# A Study on SIM box or Interconnect Bypass Fraud

دراسة حول سم بوكس أو الاحتيال التفافي البيني

**by**

**NAJMA ALGHAWI**

**Dissertation submitted in fulfilment**
**of the requirements for the degree of**
**MSc INFORMATICS**
**at**
**The British University in Dubai**

**February 2019**

# DECLARATION

I warrant that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.

I understand that a copy of my research will be deposited in the University Library for permanent retention.

I hereby agree that the material mentioned above for which I am author and copyright holder may be copied and distributed by The British University in Dubai for the purposes of research, private study or education and that The British University in Dubai may recover from purchasers the costs incurred in such copying and distribution, where appropriate.

I understand that The British University in Dubai may make a digital copy available in the institutional repository.

I understand that I may apply to the University to retain the right to withhold or to restrict access to my thesis for a period which shall not normally exceed four calendar years from the congregation at which the degree is conferred, the length of the period to be specified in the application, together with the precise reasons for making that application.

Signature of the student

# COPYRIGHT AND INFORMATION TO USERS

## Abstract

SIM box or Interconnect Bypass Fraud is one of the fastest emerging frauds in Telecom industry today, costing the industry some USD 3 Billion. Calls made through the internet are sent to SIM boxes (machines that contains SIM cards) which redirect this illegitimate VoIP traffic onto mobile networks. Fraudsters effectively bypass the inter-connect toll charging points to exploit the difference between the high interconnect rates and the low retail price for on-network calls, thus avoiding payment of the official call termination fee of an Operator or MVNO.

Keywords: SIM box, Bypass, Fraud, SIM card, VOIP, Operator, Telecom Industry, Implementation, Adoption, Payment, Inter-connect.

خلاصة

يعد سم بوكس أو الاحتيال التفافي البيني أحد أسرع عمليات الاحتيال الناشئة في صناعة الاتصالات اليوم ، حيث تكلف الصناعة حوالي 3 مليارات دولار أمريكي. يتم إرسال المكالمات التي يتم إجراؤها عبر الإنترنت إلى صناديق سيم (الأجهزة التي تحتوي على بطاقات سيم) والتي تعيد توجيه حركة مرور فويب غير الشرعية هذه إلى شبكات المحمول. يتغلب المحتالون على نحو فعال على نقاط فرض رسوم التوصيل البيني لاستغلال الفرق بين ارتفاع أسعار الربط البيني وسعر التجزئة المنخفض للمكالمات عبر الشبكة ، وبالتالي تجنب دفع رسوم إنهاء المكالمة الرسمية للمشغل أو MVNO.

الكلمات الرئيسية: مربع سيم ، تجاوز ، الاحتيال ، بطاقة سيم ، فويب ، المشغل ، صناعة الاتصالات ، التنفيذ ، التبني ، الدفع ، الاتصال البيني.

## Acknowledgement

I am grateful to Professor Cornelius Ncube for his kind support and endless motivation in my endeavor to achieve success in this thesis.

I thank the respondents from the selected Telecom Operators and Fraud Management System Vendors. I would also like to extend my thanks to my family and friends, without their support I would not have been able to achieve this

## Contents

## List of Abbreviations

| | | |
|---|---|---|
| 1. | AED | United Arab Emirates Dirham |
| 2. | ASCII | American Standard Code for Information Interchange |
| 3. | BAM | Business Activity Monitor |
| 4. | BI | Business Intelligence |
| 5. | BSC | Base Station Controller |
| 6. | BTS | Base Transmitter Station |
| 7. | CDR | Call Detail Records |
| 8. | CEP | Complex Event Processor |
| 9. | CFCA | Communications Fraud management Association |
| 10. | CLI | Calling Line Identification |
| 11. | CTR | Click-through Rate |
| 12. | CPP | Calling Party Pays |
| 13. | CUP | Current User Profile |
| 14. | DAHP | Database-Active Human-Passive |
| 15. | DAS | Data Analytics Server |
| 16. | DBMS | Database Management System |
| 17. | DDOS | Distributed DOS |
| 18. | DOS | Denial of Service |

| | |
|---|---|
| 19. DSMS | Data Stream Management System |
| 20. EDGE | Enhanced Data for GSM Evolution |
| 21. FDT | Fraud Detection Tool |
| 22. FMS | Fraud Management Association |
| 23. FIINA | Forum for International Irregular Network Access |
| 24. GPRS | General Packet Radio Service |
| 25. GSM | Global System for Mobile |
| 26. GT | Global Title |
| 27. HADP | Human-Active Database-Passive |
| 28. HSPA | High Speed Packet Access |
| 29. HTTP | Hypertext Transfer Protocol |
| 30. IDD | International Direct Dialing |
| 31. IMEI | International Mobile Equipment Identity |
| 32. IMSI | International Mobile Subscriber Identity |
| 33. ISC | International Switching Center |
| 34. ISDN | Integrated Services Digital Network |
| 35. ISUP | ISDN User Part |
| 36. LRIC | Long Run Incremental Cost Model |
| 37. MVNO | Mobile Virtual Network Operator |
| 38. QOS | Quality of Service |

| | |
|---|---|
| 39. SIM | Subscriber Identification Module |
| 40. SMS | Short Message Service |
| 41. TRs | Termination Rates |
| 42. USD | United States Dollar |
| 43. VOIP | Voice Over Internet Protocol |

**Chapter # 1**

# 1. Introduction

In this section, the foundation of the issue in which the postulation depends on will be exhibited, and in addition the exploration question and reason. Ultimately, the demarcation and the manner of the postulation will be displayed.

A SIM box (also called a SIM bank) is a tool used as part of Voice over Internet Protocol (VoIP) gateway installation. It contains a huge number of cards, which are connected to the VoIP gateway but it is kept and stored separately from it. A SIM box may or may not have SIM cards of various mobile operators present in the geographic region, allowing it to operate with several GSM gateways located in different places [1].

## 1.1.Background

Fraud is pre-existing since time immemorial and might take unlimited selection of forms. It occurs in several forms i.e. telecommunication fraud, credit-card fraud, net-dealings fraud, e-cash fraud, insurance fraud and healthcare fraud, money laundering, intrusion into computers or computer networks. The task of detecting fraud is similar in all these areas [1].

Fraud is different from revenue leakage.[1] Revenue leakage is characterized by the loss of revenues ensuing from operational or technical loopholes, wherever the ensuing losses are sometimes retrievable and customarily detected through audits or similar procedures. Fraud is characterized with felony by deception, generally characterized by proof of intent, where the ensuing losses are usually not retrievable and should be detected by analysis of calling patterns.

The Communications Fraud Management Association (CFCA) conducted a survey and determined that USD: 72–80 billion in losses are due to telecom fraud worldwide (CFCA, 2009). While many large operators have developed durable, Fraud Management Systems (FMS) to combat fraud, others are yet to develop. The Forum for International Irregular Network Access (FIINA) solely concluded that regarding 100% of operators worldwide have set in situ smart and effective fraud strategies.

The motivation behind crime is attributed to migration and demographics, penetration of latest technology, employees' discontentment, the 'challenge factor', operational weaknesses, poor business models, criminal greed, concealment and political and ideological factors [1][2].

Nowadays, there are various paths to make money. Ideology of getting richer in short span of time without big financial investments is a popular unethical practice of fraudsters. SIM boxing fraud is just one of the ways to earn quickly. This is part of the largest business in the world of telecommunications, called GSM termination. All this business is built on earnings on international calls.

It is not a secret that tariffs for international calls are always much more expensive than calls within the network. But why are international calls more expensive than the local calls. This is because in case of international calls, the telecom operator of the caller party is expected to pay interconnects charges as well as termination charges [3].

Termination rates are the charges which one telecommunications operator charges to another operator for terminating calls on its network. This model of charging these fees is known as calling party pays (CPP).

This is a part of the most important business within the world of telecommunications, referred to as GSM termination. All this business is constructed on earnings on international calls.

It's not a secret that tariffs for international calls square measure invariably far more big-ticket than calls at intervals of the network.

For example:

A customer of Operator ‘A’ mobile wants to call a friend who has an Operator ‘B’ mobile.

Operator ‘A’ can charge the client a fee per minute (the retail charge) for this decision.

Operator ‘B’ can charge Operator ‘A’ a fee for terminating the decision on its network.

This termination rate thus forms a part of Operator ‘A’'s price of providing the decision to its client. Termination rates could also be commercially negotiated or could also be regulated.

A range of approaches can be used to regulate rates. International benchmarking or cost models such as LRIC (Long Run Incremental Cost Model) or LRIC + cost models are the most common approaches to calculate the efficient levels of termination rates. In LRIC models, the termination prices square measure calculated for an economical theoretical mobile operator. The model assumes that corporations use the simplest technologies to supply mobile calls and services [3][4].

It is a protracted run model because it takes under consideration the expansion of demand, which is calculated using data on observed traffic, income and user information. It considers the period that the service supplier has to invest in capital enhancements to supply the mobile decision services [5].

Termination rates (TRs) derived from this 3 model thus calculate capability prices of every part of the network, expressed in terms of per minute use. Under a pure LRIC model, costs are calculated for an efficient hypothetical firm. The difference between both models is that while the former calculates (TRs) via the division of total costs by total demand, pure LRIC methodology calculates (TRs) through comparing a firm that gives mobile voice access and one that does not, to see the mandatory prices of providing mobile services. Historically, there was, and in some countries still is far a dialogue regarding the simplest level for interconnection rates. Some argue that approaches supported models do not take under consideration universal risks and prices. Therefore, suffer among other things, from survivorship bias (they consider that risk can be assessed by wanting solely at the returns of extant companies) and so underestimate verity level of risk [6] [7].

Another concern is based on Real Options. This considers the profit that is destroyed from the instant that a capitalist chooses to speculate and suggests that the loss of this right to speculate; ought to be taken into account once watching the expected returns on investments created.

The fundamental principle of any telecommunications network is to allow calls originating from a subscriber ‘A’ to attain a subscriber ‘B’, whether on the same network or on another network, usually known as “any to any connectivity”. In more technical terms, traffic, initiating from Subscriber ‘A’ is terminated at a point of destination. Subscriber ‘B’, and in order to permit for traffic to be routed and terminated between different operators, “interconnection” must be established.

Interconnection permits for calls placed by a subscriber in one network to achieve a subscriber in another network. Such a call is “terminated” within the destination network [7][8][9].

Motivation for the Fraud:

Detecting a fraud after the event has occurred, is not nearly as useful as catching it in real time. While CDR (Call Detail Records) is generated in real-time, immediately after related transaction is completed, most fraud detection tools use CDR for post analysis. Some of the sophisticated tools available in industry use signaling information in addition to CDRs to make detections in real-time.

However, capturing such signaling information incur additional cost as supplementary probing devices are introduced to the network. Also, these probing devices may introduce additional point of failure to the network. Therefore, relying on CDRs is most cost effective and reliable method. CDR contains all the required data to make near real-time fraud detections. But most of the business analytics tools store this data on static storage and perform batch operations on past data to calculate aggregate values such as sums and averages.

Even though such analysis gives useful insights about the past behavior, it is not sufficient in current business world, as it is not capable of exploiting the timeliness value of data and does not captures time sensitive call patterns inside CDR stream. For example, following are two use cases where real-time CDR analytic could be useful [9][10].

The subscribers' fraud motivations are:

57% from the respondents insure that fraudster's motivations to commit fraud attacks are bad experience. The subscriber inelegances, innovation and also monetary value; these findings agreed what previous studies clarified [9][10].

Greed:

The primary motivation which causes offenders to steal mobile telephones and use them to obtain services without incurring a charge to themselves is greed. Offenders may simply seek to exploit the opportunities provided by new forms of telecommunications technology to obtain calls for free (although, of course, the calls are infact only free to the offender and the legitimate subscriber has to pay unless some other arrangement can be negotiated with the service provider). Some offenders have established lucrative businesses of dealing in stolen equipment and services [11][12].

Curiosity:

If one examines the history of theft of telecommunications services, one important factor emerges which distinguishes these crimes from traditional property offences. This is the purpose for which the illegal conducts carried out. The very early cases of improper use of fixed-wire telephone services were often undertaken not for profit but out of curiosity.

Sheer interest in how systems work and the challenge of defeating security measures provides a powerful incentive which drove many individuals to commit offences against telecommunications systems. The same motivation can be seen to apply in the case of offenders who steal mobile telephone services.

The sophisticated technological procedures needed to scan security numbers and to produce counterfeit telephones obviously creates a keen challenge to technologically-minded individuals with a desire to break the law. Traditional deterrence-based sanctions which operate in respect of offenders whose motivations are primarily financial, may, therefore be inappropriate when dealing with individuals who are not intent on making a profit from their enterprise [12][13].

Envy:

In a time when new technological developments are taking place, particularly those involving attractive consumer goods such as mobile telephones which are highly publicized, there is a possibility that new social divisions could emerge based upon access to and familiarity with the new technologies. People without access to mobile telephones, for example, may feel isolated and deprived and a new environment conducive to criminality may be created in which theft of telephones and telecommunications services will become a major social problem [13].

Need:

Mobile telephones have become a readily saleable commodity in the black market making them attractive to individuals who need to obtain funds by criminal conduct. In a plea made in mitigation of sentence in the Melbourne Magistrates' Court recently, a mobile telephone thief was described as being unemployed and trying to treat a drinking problem. In another case heard before the same court in which the offender had stolen one hundred mobile telephones from motor vehicles; the offender told the police that he needed money to keep a roof over his head and to pay for food [14].

In some of the countries, the proliferation of illegal mobile telephones was such that organized groups of criminals became involved in selling telephone services to those who were unable to obtain legitimate access to services such as the indigent and illegal immigrants. In one of the cities, one such group, the 'Orchard Street Finger Hackers' became notorious. This group of offenders, who came out of the cocaine-dealing sub-culture, sold stolen long-distance telephone services in various unsavory neighborhoods to a captive clientele of illegal immigrants who were desperate to call home [15].

## 1.2.Purpose

The motivation behind this thesis is to build up a comprehension of which components that deter or cultivate SIM box Fraud. Our aspiration is to add to the exploration on of technical advancement for this type of fraud and various ways in which it can be detected and avoided [15].

## 1.3.Demarcate

The investigation of this proposition is7 delimited to the money segment and more particularly the telecom industry in United Arab Emirates (U.A.E.) and the two conventional telecom operators have contributed to this research. Conventional telecom operators in U.A.E were picked as they have comparative authoritative detection controls. There is a firm level center of the concentrate as opposed to individual, because of the novelty of the marvel inside the business [16].

## 1.4.Disposition

This thesis has concentrated on the device SIM box and its misuse is another threat to the telecom industry and what figures that are influencing potential executions of SIM box fraud. Section one of this postulation is the basic area, which presents the foundation of the issue that the proposal is concentrating on, including the reason, research inquiry and demarcations. In section two the threat is depicted together with the hypothetical structure, which the study depends on. Section three clarifies the utilized approach and how information have been gathered and investigated. In section four the outcome from the study is introduced, examined and talked about together with the hypothetical system. Finally, section five compresses the result of the theory, including suggestions, constraints and proposals for further research [17].

Chapter # 2

# 2. Literature Review

In this area, the SIM box idea, the CDR field overview, gateway structure and past exploration will be displayed keeping in mind the end goal to comprehend critical associations and discourses further on in the proposition. In conclusion, an investigation of the decision of structure and a hypothetical synopsis will be introduced.

## 2.1.SIM box

A SIM box (also called a SIM bank) is a tool used as part of Voice over IP (VoIP) Gateway installation. It contains a huge number of SIM cards, which are connected to the VoIP gateway but it is kept and stored separately from it. A SIM box may or may not have SIM cards of various mobile operators present in the geographic region, allowing it to operate with several GSM gateways located in different places.

In other words, SIM box permits you to install and manage any amount of SIM cards of different mobile operators that enables work of several GSM gateways placed in different locations. Several SIM boxes are connected in one system that has the ability to use unlimited quantity of SIMs in your system.

In competitive mobile market Operators associate degreed MVNOs have down their retail costs for occupation mobile numbers as an incentive to bring customers to their networks. Operators and MVNOs typically supply promotions or subscriptions permitting free calls to mobile numbers on an equivalent network and typically additionally to competitors' mobile networks.

Due to the difference between the interconnect rates and the retail price for on-network calls, fraudsters deploy SIM boxes to avoid paying the official call termination fee of an Operator or MVNO. This type of fraud is sometimes known as Interconnect Bypass Fraud or SIM Box fraud.

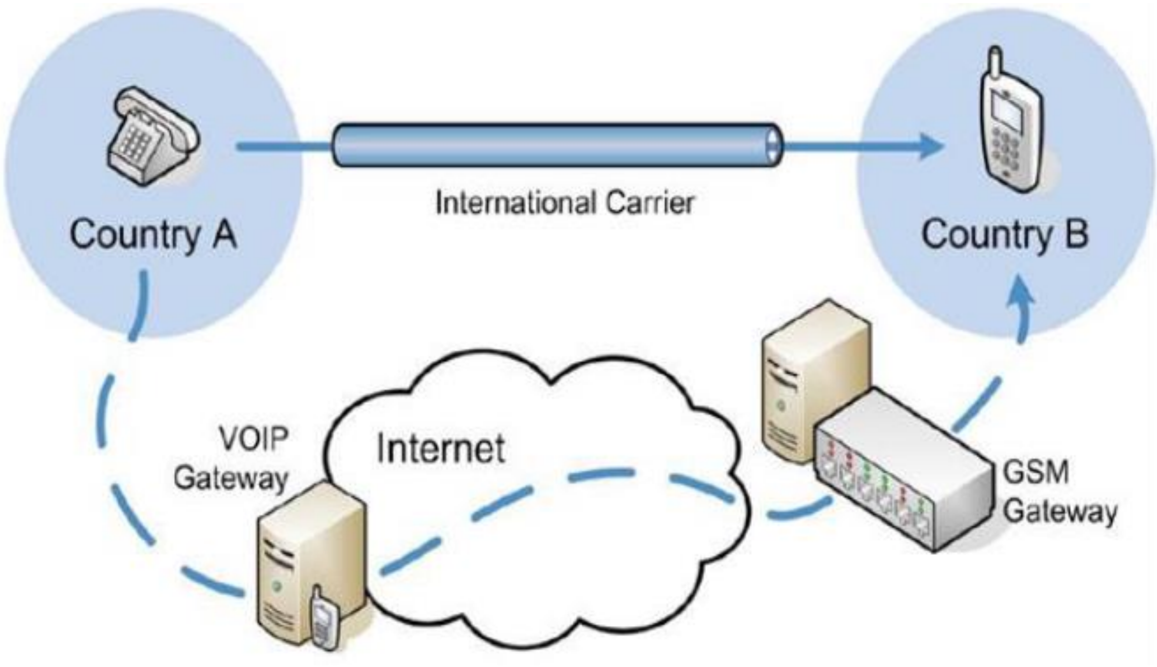


Figure 1: Basic Bypass Call Flow Architecture

## 2.2.Technology

It is a PC-based software platform providing the ability to create, modify, manage and deploy any simulation-based content – aircraft, cars, ships, weapons, e-Learning material, and more – across a multitude of domains, such as training, research & development, operations analysis, and entertainment.

- SIM box contains a wide range of software modules empowering users with infinite possibilities in creating new products and environments.
- SIM box is comprised of three main environments:
- SIM box Toolkit, a development environment;
- SIM box Server, a central management environment;
- SIM box Runtime, a delivery environment.

Several configurations of SIM box are generally available, the most widely used ones are for 60 SIMS and another one is for 120 SIMS. One system can include as many SIM boxes as you wish, which gives ability to use any amount of in the termination system.

Module-based structure of the merchandise hardware parades a good variety of potentialities to its users, such as:

- SIM cards could be placed separately from GSM modules (this option will require high-quality Internet connection between GSM gateway and SIM Box);
- SIM cards which take part in termination of voice traffic and SMS termination to different destinations/countries could be placed in the same spot, making it easy to control

their activity.

It is vital to own many GSM gateways settled in several countries or in several regions of an equivalent country.

All SIM cards that square measure concerned in termination are also settled in one or a lot of SIM Boxes placed within the same spot [20].

SIM cards rotation:

One of the optimization algorithms of the system is SIM Rotation. SIM cards among every SIM Box is divided into teams, each of these groups can be attached to a separate GSM-module of VoIP gateway. Over time, the system is ready to create changes among every cluster, changing SIM card which is responsible for making voice calls from one to another.

This not solely permits you to optimize resource consumption of each single “SIM”, however, additionally provides a clear stage to cut back their employment and, consequently, the suspicion of the mobile operators [22].

SIM cards migration:

The system is capable of registering the SIM cards on different GSM-modules with a specified frequency. If the user has numerous GSM gateways positioned in different parts of the city, system will make SIM card conduct calls from every gateway in turn, creating an illusion of subscriber’s movement. This will help the user to protect your cards from being blocked by the mobile operator [20].

Chapter # 3

# 3. Research Methodology

## 3.1.SIM box Fraud

SIM box fraud means most of the time bypass fraud or call reselling. In order to understand as to how it affects the telecom operators, we would need to understand the basics of a GSM network and the billing process first.

Let’s assign reference variables to the operator and the subscribers.

Telecom Operator = T

First Subscriber of Operator T = S1

Second Subscriber of the Same Operator = S2

**Scenario 1:** On-net Call

If a customer S1 of company T calls a friend S2 who has a subscription at the same company, the call flow will be as below:

Cellphone of subscriber S1 transmits to the nearest antenna or BTS (Base Transmitter Station) of company T. The BTS passes the call through the central computer or switch of company T, where the receiving party is recognized as being a customer of company T as well, and then the switch sends the call to the BTS where subscriber S2 has made contact fixed lines or be it glass fiber or such.  Subscriber S1 will get billed for the call. Since all the traffic is on the network of company T, they do not have to pay anyone. This is called an on-net call, where the calls are generated between customers of the same network.

**Scenario 2:** Off-Net Call

Fist Telecom Operator = T1

Second Telecom Operator = T2

First Subscriber of Operator T1 = S1

Second Subscriber of operator T2 = S2

Call flow for Subscriber S1 of operator T1 calling a friend S2 who has a subscription at the operator T2:

Cellphone of S1 transmits the network data to the nearest BTS of operator T1. The BTS passes the call through the switch of operator T1, where the receiving party is recognized as being a customer of operator T2. Switch T1 connects the call to the Switch of operator T2, that forward the call to the BTS of T2 where subscriber S2 made contact and then radio signals the call to the handset of S2. Customer S1 still gets billed for the call. As it is evident, now half of the call (the start) is on the network of T1 and the other half (the termination) of the call makes use of company T2's network. So operator T2 sends operator T1 a bill for making use of their network, which they have to maintain. This bill is called termination fee, which every telecommunications operator has to bear for off-net calls [20][21].

To bypass that termination fee, one fraudster can have a SIM box to terminate off-net traffic on the radio network of an operator. Only switch to switch traffic is charged for termination fee. With a SIM box, a fixed line call can be converted to mobile calls, using that box and activated SIM cards. The trick is that operator's offer buy off bundles for  on-net traffic, say for AED: 5.00 a month, you can call as much as you want to, customers of the same network or they have really low on-net tariffs like 5 fils per minute. They can do that since there are no costs involved for that company, as we saw in the scenario of On-net call, there are no costs for that company, as long as the calls are started and ended by their own customers [20].

**Scenario 3:** International Call

Telecom Operator in First Country = TC1

Telecom Operator in Second Country = TC2

Telecom Operator in Third Country = TC3

Telecom Operator in Fourth Country = TC4

First Subscriber of Operator TC1 = S1

Second Subscriber of operator TC4 = S2

Call flow for Subscriber S1 of operator TC1 calling a friend S2 who has a subscription at the operator TC4:

Cellphone of S1 transmits the network data to the nearest BTS of operator TC1. The BTS passes the call through the switch of operator TC1, where the receiving party is recognized as being a customer of operator TC4. Since the operator TC1 recognizes that this is an International call, it would need to transmit the call beyond the geographical limits of its own country. It is commercially and even technically not feasible for an operator to set up network all over the world. Hence, the operators will try to pass the call to its most feasible and nearest operator (TC2 in this case) who will in-turn pass the call to TC3 from another country. TC3 is expected to terminate the call in TC4's network. In technical terms, Switch TC1 connects the call to the Switch of operator TC2, which forwards the call to TC3. TC3 then connects to the switch of TC4 who passes on the call to the BTS of TC4 where subscriber S2 made contact and then radio signals the call to the handset of S2. Customer S1 still gets billed for the call.  As it is evident, now quarter of the call (the start) is on the network of T1 and the other half (the passing) of the call is through the network of TC2 and TC3. And finally the last quarter (the termination) is through company TC4's network. So operator TC4 sends operator TC3 a bill for making use of their network, TC3 sends a bill to TC2, and TC2 to TC1 which they have to maintain. Termination charges for international calls are comparatively very high as compared to local rates. This charge is called international termination fee, which every operator has to pay for international calls. Naturally, since TC1 has to indirectly bear all of the margins/costs. TC1 will pass on all the charges and their required profits to the subscriber S1. Hence, international calls are very costly [20].

To bypass that international termination fee, one fraudster can have a SIM box to terminate international traffic on the radio network of an operator. The fraudster (generally an international inter connect operator) will try to terminate the call in TC4's region via SIP/VoIP. With a SIM box you can convert VoIP calls to GSM calls, using that box and activated SIM cards. The trick is that, since the call enters through VoIP, and then it is converted to GSM through SIM box using local SIMs, it will reflect in TC4's network as local call. Hence, the interconnect operator would not need to pay the hefty international termination charges.

So the fraudsters get some SIM cards with a tariff of 5 fils per on-net call each for network TC4. He places it in the SIM box and then begins to advertise. Normally when another international operator wants to terminate a call to a customer of company TC4 they have to pay let's say AED: 2.00 per minute to company TC4. (Not the actual price, but for making it easy to understand) But they only have to pay that when traffic is connected through the switches. The fraudster then can approach company TC1 and tells them that he is able to terminate all their traffic towards customers of company TC4, but for only AED: 1.00 per minute. Company B agrees because that tariff is AED: 1.00 per minute less than if they handover the traffic via the interconnect operators. They now send their traffic to the SIM box of the fraudster that converts the traffic to mobile calls, just as if it was a giant handset with multiple SIM cards in it. Since the fraudster only has to pay the subscription fee and a tariff of 5 fils per minute while receiving AED: 1.00 per minute he is making a profit of 95 fils per minute, per SIM. He off course pays his bill right away because he wants his SIM cards open. Since the traffic is huge 5 fils per minute per SIM means, he earns a minimum of AED: 1,368.00 each day per SIM. So, if he has 10 SIMs, he is earning AED: 13,680.00 a day just by having that SIM box active.

Company TC4 then has a customer that has a monthly bill of let's say AED: 2,304.00; at first they are happy with such customer that pays his bills every month. But even though they are gaining AED: 2,304.00, they lose more than AED: 40,000.00 each month, because, if all that traffic was presented at their international switch they would have billed company TC3 AED: 43,200.00 for those calls

So the fraudsters get some SIM cards with a tariff of 5 fils per on-net call each for network TC4. He places it in the SIM box and then begins to advertise. Normally when another international operator wants to terminate a call to a customer of company TC4 they have to pay let's say AED: 2.00 per minute to company TC4. (Not the actual price, but for making it easy to understand) But they only have to pay that when traffic is connected through the switches. The fraudster then can approach company TC1 and tells them that he is able to terminate all their traffic towards customers of company TC4, but for only AED: 1.00 per minute. Company B agrees because that tariff is AED: 1.00 per minute less than if they handover the traffic via the interconnect operators. They now send their traffic to the SIM box of the fraudster that converts the traffic to mobile calls, just as if it was a giant handset with multiple SIM cards in it. Since the fraudster only has to pay the subscription fee and a tariff of 5 fils per minute while receiving AED: 1.00 per minute he is making a profit of 95 fils per minute, per SIM. He off course pays his bill right away because he wants his SIM cards open. Since the traffic is huge 5 fils per minute per SIM means, he earns a minimum of AED: 1,368.00 each day per SIM. So, if he has 10 SIMs, he is earning AED: 13,680.00 a day just by having that SIM box active.

Company TC4 then has a customer that has a monthly bill of let's say AED: 2,304.00; at first they are happy with such customer that pays his bills every month. But even though they are gaining AED: 2,304.00, they lose more than AED: 40,000.00 each month, because, if all that traffic was presented at their international switch they would have billed company TC3 AED: 43,200.00 for those calls. This loss is just through one SIM. It grows exponentially if there are multiple SIMs involved.

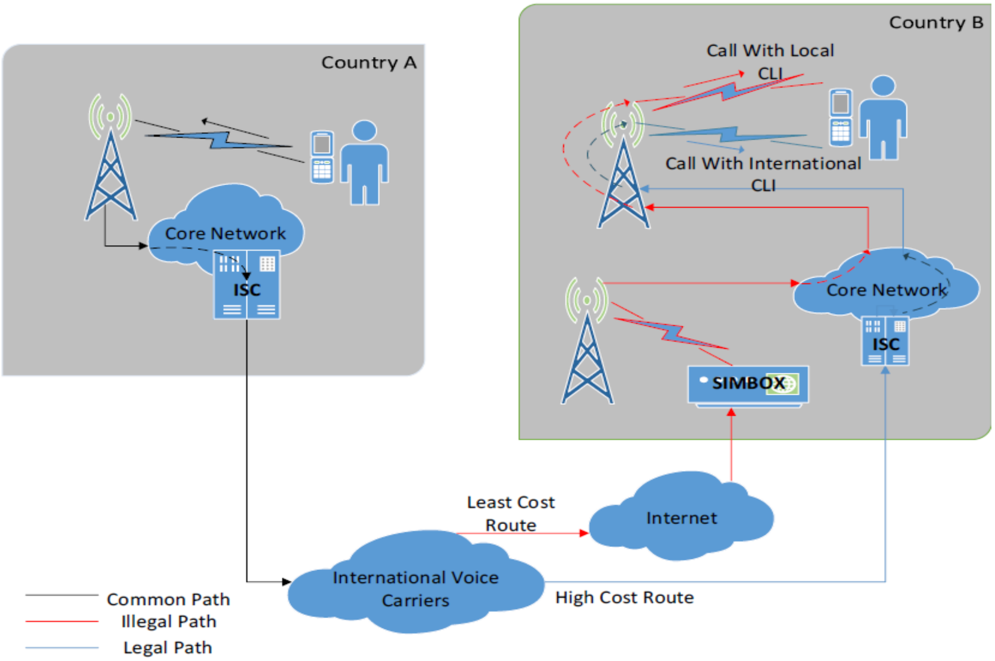


Figure 3.1: On-Net Bypass

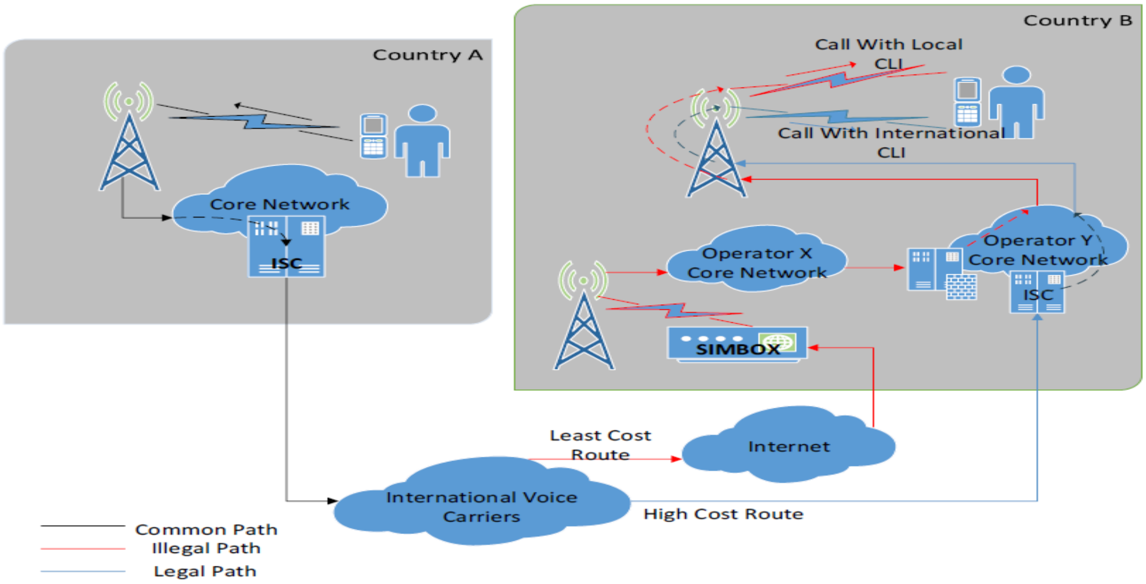


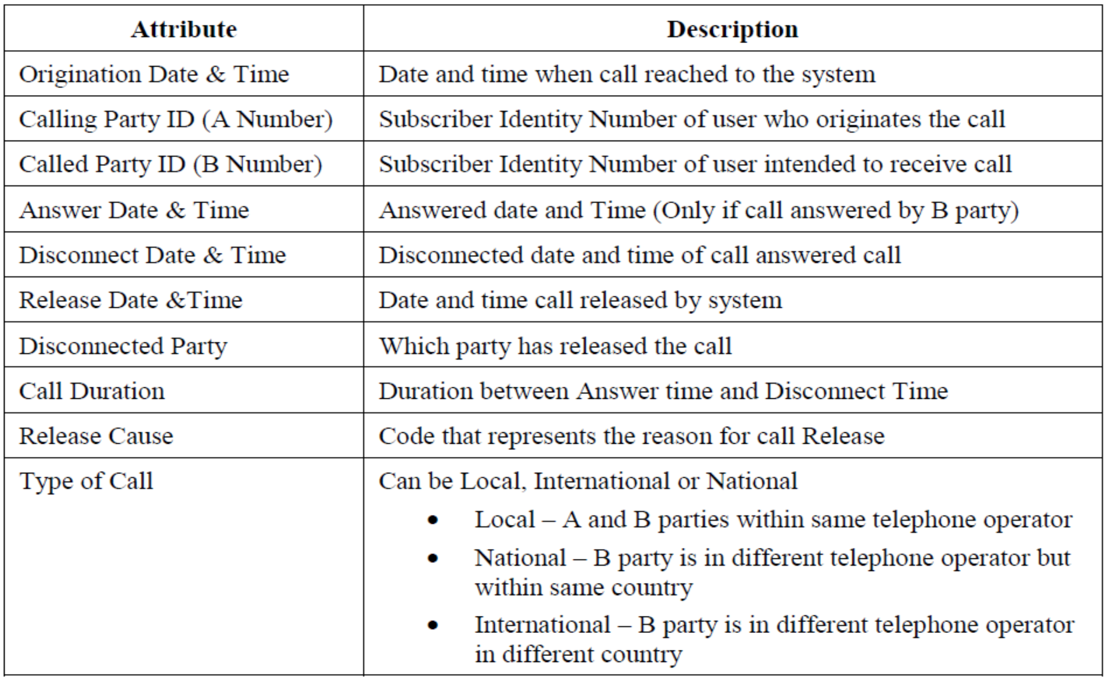
Figure 3.2: Off-Net Bypass

## 3.2.Call Detail Records (CDR)

Call Detail Record (CDR) is the call data record generated by telecommunication operator's switches. It consists of details that are specific to a single instance of a phone call or other communication transaction which is handled by that switch. The details and level of information inside the CDR varies depending on functionality of the network switch. The decoder files or functional instruction manuals of the network switch typically specify how to extract the information in CDR. CDRs are inherently used for billing purposes. In addition, it is used for troubleshooting, measuring Quality of Service (QoS), fraud detection, gain Business Intelligence (BI) and forensic investigations.

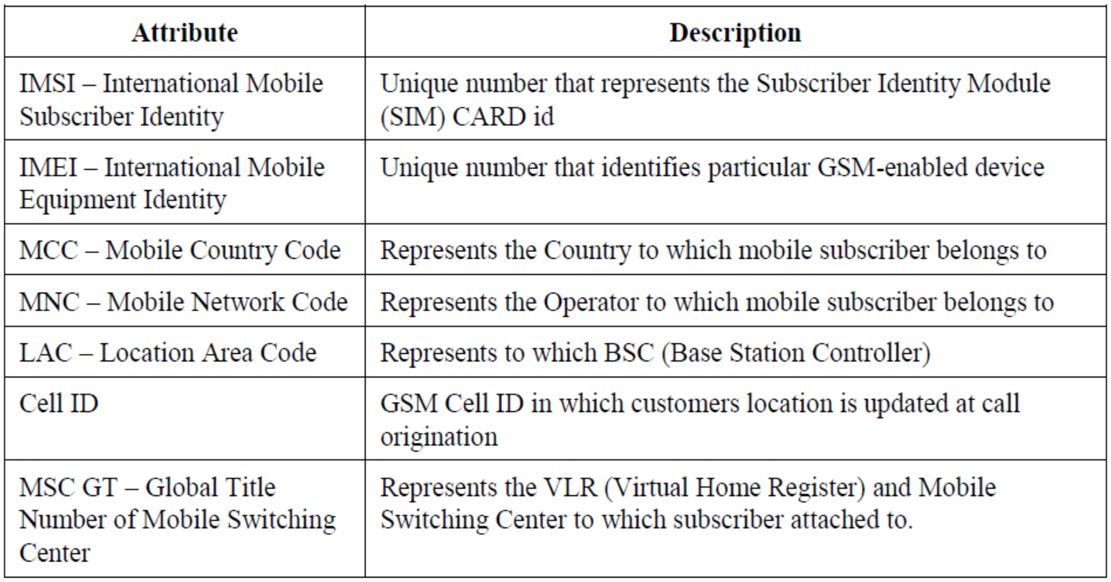
The most common details of a CDR generated for a voice call is listed in Table 3.1. In addition to these details, Global System for Mobile (GSM) telephone call can contain an additional detail that represents the information about subscriber; mobile handset and its location (see Table 3.2). CDR with these details is generated at  Class-5 switches to which actual subscribers are directly connected. Corresponding values for originating and destination parties are generated separately at Class-5 switches to which subscribers are attached. Table 3.3 list further details that are recorded at Class-4 switches to which only other switches are connected in addition to basic details mentioned in Table 3.1. These details are helpful to distinguish telecom operators in signaling interconnection. Tables 3.1 to 3.3 list only the essential details of a voice call that are helpful in this research. There are many other details that represent the QoS parameters, protocol-specific details and network switch-specific details are included in CDR. Short Message Service (SMS) and Mobile Data (GPRS, EDGE, UTMS, HSPA and LTE) technology transactions also generate CDRs [20].

Table 3.1 Common Attributes in CDR



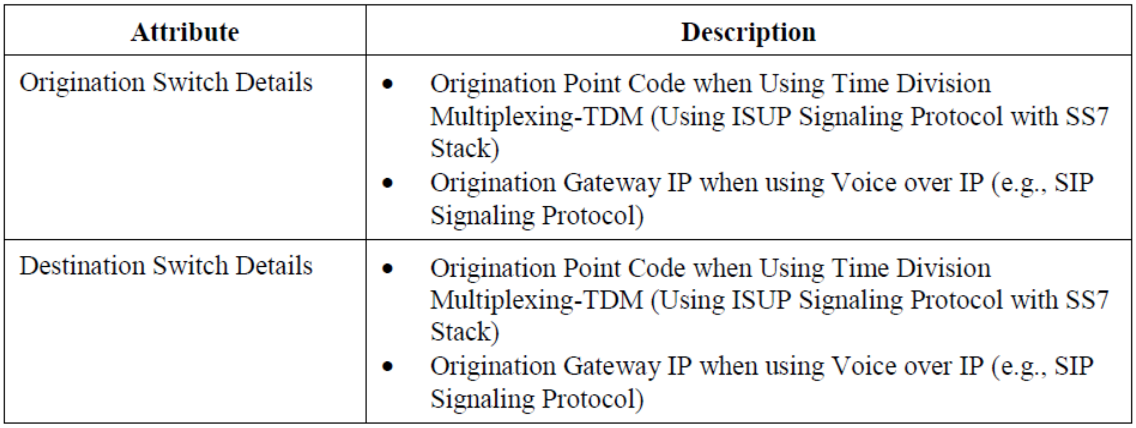
| Attribute | Description |
|---|---|
| Origination Date & Time | Date and time when call reached to the system |
| Calling Party ID (A Number) | Subscriber Identity Number of user who originates the call |
| Called Party ID (B Number) | Subscriber Identity Number of user intended to receive call |
| Answer Date & Time | Answered date and Time (Only if call answered by B party) |
| Disconnect Date & Time | Disconnected date and time of call answered call |
| Release Date &Time | Date and time call released by system |
| Disconnected Party | Which party has released the call |
| Call Duration | Duration between Answer time and Disconnect Time |
| Release Cause | Code that represents the reason for call Release |
| Type of Call | Can be Local, International or National<br>• Local – A and B parties within same telephone operator<br>• National – B party is in different telephone operator but within same country<br>• International – B party is in different telephone operator in different country |

Table 3.2 Specific Fields in CDRs from Class 5 Switches



| Attribute | Description |
|---|---|
| IMSI – International Mobile Subscriber Identity | Unique number that represents the Subscriber Identity Module (SIM) CARD id |
| IMEI – International Mobile Equipment Identity | Unique number that identifies particular GSM-enabled device |
| MCC – Mobile Country Code | Represents the Country to which mobile subscriber belongs to |
| MNC – Mobile Network Code | Represents the Operator to which mobile subscriber belongs to |
| LAC – Location Area Code | Represents to which BSC (Base Station Controller) |
| Cell ID | GSM Cell ID in which customers location is updated at call origination |
| MSC GT – Global Title Number of Mobile Switching Center | Represents the VLR (Virtual Home Register) and Mobile Switching Center to which subscriber attached to. |

Table 3.3 Specific Fields in CDRs from Class 4 Switches



| Attribute | Description |
|---|---|
| Origination Switch Details | • Origination Point Code when Using Time Division Multiplexing-TDM (Using ISUP Signaling Protocol with SS7 Stack)<br>• Origination Gateway IP when using Voice over IP (e.g., SIP Signaling Protocol) |
| Destination Switch Details | • Origination Point Code when Using Time Division Multiplexing-TDM (Using ISUP Signaling Protocol with SS7 Stack)<br>• Origination Gateway IP when using Voice over IP (e.g., SIP Signaling Protocol) |

Traditional BI techniques and fraud detections only focus about aggregate values like averages, summations, and counts. However, CDR contains the interesting patterns that reflect customer behavior. Identification of such behavior allows having revenue and margining advantage by recognizing opportunities, as well as preventing risks by unmasking threats. Rapid growth of Data Science and Machine Learning using artificial intelligence has enhanced the CDR based pattern recognition. Fraud detection is one of the major applications of CDR-based pattern recognition [20][21].

Chapter # 4

# 4. Experimental Analysis

## 4.1.SIM box Fraud Management

In order to combat SIM box or Bypass fraud, there has to be a systematic approach towards it. It needs to be driven by a process. The process to stop an ongoing fraud and reach to a level of its mitigation is known as Fraud Management.

Investigation

This is basically a technical evaluation into the threats and issues that a network faces. It also consists of analyzing marketing activities, such as providing free minutes on SIM cards for local calls or partnerships with OTT providers, as these could provide an insight to a loophole which is being exploited by fraudsters.

Detection

One of the most critical phase of Fraud Management. This is one of the phase which is extremely time dependent as well. This is essentially the primary focus of telecom fraud departments. However, the fluid nature of revenue loss means that this is more than a static activity.

Prevention

The most complex phase in anti-fraud management is ongoing protection against potential revenue and margin loss means that telecom networks can ensure they are generating pr(Vague) rather protecting all the revenues they are entitled to on an ongoing basis. This phase requires complete due diligence, exact understanding of the nature of fraud and needs to encompass all issues, new and old, to ensure minimum loss of revenue and margin [20].

Basic Principles and Requirements:

To detect fraud operators, need to have or act upon the following:

a. The operator therefore needs to have a fraud management structure that ensures that they focus on the greatest potential financial loss due to dishonesty.

b. The operator need to have a structure in place to ultimately limit the total exposure to fraud across the business and not isolated to customer airtime loss.

c. The operator need to have a clear idea of what fraud management costs, fraud losses and a formula to calculate savings and recoveries.

d. Operator got to be actively protective their customers moreover as their own network.

e. Operator needs to be progressive and forward thinking in their approach to detecting, investigating, controlling and ultimately preventing fraud [20].

Operator Fraud Team Needs to Understand:

a. What is actually being targeted and by who, what are the operator up against?

b. Understand the local culture and demography, where is the operator most vulnerable or exposed.

c. Confirm the present skills and experience, do the operator have the correct skill sets and resources.

d. Take into account any legal/regulatory legislation constraints over providing service, do operator know what they can and cannot do to prevent fraud.

e. Appreciate the categories of product or service provided.

f. Be aware of local or internationally organized crime groups, who is operator taking on?

g. What information sources are currently available to assist in fraud monitoring, detection and prevention?

h. What are the common fraud indicators that operator are using to trigger alerts, reports of illegal activity?

i. Have the operator been able to establish clear lines of communication throughout the business?

j. Development of reporting capabilities. What is in place?

k. Procedural enhancements. Who owns them and ensures compliance?

l. Education and awareness, what program is in place internally?

There are Some (KFIs) in Preventing and Detecting Fraud that should be mentioned, but are not limited to:

a. Undelivered invoices mail.

b. Returned/declined payments.

c. Un-contactable customers.

d. Changes in address information immediately after registration.

e. Roaming with very little or no home network usage.

f. International calls.

g. High usage - multiple IMEIs used.

h. Multiple accounts/SIMs [20].

In order to have a clear understanding of Fraud management, let us deep dive into its various phases:

## 4.2.Detection Phase:

Fraud detection refers to the effort to detect illegitimate practice of a telecommunication network thru detecting and informing fraud as quickly as possible once it has been committed (as cited in Nelson, 2009). Detection of SIM cards used for SIM box fraud is a challenging task for mobile operators. There are two major approaches.

- First approach is actively originating calls to the target network via test units installed in several parts of the world and scan the CLI of those calls. The device to generate calls is known as TCG (Test Call Generator). Even though this process detects SIM box numbers in real time, it is not capable of capturing majority of numbers.
- Second approach is CDR based analysis. CDRs are loaded into relational database and queries are used to identify fraudulent numbers. Rich set of attributes are needed to derive by summarizing CDR to achieve effective detection. In context of On-net bypass, operator has more details including Location and Owner Information. But detection of Off-net bypass has to be performed with limited details and strong pattern mining techniques are required.

Efficient detection process must consist minimum false positive (i.e., detect genuine customers as fraudulent) and false negative (i.e., classify fraudulent numbers as genuine) values. Moreover, number of attempts made by fraudulent SIM card before detection is another important factor. If this value is very high, fraudster can cover the cost before disconnection of SIM card. Because the Telecom industry is highly competitive, in most of the countries fraudsters can buy new SIM cards at a very little cost. Hence, traditional analysis methods fail here as that is based on past CDR analysis. Therefore, when operator disconnects, fraudsters use new set of SIM cards as they can cover profit margin. This process continues, and actual task of detection process become damage control function. However, anomaly detection remains one of the stable solutions for detection for SIM box fraud [21][22].

Limitation with traditional Fraud Management Systems and upcoming detection techniques: Traditional systems make detections by generating a set of features or aggregate values by querying static data over a large time window and make decisions based on such values. This is time consuming and by that time a fraudster can make number of successful calls before being detected. Scenarios discussed above highlight the need of a real-time CDR analysis tool, which can snoop CDR streams and generate a recent or real-time view of the telecommunication network. Also, resultant recent view should be able to integrate with past data and produce complete state of the network at a given instance with minimum latency. System should be easily customizable according to varying requirements of the operators. Lower development cost will be an extra advantage. Such a system allows operators to gain maximum advantage by exploiting timeliness of detected events and save significant amount of revenue by minimizing fraudulent activities. The solutions available in research literature lacks real-time features due to unsuitability of traditional database reliant store first process then approach for latency sensitive applications, large time windows for feature generation, shallow feature set, less awareness about context, and in ability to detect complex patterns in CDR.

Commercial systems are also based on databases uses a proprietary feature set, focuses on specific use case, and are usually expensive. Therefore, operators are unable to afford the cost of such specialized systems.

Therefore, the problem to be addressed by this research can be stated as follows:

How to detect fraudulent call patterns in real-time using CDR?

Our primary focus is to use the power of complex events to support real-time decision making in detection of grey callers and fraudulent activities which involve extreme usage [17].

Anomaly Detection:

Anomaly detection is amongst one in every area of applying knowledge for information mining and consists techniques of knowledge analysis that permit detection patterns of probability in a data set and so defines patterns of knowledge that are take into account traditional or abnormal. Associate degree of anomaly (or outlier) is an instance of knowledge that doesn't belong to any pattern predefined as traditional or really belong to a pattern thought of abnormal. In summary, the instances happiness to the pattern take into account traditional square measure thought of traditional and people UN agency don't belong square measure thought of abnormal. If those patterns square measure well outlined at associate degree early stage and therefore the model to do the anomaly detection is well engineered, this method may also be utilized in the prediction (prevention) of anomalies. Some example of anomaly detection fields are:

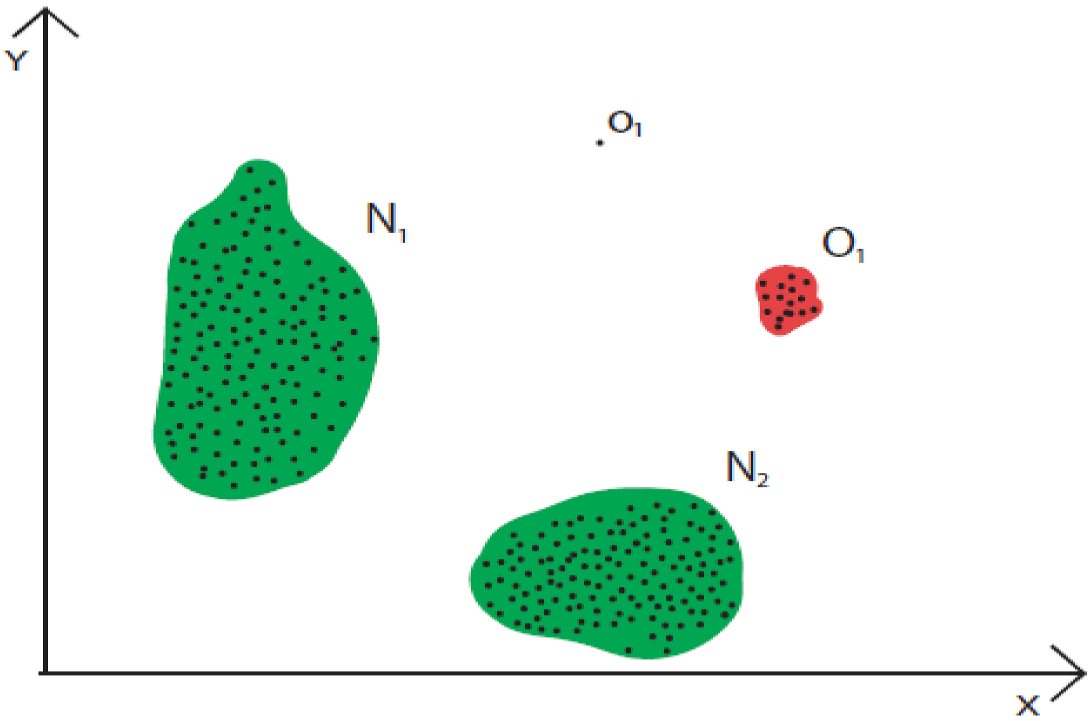
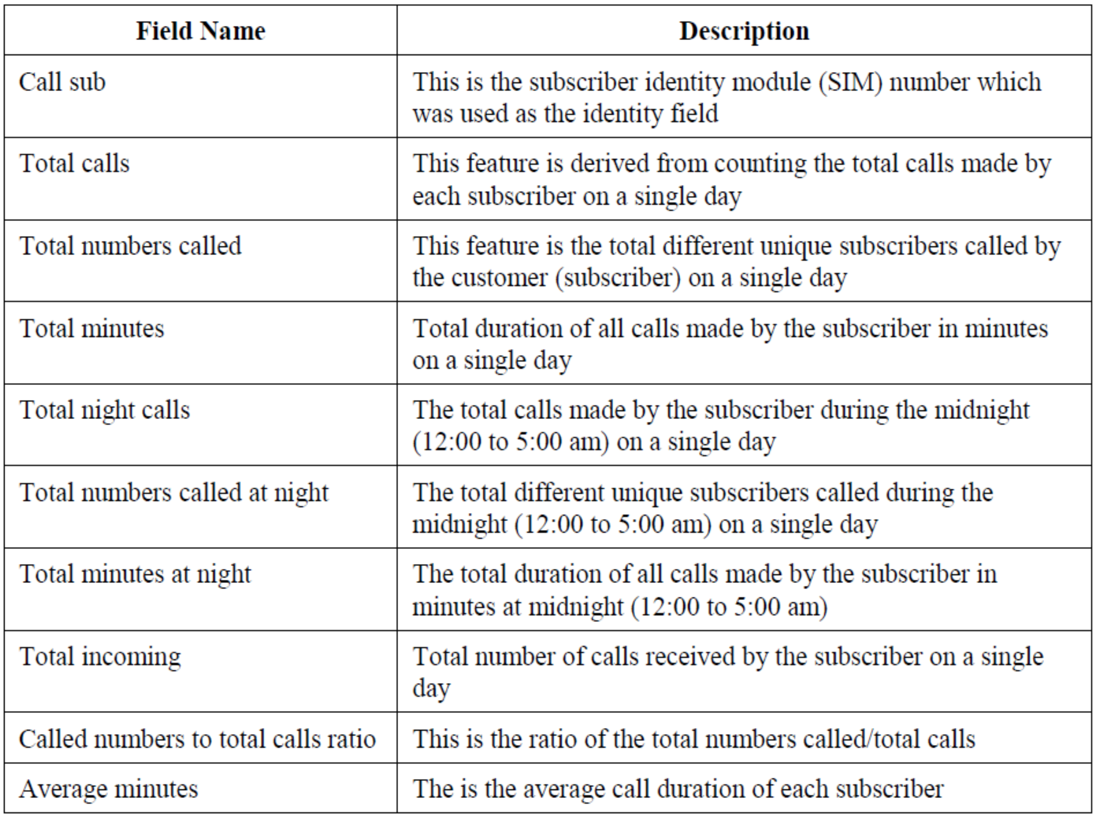


Figure 4: Outliers example in a dimensional data set

The process to build associate degree anomaly detection model consists by 2 stage, the training stage and test stage. The primary stage is that the coaching stage and is employed to make the model used for detection. The second stage is the testing stage and is employed to gauge the performance of the model. For every of this stage a knowledge set should be divided in 2 parts one for coaching and alternative for testing. These parts are generally 70% and 30% of the initial information set for coaching and testing severally. In a very visual analysis, the anomaly detection of a little set of knowledge wouldn't be sophisticated. The matter seems once the number of knowledge grows exponentially, and therefore the visual analysis starts to be less precise. Within the figure we are able to observe 2 patterns of knowledge within the cluster N1 e N2, one smaller group O1 associate degreed an instance o1. Within the case of anomaly detection, the instance o1 is without doubt and outlier, and therefore the teams N1 e N2 are not outliers, however, the smaller cluster O1 may be either a bunch of outliers or simply another cluster of traditional information [17].

Artificial Neural Network:

Artificial Neural telecommunications network is a supervised learning method with Multi-Layer Perception (MLP) as classifier. ANN is used because of its generalizing capabilities, ability to learn complex patterns and trends within noisy data and better performance records in this domain. This system derives nine details using CDR in dataset and calculated corresponding values for each calling subscriber. Table 4.1 describes the details set which was used for SIM box detection [13].



| Field Name | Description |
|---|---|
| Call sub | This is the subscriber identity module (SIM) number which was used as the identity field |
| Total calls | This feature is derived from counting the total calls made by each subscriber on a single day |
| Total numbers called | This feature is the total different unique subscribers called by the customer (subscriber) on a single day |
| Total minutes | Total duration of all calls made by the subscriber in minutes on a single day |
| Total night calls | The total calls made by the subscriber during the midnight (12:00 to 5:00 am) on a single day |
| Total numbers called at night | The total different unique subscribers called during the midnight (12:00 to 5:00 am) on a single day |
| Total minutes at night | The total duration of all calls made by the subscriber in minutes at midnight (12:00 to 5:00 am) |
| Total incoming | Total number of calls received by the subscriber on a single day |
| Called numbers to total calls ratio | This is the ratio of the total numbers called/total calls |
| Average minutes | The is the average call duration of each subscriber |

Table 4.1: Details Set Used in ANN Based Approach

In Multi-Layer Perception the ANN consists of multiple layers of computational units (neurons), connected in feed forward way. So these neurons can be categorized into input, output and hidden neurons based on layer. Connections between neurons known as edges and which has associated weights. Neurons are only connected to subsequent layers but not to the neurons in same layer. Weighted sum of multiple inputs was taken and it is fed into nonlinear activation function called sigmoid function to generate single output of neuron. This output value was passed as input to the connected nodes in next layer. Back Propagation algorithm was used to train the ANN to minimize the training errors. This algorithm calculates error value for each neuron output (difference between output of neuron and actual value) and weights of telecommunications network edges are continuously adjusted in a way that minimize errors.

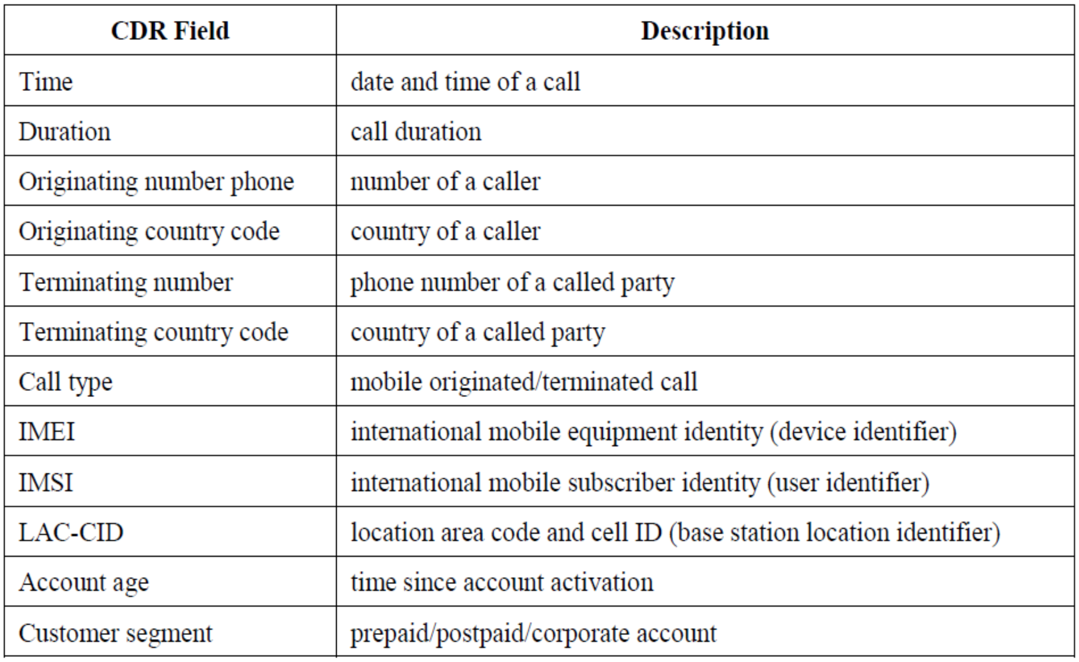
The dataset was divided to ten subsets and average error value was calculated by running experiment for each subset in turn using same model. While using one subset for testing remaining nine was combined and used for training. This is known as 10-fold cross validation. Authors have changed four parameters to find optimum ANN with highest accuracy. Number of hidden layers, number of neurons per hidden layer, learning rate and momentum are those parameters. Learning rate represents the speed at which the ANN arrives at the global minimum value for Sum Square Error (SSE). The momentum parameter represents the rate at which the ANN approaches neighborhood of optimality at early stages of algorithm. Both momentum and learning rate ranges its values between zero and one [13]. Altogether they have experimented 240 neural telecommunications networks and compared them in terms of prediction accuracy, generalization error, and time taken to build the model, precision, and recall. So they have unmasked the optimum ANN. They have identified that very high learning rates and momentum rate significantly degrade classification accuracy as it leads algorithm to overshoot the optimal configuration. A maximum accuracy of 98.7% by using lower momentum value (0.3) and moderately higher Learning rate (0.6) and using two hidden layers has been obtained. Learning and classification performed in about 17 seconds with considerably lower false positive and false negative detections.

Two years later, Support Vector Machine (SVM) based approach was released for the same dataset and compared its results with the ANN based approach. 10-fold cross validation technique has been utilized while using same details. Because this is a binary classification problem there are only two classes in the training data, in this case hyperplane is a line. But authors had to use nonlinear SVM as nonlinear curved line was required to separate boundaries. So they have used kernel functions instead of inner product and evaluated performance separately for polynomial kernels, radial basis function kernels, and linear kernels. Additionally, they have taken measurements by changing the C penalty parameter which effectively controls amount of error willing to afford in the training data. Altogether they have evaluated 40 SVM models in terms of accuracy, generalization error, and time taken to build the model, precision and recall using 10-fold cross validation method. Moreover, they have evaluated performance of both methods by changing training and testing set sizes. Above 98.5% accuracy was achieved in both ANN and SVM based approaches. Finally, they have located best SVM model and found that it performs better than ANN because of significant reduction in running time.

Even though high accuracies and lower running time were achieved in both the cases, sustainability of this approach in practical scenario is questionable due to many reasons. When we consider the dataset, its size is much smaller than typical mobile telecommunications network operator. Dataset contains CDRs of 234,324 calls made by 6,415 subscribers over two months. However, most of mobile operators, especially in Asian countries, have more than 5 million subscribers and generate more than 20 million CDRs per day. Also, they have considered CDRs from one cell id only and ratio between legitimate to fraudulent subscribers is approximately 2:1.

But in real cases more than 20,000 cell IDs need to be considered and percentage of SIM box numbers out of total customer base is very low. So we can conclude that cardinality of dataset is inferior to actual cases. Therefore, this solution's ability to achieve given performances in practical environment is not tested in. Also, CDR for two months has been considered when calculating details. But the actual requirement in SIM box detection needs to perform as early as possible. Fraudsters can cover the cost of buying new SIMs, if they successfully operate over a few hours. So there is no point of performing calculations within a few seconds as long time window is used for details calculation. Additionally, scalability of these methods with large datasets was not evaluated in both the papers.

Table 4.2 shows the attributes of the CDR used, where they have accounted important details including location details, device details and corresponding customer segment of calling party which were absent in the previous cases [13][18].



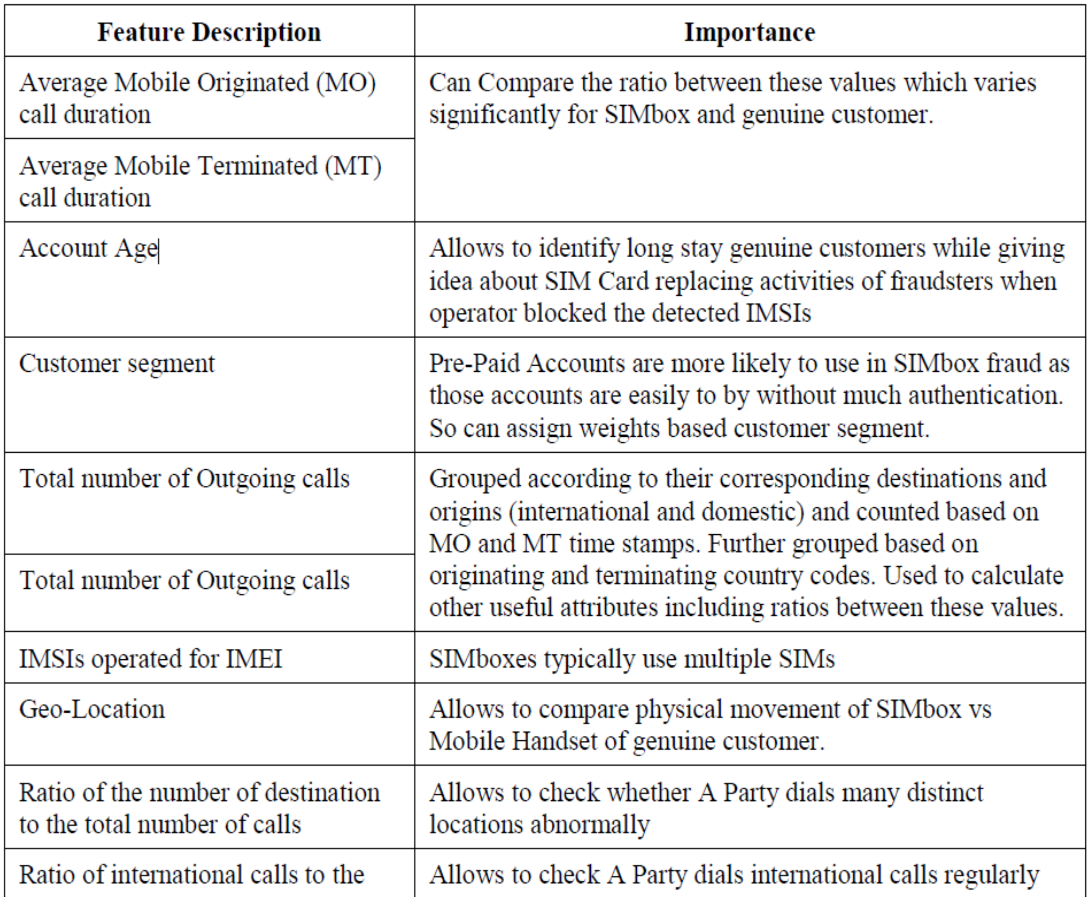
| CDR Field | Description |
|---|---|
| Time | date and time of a call |
| Duration | call duration |
| Originating number phone | number of a caller |
| Originating country code | country of a caller |
| Terminating number | phone number of a called party |
| Terminating country code | country of a called party |
| Call type | mobile originated/terminated call |
| IMEI | international mobile equipment identity (device identifier) |
| IMSI | international mobile subscriber identity (user identifier) |
| LAC-CID | location area code and cell ID (base station location identifier) |
| Account age | time since account activation |
| Customer segment | prepaid/postpaid/corporate account |

Table 4.2 CDR Fields considered in Classification based appro

Those parameters are directly used as details as described below. Using CDR with a rich set of attributes can be identified as a positive step. By considering IMEI details it is possible to block the confirmed IMEIs or detect new SIM cards that are inserted into a SIM box with a particular IMEI. But the detection logic cannot too much depend on that since advanced SIM boxes allow changing IMEIs.

Authors have derived 48 details using mentioned CDR attributes in Table 4.2. It is important to note that the details set are per IMEI basis and they have targeted to identify SIM box rather than SIM cards used for IMEI. Even though authors did not give full description about whole details set, details mentioned in Table 2.6 were highlighted. Based on these details, authors have characterized SIM box behavior. Authors have demonstrated that SIM boxes have fairly static physical behavior as they connect to a very small number of nearby base stations while a genuine customer is dynamic and moved across many base stations. This is obvious but important observation which was not presented in the previous cases. LAC-CID attribute makes this possible. Because advanced SIM boxes are capable of Location Swapping location information need to be used with care.

It has been demonstrated SIM boxes have very few Mobile Terminated (MT) calls and generate a huge number of Mobile Originated (MO) calls while genuine customers have same number of initiated and received calls. So usefulness of outgoing calls to incoming calls ratio details can be highlighted. Also, authors have presented that SIM boxes has very small duration of MT calls over the time than actual customer. Since SIM box is a machine it cannot lively answer the MT call and maintain a conversation. So it just drops the call or for forwards the call to announcement. That is the reason for this observation. Another observation is SIM box operators regularly deploy a set of new SIM cards once operator has detected and deactivated existing fraudulent SIM cards. They also tried to filter out the device called telecommunications network Probe which is used for quality measurements [13][18].



| Feature Description | Importance |
|---|---|
| Average Mobile Originated (MO) call duration | Can Compare the ratio between these values which varies significantly for SIMbox and genuine customer. |
| Average Mobile Terminated (MT) call duration | |
| Account Age | Allows to identify long stay genuine customers while giving idea about SIM Card replacing activities of fraudsters when operator blocked the detected IMSIs |
| Customer segment | Pre-Paid Accounts are more likely to use in SIMbox fraud as those accounts are easily to by without much authentication. So can assign weights based customer segment. |
| Total number of Outgoing calls | Grouped according to their corresponding destinations and origins (international and domestic) and counted based on MO and MT time stamps. Further grouped based on originating and terminating country codes. Used to calculate other useful attributes including ratios between these values. |
| Total number of Outgoing calls | |
| IMSIs operated for IMEI | SIMboxes typically use multiple SIMs |
| Geo-Location | Allows to compare physical movement of SIMbox vs Mobile Handset of genuine customer. |
| Ratio of the number of destination to the total number of calls | Allows to check whether A Party dials many distinct locations abnormally |
| Ratio of international calls to the | Allows to check A Party dials international calls regularly |

Table 4.3 Feature set used in Classification based approach

Before we look into classification algorithm it is important to highlight several concerns in details generation. Details calculation per IMEI basis has its own set of problems. Advanced SIM boxes can replace IMEIs with dummy values or other genuine IMEIs. So blocking IMEI numbers may block some genuine customers. Additionally, IMEI to MSISDN mapping give false values. Since choice of device is customer's right, operator has no control on IMEI. So applicability of this system directly in practical environment can be questioned. Better option is details calculation per MSISDN basis.

Mobile operators disconnect SIM box connections once they have detected it. So fraudsters insert many new SIM cards to SIM box frequently. Also, SIM boxes do not move the location on regular basis and attached to limited set of cell IDs. So, there is a high probability that calls originating from those cell IDs to be grey calls. Location details give sense about that. But researchers have not mentioned that they have identified such cell IDs and not presented cell ID wise SIM box distribution.

When we consider dataset, majority of the details calculated for data collected over one-week period from tier-1 cellular operator in United Arab Emirates. So dataset is considerably larger than the previous cases. But one-week period is still higher as operator loses considerable amount of revenue over that period. This dataset contained CDRs of 93,500 subscriber accounts and 500 (or 0.5%) out that is SIM boxes. Since SIM box user CDRs are mixed inside considerable amount of genuine user's data, this dataset can be considered as good mixing of SIM box users and Normal user's data than previous case's details like IMSIs operated per IMEI is calculated for data collected over five months. 66% of data labeled accounts were used as the training set while remaining 34% was used for testing. Like in previous cases, cross validation techniques were not used to increase the accuracy [17].

Classification algorithm which was used in this research is a linear combination of three classifiers associated with weight coefficients. Alternating decision tree, functional tree and random forest are the three classifiers. An alternating decision tree is derived by the combination of single question decision trees which has two types of nodes known as decision nodes and predictor nodes. Decision node contains details test condition while predicate node has single real number corresponding to negative or positive weight. Root and leaves are predictor nodes and decision node lies between two predictor nodes. So input records passed through multiple paths and output value is produced based on sign of the weighted sum. Based on training data Boosting method continuously re-weights the values in predicate node. So ultimate function of boosting method is a combination of week classifiers into strong classifiers while focusing on majority as well as outliers in the training dataset. In Random forest, multiple decision trees are generated using subset of details and prediction output is generated based majority rule. Functional tree makes decision tests for combination of the original details at decision nodes, leaf nodes, or both nodes and leaves unlike in standard decision tree in which decision test is done for single details at decision nodes.

The predictions made by Random forest algorithm provided best false positive rate of 0.0001 while offering comparably higher false negative rate of 0.16. Functional tree algorithm had done predictions with lowest false negative rate of 0.07 but false positive rate was 0.0007 which is comparably higher than value obtained for Random forest. Therefore, to increase the accuracy, multiple regression technique was used by considering prediction output of three classifiers as predictor variables and its linear combination as criterion variable. Prediction error was defined as difference between predicted data label and actual data label.

Regression weight coefficients are calculated by locating least value of square error for training dataset. Finally, they have unmasked optimum value for three weight coefficients and classified test data using novel classifier. They were able to minimize the false positive rate up to 0.0001 and false negative rate up to 0.09 and achieve 99.95% accuracy which was higher than previous cases [4], [5]. To enhance practical usage, they have filtered out accounts with less than 10 IMSIs per IMEI, probing devices and well known legitimate accounts and remaining 0.02% of accounts were used for details generation.

Even though authors gained high accuracy they have not mentioned running time of algorithm and processing requirements. To reduce computational resources, they simply used manual filtering which reduces the size of dataset. But manual filtering is not always possible. Therefore, scalability and running time of this method can be questioned [17].

Critical evaluation of above approaches reveals many areas that were not focused and thus opens up new research topics. Those facts are summarized below:

- Existing solution have only targeted the accuracy and running time of classification algorithm while considering large time window for details calculation. Those solutions did not interpret SIM box detection as time sensitive operation. So these solutions are incapable of preventing financial losses as fraudsters can make profits easily by operating safely within that time window before disconnection. Therefore, to make near real-time detections rich set of details should be generated for short-time window and classification algorithm need to be optimized according to that.
- These approaches are only capable of detecting On-net Bypass and did not pay any attention for Off-net Bypass. Both make similar kind of financial losses for many telecom operators. Details like Location, IMEI, IMSI, and Account type details may not be available for Off-

net SIM box numbers. So a rich set of novel details with additional measures is required to detect Off-net Bypass.

- Details are generated based on calling party behavior only. But by considering called party behavior a valuable set of attributes can be derived. For example, counting the subset of called party numbers which has received IDD calls and belongs to a set of called party numbers dialed by the considered calling party will be valuable details in context of grey call detection.
- Previous cases have targeted CDR data stored in static databases. But CDRs records are generated in real time and flows as streams of data. Therefore, to gain maximum advantage, a new mechanism that is capable of directly processing the multiple streams is required. Also, that mechanism should support multiple CDR streams generated by Telco nodes, as well as some static data simultaneously.
- A typical mobile service provider has a customer base of more than ten million. So data streams with very high transaction rate are generated at Telco nodes. So highly scalable and fast details generation method is required to cope up with current industry's requirements.
- Any of the discussed methods are not capable of identifying complex events masked inside CDR. Detection of complex events allow to exploit maximum situational value. This can be effectively used for SIM box fraud detection [17].

## 4.3.Investigation Phase:

During this phase, a member from the Fraud Management Team would be responsible to analyze the data and determine based on the details if it is a potential fraud case. There are few techniques an analyst may deploy to conclude his investigation

There are three common approaches for fraud investigation as below:

### 4.3.1. Decision Generation Analytics;

By analysis of the traffic that is truly received to see if this international traffic is on-net traffic or traffic from another native network.

### 4.3.2. Decision Information Analytics;

By analysis of call data records for each SIM card of telecommunication company. The subsequent Criteria are often employed in this approach in Detection method:

- Count and magnitude relation comparison of imply outgoing/incoming calls and off-net/on-net calls.
- Exclusion of calls to "allowed numbers" (e.g. client service)
- Diversity of calls, as well as total diversity and on-net/off-net diversity
- Usage of non-voice services (SMS, GPRS sessions, etc.)
- Analysis of use of number of cell sites
- Calls throughout irregular hours
- Suspicious cells, as well as automatic suspicious cell locater.

### 4.3.3. Hybrid Analysis;

By victimization each approaches, decision information analytics and decision generation analytics, so as to develop a lot of expeditiously detection system.

Various types of alert triggers can be set in the telecommunication fraud detection systems. The criteria can be set based on behavioral analysis of the fraudsters for previous cases. It can also be set based on call details and user calling pattern and any deviation from the same can be used as an alert trigger for the analyst team to investigate further [14].

Prevention Phase:

It is very important to mitigate the fraud entirely. This will help in protecting the organization's revenues and consequently improve the telecom operator's profit margins.

Prevention Measures:

a. The Fraud Team can use a number of techniques and tools in order to effectively detect, analyze, monitor, prevent and report on fraud.

b. All fraudulent attacks identified should be used to prevent future frauds if the fraud team are to reduce the company's exposure.

c. It is preferable that system controls are used instead of procedural controls in order to prevent abuse of service.

d. Analysis and identification of the techniques used and an understanding of the methodology will allow the fraud manager to determine the most effective prevention strategy to be deployed.

e. There must be an understanding as to the commercial implications to the business when developing preventative measures.

f. There are a unit variety of various prevention measures:

i. Policy related.

ii. Process and procedural related.

iii. Person related.

iv. IT system related.

v. Network system related.

vi. Physical security related.

vii. Combinations of the above [14].

Focus on Loss Prevention due to Fraud:

As bigger confidence is gained in detection ability, operators must move towards increased focus on prevention.

This must specialize in the below:

a. Involvement in the product and services development cycle - assessing the risk.

b. Root cause analysis - determining the problem, gaps or inherent weaknesses and defining the required controls.

c. Evaluation of existing processes for loss exposures - is the risk technical, procedural or people based?

d. Uncover - identify and investigate potential issues.

e. Discover - analyze, quantify and qualify issues identified.

f. Recover - implement corrective initiatives to resolve problems.

g. "Prevention is better than detection".

Blacklist/ Hotlist Management for critical and unique attributes in CDR:

a. Following the detection of fraud, distinguishing attributes of the case should be populated into a blacklist/hotlist for reference when new potential fraud cases occur.

b. Attributes such as suspected B numbers, cell sites, names, addresses, IMEI, countries etc. can be populated into the FMS to enable rapid type alarms to be generated once a subscriber matches hot listed data.

c. Many operators implement a blacklist of known fraudulent details within the network and link this to the activation process to prevent new subscribers/SIMs from being activated using known fraudulent details [14].

Case Management:

a. All previous fraud cases should be stored in order to utilize the information for intelligence purposes and to enable proactive detection of fraud in the future where fraudsters use similar names, ID numbers, addresses and calling profiles.

b. The storing of cases will also enable fraud losses to be recorded to facilitate financial reporting to management and CFO on losses. This can be attained by recording all cases in the FMS, if operators have executed one. (Changing from the manual recording practices).

c. Each fraud case when detected and confirmed must have a file created with an index holding the case reference number; the MSISDN, name, address, fraud type/source and some remarks to assist the fraud control and operations team to understand what each case is about.

d. A brief final written report should be completed detailing any corrective actions taken by the fraud control and operations team or identifying areas within the business where exposure was identified.

e. Feedback must be received as to whether the recommendations were acted upon or not and only then should a case be closed.

f. Advice should always be sought from legal as well as system owners in respect of retaining evidence for fraud case prosecution.

Reporting and Quantification:

a. The fraud control and operations team will still be responsible for providing senior management with reports and will be required to accurately measure fraudulent activity within the business.

b. To facilitate this, the fraud control and operations team should maintain and manage various daily and monthly statistics, which will be used by the fraud manager to accurately measure fraud trends and losses.

c. Predominantly the fraud control and operations team will be responsible for the quantification of fraud losses, (e. g. average fraud per case, average roaming fraud loss per case, suspected destinations (nationally/internationally), and frauds per product / service type etc.) The fraud trends reporting should be classified into the different types of fraud detected and their source (e. g. subscription, call selling, roaming, account credits (prepaid), payment fraud etc.)

Fraud Risk Assessment

a. Interview line managers - what are the perceived problems, weaknesses, opportunities for fraud in their areas.

b. Interview the "on-ground salesmen" - what is the reality, stumbling blocks.

c. Obtaining supporting data - network, billing, finances existing reporting and reconciliation.

d. System Integrity - defining security and ownership.

e. Escalation practices and incident reporting

f. Analysis and categorization - quick wins, medium term and longer term.

g. Follow up - action plan, allocated on basis of time, benefits and activity required.

Evaluating New Products and Services:

a. To ensure maximum profitability of new products and services a fraud risk evaluation is paramount as this enables both a revenue protection and fraud prevention strategy and policy to be deployed across the various business segments.

b. This practice should be designed into the business processes so that the business can be proactive to fraud and revenue management issues rather than reactive.

c. The evaluation of new products, services and systems is a vital business process that should be undertaken prior to launch and continually assessed in test environment, it should never be viewed as a single activity.

d. The resulting losses from a product or service that has not been thought through properly and the potential fraud and security fraud risks determined can result in large financial losses through process and procedure weaknesses, in addition, to losing customer confidence.

e. On some occasions, the need to get a new product or service to market will be greater than the requirement to build in fraud protection.

f. In these cases, a prior understanding of the functionality of the product will allow the fraud team to be proactive to instances of fraud.

**Gaps-motivations-cracks leading to fraud:**

1. The importance of anti-fraud section:

a. 86% from the respondents answered that the anti-fraud section is very important, this results agreed with Johan H. van Heerden, he said the mobile telecommunications industry suffers major losses due to fraud, because of direct impact of fraud on the bottom line of networks operator, the prevention and detection of fraud become apriority.

b. What Namibias Minister of Works, Transport and Communication Joël Kaapanda said, strengthen the results in the table 15, in conference held the auspices of the Forum for International Irregular Network Access (FIINA) take place in Namibia, 2005(Namibia Economist, 2005), he said" fraud management and revenue assurance are important components of all companies and all societies as stakeholders, and network security experts and fraud managers, are an integral part of effective management of telecommunications companies today. Fraud is a global problem that threatens the profits of telecommunications companies around the world. Though accurate fraud figures are nearly impossible to pin down, FIINA itself estimates a total figure of around 56 billion Euros of losses worldwide due to telecom fraud and security-related problems. It is thus clear that telecommunications fraud is one of the fastest growing industries in the world and one of the most profitable of illegal activities".

2. The most important step to stop fraud is:

a. Half of the respondents agreed that the activations step is the most important step to stop fraud, since the subscription for the service considering the window for the fraudsters to attack the operator. At Telecom operator the subscription fraud was the most popular type, before Telecom operator creation of the black list program, which considered as database contains all the disconnected customers, and did not pay their invoices, whatever the account type is individual or corporate, clarifying customer details, (Name, number, addresses, activation date, the unpaid amounts, the returned cheeks, and the deactivation date).

In addition to the formal documents the subscription form required, such as (ID, clear and stabile address, security deposit, signature), all of these are Telecom operator precautions regard the subscription fraud. These precautions deserved the attention Telecom operator pays for this type of fraud, and this agreed with the conference survey results which were held in Singapore (2007), the conference was about telecom fraud and fraud prevention, 76. 5%agreed that the subscription fraud is the most fraud type currently detected, from seventeenth conference attendance from Europe, Middle East, and south-east Asia operators (CR-X, 2007).

b. In Deticta, white paper (Deticta, 2006), detecting telecom subscription fraud, they said, subscription fraud is characterized by fraudsters using false identities in order to purchase a service from the operator for which they have no intention to pay. One of the major issues in detection subscription fraud is in difficulty in differentiating it from simple bad debt, when genuine customers are unable to pay; some estimate that nearly 30-35% of all bad debts are actually subscription fraud.

c. A 2006 survey for the Home Office (The Home Office is the lead government department for immigration and passports, drugs policy, crime, counter-terrorism and police) suggests that over 1.7 billion of identity frauds takes place annually in the UK, 372 million in the telecom operator sector alone, based on TUFF estimates that identity fraud /subscription fraud could account for 40% of all telecom fraud in the UK (Deticta, 2006).

d. It is estimated that 70% of fraud losses rates to subscription fraud which is over 728 billion a year (78 million dollars a day), (Robert and Dabija, 2009)

3. The fraud affects at:

a. 77% from the respondents agreed that fraud affects do not include only the losses from unpaid invoice; they agreed also that fraud might lead to optional loss of new and existing customers, as well as bad publicity, the above results agreed with what Deticta white paper, titled by detecting telecom subscription fraud (Deticta, 2006), they said about the impact of fraud on mobile operator and their customers, the impact of fraud is far-reaching and can affect all parts of mobile operator's business. Not only is there an obvious financial impact but there can also be serious damage to the operator's brand, customer relationship and shareholder confidence. Furthermore, network operations can be disrupted and legal and regulatory requirements can be breaches.

b. The financial losses due to fraud can be built up in several ways. Firstly, there is the direct revenue lost when fraudster make use of mobile voice and data without baying commonly compounded by having the stolen services re-sold to other subscriptions. On the top of which is the direct cost of fraud, when the operator is left to pay for fraudulently acquired service and cannot defray the cost. Common trick for fraudster is to direct calls to their own premium rate service, by tricking mobile users to call their premium rate number. The mobile operator ends up paying commission to the premium rate service owner but is not able to recover the cost.

c. To add insult to injury, fraud can result in Mobile operators breaching legal and regulatory requirements, which carries the risks of bad publicity and fines, on a wider front, bad publicity related to fraud can damage the operator brand, breach  corporate social responsibility policies, depress shareholders confidence and affect stock market performance, fraud can also cause network traffic issues and disrupt the smooth running of the network potentially affecting the quality of the service available to legitimate users [15].

d. Finally fraud is increasingly becoming a customer relations issue; it can adversely affect the service quality and directly affect customer bills, both of which can lead to disputes and possible legal actions. Since customers are increasingly aware and concerned about Mobil data security and privacy, inadequate fraud protection can therefore results in damages customer reactions and will most likely causes customers to churn to networks perceived to be more secure [15].

4. The main impact of fraud attacks consider:

a. 38% from the respondents considers the fraud losses few but serious, meanwhile the rest of the sample consider it normal or moderate losses, no one really know how much fraud is costing the industry, they can estimate the cost because Telecom operator is reluctant to admit to fraud or are not actively looking for fraudulent accounts in the bad debt.

5. Most fraud cases are discovered by:

a. 47% from the respondents believe that most of the fraud cases and acts are discovered by the fraud management system, and 23% believe that it is the sales person accuracy, in fact the reality is most fraud cases are discovered by high usage reports (billing reports) and by chance, and the results above indicates that there is a little awareness of anti-fraud sections functions and programs used to detect fraud among the employees.

b. The above facts agreed with what GRAPA, Global revenue assurance professional association, white paper presents by Ade Banjako, about fraud management methodology in developing countries (Banjako, 2009), he said the Startups tend to rely on high usage alerts based on call types, value, duration or even credit limits. This function often has close similarities to credit control and it is important to clearly define the role of the fraud team so that they can concentrate on managing fraud.

6. The bad debit resulting from fraud is:

a. There is an important difference between bad debts and fraud, bad debts concerns people with occasional difficulties in paying their invoices, this happens only once or twice per person, if the subscriber really can't pay, he or she will most probably be suspended and denied to open a new subscription in the future, but fraud always include lie and there is no intention to pay for the used service. 58% from the respondents said that more auditing in anti-fraud activities should be done, since bad debt is may hide in the subscription fraud. This is agreed with Hoath as sited in the study of Abidogun that "subscription fraud can be committed upon fixed line and mobile telephone, and it is usually difficult to distinguish from bad debt, particularly if the fraud for personal usage, both subscription fraud and bad debts are major problems to telecoms in developing and third world countries" [15].

# 5. Conclusion and Future Scope

## 5.1. Conclusion

We faced many challenges when labeling both training and test datasets for grey call detection. In On-net bypass detection operator labeling was correct. But in off-net case both operator and we faced the challenge of determining class labels. Due to security loophole in subscriber authentication in one of the wireless fixed-line operator in the country, fraudsters were able to use real subscriber devices to fraudulent activity without noticing to real customer. This issue imposed great challenge in developing rules as issue was severe in the time we acquired training dataset. So, we have used different set of rules to different operators to address this issue. Also, some of the fraudsters has used call forwarding and many other advanced techniques to replicate genuine usage behavior and mislead detection systems. Therefore, we had to go through series of verifications to decide class labels for off-net bypass detection [22].

When we consider server resource utilization, grey call detection application consumed more resources than extreme usage detection. Initial plan was to implement aggregation for recent one-hour sliding window on CEP itself. But that approach was not feasible due to complexity of join queries and number of records involved in join queries. When system ran on this configuration, processing of some streams were lagging related to others. Pattern queries on

CEP were affected due this lagging. So, we have performed calculations based on one-hour sliding window on DAS. Calculations based on 24-hour sliding window were also done in DAS. But in this case, we have done those 24-hour calculations offline and merged with real-time view. This is equaled same as performing 24-hour based calculations in separate DAS server. Ideally this can be done in separate physical server in parallel to real-time calculations. Calculations related to extreme usage scenario was done on CEP only and resource utilization is comparably lower in this case.

Due to privacy concerns, operators were unwilling to expose these data to outside parties, so we identify the limitation of reproducibility. Also, we were not authorized to bring CDR details outside and experiment the system with better computing resources due to privacy concerns of operator. System was tested on server available in operator premises. With better computing resources we may able to test this system on higher data rate and evaluate system performance. Also, operator did not provide CDRs for full customer base due to privacy reasons. We have gained access to the CDRs of subset of customer base after operator has made some precautionary actions to preserve privacy. Therefore, we were unable to perform analysis on full customer, but the dataset provided was sufficient to implement comprehensive solution [22].

## 5.2. Future Scope

Inclusion of machine-learning techniques and using Neural Network or Tree-based classifier on derived feature set is interesting future work of this project. But this will be challenging task as some of the grey call instances replicate genuine behavior and that may corrupt learning process. Using machine-learning approach for Off-net bypass detection will be more challenging as numbers belong to different operators show different behaviors. So, hybrid method of rule-based and machine-learning based classification will be a fitting approach. Integration of WSO2 Machine learner which is a WSO2 module for predictive analytics will be another interesting future enhancement of this project.

Also, we can expect significant performance enhancements if this system can be run one clustered environment with high processing power. With high computing resources, more calculations can be moved to CEP and detection speed can be further increased. Also, scaling the proposed system to handle CDRs of full customer base of the operator is another challenging future work. Additionally, this system can be extended to detect handset theft scenario in future based on operator’s requirement. Developing CEP queries to detect network bypass and SIM box attacks on voice network could be value addition to the proposed solution. However, this system landmarks the good initiative in near real-time fraud detection in telecom operators by deviating from traditional database reliant approach [17][18][22].

## References

[1] R., Arnoff (2013). *Global Fraud Loss Survey 2013 by Communications Fraud Control Association*. [online] Amdocs. Available at: http://www.cvidya.com/media/62059/ [Accessed 9 Jul. 2018].

[2] Sallehuddin, R., Ibrahim, S., Mohd Zain, A. and Hussein Elmi, A. (2015). Detecting SIM Box Fraud by Using Support Vector Machine and Artificial Neural Network. *Jurnal Teknologi*, 74(1).

[3] Gent, A. (2017). Fighting fraud on mobile networks. *Computer Fraud & Security*, 2017(2), pp.10-13.

[4] H. Grosser, P. Britos and R. García-Martínez. Detecting fraud in mobile telephony using neural networks. *in Springer-Innovations in Applied Artificial Intelligence Lecture Notes in Computer Science*, (2005) 3533.

[5] H. Elmi, S. Ibrahim, and R. Sallehuddin. Detecting SIM Box fraud using neural network. *IT Convergence and Security-2012. Springer*. vol. 215

[6] Sallehuddin, R., Ibrahim, S. and Hussein Elmi, A., 2014. Classification of sim box fraud detection using support vector machine and artificial neural network. *International Journal of Innovative Computing*, *4*(2).

[7] Murynets, M. Zabarankin, R. P. Jover, and A. Panagia. Analysis and detection of SIM box fraud in mobility networks. *in Proc. IEEE INFOCOM* '14

[8] E. Okutoyi (2012). *SIM Box Fraud - New Headache for Africa's Mobile Operators*. [Online] Available: http://www.humanipo.com/news/142/sim-box-fraud-new-headache-for-africas-mobile-operators/

[9] Intec. *What is SIM box fraud.* [online] Available: https://www.xintec.com/fraud-management/what-is-simbox-fraud-and-why-is-it-so-hard-to-beat/

[10] N. Marz and J. Warren (2014). A new paradigm for Big Data. *in Big Data. Principles and best practices of scalable real-time data systems.*

[11] Fraudforthought.com. (2019). *Why fraud evolves, and how to address it | Fraud for Thought*. [online] Available at: http://fraudforthought.com/index.php/why-fraud-evolves-and-how-to-address-it/

[12] Fraudforthought.com. (2013). *TalkTalk customers are victims of fraud – AGAIN | Fraud for Thought*. [online] Available at: http://fraudforthought.com/index.php/talktalk-customers-are-victims-of-fraud-again/ [Accessed 26 Feb. 2019].

[13] Elmi, A.H., Ibrahim, S. and Sallehuddin, R., 2013. Detecting sim box fraud using neural network. In *IT Convergence and Security 2012* (pp. 575-582). Springer, Dordrecht.

[14] Reaves, B., Shernan, E., Bates, A., Carter, H. and Traynor, P., 2015. Boxed out: Blocking cellular interconnect bypass fraud at the network edge. In *24th {USENIX} Security Symposium ({USENIX} Security 15)* (pp. 833-848).

[15] Sahin, M., Francillon, A., Gupta, P. and Ahamad, M., 2017, April. Sok: Fraud in telephony networks. In *2017 IEEE European Symposium on Security and Privacy (EuroS&P)* (pp. 235-250). IEEE.

[16] Rosas, E. and Analide, C., 2009. Telecommunications fraud: problem analysis-an agent-based KDD perspective. *Aveiro: EPIA*, *2009*.

[17] Kazerooni, M., Zhu, H. and Overbye, T.J., 2014, February. Literature review on the applications of data mining in power systems. In *2014 Power and Energy Conference at Illinois (PECI)* (pp. 1-8). IEEE.

[18] Ighneiwa, I. and Mohamed, H., 2017. Bypass Fraud Detection: Artificial Intelligence Approach. *arXiv preprint arXiv:1711.04627*.

[19] Sahin, M. and Francillon, A., 2016, October. Over-The-Top bypass: Study of a recent telephony fraud. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1106-1117). ACM.

[20] Ayamga, D., 2018. TELECOMMUNICATION FRAUD PREVENTION POLICIES AND IMPLEMENTATION CHALLENGE.

[21] Tawashi, E. and Omer, H.A., 2011. detecting fraud in cellular telephone networks" jawwal" case study. *detecting fraud in cellular telephone networks" jawwal" case study*.

[22] Richards, C., Richards, P. and Ramachandran, H., Network Kinetix, LLC, 2018. *System and method for an automated system for continuous observation, audit and control of user activities as they occur within a mobile network*. U.S. Patent Application 15/783,436.

# Appendix -1

## Interview Questions

1. What does SIM box Fraud mean to you?
2. What are the motivations SIM box or bypass fraud?
3. How do you find the impact of SIM box fraud to Telecom Operators?
4. Do you find any obstacles either technically or managerially that preventing the detection of SIM box fraud?
5. Based on your experience, what are the disadvantages of the SIM box fraud?
6. Which type of detection technique will be suitable for telecom operators in middle-east?
7. What are the benefits of having fraud detection system by Telecom Operators?
8. How do you find the proposed solutions to mitigate SIM box fraud?
9. How do you see the future of SIM box fraud and its impact with the new upcoming technologies?