iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.37011

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Revenue Assurance and Fraud Detection for Telecom Operators – Combating Bypass Fraud

Srikanth Aravamuthan

I. INTRODUCTION

Revenue Assurance and Fraud Detection (RAFD) is a combination of organizational structure, processes, and technology, responsible for understanding and monitoring the entire revenue process which includes:

1) Ensuring revenue risks are identified and addressed throughout the RAFD process
2) Assisting operations with the development of their processes and controls while ensuring the impact on other processes is considered and minimalized
3) Providing overall monitoring to ensure that process performance is measured and accurately reported

The primary objective of RAFD is to identify and erase leakages and therefore to maximise revenues. State of the art RAFD Management covers residential, business and wholesale customer as well as the partner/ supplier side to maximize revenues and minimize spend

II. BACKGROUND

Telecommunication fraud already costs the communications industry up to $28.3 billion per year, according to the Communications Fraud Control Association (CFCA). As per the CFCA survey report only 50% of the RAFD departments within the organization are monitoring more than 50% of their revenues*. The survey also highlights that approx. 77% of bad debts in an organization are because of fraud.

Most traditional Revenue Assurance and Fraud Management systems are static systems. They extract data from pre-defined data sources and follow a fixed rule library which is not self-updating. Given that fraudsters are constantly updating their methods, hardware and software technologies, fraud management systems in the organization become outdated quickly and requires constant high-cost interventions

Top 10 Fraud types*

Fraud Type	Global revenue impact
International Revenue Share Fraud (IRSF)	$5.04 billion
Arbitrage	$3.28 billion
Interconnect bypass fraud (e.g., SIM Box)	$2.71 billion
Domestic Premium Rate Service (In Country)	$2.27 billion
Traffic Pumping (includes Domestic Revenue Share, TFTP)	$2.0 billion
Commissions fraud	$1.76 billon
Device / hardware reselling	$1.76 billion
Theft / stolen goods	$1.49 billion
Friendly fraud	$1.17 billion
Wholesale SIP trunking fraud	$.98 billion

* https://cfca.org/wp-content/uploads/2021/02/CFCA-2019-Fraud-Loss-Survey.pdf

III. BYPASS FRAUD

As seen by the survey the global impact (only the detected and reported loss) on account of the bypass fraud alone is $2.71 billion and this number has been a big problem for many service providers for a long time

Use of SIM boxes to terminate international traffic is one of the most prevalent forms of bypass fraud. The risk of revenue loss from interconnect bypass using SIM boxes and the International Revenue Share Fraud (IRSF) can be managed using AI – ML based architecture described herein.

IV. AI ENABLED SELF-LEARNING RAFD MODEL FOR COMMUNICATION SERVICE PROVIDERS (CSP)

CSPs can setup an intelligent, self-learning and healing RAFD system to combat the various frauds faced by them. The self-referential "learning" should ensure that the rule sets are dynamic, up to date and capable of tackling the newer forms of fraud and possible revenue leakages. Another differentiation a CSP would need is that the system should be easily scalable to meet real time traffic demands and also need to have the capacity to accommodate different types of vendors, operators, carriers and partners.

A range of data modelling, analytical and statistical techniques can be deployed using this architecture to effectively detect the frauds and revenue leakages. This modelling and analytical tool is designed will have preconfigured rules and best practices. The model is broadly classified under basic heads as shown in the figure below.

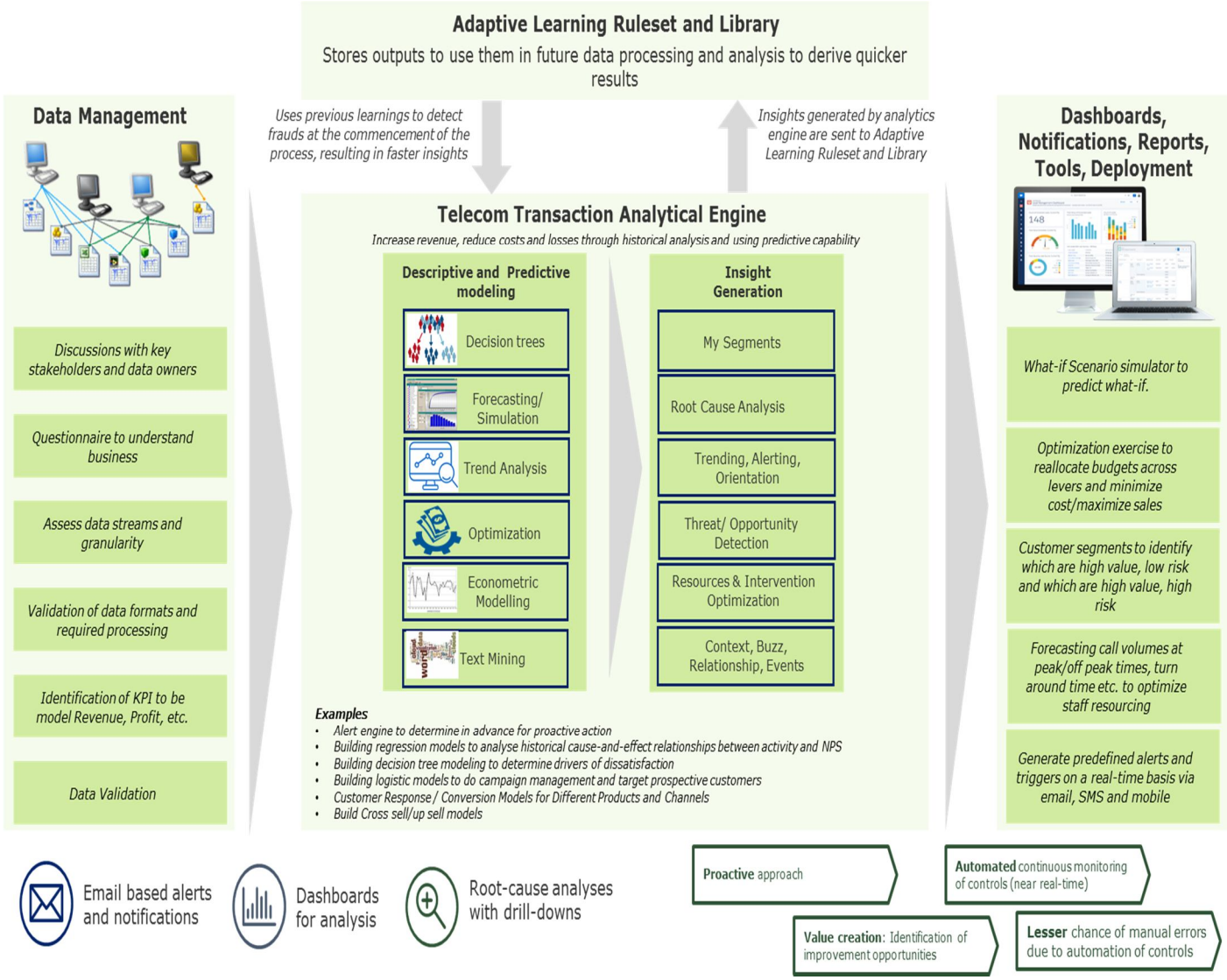


Fig 1(a): Technical Architecture – Transaction Analytical Engine – Reporting & Alerts

Once the analytical models are built, the insights derived based on the rules defined are fed back to the model for adaptive training. Post successful testing they can provide a real time and proactive identification of potential fraud scenarios and revenue leakages. The ideal data flow for effective real time analysis and action is given by the figure below

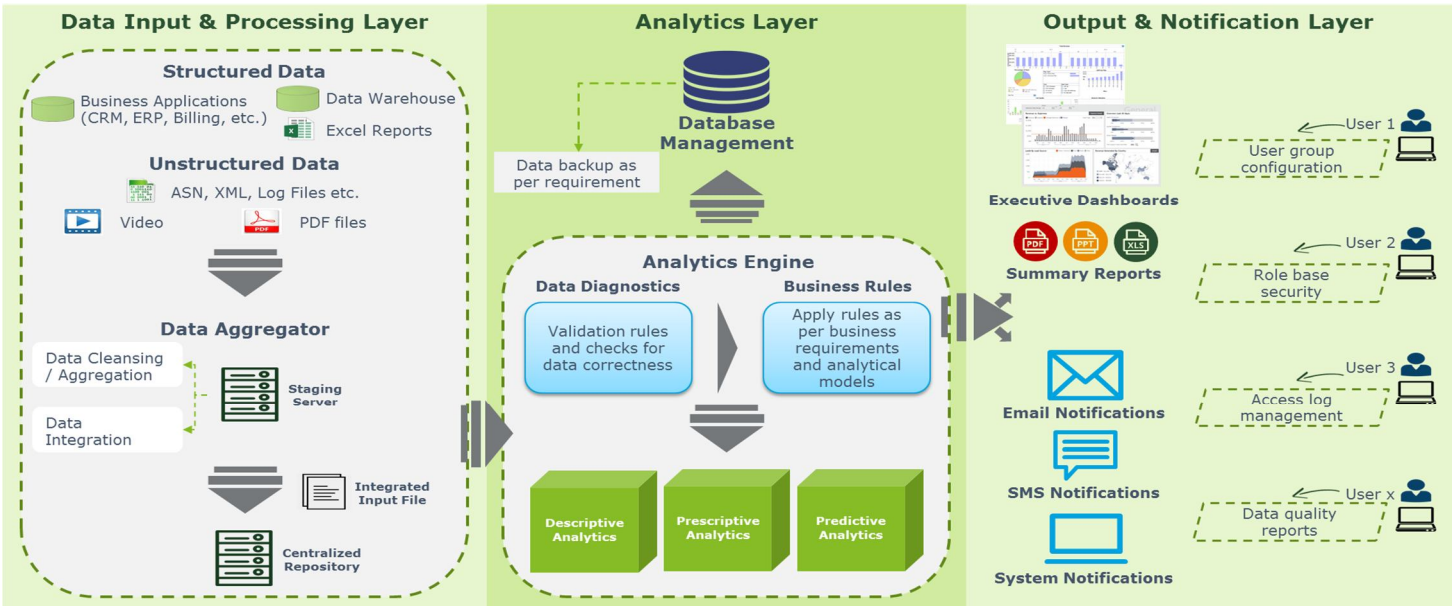


Fig 1(B): Transaction Analytical Engine – Business Flow– Reporting & Alerts

Key Benefits of this model are:

1) Adaptive Learning Capability – With an adaptive ruleset and library, this model is constantly adjusting itself to effectively identify potential fraud basis newer information inputs
2) Faster root cause investigation
3) Finding hard-to-detect revenue leaks
4) Continual savings through real-time KPI tracking
5) On-Going revenue stream protection

V. BYPASS FRAUD (SIM BOX) DETECTION APPROACH

It has been commonly observed that bypass fraud is more prevalent in the countries where the cost of terminating international calls exceeds the cost of origination of national call by a considerable margin. This difference in rates provides profit margins for the fraudsters which serve as the key motivating factor for them to invest in obtaining the technology/equipment required for conducting bypass fraud on a large scale. The below techniques will enable this model to evaluate the likely hood of SIM boxes:

A. Comparison of International terminating cost vis-a-vis local calling tariffs

The international terminating cost will be compared to the cheapest outgoing local call tariff on a predefined frequency within the model to establish if there is enough motivation to bypass international traffic. During this analysis local calling rates for the telecom operator as well as the local calling rates offered by competition are to be fed as data to the model. Higher the difference between the international termination costs and local origination costs higher the probability of bypass frauds.

B. Evaluating New Products

Likelihood of bypass fraud is rare in countries where differences between international terminating rates and national outgoing charge are low. However, the model also assists in identifying if there may be specific products in the market that may change the dynamics making it favourable for SIM box operators.

For example: Product offering unlimited voice calls to on-net and off-net numbers for a flat monthly fee. Introduction of this product encouraged SIM box operators to use this promotion to their benefit

Promotions involving bonus talk time or discounted tariffs launched by operators in the country need to be regularly scrutinized to eliminate the possibility of bypass frauds.

C. Trend Analysis

The Model also assist in trend analysis of terminating international calls by country and operator which will help in prediction on probable existence of SIM boxes.

D. Device Intelligence

SIM box is a functioning gateway that has the ability to install an array of SIMs. Each SIM card can be connected to VoIP/GSM gateway via internet. The model based on this architecture will have scripts built with logic to identify their behaviour to analyse data records. SIM boxes exhibit a distinctive usage pattern that enables statistical profiling. This logic can be built in as device intelligence using a self-learning and adaptive predictive model which proactively identifies SIM Boxes. The model parameters and rules can be tweaked in response to fraudsters' by adapting to their changing behavioural patterns. The parameters highlighted below will be used for creating a risk profile for the MSISDN and compensate usage analysis with call testing. Call testing scenarios will be developed based on the learning from usage analysis. These scenarios can also be built in statistical data model, an adaptive learning library (as shown in model above), which constantly learns, adapts, and adjusts itself.

E. High percentage of outgoing to incoming call ratio

The model will identify high ratio of outgoing calls to incoming calls though data analytics as probable SIM boxes. This will allow detection of termination of calls received through IP network on to GSM / Fixed line network.
Outgoing call ratio is defined by the following formula:

$$Outgoing\ Call\ Ratio = \frac{Total\ Outgoing\ Duration}{Total\ Incoming\ Duration + Total\ Outgoing\ Duration}$$

Alternatively, the model will also use call duration instead of call count as SIM boxes are modified to receive calls from customers who notice a missed call on their mobile to improve the ratio of incoming to outgoing call count. There is also increase in trend where SIM boxes use IVR and recorded messages to customers calling them to improve the incoming to outgoing call duration ratio. It is also noticed that SIM boxes are programmed to call each other at predefined frequencies to improve the incoming to outgoing duration ratio. However, trends and patterns around these calls can be identified through our analytical and statistical model

F. High Percentage of on-net Calls

The model can be used to detect high ratio of on-net calls to associate with SIM boxes. Originating call costs for on-net numbers tend to be cheaper than cost for calling off-net numbers, considering that local interconnect charges are to be paid. To gain a better margin SIM boxes are programmed to choose SIMs from the same operator from the array of SIMs available in the SIM box.

Formula used in detecting this scenario is:

$$\frac{Total\ Duration\ of\ Outgoing\ Calls_{Onnet}}{Total\ Duration\ of\ Outgoing\ Calls}$$

G. Diversity

Fraudsters are constantly adapting to new security measures and learning how to counter detection techniques that are based on analysis of mobile subscriber's usage pattern. For e.g., SIMs involved in bypass fraud may be programmed to call each other at a specific time period to avoid detection by scripts. In this scenario, the logic defined in Model can be tweaked to exploit other weaknesses of a SIM box device like No SMS sent or received, no data session carried on the SIM.
Diversity is defined as the ratio of count of total outgoing voice activity vis-a-vis count of total activity on the network. It is defined by:

$$\frac{\text{Count of Total Outgoing Voice Calls}}{\text{Count of total O/G voice calls + Count of total I/C Voice + Count of total O/G SMS + Count of total I/C SMS}}$$

An additional condition can be built, to include only MSISDNs without data usage based on the number of exceptions obtained from the previous step.
As fraudsters continue to manipulate usage behaviour to avoid detection, we will also have to incorporate variable thresholds and rules in the adaptive model based on the new usage pattern.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

H. High Number of calls to Distinct B party Numbers (Uniqueness)

Interconnection bypass usage pattern is also characterized by a high number of voice calls to distinct phone numbers as the usage pattern of SIM boxes would be different to that of individual users. SIM boxes would have a huge spread of called numbers when compared to individual callers.

Uniqueness is defined in the model as:

$$\text{Uniqueness} = \frac{\text{Count of distinct B} - \text{Number}}{\text{Count of total OutGoing voice calls}}$$

MSISDNs used in outbound call centres also tend to show similar characteristics as above and further investigations and test calls may be necessary to narrow down on the SIM boxes

I. Cell Site Analysis

The Model will identify SIMs which do not have handoffs between cell sites during travel but have dis-appear and reappear pattern. The typical characteristic of SIMs used in illegal call termination will be very low mobility or no mobility at all (may latch on to other cells sites in surrounding areas from time to time based on signal strength). Hence the model will detect such patterns.

Instances of fraudsters operating SIM boxes in vehicles (mobile units) to avoid this fraud pattern have been reported in certain countries. However, lack of good quality internet services on the move in most places makes this proposition a daunting task for the fraudsters.

It has been noticed that fraudsters use more than one location to house SIM boxes and swap the SIM trays between these to avoid detection through SIM box analysis.

J. Time Difference Between Successive Calls

The model can be designed to identify SIMs where calls are initiated in less than 2 minutes after the termination of the previous call. A similar calling pattern for the entire day including off-peak hours indicates a strong possibility of the MSISDN being used in SIM boxes.

The volume of traffic bypassed via SIM boxes translates into revenue for the fraudsters. Hence, in addition to the high count of outgoing calls, the time difference between successive calls is also minimal.

Connections used for business purposes (PBAX) and call centres may also show similar characteristics and further investigations may be required to wade off false alarms.

K. IMEI-MSISDN Analysis

To avoid detection, fraudsters often regularly rotate the SIMs. There are a wide variety of SIM boxes in the market. The older versions have one IMEI for the entire SIM slots whereas the later versions have programmable IMEIs for individual SIM slots. The programmable version has the capability to assign a different IMEI to the same SIM slot for each outgoing call.

Although SIM box technology has made significant advances, the AI enabled model can continue to detect patterns relating to IMEI-MSISDN pairing. Instances of one MSISDN associated to several IMEIs and vice-versa is still an indicator of possible SIM box behaviour.

Multiple MSISDN with same IMEI and MSISDNs with multiple IMEI on any given day are indicators of SIM boxes

The results of this test can be skewed due to handsets with no IMEI (000000 IMEI) or PBX. This can be remarkably effective where there is a regulation banning the use of handsets without IMEI and where PBX is not sold as a service.

L. Identifying Instances of off-net Bypass

Fraudster may use competitor's connections or any other means of terminating other than correct SIMs; it is identified as off net bypass fraud. Competitor's connections suspected to be used for this fraud can be identified by the combination of following logic:

1) High percentage of incoming calls from the MSISDN with very few outgoing calls to those MSISDN
2) High count of unique called party numbers by off-net MSISDN
3) Short duration between successive incoming calls from off-net MSISDN

Since several relevant information pertaining to SIM box characteristics like IMEI number, cell site information may not necessarily be available for incoming records at the home operator's switch, applying all the logic sets designed for detecting home operator's SIMs used in SIM boxes, will not be applicable.

M. Recharge Analysis

Due to stringent verification for post-paid subscribers, SIMs used for SIM boxing is usually prepaid. A committed fraudster will have several SIMs (usually running into several hundreds) and recharging these SIMs using conventional USSD/IVR methodology will become cumbersome. Therefore, the preferred mode of recharge for SIMs used in SIM boxes is 'Electronic Vouchers' or 'Airtime Transfers'.

This mode in itself will not throw indicators of SIM boxes, but can be very effectively used to corroborate evidences for suspected SIMs boxes

N. Reports from 'Device Manager'

Most operators have a system (Device Manager) that captures information of subscriber handsets that includes model number, manufacturer, IMEI etc. Apart from providing details of the handset the SIM card is currently associated with the system also has a capability to generate reports for MSISDNs that have initiated a device change in a particular time period.

Comparison of this report with the list of possible SIM box suspects identified via analytics can also help in corroborating the findings.

VI. TEST CALL GENERATION TO MINIMIZING FALSE POSITIVES

One of the greatest challenges in SIM box detection is to avoid false positives. Apart from the aspects discussed in the above section to prevent false positive in SIM box detection some of the other measures used by this model are detailed below:

VII. USAGE CHARACTERISTIC SCORING

In order to lower the risk of false positives, it is necessary to consider multiple fraud characteristics in analysis. If only, one characteristic is used to detect interconnection fraud, such as subscriber making high count of outgoing calls, the analysis would certainly be ineffective. An example of usage characteristic scoring mechanism is explained below:

Usage characteristics	Uniqueness to SIM box - A	Difficulty to escape detection - B	Fraud scale = A + B
High percentage of outgoing to incoming call ratio	High	High	6
Distinct B party numbers	Medium	High	5
Limited or zero mobility	High	Very high	7
Successive calls in short duration	Very high	Very high	8
Voice only usage	Low	Low	2
Recharge (for prepaid) analysis	Low	Low	2
IMEI -MSISDN analysis	Medium	Medium	4

Scoring methodology: Very High = 4, High =3, Medium = 2, Low =1

A simple scoring mechanism will involve assigning weight age to the usage characteristic based on the uniqueness to SIM box usage pattern and difficulty in avoiding detection.

Based on the scoring matrix, any subscriber with a usage characteristic matching a fraud score above a predetermined fraud score can be considered as a genuine SIM boxing suspect.

VIII. HYBRID DETECTION TECHNIQUES

We can perform this using a SIM box detection tool (Test call generator) **to proactively detect SIM boxes on the network** embedding Analytics. The output of this tool, whoever are the false positives, will be fed into the Adaptive Learning Library. The advantages of this approach is an extremely high fraud hit ratio and proactive identification of SIMs used in SIM boxes even before the SIMs start exhibiting classic SIM box behaviour pattern. However, this approach does not ensure complete coverage and does not provide insights into the fraudster's usage pattern. Hence there is a need to bring both the above approaches under a single complementing solution which will eradicate the deficiencies faced by each of them individually.

The test cases can be fed into the test call generator which can highlight MSISDNs that are used in SIM boxing. The usage pattern of these MSISDNs are then analysed from CDRs and appropriate thresholds are identified. These thresholds are then configured in call profile and pattern-based detection methodology to cover a wider base, resulting in faster identification of possible suspects.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

IX. PERFORMING ADDITIONAL TEST CALLS

SIM boxes usually do not have the ability to receive incoming calls. In most cases, an attempt to call a MSISDN used in SIM boxes ends up being 'not reachable' or is directed to voice mail. Certain SIM boxes have the ability to receive incoming calls; however, calls to MSISDNs configured in such SIM boxes are usually never answered.

By performing test calls to MSISDNs suspected to be involved in SIM boxing, the analyst can reasonably confirm if the observations are accurate.

X. BEST PRACTICES TO DISCOURAGE SIM BOXES

Although detective controls will minimize the revenue exposure to operators on account of bypass frauds, the following practices are followed in many countries which may deter the motivation of fraudsters to be involved in SIM boxing racket.

A. Migration to Higher Tariff Band

Most operators suspend the services of MSISDNs confirmed to be involved in SIM boxing. Although this eliminates the possibility of any subsequent usage, operators must explore the possibility of migrating these MSISDNs to a separate rate plan with higher cost for terminating calls. Since the fraudster is unaware of the change in tariff, his usage will continue to remain constant (until they realize the change) but the operator will stop leaking revenue on account of bypass fraud.

B. Increase Customer Awareness

Customer education and incentives for customers to report SIM box suspects based on CLI received can act as a highly effective tool for early detection of SIM boxes. In case of SIM boxes used for termination of international calls, customers would receive a local call CLI for an International Incoming call.

A similar endeavour to educate customers on the ill effects of bypass frauds (with emphasis on reduced QoS, missing call backs etc.) will substantially empower the operator in the fight to eliminate SIM boxes.

C. Obtain support of the Regulator and other Telcos

Bypass fraud is a menace that affects all Telecom Operators in the country. It is recommended that a partnership be formed among all Operators to fight Bypass Frauds. Regular exchange of SIM box reports among the operators will enable each other to leverage on the capabilities existing in the other operators and to pro-actively stop SIM boxing.

Operators in tandem with regulatory body can pass laws that awards stringent penalties for fraudsters. Operators can also rely on the regulatory body to facilitate implementation of laws banning devices that does not have or has an invalid IMEI. TRAI in India played a crucial rule in purging all devices without valid IMEIs from the network of all Indian Operators.

XI. SUGGESTED IMPLEMENTATION APPROACH TO THE SOLUTION AT A CSP

Highlighted below is our approach to implementation of Revenue Assurance and Fraud Detection management:

1) Phase 1: Understand

Key activities in this phase can include:

- Introduction to the existing Fraud Management team and processes (if any)
- Obtain a clear understanding of the AS-IS day to day operations currently followed by the Fraud Management Function (if any)
- Study existing Revenue Assurance and Fraud Management tools (if any)
- Study and understand the AS-IS systems
- Study end-to-end operations
- Study key vendor/partner processes and interfaces
- Understand and document current control points
- Understand the various data sources and review current data sets for effectiveness
- Understand key challenges and risk faced by the business
- Understand future state requirements and expectations

Key outcomes of this phase include:

- High level business understanding document on the As-Is state

2) Phase 2: Analyse& Identify

Key activities in this phase includes:

- Assess current state of governance, processes and systems
- Assess future state of governance, processes and systems
- Determine gap between as-is and future state
- Identify major areas for revenue leakage and fraud risk
- Identify products & services with potential risks for revenue leakage and fraud
- Infrastructure design for the RA and FM tool
- Sizing the infrastructure requirement for deploying the RA and FM
- Define transformation strategy
- Review of rules and thresholds to minimize false exceptions and to achieve continuous improvement

Key outcomes of this phase include:

- Identified revenue leakage risks
- Gap analysis

3) Phase 3: Design &Build

Key activities in this phase includes:

- Design the right-fit system for the current operations
- Design the business architecture for the planned future state operations
- Design the data flow requirements as per the planned architecture
- Assess key vendor/partner processes and interfaces
- Define the logic and the rules in the system that will be deployed
- Building the right statistical model
- Building the adaptive learning algorithm for the operations on a test environment
- Testing the statistical model and algorithm on a test environment
- Finalize implementation plans in discussion with the key stakeholders

Key outcomes of this phase include:

- Model Building
- High-level implementation and transformation plan
- RA and FM infrastructure identification

4) Phase 4: Implement &Operate

This is the steady-state phase of the project. Key activities in this phase includes:

- Model deployment on the identified infrastructure
- Adaptive learning algorithm modified and deployed
- Execute and monitor implementation
- Revising of the logic and rules of the system on an on-going basis
- Testing of the re-configured rules are working as per plan
- Continuous improvement based on periodic reviews

Key outcomes of this phase include:

- Fully implemented and tested RAFD Management system deployed

A. Implementation Plan

High Level Implementation plan phase wise along with key activities has been displayed in the below figure.

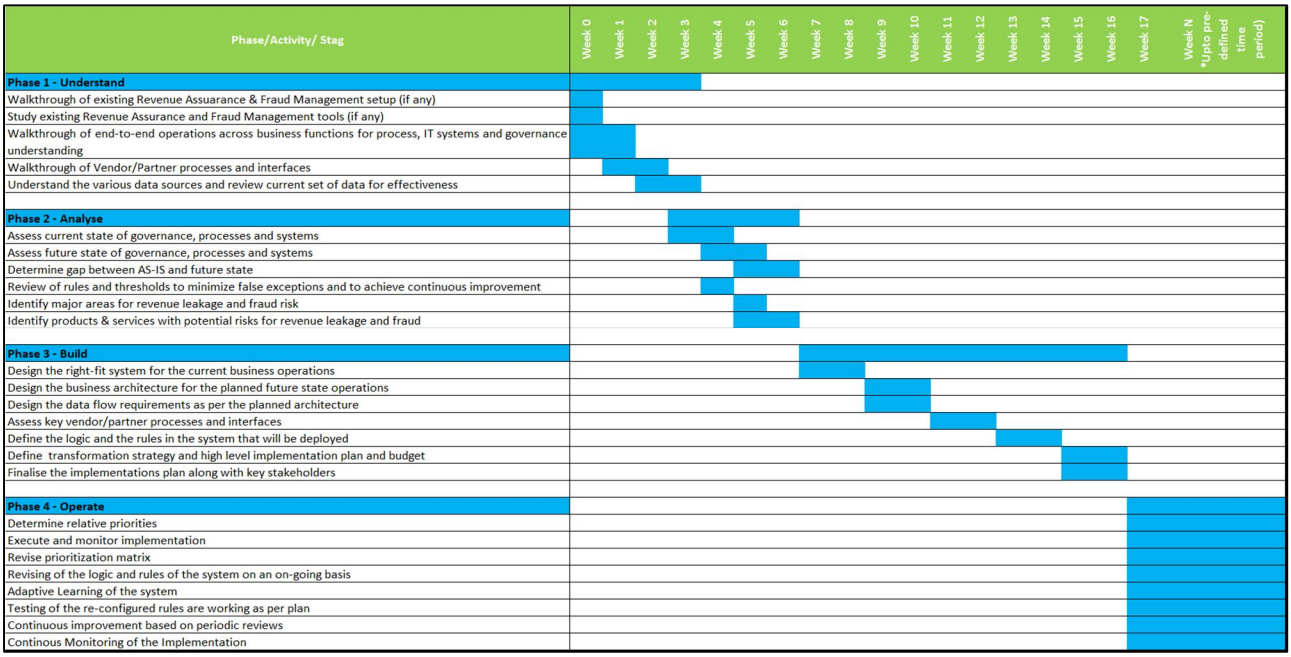


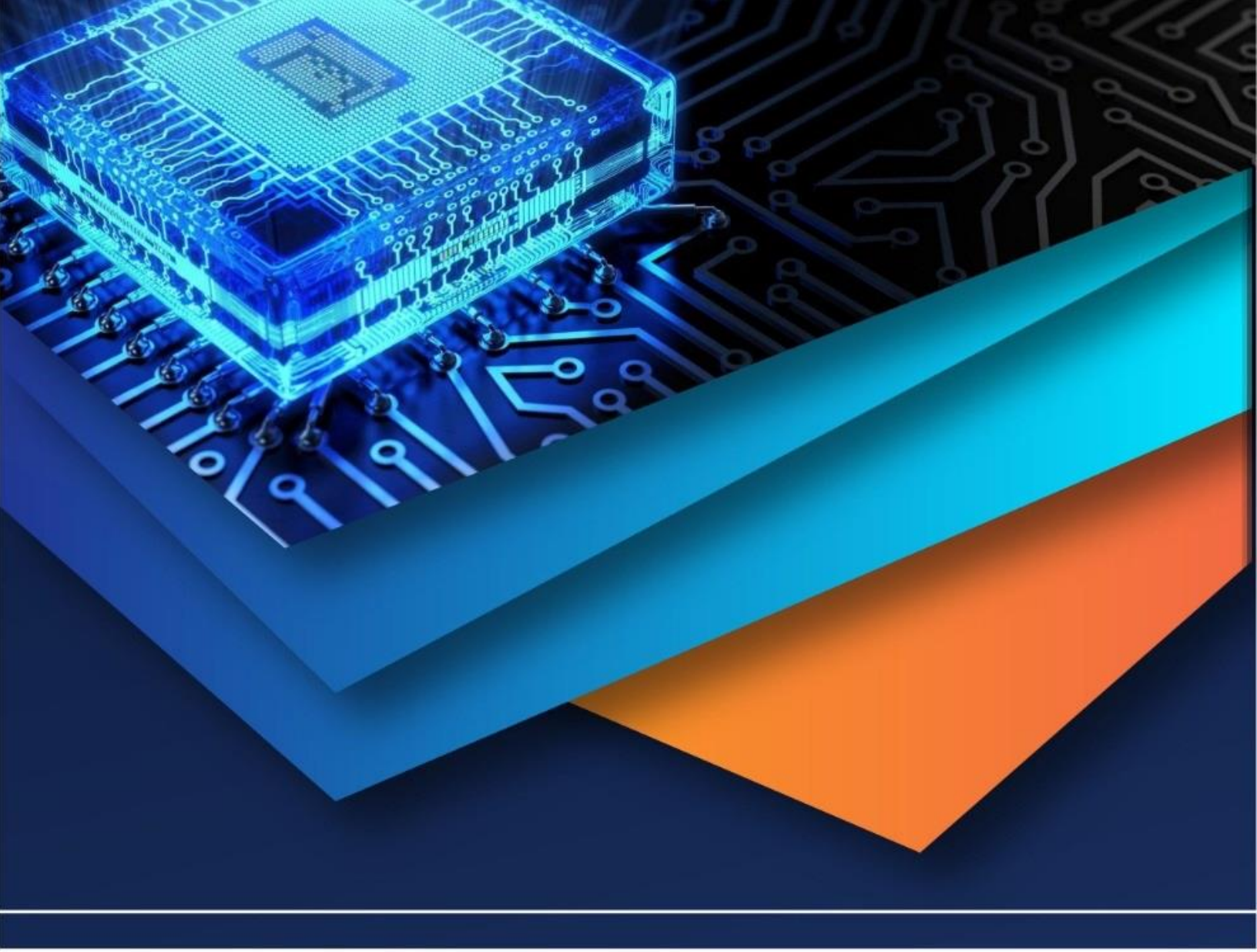
Figure 2: High level Proactive RA and FM Execution Plan

XII. CONCLUSION

The implementation of AI enabled, self-learning RAFD architecture with the proactive action mechanism will greatly reduce the revenue loss due to fraud faced by the CSP. This architecture can be easily adopted to combat all types of revenue leakage and fraud in a CSP.

XIII. LIST OF ABBREVIATIONS

A. AI – Artificial Intelligence
B. CDR – Call Detail Record
C. CLI - Calling Line Identification
D. FM – Fraud Management
E. GSM – Global System for Mobile communications
F. I/C – Incoming
G. IMEI - International Mobile Equipment Identity
H. IVR - Interactive voice response
I. KPI – Key Performance Indicators
J. ML- Machine Learning
K. MSISDN – Mobile Station International Subscriber Director Number
L. O/G – Outgoing
M. PBAX - Private Branch Automatic Exchange
N. PBX - Private Branch eXchange
O. QoS – Quality of Service
P. RA – Revenue Assurance
Q. SIM – Subscriber Identity Module
R. SIP – Session Initiation Protocol
S. SMS – Short Message Service
T. TRAI – Telecom Regulatory Authority of India
U. USSD - Unstructured Supplementary Service Data
V. VoIP – Voice over Internet Protocol



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY