikar, W.B., Gerhana, Y.A, and Rahmania, A.F. (2018). An Approach to Classify Eligibility Blood Donors Using Decision Tree and Naive Bayes Classifier. *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, Parapat, Indonesia, pp. 1-5.

APPENDIX

How to Create CDR for Roaming Voice (Outgoing):

CDROMMFITAOM76291.txt

a. File Name format:-

The File Name should be start with ‘CD+Country Source Name+ITAOM+any unique 5 Digit’
How to find Country Source name:
The Country source name will be given by using below query.
Query: select * from mpdpltab where plmnname like'%Romania%'
Example: - CDROMMFITAOM76291

b. (Row No: -3)

In the CDR file mention the ‘country source name’ in Sender (Row No: -3)
Example: Sender: ROMMF

c. (Row No: -4)

In the CDR file mention the “ITAOM” as it is in Row No:-4
Example: Recipient : ITAOM

d. (Row No: -5)

In the CDR file mention the any five unique digits (Which we have given in filename) in ‘File Sequence Number’ (Row No: -5)
Example: File Sequence Number: 76291

e. Time Stamp in the CDR file

All mentioned timestamp should be different from each other or should not be same.

f. (Row no: -37, 38, 85, 86)

Mention the IMSI and MSISDN no in Row no: -37, 38, 85, 86

g. Calling MSISDN Writing format for CDR

For MSISDN add prefix as 39(as its Italian number)
Example; Msisdn: 393462961646

h. (Row no-40, 41 and 88)

Insert the destination country no (Example: -Italy) in called number (Row no-40) and Dialled Digit (Row No-41) and calling number (Row No-88) place with country code.
Example: Called Number : 390269831
Dialled Digits: 390269831
Calling Number : 390269831

i. (Row No-31)

Insert the country code of destination country in Row No-31
Example: Country Code: 39

j. (Row No-44 and 90)

Insert the duration in second in the Row No-44 and 90.
Example: Total Call Event Duration: 600

k. (Row No-80) –Very important

Insert the value -1
If called no. is fixed or Mobile no. of other operator --- ALL countries (Except Italy)

Insert the value 322
If called no is of other operator in Italy (Mobile+Fixed Line) – country - only Italy

Insert the value -220
If called no. is fixed line (Fixed Line) --- ALL countries
Insert the value 341
If called no is of international (Mobile only) ---- All countries

How to Create CDR for Roaming Voice (Incoming):

CDIRLECITAOM68266.txt

a. File Name format:-

The File Name should be start with 'CD+Country Source Name+ITAOM+any unique 5 Digit'
How to find Country Source name:-
-The Country source name will be given by using below query.
Query: select * from mpdpltab where plmnname like'%Irlanda%'
Example: - CDIRLECITAOM68266

b. (Row No:-3)

In the CDR file mention the 'country source name' in Sender (Row No:-3)
Example: Sender: IRLEC

c. (Row No:-4)

In the CDR file mention the "ITAOM" as it is in Row No:-4
Example: Recipient : ITAOM

d. (Row No:-5)

In the CDR file mention the five unique digits (Which we have given in filename) in 'File Sequence Number' (Row No:-5)
Example: File Sequence Number: 68266

e. Time Stamp in the CDR file

All mentioned timestamp should be different from each other or should not be same.

f. (Row no:-37, 38, 85, 86)

Mention the IMSI and MSISDN no in Row no:-37,38,85,86

i. Called MSISDN Writing format for CDR

For MSISDN add prefix as 39(as its Italian number)
Example: Msisdn: 393462961646

j. (Row ID -40, 41, 88)

Insert the Source country no (Irlanda) in called number (Row no-40) and Dialled Digit (Row No-41) and calling number (Row No-88) place.
Example: Called Number : 35312958900, Dialled Digits: 35312958900
Calling Number : 35312958900

k. (Row No-31)

Insert the country code of Destination country (Romania) in Row No-31
Example: Country Code: 40

l. (Row No-44 and 90)

Insert the duration in second in the Row No-44 and 90
Example: Total Call Event Duration: 600

m. (Row No-80) –Very important

Insert the value -1
If called no. is fixed or Mobile no. of other operator --- ALL countries(Except Italy)

Insert the value 322
If called no is of other operator in Italy (Mobile+Fixed Line) – country - only Italy

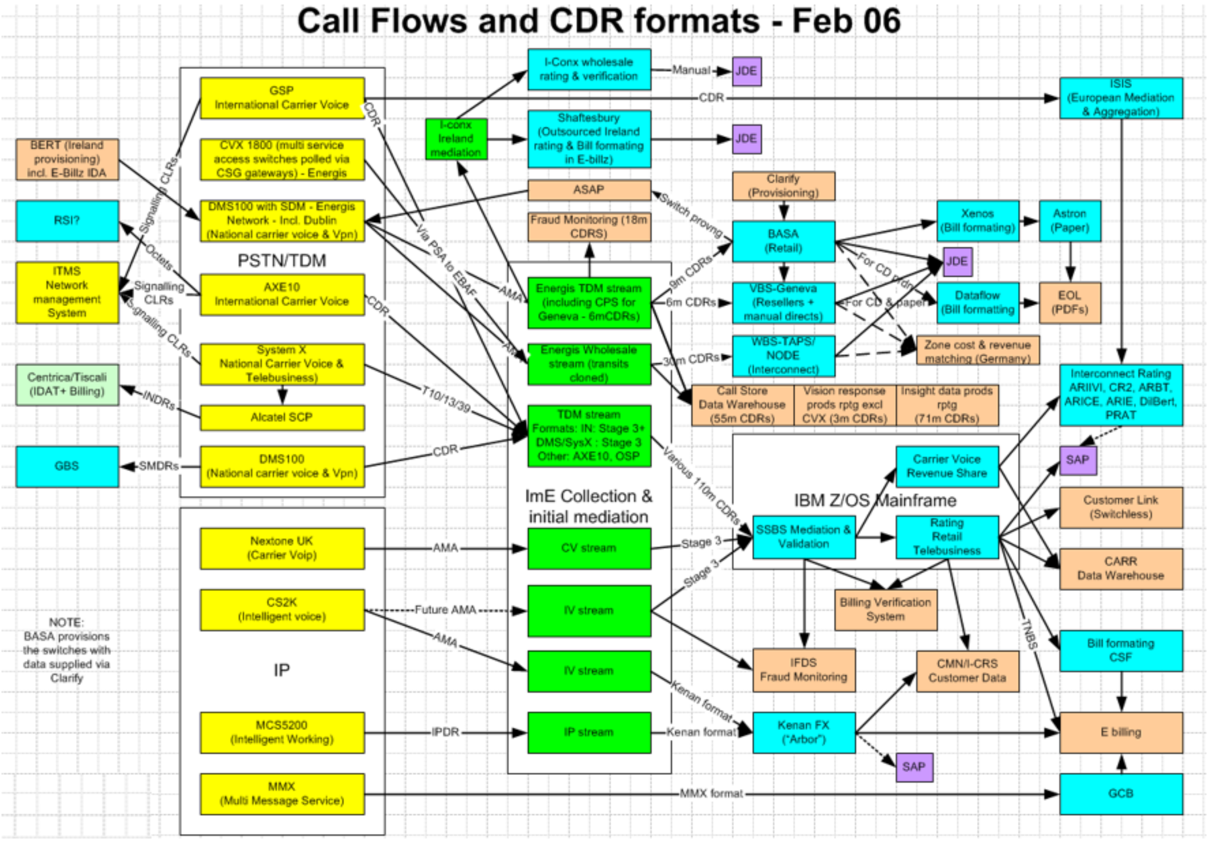
Insert the value -220
If called no. is fixed line (Fixed Line) --- ALL countries

Insert the value 341
If called no is of (Mobile only) ---- All countries

Billing Process/Testing
The billing process starts where a call is received into the network. This is either from a directly hosted customer initiating a call, or where received from an interconnected carrier. The billing process can be split into three main sections;

- CDR generation
- Mediation
- Downstream Processing

CDR information may not be the only information processed, or used, by the billing systems. The following diagram is intended as an illustration as to how complex the billing system can be. The point here is to understand that when dealing with an individual enhancement etc then the holistic view must be taken to ensure that a change in one place doesn’t affect existing processing or products.


Call Flows and CDR formats - Feb 06
PSTN/TDM
GSP International Carrier Voice
CVX 1800 (multi service access switches polled via CSG gateways) - Energis
DMS100 with SDM - Energis Network - Incl. Dublin (National carrier voice & Vpn)
AXE10 International Carrier Voice
System X National Carrier Voice & Telebusiness)
Alcatel SCP
DMS100 (National carrier voice & Vpn)
BERT (Ireland provisioning) incl. E-Billz IDA
RSI?
ITMS Network management System
Centrica/Tiscali (IDAT+ Billing)
GBS
Signalling CLRs
Octets
Signalling CLRs
Signalling CLRs
INDRs
SMDRs
CDR
Via PSA to EBAF
AMA
AMA
CDR
T10/13/39
CDR
I-conx Ireland mediation
I-Conx wholesale rating & verification
Manual
JDE
CDR
Shaftesbury (Outsourced Ireland rating & Bill formating in E-billz)
JDE
ASAP
Fraud Monitoring (18m CDRS)
Switch provng
Energis TDM stream (including CPS for Geneva - 6mCDRs)
Energis Wholesale stream (transits cloned)
TDM stream Formats: IN: Stage 3+ DMS/SysX : Stage 3 Other: AXE10, OSP
ImE Collection & initial mediation
9m CDRs
6m CDRs
30m CDRs
Clarify (Provisioning)
BASA (Retail)
VBS-Geneva (Resellers + manual directs)
WBS-TAPS/ NODE (Interconnect)
For CD ordn
For CD & paper
Xenos (Bill formating)
Astron (Paper)
JDE
Dataflow (Bill formatting
EOL (PDFs)
Zone cost & revenue matching (Germany)
ISIS (European Mediation & Aggregation)
Call Store Data Warehouse (55m CDRs)
Vision response prods rptg excl CVX (3m CDRs)
Insight data prods rptg (71m CDRs)
Interconnect Rating ARIIVI, CR2, ARBT, ARICE, ARIE, DilBert, PRAT
SAP
Various 110m CDRs
IBM Z/OS Mainframe
Carrier Voice Revenue Share
SSBS Mediation & Validation
Rating Retail Telebusiness
Customer Link (Switchless)
CARR Data Warehouse
IP
Nextone UK (Carrier Voip)
CS2K (Intelligent voice)
MCS5200 (Intelligent Working)
MMX (Multi Message Service)
AMA
Future AMA
AMA
IPDR
MMX format
CV stream
IV stream
IV stream
IP stream
Stage 3
Stage 3
Kenan format
Kenan format
Billing Verification System
IFDS Fraud Monitoring
CMN/I-CRS Customer Data
Kenan FX ("Arbor")
SAP
TNBS
Bill formating CSF
E billing
GCB
NOTE: BASA provisions the switches with data supplied via Clarify

Testing is done to prove the accuracy of the information a given platform is stamping in its CDR information. This statement infers that information regarding the test call must be carefully noted in order to provide a means to check the subsequent CDR.

New software loads on existing platforms need to be tested to ensure.

- The load remains compliant to the requirements
- Billing output format remains the same as the previous load
- Field content remains the same as the previous load

New platform types always need to be tested for billing output to ensure.

- To fully understand the field content (matches vendor documentation) for each field
- The platform is compliant to any BABT obligations

Intermittent testing will take place from time to time as new product requirements are introduced, or otherwise how information manifests in specific fields will need to be proven to a level that provides sufficient 'certainty' to exploit that field.
Test cases for each platform are listed to this document.

a. Test Call Information

The information required for billing testing must be accurate as it is the billing information itself that is in part being checked for accuracy (the checks also prove processing). Therefore all the relevant call details must be captured as the call is made in order to provide the basic information to check the billing output against. It is then imperative that the following information is captured manually for each test call;
Date of test

- Platform type

- Ingress signalling type
- Dialled number (the B number as received by the platform)
- Calling Party Number (billing number)
- Presentation Number CLI (if different from Billing Number)
- Time of Seizure
- Time of Answer
- Time of re-answer
- Time of transfer
- Time of transfer answer
- Time of A party Release
- Time of B party release
- Time of C party release
- Incoming route/sending gateway IP address
- Outgoing route/receiving gateway IP address

Timings must be captured in a manner that cannot be confused. A time interval of at least 10 seconds between party call, answer, and any releases must be provided. This includes a period of 10 seconds between each party releasing from the call.

The timings noted for billing test calls are to be as accurate as possible in order to provide a means for CDR accuracy to be checked. Either the testing engineer is to use accurate timing monitor applications, eg ITMS, Wireshark, or is to manually capture timings by setting their watch/laptop to the speaking clock. Note, there will be a slight amount of manual error (within a second) but it is important the calls cannot be confused with other test calls.

Additional timings may be captured from external monitoring equipment, eg ITMS is a good source as it provides timing down to milli seconds and is synchronised by an NTP time source.

Test calls are to be completed on the following occasions.

- Implementation of a new platform type
- After any data fill development
- After a new feature has been implemented

This ensures any adjustment required during development does not render as null and void any billing testing that had thus far been completed.

Test calls are to be run on a stable platform where no software development or patching is taking place with no datafill changes except by those of the test engineer that are necessary to complete specific test scenarios. Billing tests are to be carried out on a platform that matches the software load of the live network platforms except where specific features are being tested for live deployment. The Finance department must sign off that billing tests are complete before live deployment.

Tests to be run include the following.

All the following basic test call scenario's must be run when running billing tests for a new platform, otherwise a suitable sub-set of tests (as agreed with Billing Change) are required;

#	Test Scenario Description

1	Line ‘A’ (protocol A) calls Line ‘B’ (also protocol A). Both lines are hosted on the test switch. Line ‘B’ is busy. The call is PSTN number dialed and is not INDP triggered.
2	Line ‘A’ (protocol A) calls number ‘B’ (also on this switch). Both lines are hosted on the test switch. Number B is within a working number range, but is not presently allocated to a working line. Number B dialed as an incomplete address. The call is PSTN number dialed and is not INDP triggered.
3	Line ‘A’ (protocol A) calls Line ‘B’ (also protocol A). Both lines are hosted on the test switch. Line ‘B’ does not answer. The call is PSTN number dialed and is not INDP triggered.
4	Line hosted on this switch (protocol A) to another line hosted on this switch (also protocol A). Call does not route off the switch, around a loop, or trigger to the IN. B subscriber answers (would result in ‘Answer/Charge’ message). Call held up for at least 10 seconds and then subscriber ‘A’ clears. The call is PSTN number dialed and is not INDP triggered.
5	Line hosted on this switch (protocol A) to another line hosted on this switch (also protocol A). Call does not route off the switch, around a loop, or trigger to the IN. B subscriber answers (would result in ‘Answer/No Charge’ message). Call held up for at least 10 seconds and then subscriber ‘A’ clears. The call is PSTN number dialed and is not INDP triggered.
6	Line hosted on this switch (protocol A) to another line hosted on this switch (also protocol A). Call does not route off the switch, around a loop, or trigger to the IN. B subscriber answers. Call held up for at least 10 seconds and then subscriber ‘B’ clears and ‘A’ awaits for call to clear down. Subscriber ‘B’ re-answers before call cleared (if not cleared immediately). The call is PSTN number dialed and is not INDP triggered.
7	Line hosted on this switch (protocol A) to another line hosted on this switch (also protocol A). Call does not route off the switch, around a loop, or trigger to the IN. B subscriber answers. Call held up for at least 10 seconds and then subscriber ‘B’ clears and ‘A’ awaits for call to clear down. Subscriber ‘B’ does not re-answer. The call is PSTN number dialed and is not INDP triggered.
8	Line hosted on this switch (protocol A) to another line hosted on this switch (also protocol A). Call does not route off the switch, around a loop, or trigger to the IN. Call dialed as a locally dialed number. The call is not INDP triggered.
9	Line ‘A’ hosted on this switch (protocol A) to calls line ‘B’ hosted on this switch (also protocol A). Call does not route off the switch, around a loop, or trigger to the IN. Call dialed as a locally dialed number. ‘A’ subscriber then conferences in line ‘C’ hosted on this switch (also protocol A) and drops our of the call. Ensure two records with subscriber ‘A’ as the originating caller, and the duration for each call correctly details when the call between B and C dropped, not when ‘A’ dropped from the call. The call is not INDP triggered.
10	Line ‘A’ hosted on this switch (protocol A) to calls line ‘B’ hosted on this switch (also protocol A). Call does not route off the switch, around a loop, or trigger to the IN. Call dialed as a locally dialed number. Call forwarded from Line ‘B’ before answer to Line ‘C’ (also protocol A). Ensure there is a record for line ‘A’ to line ‘B’, and another record for the call line ‘B’ to line ‘C’. The call is not INDP triggered.
11	Line ‘A’ hosted on this switch (protocol A) to calls a number not hosted on this switch. Called Number is busy. The call is PSTN number dialed and is not INDP triggered.
12	Line ‘A’ hosted on this switch (protocol A) to calls a number not hosted on this switch. Called Number does not answer. The call is PSTN number dialed and is not INDP triggered.
13	Line ‘A’ hosted on this switch (protocol A) to calls a number not hosted on this switch. Called number is incomplete. The call is PSTN number dialed and is not INDP triggered.
14	Line ‘A’ hosted on this switch (protocol A) to calls a number not hosted on this switch.

	Called subscriber answers. A party clears, B party awaits call to clear before resetting handset/clearing their side of the call. Ensure call clears down, and CDR duration is correct. The call is PSTN number dialed and is not INDP triggered.
15	Line ‘A’ hosted on this switch (protocol A) to calls a number not hosted on this switch. Called subscriber answers (Answer/Charge). A party clears. Ensure call clears down immediately, and CDR duration is correct. The call is PSTN number dialed and is not INDP triggered.
16	Line ‘A’ hosted on this switch (protocol A) to calls a number not hosted on this switch. Called subscriber answers (Answer/No Charge). B party clears for 5 seconds, and then re-answers. Call held up over 15 seconds Ensure CDR duration reflects call from Answer to final release. The call is PSTN number dialed and is not INDP triggered.
17	Line ‘A’ hosted on this switch (protocol A) to calls a number not hosted on this switch. Called subscriber answers. B party clears and does not re-answer. A party waits for the call to time out before releasing. Ensure CDR duration reflects call from Answer to final release (will need to corroborate CDR duration with received signalling). The call is PSTN number dialed and is not INDP triggered.
18	Line ‘A’ hosted on this switch (protocol A) to calls number ‘B’ hosted on another switch (also protocol A). Call does not route around a loop, or trigger to the IN. Call forwarded from Line ‘B’ before answer to Line ‘C’ (also protocol A). Ensure CDR reflects call from line ‘A’ to line ‘B’. The call is PSTN number dialed and is not INDP triggered.
19	Line A hosted on this switch (protocol A) calls a UK fixed national number, ie an ‘01xxx’ or ‘02xxx’ number. Ensure resultant CDR reflects the correct destination. The call is PSTN number dialed and is not INDP triggered.
20	Line A hosted on this switch (protocol A) calls an International number, ie an ‘00(1-9)xxxxx’ number. Ensure resultant CDR reflects the correct destination. The call is PSTN number dialed and is not INDP triggered.
21	Line A hosted on this switch (protocol A) calls a UK mobile number, ie an ‘07(5-9)xxx’ number. Ensure resultant CDR reflects the correct destination. The call is PSTN number dialed and is not INDP triggered.
22	Line A hosted on this switch (protocol A) calls a UK personal number, ie an ‘070xxx’ number. Ensure resultant CDR reflects the correct destination. The call is PSTN number dialed and is not INDP triggered.
23	Line A hosted on this switch (protocol A) calls a UK 03xxx number. Ensure resultant CDR reflects the correct destination. The call is PSTN number dialed and is not INDP triggered.
24	Line A hosted on this switch (protocol A) calls a UK 05xxx number. Ensure resultant CDR reflects the correct destination. The call is PSTN number dialed and is not INDP triggered.
25	Line A hosted on this switch (protocol A) calls a UK 08000xx number. Ensure resultant CDR reflects the correct destination. The call is PSTN number dialed and is not INDP triggered.
26	Line A hosted on this switch (protocol A) calls a UK 0845xx number. Ensure resultant CDR reflects the correct destination. The call is PSTN number dialed and is not INDP triggered.
27	Line A hosted on this switch (protocol A) calls a UK 0871xx number. Ensure resultant CDR reflects the correct destination. The call is PSTN number dialed and is not INDP triggered.
28	Line A hosted on this switch (protocol A) calls a UK Premium Rate number, e.g. an 090xxx number. Ensure resultant CDR reflects the correct destination. The call is PSTN number dialed and is not INDP triggered.
29	Line A hosted on this switch (protocol A) calls a UK Specific number, e.g. ‘118118’. Ensure resultant CDR reflects the correct destination. The call is PSTN number dialed and is not INDP triggered.

30	Line A hosted on this switch (protocol A) calls 999. Ensure resultant CDR reflects the correct destination. The is not INDP triggered.
31	Line A hosted on this switch (protocol A) calls 112. Ensure resultant CDR reflects the correct destination. The call is not INDP triggered.
32	For a line/trunk hosted on this switch (protocol A), with default CLI Classification set to ‘Available’, dial a fixed PSTN number prefixed ‘141’, e.g. ‘14101xxxx’ The call is not INDP triggered.
33	For a line/trunk hosted on this switch (protocol A), with default CLI Classification set to ‘Available’, dial a fixed PSTN number prefixed ‘1471’, e.g. ‘147101xxxx’ The call is not INDP triggered.
34	For a line/trunk hosted on this switch (protocol A), with default CLI Classification set to ‘Withheld (Temporary Mode)’, dial a fixed PSTN number prefixed ‘141’, e.g. ‘14101xxxx’ The call is not INDP triggered.
35	For a line/trunk hosted on this switch (protocol A), with default CLI Classification set to ‘Withheld (Temporary Mode)’, dial a fixed PSTN number prefixed ‘1470’, e.g. ‘147001xxxx’ The call is not INDP triggered.
36	For a line/trunk hosted on this switch (protocol A), with default CLI Classification set to ‘Withheld (Permanent Mode)’, dial a fixed PSTN number prefixed ‘141’, e.g. ‘14101xxxx’. Call should complete and ‘141’ prefix will be ignored. The call is not INDP triggered.