Addis Ababa University

College of Natural Sciences

Developing a Detection Method for Interconnect Bypass Frauds Using Fuzzy Logic

Tadele Ayalew Degu

A Thesis Submitted to the Department of Computer Science in Partial Fulfillment for the Degree of Master of Science in Computer Science

Addis Ababa, Ethiopia

July 2021

Addis Ababa University

College of Natural Sciences

Tadele Ayalew Degu

Advisor: *Dagmawi Lemma (PhD)*

This is to certify that the thesis prepared by Tadele Ayalew, entitled: *Developing a Detection Method for Interconnect Bypass Frauds Using Fuzzy Logic* and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Name Signature Date

Advisor: _____

Examiner: _____

Examiner: _____

ABSTRACT

An interconnect bypass fraud is a telecom fraud that manipulates technological advancements and realized over the existing cellular networks with the intention of gaining illegal benefits. It results a degraded quality of service and financial loss.

Existing prevention mechanisms collect call detail records to detect the fraud by analyzing various predefined behaviors. Hence, such systems play the role of intrusion detection by recording known behaviors. Thus, illegal accesses of a cellular network would be detected if the activity is similar with previously identified suspicious act, this further is a major concern of having a higher rate of false positive and/or false negative alarms.

As interconnect bypass fraudsters are basically attacking the cellular network from a stationed location through a series of fixed network elements, a mobile subscriber, yet stationed is a suspect to be fraudster. In order detect new fraudulent act by studying the activity with respect to the natural set of mobile users (i.e., mobility) and mitigate the false negative and false positive rates, we have introduced a new detection method through a design science approach. We intend to trace mobile subscription but operates from fixed location. Our method gets inputs from home location register and monitors the mobility history of cellular network users by applying a fuzzy logic. We have tested the method by logging the location histories of 1037 randomly selected users. We have detected the fraudulent users with 1.92% up to 5.88% false positive rate and 0.88% up to 5.88% false negative rate.

Keywords: *interconnect bypass fraud, home location register, fuzzy logic, cellular network*

ACKNOWLEDGMENTS

I would like to thank Dagmawi Lemma (PhD) for the guidance and advise he had been giving me from the beginning to the completion of this thesis. I am also thankful to Yenalem Ayalew (PhD) and Fitsum Yitbarek for their roles that they've played on. They had been encouraging me during the course of this thesis.

Information and logged files were very important throughout this thesis. In this regard, there were people whose contributions were helpful. Specially, Habtamu, Gizachew, Selam, Alamir and Hayat, they deserve more than an appreciation.

Tadele Ayalew Degu

Table of Contents

List of Figures

List of Tables

ACRONYMS

ANI	- Automatic Number Information
ANN	- Artificial Neural Network
BSC	- Base Station Controller
BTS	- Base Transceiver Station
CDMA	- Code Division Multiple Access
CDR	- Call Detail Record
CLID	- Caller Identity
CNP	- Cellular Network Provider
CS	- Circuit Switch
EIR	- Equipment Identity Register
EPC	- Evolved Packet Core
FDMA	- Frequency Division Multiple Access
FMS	- Fraud Management System
GGSN	- Gateway GPRS Support Node
GSM	-Global Stations for Mobile communications
HLR	- Home Location Register
IDS	- Intrusion Detection System
IMSI	-International Mobile Subscriber Identity
ISDN	-International Subscriber Directory Number
ISP	- Internet Service Provider
LRN	- Location Routing Number
LTE	- Long Term Evolution
MIMO	- Multi-input Multi-output
MSC	- Mobile Switching Center
MSRN	- Mobile Station Roaming Number
OFDMA	- Orthogonal Frequency Division Multiplexing
OTT	- Over-the-Top
PBX	- Private Branch Exchange
PLMN	- Public Land Mobile Network
PS	- Packet Switch
QoE	- Quality of Experience
QoS	- Quality of Service
RAN	- Radio Access Network
RNC	- Radio Network Controller
SGSN	- Serving GPRS Support Node
SVM	- Support Vector Machine
TDMA	- Time Division Multiple Access
TDoS	- Telecom Denial of Service
VLR	- Visitor Location Register
VoIP	- Voice over IP

CHAPTER ONE

INTRODUCTION

In this chapter, we have introduced about telecom frauds in general and interconnect bypass fraud in detail in addition to indicating the problem statement, the questions that we intended to address, the methods applied, the reason that motivated us, the scope and limitations of this thesis.

Telecom networks are mostly classified as computer networks, public switched telephone networks, IP networks and cellular networks [1].

- In the first type, there are computers that apply a set of communication protocols over digital interconnections to share resources which are provided by various network nodes. The commonly used nodes are routers, switches, servers, bridges, repeaters, hubs, network interfaces and firewalls.
- Public switched telephone networks (PSTN) are those which rely on circuit switched networks and consists of interconnected telephone lines, fiber optic cables and communication satellites. The routing of connections is obliged to pass through several switching centers.
- Packet switched networks, also known as IP networks, are those which lets a data to be transmitted in the form of small blocks (packets) over a specified channel and duration.
- Cellular networks, also known as mobile or radio networks, are those in which the network is distributed into cells. A cell in this case is the geographic area, served by one or more transceiver (transmitter plus receiver).

Nowadays, the boundary between each of the telecom network types is being broken. For example, since the cellular and IP networks are merging, it became possible for voice to pass over Internet. This advancement has brought the concept of voice over Internet protocol (VoIP). VoIP allows for users to make or receive a call even from or to a personal computer

through the IP network. Thus, voice is flowing through IP network in addition to a dedicated line.

If it is subscribed in a legal way, VoIP has benefits like usage cost reduction, ease of use and provided many feature that the traditional cellular network does not have [2]. It could also support to transmit telephone calls over a high-speed Internet connection instead of traditional fixed telephone lines. These calls do not travel directly from a caller to a recipient's computer but rather, in the middle, there would be layered computers which belong to VoIP service providers. However, by mentioning its major role for telecom fraud, many researches such as [3, 4] argued as there are exceptions that also makes VoIP disadvantageous.

Telecom fraud in general and interconnect bypass frauds in particular are sorts of cybercrime. The intention and characteristics of an interconnect bypass fraudster could be assumed as one among the black hat hackers of a cyber. Indeed, the victims are not only mobile service providers but also all stakeholders including security regulators and mobile subscribers. For that matter, the loss that it brings is measured not only in cash but also in quality of service and correctness of call records.

On the other hand, every cellular network provider (CNP) has two databases in its core unit, known as home location register (HLR) and visitor location register (VLR). HLR is a static central database while VLR is dynamic and many in quantity. Both of these databases stores the entire information of mobile subscribers [5]. This information is all-inclusive and record detailed information such as the international mobile entity identity (IMEI) of the device in which the SIM is inserted on, call processing, authentication and many other attributes. In addition, HLR stores the location area identity (LAI) that is a unique identifier for the purpose of updating the current location of mobile subscribers.

HLR and VLR have a joint communication interface which lets them to share the information. Specially, when a mobile user changes its LAI (i.e., from a specific geographical area to another), HLR will update by getting the current information from VLR through the joint interface.

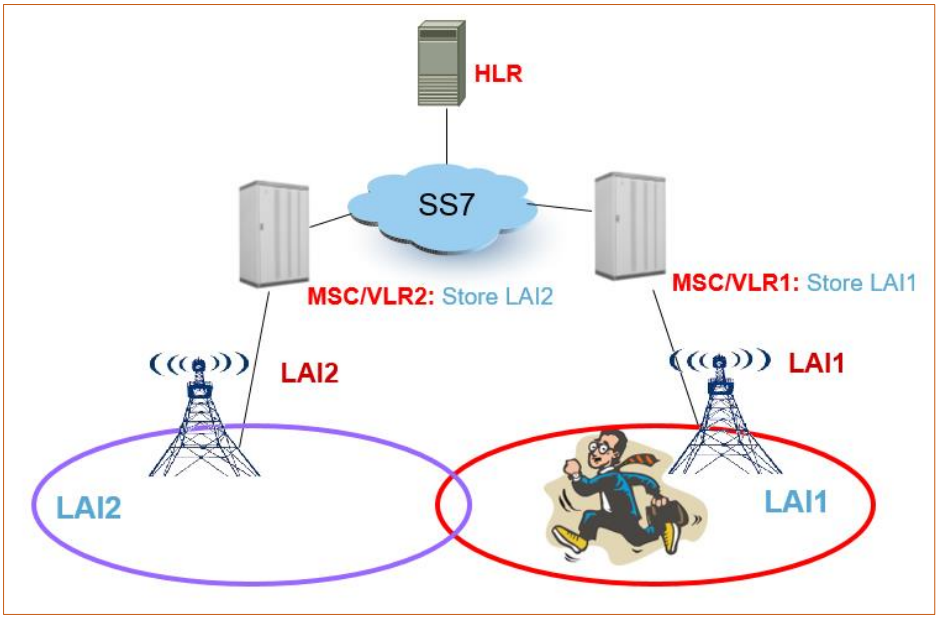


Figure 1.1: *Normal Location Update Situation*

As shown in Figure 1.1 [6], LAI_1 is found in the area coverage of VLR_1 so that the original information of the user is kept in VLR_1. On the other hand, VLR_2 is responsible for keeping information about subscribes in LAI_2. Suppose the user change location from LAI_1 to LAI_2, HLR will update the information of that specific subscriber at least once. There is also a possibility of update to be more than once even in a single VLR. (e.g., if the user moves from one network cell to another cell).

1.1 Motivation

In 2019, ethio telecom publicly announced that the annual revenue loss of the company had become weighty. It was also lately declared in the first row of potential risks list [7]. That was an influential fact that leads us to analyze more about telecom frauds by getting real information from real fraudulent groups and looking for the practices of different telecom operators. In every year, interconnect bypass fraud is found among the "major challenges" list of different reports [8] [9]. There seems nothing other than fraud that worries the telecom industry [9].

The primary intention of the fraudster is to get illegal revenue in addition to other issues like revenge and identity theft. According to [9], in 2019, the worldwide annual loss was estimated as it could exceed US$38.1 billion. As the vulnerability they left and the counter measure approaches they followed may vary, the impact level may also vary from one service provider to another. Due to the dependency of telecom operators on third party vendors, especially in the developing countries, it is afraid not to be a cause of multidimensional (social, economic and security) crises as well as a means for complicated crimes [9, 10].

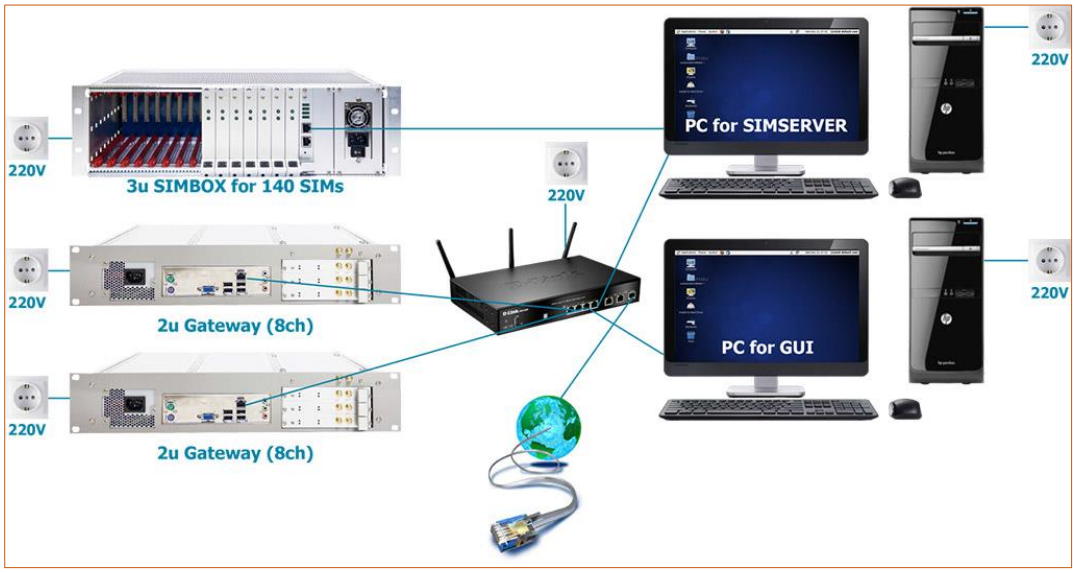


Figure 1.2: *Interconnect Bypass Setup*

Even though the targets of fraudsters are cellular networks, in most scenarios, interconnect bypass frauds are subjected to a fixed network. As illustrated in Figure 1.2 [11], behind the interconnect bypass gateway, there is Internet where it is basically provided by a series of stationed (fixed) routers or network elements. For this reason, a voice call which was generated from a mobile (moveable) device and a voice call which was generated from an interconnect bypass gateway (a one which has a non-moveable network element on behind) seems to be separated simply by recognizing the mobility. This was another insight that increased our initial motivation. Normally, if the device of a caller is really mobile, it should change its location at least once in its call history. Oppositely, if the call was rerouted by SIMBOX that holds for example 240 SIMs, all calls which are associated to these SIMs will be in a fixed location. We took this as one signature of the attack, among others.

Thus, with these deriving reasons, we became motivated to partake for this hot issue and contribute a detection method.

1.2 Statement of the Problem

Fraudsters use various techniques to bypass the security system of CNPs for exercising unauthorized activity. One of the techniques is by deploying and configuring devices between two or more CNPs to reroute the traffic. By doing so, they aim to gain financial benefit by faking the international terminations as local calls; and the subscriber could be charged below the legal fee, yet the CNP is not benefiting from the services running over its own infrastructure [8].

Since the setup that is made by fraudsters is very complicated and the traffic comes through the support of VoIP [10, 11], it is difficult to detect the fraudulent action/traffic. Here, it is important to mention that not all VoIP services are interconnect bypass frauds since there are legal use of VoIP applications (e.g., Viber-to-Viber, Viber-to-None-Viber, Skype-to-Skype, Skype-to-None-Skype and the like). Interconnect bypass fraud can be of three sub types – SIMBOX interconnect, global stations for mobile (GSM) gateway termination and over-the-top (OTT) bypass [10, 12]. Basically, VoIP applications are technical instances of OTT fraud in which case the end users (callers) became benefited in terms of cost. However, in the case of SIMBOX interconnect or GSM gateway termination, the benefit is for the mediator - not the end user. The one to be called and/or the caller may not have even the information as they are talking through a fraudulent line. For that matter, there may not be even a caller but rather, the fraudster by him/herself might generate a fake call traffic on behalf of the caller.

On the other hand, with the discovery of these frauds, several defensive approaches have been proposed and some of them are already implemented. However, most of these solutions are configurable often manually by intrusion detection system (IDS) [13]. These can detect abnormal network usage which is an illegal access of CNPs' cellular network for any intention. However, in most cases, there is a major concern of having a high rate of false positive and/or false negative alarms [3, 7, 12]. Besides, these systems can only be as effective as its configuration. According to ethio telecoms' report in 2020 [7], for example, "significant number" of detected fraudsters had complained and confirmed as they were legitimate, yet the

system flagged them false positive. In the case of false positives, a normal usage of the cellular network may be mistakenly identified as a fraud so that it would bring denial of service (DoS) even for legitimate telecom subscribers [14, 15]. On the other hand, in the case of false negatives, the system assumes a fraud of telecom networks as legitimate one.

The manual configuration of IDS also demands carefulness and honesty of the human operator. In other words, the success of an IDS depends on the human behavior. On the contrary, humans by themselves can be one point of security threat [13]. Therefore, in addition to configuring the IDS based on predefined signature and/or anomalies, it is important to further investigate the opportunities we can have in the existing architecture of the CNP infrastructure to find an optimal and effective solution. One way to identify the legitimacy of services could be by tracing and learning the expected feature of the service. For example, the system should be suspicious about a user who is making a call from a mobile phone but always from the same location.

Even though HLR and VLR support real time modifying and querying of user location information, location registers have not been yet significantly associated with telecom fraud handling. Literally, a switched-on mobile is supposed to be in one location area in which that location area covers a group of radio network cells. If the user device moves, the location area will be changed and that would be stored in VLR. This enables the system to detect the mobility of a users' device in real time.

Though VLR can be used to trace the location of a mobile user, one VLR has a wide coverage as a single location area contains a group of cells. This may lead to a wrong conclusion that a user device is mobile (moveable) if and only if it gets out from one location area and gets into another location area. The conclusion could be a Boolean value (yes or no). As a result, since there are also users who are less mobile, the decision we are going to make will be neither logical nor realistic.

Therefore, the context of the result should be taken into consideration before making decision. Hence, to narrow the scarcity of VLR/HLR, we need to apply a fuzzy logic. In a fuzzy logic, unlike to Boolean logic, there is no absolute true or absolute false value [16]. In our case, it will give us a chance to consider a user device as mobile if the user moves, no matter how far.

Therefore, in order to effectively determine and learn about the true characteristics of a cellular network usage, it is important to investigate the following research questions.

a) While the HLR and VLR are used to maintain user location, why not associated to detect interconnect bypass frauds by customizing the mobility management of cellular network users?

b) Since there would be a location update on HLR if the user enters from one location area to another, how can we use this information to detect interconnect bypass frauds and what other attributes and logics can be used to identify the specific location of a cellular network user?

c) How do we drive a more correct detection of interconnect bypass frauds by using the value of these attributes? And how effective would that be?

Accordingly, we conduct this research to answer the above research questions. In subsequent sections, we present our findings from literatures showing types of frauds, defense mechanism related works. Then, the proposed solution is presented followed by experiment results and discussion.

1.3 Objective

- General Objective

Our main objective is to develop a method of interconnect bypass fraud detection by applying a fuzzy logic over mobility management.

- Specific Objectives

For developing the method, we have the following specific objectives.

i. To record sample profile histories of cellular network users from HLR.

ii. To identify basic attributes through abstraction to store some of the last movement history of each user device in a simple database with basic attributes.

iii. To design a method that applies a fuzzy logic on the recorded data.

iv. To identify the mobility of each user device in a cellular network, at least if it moved sometimes throughout its call history.

v. To categorize each user based on their resulted flag (i.e., highly trusted, likely trusted, trusted, suspected, likely suspected and highly suspected).

1.4 Methods

We started the process of designing by observing the real network and pass through the following processes.

i. ***Observation on the Real Network:*** in order to have a deeper knowledge on how the cellular network is structured, which network element is next and behind to the user device, where did the cellular network and the Internet intersect. Actually, it has been taken into account as one CNP may have a minor difference with the other in deployment of network.

ii. ***Database Design:*** a simple standalone database that has fundamental attributes like the calling number, location and the like. This will be the primary input for the Fuzzifier then for the mobility checker.

iii. ***Detection Method Development:*** based on fuzzy logic and mobility management.

iv. ***Set and Select Sample User Devices:*** samples will be selected randomly and these will be connected to developed method. We will handle this not by physically connecting the user devices, but rather, taking the history logs from HLR.

v. ***Apply the Method for the Samples:*** the fuzzy logic and mobility checker will be employed on selected samples.

vi. ***Analyze the Output Data:*** after applying the method, we will analyze the output and answer our research questions.

1.4.1 Sampling

Our population was targeted to all users of a cellular network in a particular service provider (i.e., ethio telecom). While we were determining the sampling method and sample size, at the

beginning of 2020, ethio telecom reported as it has 44.53 million mobile subscribers [7]. This implies that we had a large population and it would be tough to identify each of the mobile subscribers. For this reason, we used a random probability sampling technique so that each subscriber would get an equal chance of being selected.

In order to determine the size of the sample, we have considered the following inputs as per the common trends of sampling [17].

Table 1.1: *Considered Attributes for Sample Size Determination*

Input	Value	Explanation
Population Size	44,530,000	According to ethio telecoms report [7]
Margin of Error (Confidence Interval)	±4%	The level of precision that we required to be
Confidence Level	99%	The possibility that the confidence interval holds the true proportion
Sample Proportion	50%	It is the expected result. Since we didn't conduct a pilot study, we simply took the conservative value

As presented in Table 1.1, the required sample size can be calculated by using a simple calculator from [17] to be **1,037**.

1.5 Scope and Limitations

It is obvious that emulating it in a real network would bring a more realistic result. However, there is no any legal option for importing SIMBOX terminals or GSM gateways in addition to their expensive cost. Therefore, getting a permission and cooperation for accessing the active network and its elements from ethio telecom was the only option that we got for experimenting our method. Fortunately, the author of this thesis is an employee of ethio telecom, having an opportunity at least to know the over-all network topology and privileged administrators for each network element. However, due to privacy worries defined by the respected departments, the opportunity we got was limited to a secondary data from HLR and testbed from ethio telecoms' telecom excellency academy (TExA).

1.6 Application of Results

Because of the standards which are regulated by the International Telecommunication Union (ITU), most CNPs have similar network traffic transportation. As a result, the solution that we have proposed would be applicable to all Internet/telecom service providers. On the other hand, for a better certainty, the proposed method could also be easily integrated with existing and coming detection approaches.

The computational cost of detection of interconnect bypass fraud will be reduced. The main ground for this prediction is because it is not required to store large-sized history data (e.g., call recordings) for the purpose of analysis. In addition, detecting the fraudsters of interconnect bypass fraud will be as simple as recognizing the mobility practice of the user device. Hence, this is all about introducing a significant solution not to be attacked by interconnect bypass fraudsters.

1.7 Organization of the Thesis

Our thesis shows how to apply a fuzzy logic on the mobility information which could be found from HLR for the purpose of detecting interconnect bypass frauds. The detailed parts of our thesis is organized in the following chapters.

Chapter Two: Literature Review

We will assess and discuss literatures about interconnect bypass frauds. In addition, we will discuss about various types of telecom frauds, how our approach will help, the role of VoIP in cellular network as well as in interconnect bypass fraud.

Chapter Three: Related Work

We will discuss and indicate the gaps in the previous related works which were researched about detection and/or prevention approaches of interconnect bypass frauds. Considering their distinct approaches, a few number of related works will be presented selectively.

Chapter Four: Interconnect Bypass Fraud Detection (IBFD) System

The method that we have proposed, named as IBFD system, will be discussed in chapter four. The architecture of the mobility checker, different components of the central system and the database design will be presented in a detailed kind.

Chapter Five: Prototype and Evaluation

In chapter five, the experimental settings, procedures and results, the tools which were used for prototyping and their justification will be discussed.

Chapter Six: Conclusion and Future Work

Finally, we will complete it by summarizing all activities which we did in major. In chapter 6, we will also indicate our future work.

CHAPTER TWO

LITERATURE REVIEW

In this Chapter, we will discuss about the overview of fraud in the telecom environment, types of telecom frauds and fraud detection trends in general. Then, by focusing on its sub kinds and practices, we will describe more about one of the major types of a telecom fraud known as an interconnect bypass fraud. In addition to these, how the fraudsters are taking advantages from VoIP to make frauds specially in cellular networks will be discussed as well. As a final point, we have also a briefing about fuzzy logic as it is the approach that we chose to make our method effective.

Telecom fraud is one of the major cyber threats targeting individuals, companies or governments aiming to access information or control the telecom infrastructure without proper authorization having the intention of economic gain in addition to other motives [9]. As the cyberspace embraces the telecom ecosystem, telecom frauds may arise due to one or more of the following security threat sources [15, 18].

- Weakness in the telecom network infrastructure and communication protocols,
- The insider effect (personnel),
- Rapid growth of hacker (fraudster) community,
- Openness of the cyberspace (Internet),
- Social engineering,
- Limitation in effectiveness of reactive solutions

2.1 Most Common Types of Telecom Fraud

Fraud is becoming a never-ending risk to telecom network operators, and it has become challenging and difficult to identify how, when, or where new fraud settings will attempt to attack services. Nowadays, in the telecom ecosystem, fraud is an increasing threat. Telecom fraud includes a variety of illegal activities on telecom operator network. Telecom frauds exist

and adversely affect the carrier providers financially as well as in terms of extensive voice bandwidth, degrading service quality and network resources [3, 9].

These telecom frauds are various in kinds. Some of these are briefly described hereunder.

a) Telecom Denial of Service

It is another form of attack [18], similar to the traditional data network denial of service (DoS) attacks. In this form of attack, fraudsters make a huge number of phone calls and hold it for long durations by targeting to attack network's availability.

In a DoS attack, unauthorized users flood a system with too many access requests. Due to this, legitimate users will not be able to access the network. Similarly, TDoS attacks can impair a voice network's availability. It can also be used as a tool for extortion. They are being considered as a threat to public safety. Fraudsters of this kind have taken triggering attacks against hospitals, security offices, and other public services [8].

b) Vishing

Vishing is a synonymous of phishing. Phishing is a kind of cybercrime that uses email messages with phone addresses, websites, or window pop-ups. Hackers of this type have an intention to gather personal information, then, for easily making an identity theft. Likewise, a kind of phishing that uses the phone instead of email is known as Vishing (VoIP Phishing). Hackers of this type act as a legitimate user to attempt and gather information for identity theft or other forms of fraud.

Spoofing is a common technique used by attackers of this type. Specifically, spoofing by altering the caller identity (CLID) so that the victim would assume as called from his/her bank, government office, reputable private company or someone important [9, 15].

Therefore, mostly the victims of phishing or vishing are careless users, who post personal information on social networking websites. Fraudsters are taking those details as an advantage to act as trustful to collect significant personal information.

c) Interconnect Bypass Fraud

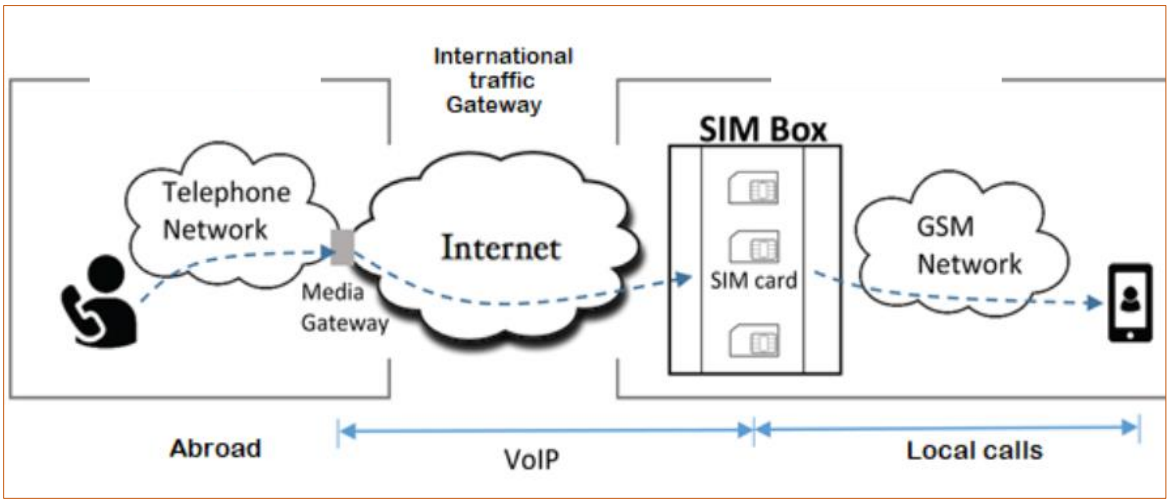


Figure 2.1: *Interconnect Bypass Fraud Traffic Flow*

Interconnect bypass fraud, as described by different scholars [10, 19, 20], is one form among many frauds that runs over VoIP to change the route of international incoming calls by differing traffic away from the legal interconnect gateways. As the traffic route has been indicated in Figure 2.1 [11], SIM-Box is used in between networks so that incoming calls will be changed to appear as they were originated domestically. The fundamental gateway equipment that is used by this fraud is called SIM-BOX or VoIP GSM Gateway [10].

Due to interconnect bypass fraud alone, in 2019, the global loss has been estimated to be at about 32.7 billion USD. According to some surveys, this kind of fraud is found in the two most top lists of fraud types worldwide [20]. The problem is also beyond financial losses. Every bypassed call degrades the quality of service and experience (QoS/ QoE). The voice quality of SIM-BOX calls is quite poor, call setup time is longer, and the call success rate is considerably lower. Most badly, the privacy of users' communication will be compromised as it will pass through the ears of fraudsters/SIM-BOX facilitators [9, 20].

Interconnect bypass fraud by itself could be classified into three sub types: GSM gateway termination, SIM-boxing fraud and OTT fraud.

2.1.1 GSM or VoIP Gateway Termination

GSM gateway termination got its first name from the cellular network standard (i.e., second generation (2G). However, after the introduction of 3rd, 4th and 5th generations, the termination tool has been also advanced, respectively. It is also known as VoIP gateway. Gateways, in this case, are interconnect systems allowing voice interoperability between other incompatible radio communications (e.g., incompatible radio frequency bands).

Figure 2.2: *GSM gateway Termination Route*

Most gateway devices are portable, but they require a permanent configurations system. As shown in Figure 2.2 [21], calls might be terminated to or originated from GSM gateway. In addition to its role for interoperability, the termination device can also connect trunked talk groups, encrypted networks, public telephone systems, and cellular or satellite phone connections.

A single GSM/VoIP gateway could have many channels. Each channel gave an opportunity to the fraudster to terminate many international calls. The device could be used in combination with a SIM-Box or SIMs might be directly attached to it.

2.1.2 OTT Fraud

It is a technique of interconnect bypass frauds that differs from SIM-Boxing and GW termination as there would be an application between the two ends of a traffic instead of a device. OTT applications are similar to those we commonly use, such as applications are Viber, Skype, Tango, Telegram, Snapchat, WhatsApp, IMO, Facebook messenger, Google

due, Google voice and the like. In many CNPs, less to none of these familiar VoIP based applications are not officially classified as illegal. Because, service providers of these applications do not require a payment for such services, specially, audio/video calls between applications has been kept in the grey eyes of many CNPs [12]. However, for calls which are to be made between the application and the traditional cellular network, some of the application providers have set a relatively minimum price for outgoing calls (e.g., Viber to None-Viber) [22].

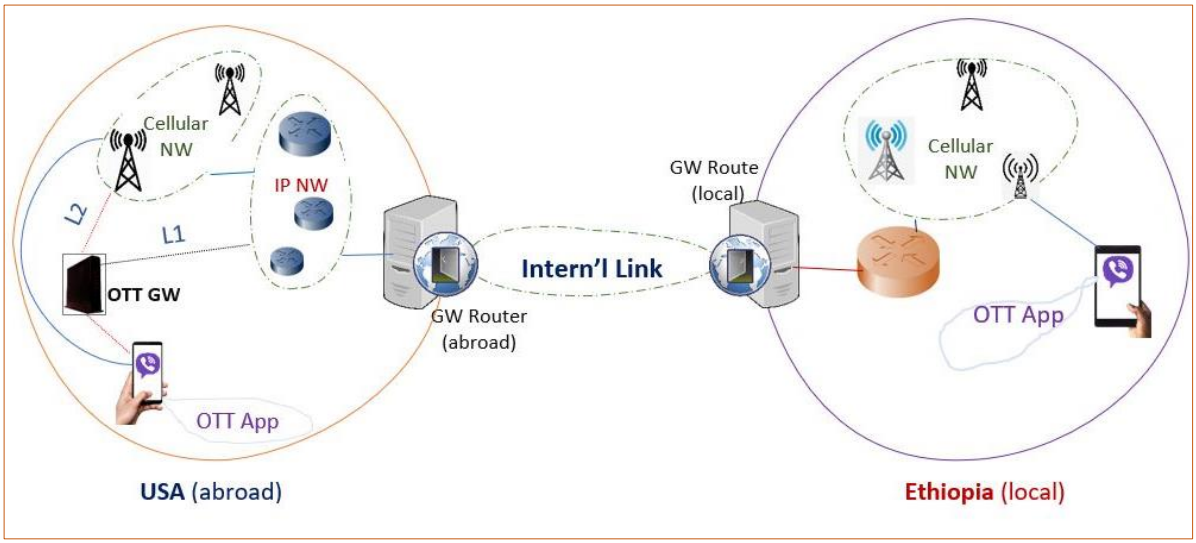


Figure 2.3: *Over-the-Top (OTT) Traffic Flow*

As shown in Figure 2.3, the name "over-the-top" comes from its nature as providers of these applications do not have their own network infrastructure. But, their services run over existing telecom infrastructure. They fraudulently use the network of other operators. Here, it would be better to briefly discuss about the following two scenarios.

a) The grey scenario: in this case,

- Both end users (caller & the one to be called) would be equally benefited though the application provider may gain indirectly (e.g., from advertisements).
- Both end users should have the same installed application so that the traffic would be application-to-application (e.g., Viber-to-Viber, Skype-to-Skype, Messenger-to-Messenger, WhatsApp-to-WhatsApp, Tango-to-Tango, Telegram-to-Telegram, etc.)

- Not officially labeled as fraud or authorized by many CNPs. However, some CNPs forbid such service, and all OTT applications are blocked in their network. Other CNPs does not care about that and they themselves are even official customers to either of those applications [8, 18]. For example, in Figure 2.3, end users may have different device types (mobile, PC or tablet) but the same application (i.e. Viber). The maximum payment that is expected from the end users is only the Internet connection fee.

b) The black scenario: in this case,

- The caller and the application provider would have a shared opportunity. The provider would oblige the caller to directly pay for the service.
- If the caller has already installed either of the OTT applications, it is enough to make a call. The traffic would be between application-to-non-application (e.g., Viber-to-Non-Viber, Skype-to-Non-Skype, etc.)
- OTT service providers have set their own gateway. This gateway will have a role to interconnect the application-generated traffic to the unleased mobile network without legal authorization. Recalling Figure 2.3, let us say the caller is in Ethiopia and s/he had subscribed for an OTT application by paying relatively low cost than the legal price. It doesn't matter whether the one to be called in USA is subscribed for that particular OTT application or not. The caller will simply turn on her data so that s/he will be connected to the cellular network of ethio telecom. Since a mobile network is connected to an IP core network, the call will be treated as a data traffic. Then, gateway of the OTT application provider will receive that traffic and simply terminate it to uncontrolled voice call. This implies, payment would be issued to neither side of the CNPs even though the infrastructure of both was used. On the other hand, only the provider of an OTT application and caller will be advantaged.
- It is just like trading in someone's shop without renting it or any deal. This is an absolute fraud in which every CNP is fighting for.

OTT services work on top of data links without the control of ISPs to deliver voice, text, and video contents using packet switching technique. Compared to other types of interconnect bypass frauds, in OTT frauds, end users are also intentionally participating to make it happen and will have a shared benefit. However, in the case of SIM-Boxing or GSM gateway termination, the benefit is only for the provider - not the end users. The one to be called and/or the caller may not have even the information as they are talking through a fraudulent line.

2.1.3 SIM-Boxing Fraud

This is a wide spreading sub type of interconnect bypass fraud in which the bypass is done by connecting incoming VoIP calls to the telecom operator's cellular network using a special device called a SIM-Box [20]. Fraudsters will set this device between the radio network of a local CNP and the international gateway. Commonly, on the other side (abroad), there will be someone who is cooperative to the fraudster. That cooperator has a role either to generate artificial international call traffics or to reroute legitimate traffics to the SIM-Box.

A SIM-Box, also known as a SIM card reader or a SIM bank, is an advanced mobile device that connects VoIP calls to a mobile voice network. It could have a capacity to hold hundreds of SIM cards as shown in Figure 2.4 [9, 12]. For example, if it holds 300 SIM cards, that device alone could reroute about 300 international calls to appear as local. By doing so, telecom operators will be highly impacted (e.g., revenue loss, availability, and reliability reduction) while fraudsters gain. It also reduces QoS/QoE rates of legitimate end users since it injects a huge number of audio-calls at once to a cellular network which was supposed to handle a limited number of calls.

Besides to the capacity of a single SIM-Box to hold many SIMs, a single fraudster has also an option to connect as many SIM-box devices as he/she wants. This implies, if the fraudster is using 100 SIM-Boxes holding 300 SIMs, the impact would be 100 * 300 at a time. Again, if there are many fraudsters in the territory of a single telecom operator (also known as an autonomous system (AS)), the impact would also be multiplied.

Figure 2.4: *Sample Device of SIM-Box (SIM bank)*

Neither the caller from abroad nor the one who is going to be called are not beneficiary from the process. In most cases, none of them may not be aware of even the existence of a fraudster in the middle of their communication. As shown in Figure 2.1, the immediate upper link of users at both ends are legitimate mobile networks. There might be many motives for fraudsters to do this fraud, but, revenge and financial gain are the major.

- ***Revenge:*** some former employees of operators use their information to attack their previous organization, by believing that they lost their job due to an ill-treatment. On the other hand, almost in all countries, there is at least one telecom operator owned by respective governments [8]. As a result, some opponents might use it for indirectly weakening the government.
- ***Financial gain:*** suppose the caller is from operator/Country-A and the receiver is in operator/Country-B (e.g., Ethiopia). Somewhere in Country-B (town-y or else), the fraudster will setup his SIMBOX tools (as shown in Figure 1.2, 2.1 & 2.4). If the termination cost of A to B was $1 per minute, operator in country-A would charge the caller $1 per minute. Since the call will be fraudulently re-routed as local, Operator in Country-B will charge operator-A by $0.05. This implies, Operator in Country-B will lose at least $0.95 (95%) in each minute for each fraudulent SIM. On the other hand,

the fraudster in Operator-B has an agreement with another fraudster in Operator in country-A (most possibly, directly with CNPs and/or ISPs) to share the difference.

As common to all interconnect bypass fraud types, behind to a SIM-Box device also, there is a VoIP technology.

2.2 Existing Anti-Fraud Approaches

The common anti-fraud mechanism is known as fraud management system (FMS). Even though it has not uniquely standardized, this system has a significant role for detecting and reacting to the fraudulent events. Different CNPs have adopted their FMS based on the nature of the threat which they thought potential.

Since there are various forms of telecom frauds, the types of the threats that FMS is trying to overcome are also different. In the case of interconnect bypass frauds, the consequence touches the two classes of threat (i.e., unauthorized disclosure & usurpation), at least partially. Because, it is a condition whereby a fraudster interconnects to the cellular network (unauthorized disclosure) and also, virtual control of a network infrastructure (usurpation).

Most FMSs took call detail records (CDR) as an input from a customer relation management system (CRM). Actually, there are also FMSs which integrates CRM with other supplementary systems (such as billing system, operation support system). Then, if a fraudulent activity is detected, the counter measure will be taken manually. This implies, FMSs significantly shared the characteristics of intrusion detection systems [19].

Based on their goal, the existing anti-fraud approaches would lay in either of the following classifications.

i. **Detection and Prevention:** CNPs have adopted their FMS by using a data mining technique which is the process of inspecting, cleaning, transforming, and modeling data with the aim of discovering useful information for making conclusions and decisions. Besides to the considerable challenges (e.g., volume of data, velocity of real time fraud, variety of data pattern), many CNPs prefer this approach to detect and prevent frauds [18].

There are also few specific anti-fraud tools which were developed by applying data mining techniques. These includes “FraudBuster”, “AT first”, “araxxe” and “SIGOS”).

- ***FraudBuster*** was created in 2010 and its practicality is on interconnect bypass frauds detection for mobile networks. By using a disruptive big data technology, it continuously monitors the CDR of a mobile traffic to detect fraudsters in real time. Its detection module is assumed to be flexible, allowing for rules to be created or updated at any time to counter the new tricks that fraudsters may use. It has also its own integrated control layer for triggering an update or creation of detection rules [23].
- ***SIGOS*** is an automated end-to-end active testing, application experience testing and fraud detection of telecom networks, services, and applications. Unlike FraudBuster, SIOGS focuses also on QoS besides to handling interconnect bypass frauds. To detect such frauds, it has tried to embed an artificial intelligence and machine learning.

ii. **Analysis:** there is no an absolute detection and/or prevention approach for interconnect bypass frauds (telecom frauds in general). For this matter, most CNPs do a certain analysis to identify the damage faced and design a recovering way, including the following.

a. ***Hotline:*** many telecom operators have designed their confidential channel for real time assistance and responses to any fraud kinds. In the center (telecom service provider), there would be a personal agent in addition to an interactive voice response (IVR). For a predefined characteristic of a fraud, victims or those who had the information of fraudulent could directly report to the IVR system. Otherwise, the IVR will lead them to the dedicated agents. In either of the two options, known potential frauds and misconducts will be collected and communicated to the analyzers. Then action to be taken might be disabling the reported service or conducting further analysis.

b. **Auditing Sub-Systems:** such systems are reasonably designed and developed based on the risks identified in advance. All activities are dealt in depth to the nature and degree of the risk involved.

2.3 Cellular Networks

Cellular networks are victims of interconnect bypass frauds. Normally, VoIP has a fundamental role to most of the telecom frauds especially to the three interconnect bypass techniques (SIM-Boxing, OTT, and gateway terminations).

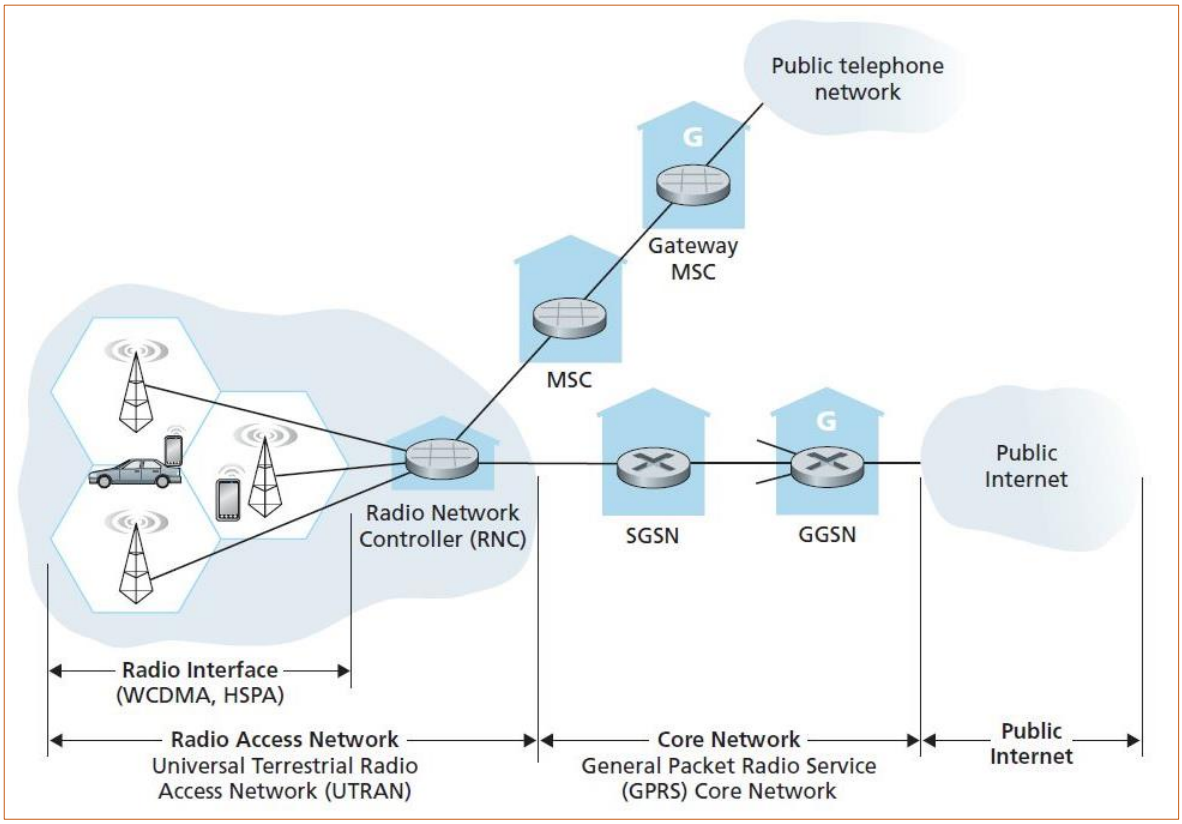


Figure 2.5: *Cellular Network (3G) System Architecture*

Even though there are different way of classification, cellular network elements could be classified as core and access network based on their functionality. The network elements of the two parts are shown in Figure 2.5 [14].

a) The Core Network Unit

The core network fundamentally processes the switching and routing of all voice calls and data. It interoperates with components of the existing cellular voice network. The core network by itself could be divided into packet switch (PS) and circuit switch (CS). The PS network unit, as shown in Figure 2.5 [14], includes a Serving GPRS Support Node (SGSN) & a Gateway GPRS Support Node (GGSN) while the CS network unit includes a mobile switching center (MSC) & a visitor location register (VLR). In common, both the CS and PS have a home location register (HLR) and equipment identity register (EIR).

- **SGSN:** traces the location of a mobile station and performs network access control, packet routing and transfer, mobility management, logical link management, radio resource management, network management and security.
- **GGSN:** is a gateway that connects the PS with external packet-based network (i.e. the Internet).
- **MSC:** is an interface between the mobile communication system and the CS network (e.g., PSTN, ISDN) with functions: user authorization, setup calls and handoff. It is co-sited with VLR.
- **VLR:** is a dynamic user database, containing an entry for each mobile user that is currently in the portion of the network. It obtains all necessary information from HLR. If the mobile subscriber leaves the control area of one VLR, that subscriber will be registered to another VLR. Then, the temporary data of that user will be deleted from the first VLR.
- **HLR:** a central static database of all mobile subscribers' complete information. It also supports online modifying & querying of users' location information in addition to location updating, call processing, authentication, supplementary service, and the like.

b) The Radio Access Network (RAN)

RAN is found between mobile subscribers and the core network in which the included elements are different based on the so-called mobile network generations (Gs').

≡ **In 2G:** a base station controller (BSC) that assigns base transceiver station (BTS) radio channels to mobile users, performs paging and handoff. Paging is the process of finding the mobile users' current cell while handoff is a situation where a mobile station maintains its association during a call (between channels in the same cell, cells in the same/different BSC, cells under the control of the same/different MSC and also between 2G, 3G, 4G, and 5G).

≡ **In 3G:** radio network controller (RNC) which controls several BTS (NodeB).

≡ **In 4G:** evolved NodeB (eNodeB) that manages the radio resources including radio bearer control, radio admission and connection mobility control.

Table 2.1: *RAN in Different Standards and Generations'*

	2G	3G	4G	5G
Basic RAN elements	BSC BTS	RNC NodeB	eNodeB	
Access methodology	TDMA	CDMA	FDM + TDM (OFDMA)	FDMA + TDMA (OFDMA)
Services	Voice	Voice + Data	Voice + Data	Voice + Data

Table 2.1 summarizes progress of radio access network ends, access methodologies and services with regard to different network generations (G). Since the middle of the 3rd generation, the PS and CS have been merged to one called "Evolved Packet Core" (EPC). The major concern of EPC is to manage network resources for a higher QoS, to make a solid line between the network control and user data planes. Thus, the core unit of a cellular network has become IP based so that both voice and data will be carried only in IP datagrams [14].

This tends us to discuss about the voice over Internet protocol (VoIP) in a cellular network.

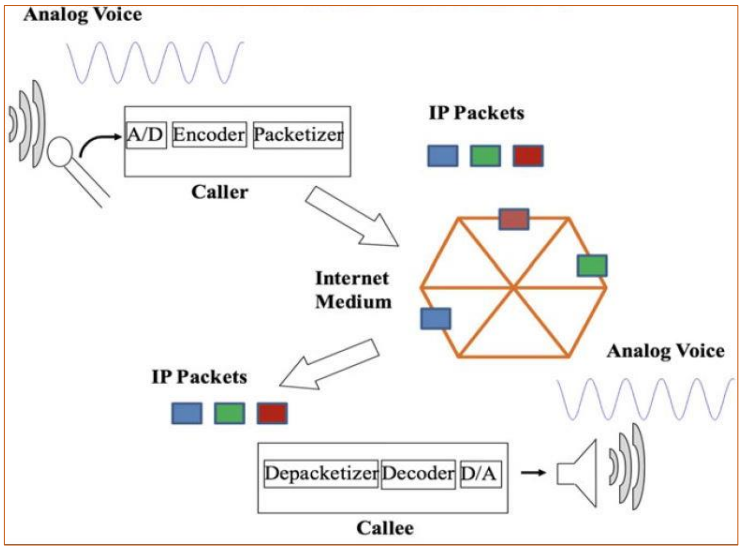


Figure 2.6: *End-to-end VoIP Communication*

VoIP is one result of technological advancement in communication protocols and techniques of transmission used for conversational voice to through any IP based network for the purpose of: cost reduction, value added services, anytime-anywhere communication, ease of deployment and simplification of transport protocol [2, 14]. It basically requires, as shown in Figure 2.6 [2], digitizing the analog voice and sending it in the form of IP packets over the IP-based network. Recalling that IP provides a best-effort service, besides to the previously mentioned advantages, VoIP also inherits IP known limitations - packet loss, jitter and end-to-end delay [14].

Therefore, we should take in mind here is that implementing the techniques of VoIP by itself is not a fraud. Because, CNPs and respective operators themselves are re-deploying their infrastructure by merging the circuit switched and packet switched network unit into one. Any type of content (text, video, image, voice) could be communicated via the all-purpose packet switched network. However, as discussed in Section 2.2, if it is used to access CNPs' network in unauthorized manner for any intention, it becomes an absolute fraudulent.

2.4 Fuzzy Logic

In this thesis, we prefer to investigate the use of fuzzy logic for improving decision-making in detecting interconnect bypass fraud. Because, Boolean logic might worsen the rate of false positive or false negative results. Therefore, it is significant to discuss about the fuzzy logic and how it could be associated with our method.

In the classical or Boolean logic, the truth values of variables are either 0 (false) or 1(true). Fuzzy logic, on the other hand, is a many-valued logic which is supposed to handle even uncertainties so that all real numbers between 0 and 1 could be considered. For this reason, fuzzy sets are said to be mathematical representations of vagueness and inaccurate information [16]. For example, assume we want to evaluate the quality of an IP traffic, in the case of classical logic, based on some measurements (jitter/ packet loss/ latency), the result would be 1 to imply "the quality is good" or 0 to mean "it is bad".

But, how about if the quality was partially good (between good = 0 and bad = 1)? Will we make the right decision? Fuzzy logic has solved this [16].

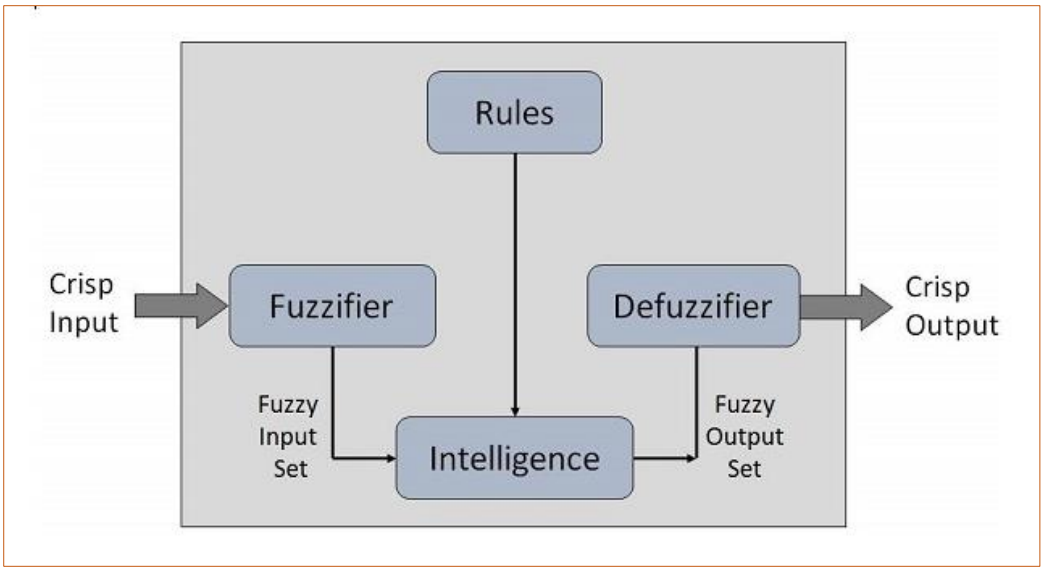


Figure 2.7: *Architecture of a Fuzzy Logic*

In a fuzzy logic, partially ordered scale of truth values (degrees) are allowed with a graded approach in which the smallest degree being 0 and the largest one being 1. It also respects the possibilities of human decision-making levels: certainly yes, possibly yes, cannot say,

possibly no, certainly no [16, 24]. Figure 2.7 [16] shows components of a fuzzy logic system, where the basic operations at each part are briefed as follows.

1. **Fuzzifier:** the raw value which is taken directly from sensors or measurements is a crisp value. For a fuzzy logic process, this value needs to be converted into fuzzy sets. The set holds, mostly, five variables (i.e., values of large positives, medium positives, small value, medium negatives and large negatives).
2. **Rules:** this component consists of asset of "IF-THEN" rules. This will govern the decision to be made by a particular system.
3. **Intelligence:** it is also known as a knowledge base that helps to simulate the reasoning process of humans.
4. **Deffuzifier:** fuzzy sets are confusing to understand. Therefore, the set that was formed in the proceeding components should be converted back into a crisp value once after the point value decision-making is known.

CHAPTER THREE

RELATED WORK

Many approaches, recommendations and algorisms have been proposed for the sake of interconnect bypass fraud detection. Commonly, most of them are dependent on CDR. Most of them take collection of CDR as input and came up with their conclusions. Mostly, these conclusions were drawn by applying some of the data mining techniques on the input data, but after the attack is done. The other common feature that many of the existing fraud management systems and trends shared is generalization. However, it seems impractical to bring a unified algorithm for various forms of telecom fraud. For example, the mechanism which was effective to handle SMS related frauds may not prevent from PBX hijacking or phishing. However, there are also some previous works which have dealt more specifically on interconnect bypass frauds.

As described in [14], an intrusion detection system (IDS) is the process of monitoring the events occurring in a network and analyzing them for signs of possible threats of violation of access policies. This includes detecting malwares, unauthorized access of attackers and attempts to gain additional privileges. Having this in mind, in our review, most of the previous related studies are just imitations of a network IDS, specifically falling in a signature-based approach.

As one methodology of IDS, signature-based approach is the simplest approach which takes some pattern or behavior which could lead us to threat identification [14]. For interconnect bypass fraud detection/prevention also, prior solutions took a dataset pattern as signature and followed that certain pattern to detect suspicious activities.

In order to detect interconnect bypass frauds, the authors of [19] began from the approaches they thought popular; “call generation analytics”, “call database analytics” and “hybrid analytics”. Then, by assuming as there will be the same person at least in one side of a call, they proposed to come up with a detection system based voice recognition of the caller. The system basically has two phases: training phase and verification phase. In its training phase, it is expected to build a database that contains voices of each speaker (caller). These sample

voices will be taken from any time of each call. This makes a process to build the database and storing all the speakers on it. Once if the caller is already registered, the system will increase the usage number of a specific number (SIM).

In its verification phase, the system considered three variables on a specific SIM card: M, T and F. Where M represents the speaker having largest number of calls, T represents the total number of calls by all speakers who used that SIM and F represents the value determined by implementing speech recognition procedure ($0<F<1$). If M is less than F*T, the authors thought that SIM is an irregular (fraudster). Else, it is regular (not a fraudster).

However, we have observed three major limitations.

- Privacy will be totally disregarded. Normally, the authors have recommended the providers to implement an encryption even though they did not specify how to do so. By encrypting the record even, it might keep away from outsiders (but not from CNPs).
- Due to hugeness of records in the database, the performance is under doubt.
- By using a single number (SIM) and even in a single call, there may be more than one speaker. For example, all family members could have only one mobile and if a member make a call to their relatives, that mobile could rotate to other members. This implies, according to [19], this family will be considered as fraudsters.

On the other hand, the authors of [25] applied both Artificial Neural Network (ANN) and Support Vector Machine (SVM) as classification techniques by extracting nine features of data from CDR (total calls, total called numbers, total minutes, total night calls, total numbers called at night, total minutes at night, total incoming calls, called numbers to total calls ratio and average minutes). Then, they developed ANN and SVM models from all possible combinations of the nine parameters. Finally, they experimented both models and concluded as SVM approach was 99.06% accurate for classification which is supposed to be vital for detection.

However, here also, the dataset was collected only from CDR so that we have doubts on the result for the following reasons.

- CDR by its nature has many user groups and interfaces (e.g., privileged customers, security offices, salesperson, customer care offices and many others can access it) which makes exposed it to be easily manipulated by fraudsters.
- False negative result may occur. Because, all or some of the applied parameters could be fabricated by the fraudster. For example, fraudsters could falsify the system and hide themselves by making fake calls to each other (i.e., member SIMs of a SIM-Box may call each other).

On the other hand, in [26], a passive detecting system was proposed to block bypass frauds at the network edge instead of access edge elsewhere. The authors called it as "ammit" and it took raw voice which was generated from caller and being transmitted through cellular networks. Immediately after receiving the audio, a measurement of its degradation would follow by using a call fingerprinting system known as "Pindr0p". All audios are supposed to be characterized and that will be used as a base for detection architecture they have designed. As a result, both the concealed and unconcealed packet losses will be detected.

At this stage, detection of SIM-Boxers is possible but with significant false positives. To minimize the rate, the authors iteratively applied an Algorithm called Fast Fourier Transform (FFT), having a capability to analyze audio on a real time. Finally, the authors did an experiment on real SIM-Boxes which results 87% of effectiveness with no false positive.

The main strength of this solution, proposed in [26], is its consideration as fraudsters may be aware of its detection techniques. If fraudsters knew it, they would not have any excuse to escape it. So, the authors placed it at the core network edge such that it would be unreachable. However, as noted also in [19], the computational cost will be a critical issue.

The authors of [27] proposed an approach, known as mobile phone detection system, that analyzes the user's calling behavior by SVM in addition to fuzzy clustering. Here also, the main input was CDR from which the authors took five relevant features. Then, by taking this five attributes, they have applied a fuzzy clustering technique to develop the SVM classifier.

This approach has a profile builder module and a fraud detector module. In the profile builder, the relevant features (i.e., cell ID, date and time of the calls, duration of the calls, type of the call, frequency of a call) will be filtered out from CDR by employing principal component

analysis. Once they have extracted the attributes, they have classified the call records into training and testing datasets. Then, in the fraud detector module, they have employed different types of SVMs (e.g., library for SVM, least square SVM) and different fuzzy clustering methods (e.g., fuzzy c-means and fuzzy k-means).

Consequently, the authors of [27] has an algorithm that takes CDR as an input and passes through the following basic steps in order to get the results of fraud detection.

i. Perform dimensionality reduction
ii. Perform feature extraction to build a user profile
iii. Partition the data-set into two groups – training data and testing data
iv. Perform c-means and k-means clustering on training data
v. Choose a sub-cluster of the clusters and denote this as SUB
vi. For each data point x that belongs to SUB, set its fuzzy membership to 1
vii. For a data point that does not belong to SUB, find out the cluster whose center is closest to x and calculate fuzzy membership of x with this cluster.
viii. Assign fuzzy membership values to the clusters
ix. Determine the tuning parameters C and σ
x. Train the SVM classifier using the clusters
xi. Test it with the trained SVM classifier model

In general, the following gaps in the existing approaches need to be addressed.

- ✓ Commonly, CDR was taken as an input even though it can't be enough as the fraud is supported by VoIP features for easier fabrication [12].
- ✓ Prevention requires manual operation for reacting even if the detection succeeded.
- ✓ Unreasonable computational cost.
- ✓ Rate of false positive and false negative is high.

CHAPTER FOUR

PROPOSED SOLUTION: IBFD SYSTEM

In this chapter, the structure and behavior of the proposed solution (i.e., Interconnect Bypass Fraud Detection (IBFD) System) is discussed. Identified system components, data flow and the interaction between identified components are described in addition to the nature and management of persistent data.

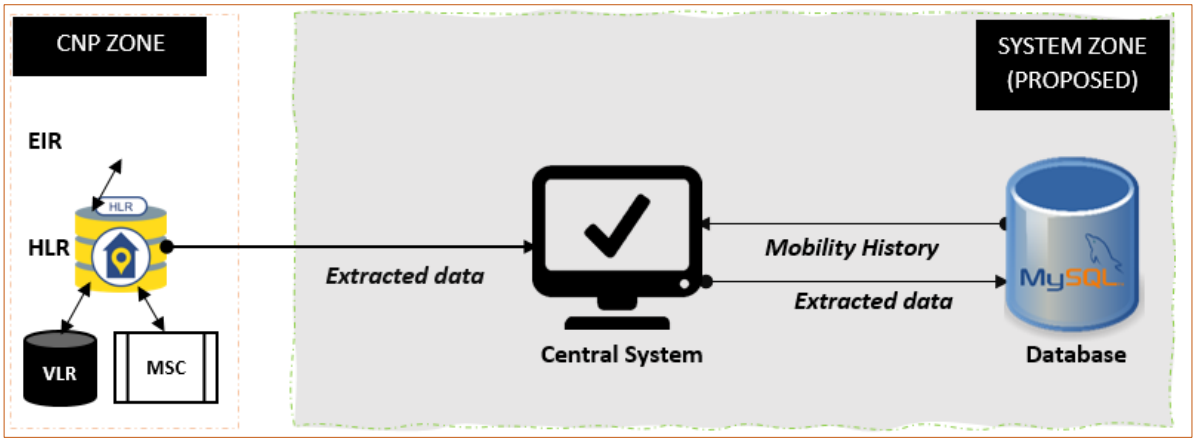


Figure 4.1: *Proposed System Architecture*

For detecting interconnect bypass frauds, we have proposed a new method which has a direct interaction with HLR and indirectly with VLR and MSC. As shown in Figure 4.1, the system extracts some relevant attributes (which will be described later in this chapter) and stores it in a separately designed database. The database is used to keep records of mobility history of each mobile user, and the system components should maintain a constant interaction with the HLR and the database as required. To identify whether a user is mobile or immobile, a fuzzy logic is applied by selectively taking the latest histories of location updates and considering the observable location variations. Then, based on the result from the logic, the activity would be then flagged (suspected, legitimate or fraud), where further reactions will take place based on the flag value.

In addition to the elements of a cellular network which were briefed in Section 2.3, area locating parameters (such as cell, routing area, MSC/VLR area, PLMN) and identifiers (such as IMSI, MSISDN & MSRN) are considered.

Current location of a mobile subscriber is maintained by HLRs and VLRs [14]. This implies, it is possible to get a raw input from the CNP zone containing identities, authentication data, location updates and all subscription information (subscribers' name, subscribed services, charging agreements, and so on). Hence, the system can extract significant data to store the location updates dynamically each time when a call is made. On the other hand, other parameters (e.g., IMSI, IMEI) are statically recorded only once.

Major components of the proposed solution are briefly shown in Figure 4.1 while the details, interfaces and subcomponents are presented next in Sections 4.1 and 4.2.

4.1 Central System Architecture

As shown in Figure 4.1, our central system is the major component in the solution that we proposed. It will be interacting both with HLR and the interconnect bypass fraud detection database (IBFD-DB). It would get relatively a rough input from the HLR and transforms it into meaningful records. Then, by applying a fuzzy logic, these would make the mobility management possible.

Mobility management by itself is an existing trend in cellular networks since it is important to make the access and handover processes as ubiquitous as it must be. For example, handover or delivering calls will not be successful unless the current location of the targeted mobile was tracked. Normally, the existing mobility management is used to identify only the current location. In our case, we have recorded each of the location histories, as an opportunity to determine mobility.

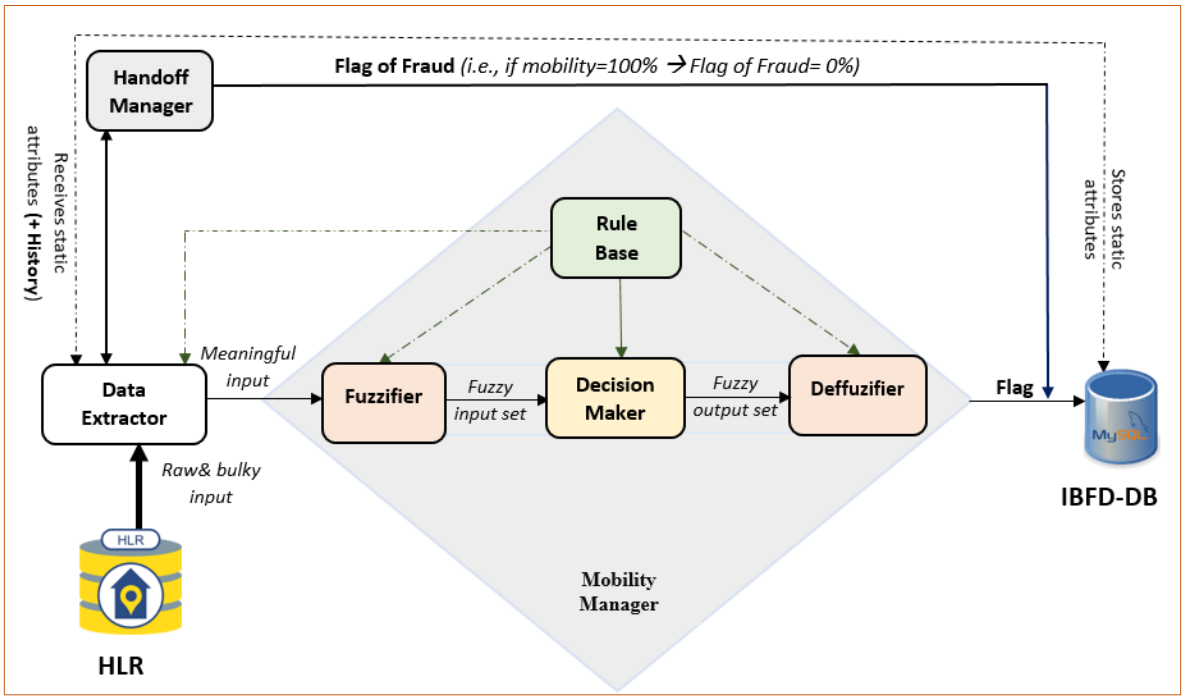


Figure 4.2: *Components of the Central System and Flowing of Data*

The detailed modules in the central system architecture, which are manipulating to the location information with the intention of identifying mobility, are presented in Figure 4.2 where the initial input is taken from the HLR (i.e., indirectly from MSC and VLR) and the final decision would be returned also to the HLR. In general, based on its mobility history, our system will make the HLR to be intelligent whether the caller has to be allowed for using the cellular network or not.

4.1.1 The Data Extractor

In a single telecom service provider alone, there are millions of users with various kinds of service subscriptions. HLR is responsible also for storing all that, and by its nature, it is expected to interact with many databases and network elements. However, HLR stores much information than what is relevant for the proposed system - even though, relevant for the cellular service. Hence, we included the data extractor to filter out and make the input as purposeful as required in our mobility checker module.

```
1  check system connection to HLR;
2  if connected
3     receive call setup alert;
4        check if initiated call is online;
5              if online
6                 until the initiated call ends
7                    remark = handoff-manager (MSISDN);
8                       if (remark = "extremely mobile")
9                          update fraud flag of MSISDN 0%;
10                         end;
11                      else
12                         check if HLR is connected with respective VLR and MSC;
13                            if connected
14                             read MSISDN, MSC Number, LAI, time, IMEI state from HLR;
15                             check if subscribers' profile was stored into the DB;
16                               if already profiled
17                                register only current IMEI and LAI aside to its history;
18                               else
19                                register MSISDN, MSC Number, LAI, time, IMEI state to DB;
20                             end;
21                else
22                    send notification for the network operator;
23                    goto 7;
24                else
25                   goto 1;
26 else
27    check if connection interface is up;
28       if down
29          alert for interface maintenance;
30          goto 1;
31       else
32          re-establish a connection;
33          goto 1;
```

Algorithm 4.1: *Algorithm for Extraction of Relevant Parameters from HLR*

The data extractor module, uses Algorithm 4.1 to filter only the fields which are relevant for our system (i.e., MSISDN, MSC Number, LAI, time, IMEI Status) from hundreds of the attributes in HLR. The data extractor will be executed for each mobile subscriber and for each call. Even in a single call, most of the time a handoff or handover may occur as the caller is possibly moving while a call is in progress. This should be checked iteratively as indicated in line 7 of Algorithm 4.1. The specific type of occurred handoff will be identified by another subcomponent which we named it as a "handoff manager". For example, if it was occurred because of a call transfer between channels (time slots) which means in the same cell, we should skip since it has nothing for detecting the users' mobility.

```
1    receive MSISDN and area number;
2    extract MSC Area, BSC area, cell from the area number;
3    compare the current BSC area or MSC area to the previous;
4         if changed
5              remark = "extremely mobile";
6         else
7              remark = "unknown"
8    Return remark;
```

Algorithm 4.2: *Handoff Manager*

On the other hand, as indicated in line 7 and 8 of Algorithm 4.2, if it was occurred since there was a handover between two cells in the same MSC a fuzzy logic must be employed. Because, cells could be changed even the user was stationed in the same place, due to many possible reasons like:

- If that cell fails
- If the number of users using that cell reaches its maximum capacity
- The user is in the overlapping coverage area of two or more cells and if there is a better quality in the other cell

4.1.2 The Fuzzifier

As shown in line 17 and 19 of Algorithm 4.1, histories of unprocessed locations are extracted and stored in the database. In the fuzzifier, raw inputs are going to be transformed into fuzzy sets. This will be the beginning point of the fuzzy logic employment to our system by considering the history of location area IDs as captured inputs.

From the list of LAIs (i.e., line 7 of Algorithm 4.3), by comparing each of consecutive LAIs of a specific MSISDN, it is possible to calculate how much was the percentage of location variance. It would be 100% if all the LAIs that the user was positioned while making calls are completely different, and, 0% if totally the same. If the resulted percentage (line 17 of Algorithm 4.3) is between 0 and 100, the user is somehow mobile, implying a fuzzy logic is required to be applied.

```
1    LAI_history = query LAI history of MSISDN from database;
2    CallCount= count of LAI_history;
3         if (CallCount <= 50)
4              flag = "white";
5              Variation_Percentage = 100%;
6         else
7              randomSelection (50, LAI_history);
8              Variation = 0;
9              i = 1;
10                  for j = i+1 upto 50
11                       if (randomSelection[i] = randomSelection[j])
12                            Variation = Variation + 0;
13                       else
14                            Variation = Variation + 1;
15                  i = i + 1;
16                  j = j + 1;
17             Variation_Percentage = (Variation / 50) * 100;
18             flag = "unknown";
19   return (flag, Variation_Percentage);
```

Algorithm 4.3: *Location Variance Calculation and Input Setting for the Fuzzifier*

A phase where the inputs of the fuzzifier are expected be achieved by Algorithm 4.3. In this Algorithm, all locations (LAIs) of a given MSISDN is searched from the database shown in Figure 4.2. Then, the records are counted and assigned to a variable named as CallCount.

For example, Figure 4.3 shows LAI history of user for an MSISDN (251901008029), which is extracted as the sample to explain the case.

```
310123124, 310123124, 310124357, 310123124, 310123124, 310121233, 310121233,
310121233, 310121233, 310125509, 310125509, 310123124, 310091190, 636011101,
636011101, 636011101, 636011101, 636012336, 636012336, 636013900, 636011102,
636010009, 636010009, 636019282, 636019282, 636016744, 636016744, 636010451,
636012202, 636012202, 636019110, 636019110, 636019110, 636019110, 636019110,
636013903, 636012129, 636011110, 636014550, 636014550, 636014550, 636014550,
636014550, 636014550, 636011105, 636011105, 636011101, 636011101, 636016660,
636013200
```

Figure 4.3: *Sample Content of LAI history for a Sample MSISDN*

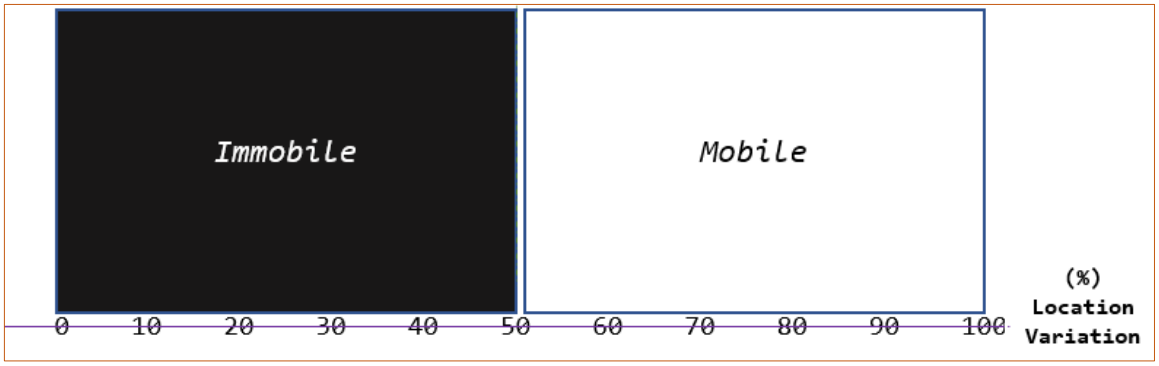
Using the sample shown in Figure 4.3 and applying Algorithm 4.3; 24 location variation can be detected while the number of LAIs is 50. Therefore, the variation percentage became 48%,

implying its fuzziness (in order to identify the users' mobility, a fuzzy logic needs to be applied).

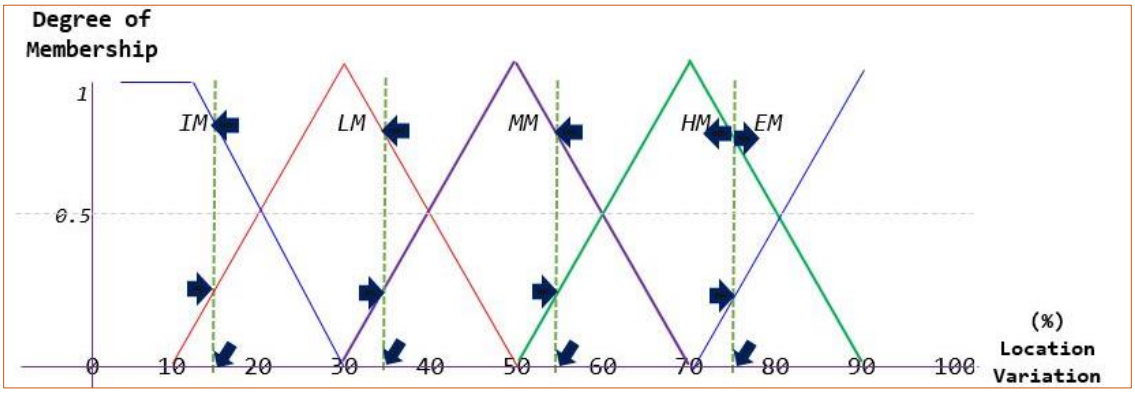
At this stage, a fuzzy set will be defined. A fuzzy set is fundamentally characterized by its membership function and it is a pair of U and m where: -

- U is the universal set of all possible location variations (0% to 100%),
- $m: U \rightarrow [0, 1]$ is a membership function,
- For each x ε U, the value of $m(x)$ is known as degree of membership of x in (U, m),
- The function $m = \mu_A$ is called the membership function of the fuzzy set $A = (U, m)$.

Each element of U is mapped to a value between 0 and 1, which indicates the membership value and quantifies the degree of membership of the element in U to the fuzzy set A.



(a) *Membership Function of Classical Set Ū*



(b) *Membership Function of Fuzzy Set U*

Figure 4.4: *Comparison Between Membership Function of Classic and Fuzzy Sets*

If our choice was a classical logic (instead of a fuzzy logic), the decision would be as Boolean as the one shown in Figure 4.4(a).

Based on our discussion on Section 2.5, we assumed fuzzy logic imitates a human like decision since it includes a range of possibilities between YES and NO. Commonly, based on their level of sense, decisions of humans will lay in one of the five possibilities. These are: *"absolutely yes", "possibly yes", "cannot say", "possibly no"* or *"absolutely no"* [16]. By taking the parallels of these possibilities, as shown in Figure 4.4(b), we have five linguistic variables which are mapped from the percentages of location variations where IM, LM, MM, HM and EM are to mean immobile, less mobile, moderately mobile and highly mobile and extremely mobile respectively. The vertical line in Figure 4.4(b) represents a specific percentage of variation that the three truth values indicate (the right, left and down indicating arrows in Figure 4.4(b)). So, each of the location variations are evaluated as per their membership to the five functions (IM, LM, MM, HM & EM). For example, in the first broken vertical line where the location variation is at 15%:

- ✓ The down indicating arrow on the bottom of Figure 4.4(b) (at 0) implies it has a zero membership in the fuzzy sets MM, HM and EM.
- ✓ The right indicating arrow (at 0.2) implies this variation (i.e., 15%) has a slight membership to LM.
- ✓ The left indicating arrow (at 0.8) implies this is visibly IM.

Therefore, if the LAIs of MSISDN varies by 15%, it has 0.2 membership in the fuzzy set "less mobile" and 0.8 in "immobile". The same is true for others.

As it became a common technique [16, 24] and as defined in equation (1), we have defined the fuzzy sets by using a triangular membership function so that each value will contain: -

- ➲ A slope where the value is increasing (lower limit),
- ➲ A peak where the value is equal to 1 (peak value),
- ➲ A slope where the value is decreasing (upper limit).

$$\mu A(Xi)\begin{cases} 1 & if\ Xi < \alpha \\ \frac{Xi-\alpha}{\beta-a} & if\ \alpha \le Xi < \beta \\ \frac{c-Xi}{c-\beta} & if\ \beta \le Xi < c \\ 0 & if\ c \le Xi \end{cases} \quad (1)$$

where α (lower limit), β (peak value), and c (upper limit) are elements of U and they are the parameters of the membership function while μ_A (X_i) is the membership function of X_i in A and μ_A is the degree of membership of X_i in A. For example, from Figure 4.4 (b), let us see the first triangle in which the values of α, β and c are 10%, 30% and 50% respectively.

► Let the value of X_i be 20%.

► Based on equation (1), since the value of Xi fall between α and β,

$$\mu A(Xi) = \frac{Xi-\alpha}{\beta-a}$$

$$= \frac{20-10}{30-10}$$

$$\therefore\ \mu_A(X_i) = \underline{\mathbf{0.5}}$$

For any fuzzy set A, especially in a triangular membership function, any x which is an element of U that satisfies $\mu_A(x) = 0.5$ is called a crossover point [16].

```
1     /* NOTATION: TMF() = Triangular Membership Function ()
         a = Lower Limit, b = Peak Value, c = Upper Limit */
3     TargetVariation = calculateLocationVariation(LAI₁ upto LAIₙ);
4     x = TargetVariation;
5     for i = 10 upto 90
6           a = i;
7           b = a + 20;
8           c = b + 20;
9           if x < a
10                TMF(x; a,b,c) = 1;
11          else if a <= x < b
12                TMF(x; a,b,c) = (x-a)/(b-a);
13          else if b <= x < c
14                TMF(x; a,b,c) = (c- x)/(b-a);
15          else if c < x
16                TMF(x; a,b,c) = 0;
17    Return TMF(x; a,b,c);
```

Algorithm 4.4: *Membership Function of the Fuzzy Set*

The degree of membership is mapped to a value not less than 0 and not greater than 1. By taking the value of variation, in which its degree of membership was above 0.8, the following rationalizations will be the result:

I. If variation is above or equal to 77%, extremely implies mobile (EM)

II. If variation is less than 77% but above 57%, highly implies mobile (HM)

III. If variation is between 36% and 56%, moderately implies mobile (MM)

IV. If variation is above or equal to 15% and less than 36%, less mobile (LM)

V. If variation < =15%, the user is immobile (IM)

4.1.3 The Rule Base

CNPs also have an equipment identity registry (EIR) which records the current status (i.e., blacklisted, whitelisted or unknown) of all the mobile devices. Each time when a call is made, the MSC requests the IMEI of the mobile station, which is then sent to the EIR for authorization and the status will be confirmed to HLR. Therefore, we let the rule base component to obtain the IMEI status from HLR (indirectly, from EIR) for improving the accuracy of our decision. So, in the rule base component, rules will be constructed based on the fuzzy set combined with IMEI status (by extracting it in the first module).

Table 4.1: *Matrix of Fuzzy Sets and IMEI Status*

Mobility / IMEI Status	EM	HM	MM	LM	IM
Blacklisted	Rule-3	**Rule-4**	**Rule-5**	**Rule-5**	**Rule-6**
Gray listed	Rule-2	Rule-3	**Rule-4**	**Rule-5**	**Rule-5**
Whitelisted	Rule-1	Rule-2	Rule-3	**Rule-4**	**Rule-5**

All possible combinations between the IMEI status and mobility status we considered are shown in Table 4.1.

Table 4.2: *List of Rules*

Rule	Condition	Action Flag
1	If mobility is extremely high ***AND*** IMEI Status is in a whitelist	(Highly Trusted)
2	(If mobility is high **AND** IMEI Status is in a whitelist) ***OR*** (If mobility is extremely high ***AND*** IMEI Status is unknown)	Trusted
3	(If mobility is moderate ***AND*** IMEI Status is in a whitelist) ***OR*** (If mobility is high ***AND*** IMEI Status is unknown) ***OR*** (If mobility is extremely high **AND** IMEI Status is in a blacklist)	Likely Trusted
4	(If mobility is less *AND* IMEI Status is in a whitelist) ***OR*** (If mobility is moderate AND IMEI Status is unknown) ***OR*** (If mobility is high **AND** IMEI Status is in a blacklist)	Likely Suspected
5	(If user is immobile ***AND*** IMEI Status is in a whitelist) ***OR*** (If mobility is less **AND** IMEI Status is unknown) ***OR*** (If mobility is high ***AND*** IMEI Status is in a blacklist) ***OR*** (If mobility is moderate **AND** IMEI Status is in a backlist)	Suspected
6	(If user is immobile ***AND*** IMEI Status is in a blacklist)	(Highly Suspected)

The rules that could be drawn from the matrix are shown in Table 4.2. Based on the matrix which was defined earlier, we would have a total of 6 rules (i.e., highly suspected, suspected, likely suspected, likely trusted, trusted and highly trusted).

4.1.4 Decision Maker

```
1     Connect to the rule base;
2     Get MSISDN, IMEI status;
3     Perform fuzzification to get fuzzy sets;
      /* notations:
      EM= "extremely mobile", HM= "highly mobile", MM= "moderately mobile",
      LM= "less Mobility", IM= "immobile", W= "whitelisted", U= "unknown",
      B= "blacklisted";
      */
9     If (Mobility=EM && IMEI_Status=w)
10          Flag= "highly trusted";
11    Else if ((Mobility=HM && IMEI_Status= W") || (Mobility=EM && IMEI_Status= U"))
12          Flag= "trusted";
13    Else if ((Mobility=MM && IMEI_Status= W") ||
        (Mobility=HM && IMEI_Status= U")) || (Mobility=EM && IMEI_Status= B"))
15          Flag= "likely trusted";
16    Else if ((Mobility=LM && IMEI_Status= W") ||
        (Mobility=MM && IMEI_Status= U")) || (Mobility=HM && IMEI_Status= B"))
18          Flag= "likely suspected";
19    Else if ((Mobility=IM && IMEI_Status= W") ||
        (Mobility=LM && IMEI_Status= U")) || (Mobility=HM && IMEI_Status= B")) ||
        (Mobility=MM && IMEI_Status= B") || (Mobility=IM && IMEI_Status= U")
22          Flag= "suspected";
23    Else if (Mobility=IM && IMEI_Status= B")
            Flag= "highly suspected";
```

Algorithm 4.4: *Algorithm Used by Decision-making Component*

The decision-making component is the component that applies the rule base in decision making. Based on the rules set on the rule base, as indicated on Algorithm 4.4, the decision-making component determines a flag to the user. The result at this stage will be a set of fuzzy outputs.

4.1.5 Deffuzifier

The fuzzy sets which are in the form of linguistic terms must be converted to a single crisp valued output. Though there are various possible deffuzification methods (e.g., center of sums, center of gravity, maxima methods, weighted average and so on), we have chosen to apply center of sums since it is commonly appropriate to fuzzy sets [28] and it can be easily applicable to our membership function (i.e., triangular).

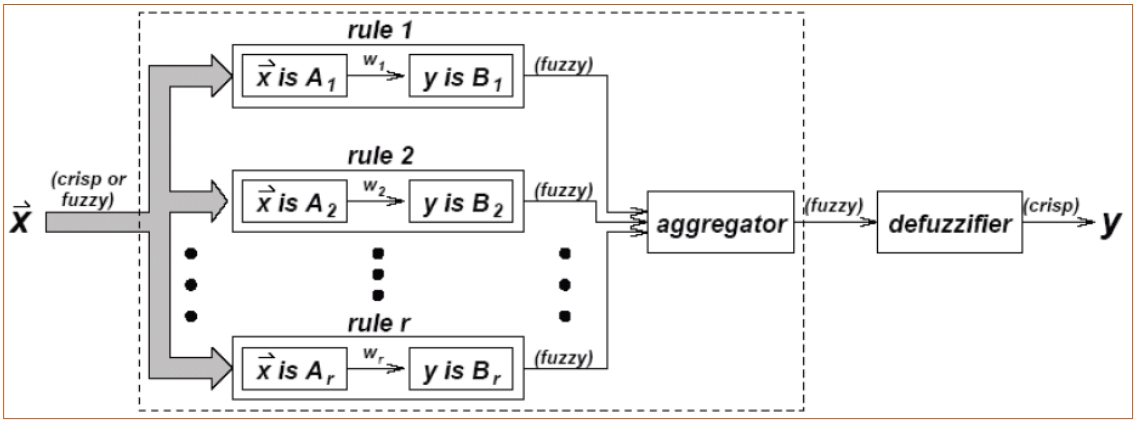


Figure 4.5: *Input/ Output Value Type Before/ After Deffuzification Module*

In the previous component, we have stated six rules as shown on Table 4.2. As Figure 4.5 [16] shows, the Deffuzifier needs all fuzzy results of rules to be aggregated in a single fuzzy value. We have replaced the OR & AND Boolean operations to MAXIMUM & MINIMUM values respectively.

- **Rule 1:** EM (mobility) **^** WL (IMEI_Status) ⇒ Flag (Highly Trusted)
  ⇒ **Highly Trusted =** MIN (0.77, 1) = **<u>0.77</u>**
- **Rule 2:** (HM(mobility) **^** WL (IMEI_Status)) **V** (EM(mobility) **^** GL (IMEI_Status))
  ⇒ Flag (Trusted)
  ⇒ **Trusted =** MAX (MIN (0.57, 1), MIN (0.77, 0.5)) = **<u>0.57</u>**
- **Rule 3:** MM(mobility) **^** WL(IMEI_Status)) **V** (HM(mobility) **^** GL(IMEI_Status)) **V** (EM(mobility) **^** BL(IMEI_Status))
  ⇒ Flag (Likely Trusted)
  ⇒ **Likely Trusted =** MAX (MIN (0.36, 1), MIN (0.56,0.5), MIN (0.77, 0))
  = **<u>0.5</u>**
- **Rule 4:** (LM(mobility) **^** WL(IMEI_Status)) **V** (MM(mobility) **^** GL(IMEI_Status)) **V** (HM(mobility) **^** BL(IMEI_Status))
  ⇒ Flag (Likely Suspected)
  ⇒ **Likely Suspected =** MAX (MIN (0.15, 1), MIN (0.36, 0.5), MIN (0.56,0))
  = **<u>0.36</u>**
- **Rule 5:** (IM(mobility) **^** WL(IMEI_Status)) **V** (LM(mobility) **^** GL(IMEI_Status)) **V** (MM(mobility) **^** BL(IMEI_Status)) **V** (IM(mobility) **^** GL(IMEI_Status))
  ⇒ Flag (Suspected)

⇒ **Suspected =** MAX (MIN (0, 1), MIN (0.15, 0.5), MIN (0.36, 0), MIN (0, 0.5))
= **0.15**

- **Rule 6:** IM(mobility) **^** BL(IMEI_Status) ⇒ Fraud (Flag)
  ⇒ **Highly Suspected =** MIN (0, 0) = **0**

$$x^* = \frac{x \sum_{i=1}^{N} x_i \ . \ \sum_{k=1}^{n} \mu_{A_k}(x_i)}{\sum_{i=1}^{N} \sum_{k=1}^{N} \mu A_k \ (x_i)} \qquad (2)$$

Where n is the number of sets, N is the number of fuzzy variables, and $\mu A_k(x_i)$ is the membership function for the k^{th} fuzzy set. By using Equation 2 [28], our crisp value can be calculated as follows.

Flag of Fraud = $\frac{(0.2*(15+35+50+55+75))+(0.8*(15+35+50+55+75))}{5}\%$

= 46%

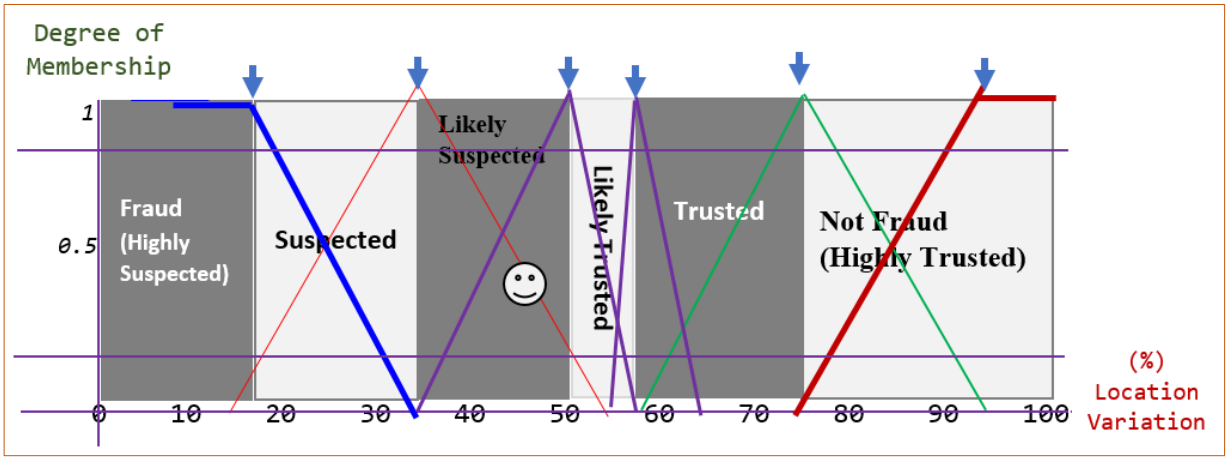


Figure 4.6: *Crisp Output of Deffuzification Module*

Figure 4.6 shows the resulted crisp value and deffuzification process. The output (46%) is the exact value in which the location varies. It was found by applying the center of sums method where the defuzzified value is defined as follows.

4.2 Database Design

As shown in Figures 4.1 and 4.2, the system incorporates two databases - one which already exists in telecom service providers (HLR), and other is the one that we introduced as IBFD-DB. Let us discuss in detail how HLR and IBFD-DB would have a role to our method.

4.2.1 HLR /VLR

This is a static database in which it is used to be as a heart of mobile network. Since it holds all the information, call routing will not succeed unless HLR exists in the middle. This implies, an outage problem with the HLR typically consequences a significant outage/problem for a mobile network. It is used to manage which types of calls is permitted for each user and how they are handled. Also, it could control access to additional services such as data and traces separate devices. Most probably, larger network providers could have multiple HLRs, but, each subscriber belongs to only a single HLR. As we have discussed briefly in the previous sections (particularly in Section 4.1, HLR by itself is almost irrelevant without VLR and MSC even though it is key to any mobile carrier's network. Therefore, HLR is a database containing hundreds of attributes including those which we took as essential to our system as a result.

4.2.2 IBFD-DB

This is the database that we had introduced in which attributes and their respective records will be extracted directly from HLR. It is a standalone relational database having its own tables and holding the following basic attributes.

Attribute	IMSI	MSISDN	MSC Number	LAI	Time	IMEI	IMEI Status
Data Type	Float	Varchar	Float	Float	Timestamp	Float	Varchar

Where: -

a) IMSI: international mobile subscriber identity which is composed from the following.

- *MCC* - Mobile Country Code
- *MNC* - Mobile Network Code
- *MSIN* - Mobile Subscriber International Number

b) MSISDN: Mobile station international subscriber directory number which is primary key to the HLR record and with the following components.

- *MCC* - Mobile Country Code
- *NDC* - National Destination Code
- *SN* - Subscriber Number

c) MSC Number: indicating which MSC is responsible for a specified MSISDN.

- *MCC* - a country code that specifies the country where an MSC is located.
- *NDC* - National Destination Code
- *LSP* - is a locally significant part that is assigned by service providers.

d) LAI: a unique identifier for a specific location area within a known PLMN and it is structured as follows.

- *MCC* - Mobile Country Code
- *MNC* - Mobile Network Code
- *LAC* - Location Area Code.

e) Time: is a time stump indicating a moment of recording.

f) IMEI: is a 15-digit serial number that is assigned by the manufacturer of the device and it is a concatenation of three things.

- *Type Allocation Code (TAC):* is an 8-digit number which is a composition of reporting body identifier (2digits) and manufacturer identifier (6digits). It is assigned by the international GSM association based on the request from device manufacturers.
- *Serial Number (SNR):* is a 6-digit serial number to uniquely identify the unit of a specific model
- *Checking Digit (CD):* is used to assure that IMEI number is not adjusted on the network.

g) IMEI Status: a one that is expected to hold a remark indicating if the mobile device is blacklisted or not.

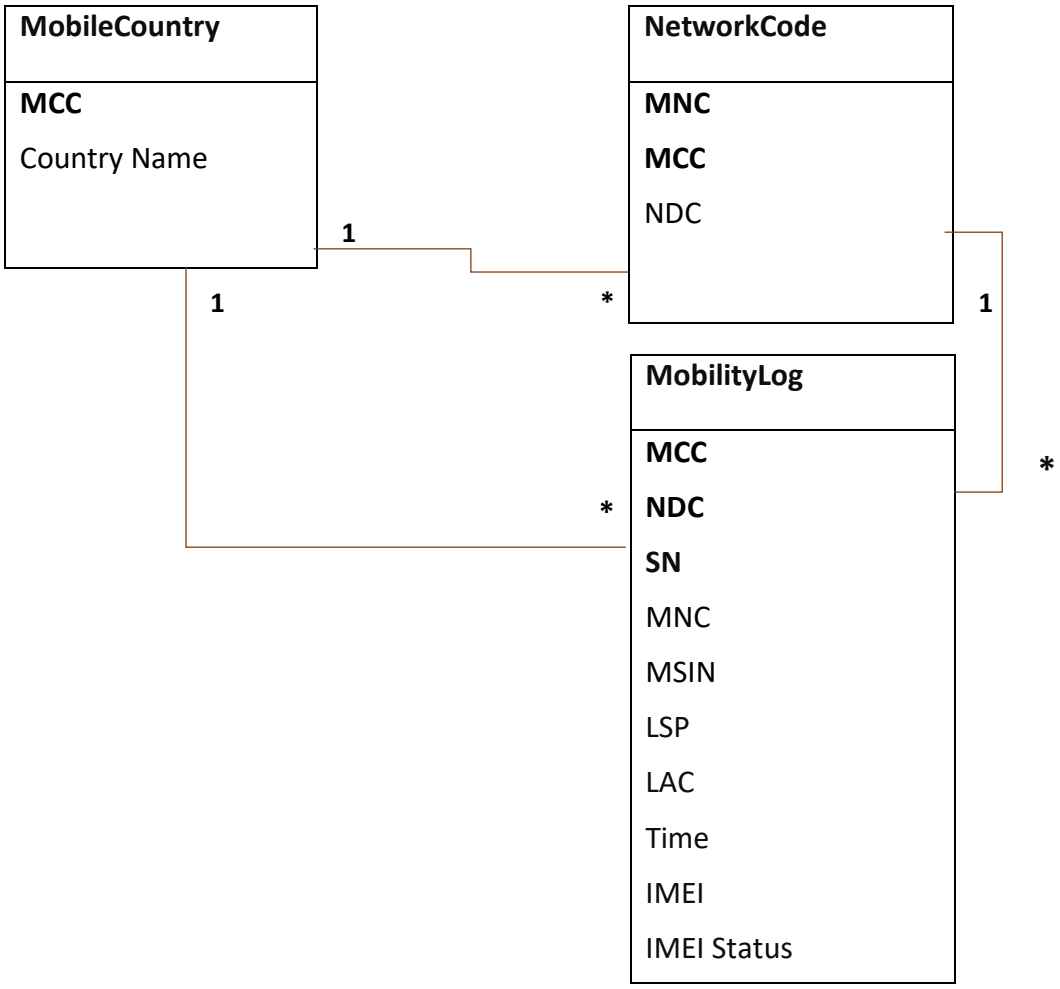


Figure 4.7: *IBFD-DB Design*

Therefore, even the extracted data contains a data redundancy and anomalies. In order to reduce that, normalization should be implemented. After normalizing it into a first, second and third normal forms, we got three tables with respective attributes as shown in Figure 4.7.

CHAPTER FIVE

PROTOTYPE AND EVALUATION

In order to evaluate the IBFD method, a prototype has been developed and evaluation is done using sample data from location registry. In this chapter, the tools that we have used for prototyping with respective justifications are described first. Then, the experimental settings, procedures and results are discussed.

The high level architecture of the system environment where IBFD method is exercised, which is shown in Figure 4.1 of Chapter 4, consists the HLR. The existing database (HLR) is simulated for evaluation purpose. In addition, almost all of the components useful for the proposed solution ((i.e., IBFD method)) shown in Figure 4.2 are developed and experimented.

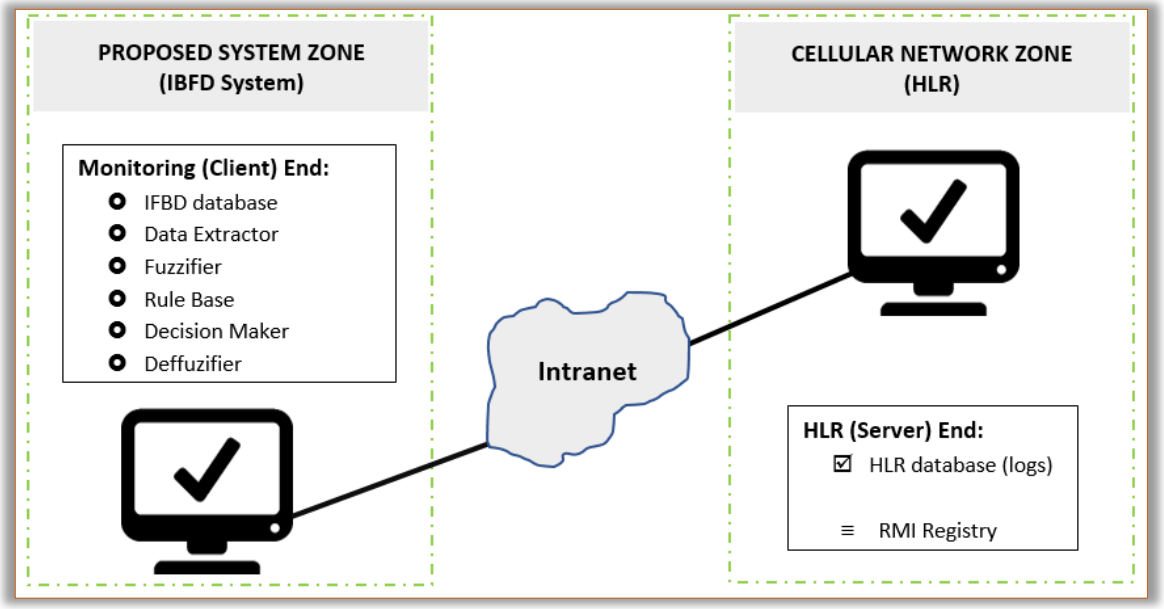


Figure 5.1: *Experimental Setup IBFD System Using RMI Model*

Since handoff occurs only in an active cellular network, we could not include the handoff manager component in the simulation process. From an active HLR of ethio telecom, 1037 logged data are sampled, and used in the simulated HLR. And, a remote method invocation (RMI) is used to run the HLR-like and our IBFD method from separated computers as shown in Figure 5.1.

5.1 Tools & Programing Languages

As listed in Table 5.1, we have used the various tools and programing languages to evaluate the method.

Table 5.1: *Hardware and Software Selections for Simulation*

Software: Tools & Programing Languages		
#	**Tool/Programing Language**	**Purpose/Reason of Selection**
1	NetBeans IDE 8.2	It is an integrated development environment (IDE) that we have used to develop our main system. We have selected it from other related IDEs for the following reasons. ▪ Free and open sourced, ▪ Easy to manage and to build the graphical user interface, ▪ Extensible platform and supports even for non-java codes
2	MS Excel 2016	To export randomly selected sample records from real HLR database.
3	Microsoft SQL Server Management Studio 2018	→ To import the sample data (from MS Excel) → To develop our main database which we named it as IBFD-DB → We chose it from other database development tools because: ▪ It enables to import data easily from MS Excel ▪ It is a free, familiar and streamlined one having a built-in transparent data compression feature besides to its encryption
4	RapidMiner	It is a data mining tool which we have used it to evaluate our solution by analyzing the system outputs.

Hardware: Experimental Settings		
Device Name	**Specification/Parameters**	**Description and Reason of Selection**
Personal Computer (Quantity 2)	*OS:* MS windows 10, *RAM:* 8GB *Processor:* 2GHz	Used for developing and running our newly introduced database and central system.

5.2 Experimental Procedure

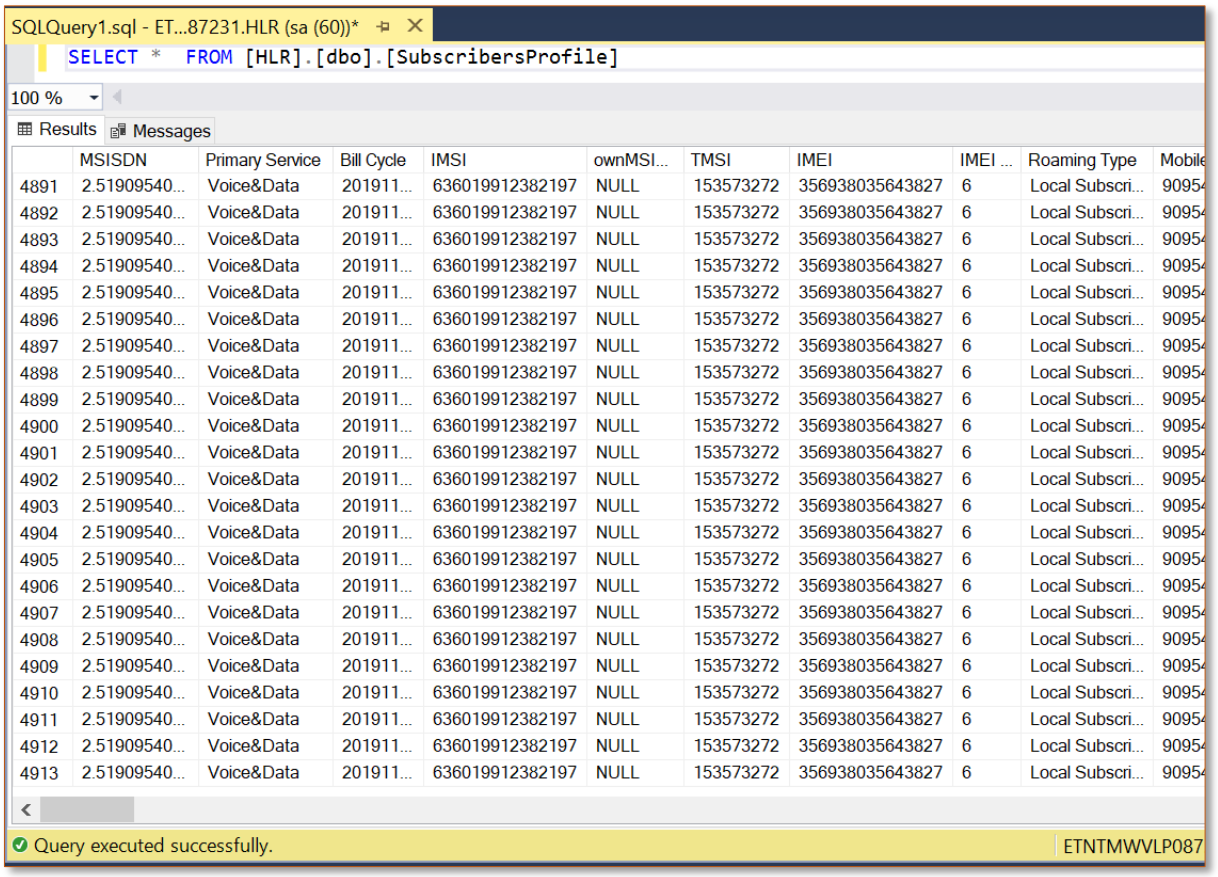
There were many free virtual environments to experiment our solution (e.g., MAPS, N3 and others to simulate the cellular network). Also, Telecom Excellency Academy (TExA) of ethio telecom has a deployed testbed for all cellular network elements (e.g., Huawei HSS9860 for

HLR). However, even though these devices have the functionality of cellular network elements, they do not carry existent (live) subscribers. For this reason, we used HLR testbed only for partially confirming the integration of our system and we did the actual experiment by using a logged file.

We began the core experiment by collecting 1037 randomly sampled static profiles of unique MSISDNs. Then, we recorded 50 sample history logs for each at a random time. Hence, the total logged sample is 51,850.

The size of sample record history (i.e.,50) and size of sample HLR entry (i.e., 1037) are randomly selected with the following assumptions and justifications.

- ≡ As far as it is a significant user (its call history is constructed from more than 50 calls), a mobile subscriber is assumed to change its location at least once in his/her call history.
- ≡ Random sampling is used so that subscribers have no information as their location history is being logged. For this reason, they don't have a chance to compromise our assumption instead of accessing the cellular network with their actual intentions.
- ≡ Since most of the subscribers are supposed to reside at home or one place (i.e., no location update would be observed), logs of the night-time (6:00PM – 6:00AM) were not included in the sample.



	MSISDN	Primary Service	Bill Cycle	IMSI	ownMSI...	TMSI	IMEI	IMEI ...	Roaming Type	Mobile
4891	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4892	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4893	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4894	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4895	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4896	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4897	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4898	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4899	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4900	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4901	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4902	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4903	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4904	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4905	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4906	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4907	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4908	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4909	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4910	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4911	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4912	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095
4913	2.51909540...	Voice&Data	201911...	636019912382197	NULL	153573272	356938035643827	6	Local Subscri...	9095

Figure 5.2: *Some of Sample Histories Recorded from Existent HLR*

Since it is not permitted (i.e., for the sake of security) to directly connect our system with the existent HLR, our implementation was done by making it standalone. Inputs that we intended to get from HLR was stored in Microsoft Excel spreadsheet, and then, the records are imported into Microsoft SQL Server Management Studio. As Figure 5.2 partially shows, the recorded samples that we hold in spreadsheet and imported into SQL Server Management Studio.

```
public class IBFPSysImplmentation extends UnicastRemoteObject implements IBFPSymInterface{
  Connection con;
  public IBFPSysImplmentation() throws RemoteException
  {}
                                                                                                    /* Data Extraction */
  public String dataExtraction (String ServiceNum) throws RemoteException
  {
   String remark="";
    Float VLR, MSCNumber, LAI,  IMEI;
    String IMEIStatus, timeOfLog;
    int counted=0;
    try
    {
                                                                                  /* Step-1: Check system connection to HLR  */
    Class.forName("com.microsoft.sqlserver.jdbc.SQLServerDriver");
    con = DriverManager.getConnection("jdbc:sqlserver://localhost:1433;databaseName=HLR;user=sa;password=TAD;");
                                                          /* Step-2: Check if HLR is connected with responsible VLR and MSC    */
    PreparedStatement CountData=con.prepareStatement("SELECT * FROM SubscribersProfile WHERE MSISDN=?");
    CountData.setString(1, ServiceNum);
    ResultSet CountDataResult=CountData.executeQuery();
    while(CountDataResult.next())
      {          counted=counted+1;         }
    if(counted==0) {  remark="not";  /* Implies, not connected to the responsible VLR/MSC */    }
    else
      {
                        /* Implies, connected to the responsible VLR/MSC. So, Proceed to extract all relevant information from HLR */
        PreparedStatement HLRData=con.prepareStatement("SELECT * FROM SubscribersProfile WHERE MSISDN=?");
        HLRData.setString(1, ServiceNum);
        ResultSet HLRDataResult=HLRData.executeQuery();
        while(HLRDataResult.next())
        {
          VLR =  HLRDataResult.getFloat(18);
          MSCNumber = HLRDataResult.getFloat(16);
          LAI = HLRDataResult.getFloat(12);
          IMEI = HLRDataResult.getFloat(7);
          IMEIStatus = HLRDataResult.getString(43);
          timeOfLog = HLRDataResult.getString(17);
                                                                                    //inserting it into Mobility_History table
          PreparedStatement ExtractedData=con.prepareStatement("INSERT INTO History VALUES (?,?,?,?,?,?,?)");
          ExtractedData.setString(1, ServiceNum);
          ExtractedData.setFloat(2, VLR);
          ExtractedData.setFloat(3, MSCNumber);
          ExtractedData.setFloat(4, LAI);
          ExtractedData.setFloat(5, IMEI);
          ExtractedData.setString(6, IMEIStatus);
          ExtractedData.setString(7, timeOfLog);
          ExtractedData.executeUpdate();
        }
        remark="yes";
      }
     con.close();
    }
    catch(Exception ex)
    {
      System.out.println(ex.getMessage());
    }
    return remark;
  }
```

Figure 5.3: *IBFD System Implementation [Partial View]*

Our system, as discussed in Chapter 4, has six basic components. Each of the components have their own algorithm (as show in Algorithms 4.1, 4.2, 4.3 and 4.4). Consequently, Figure 5.3, partially shows the implementation code of those algorithms.

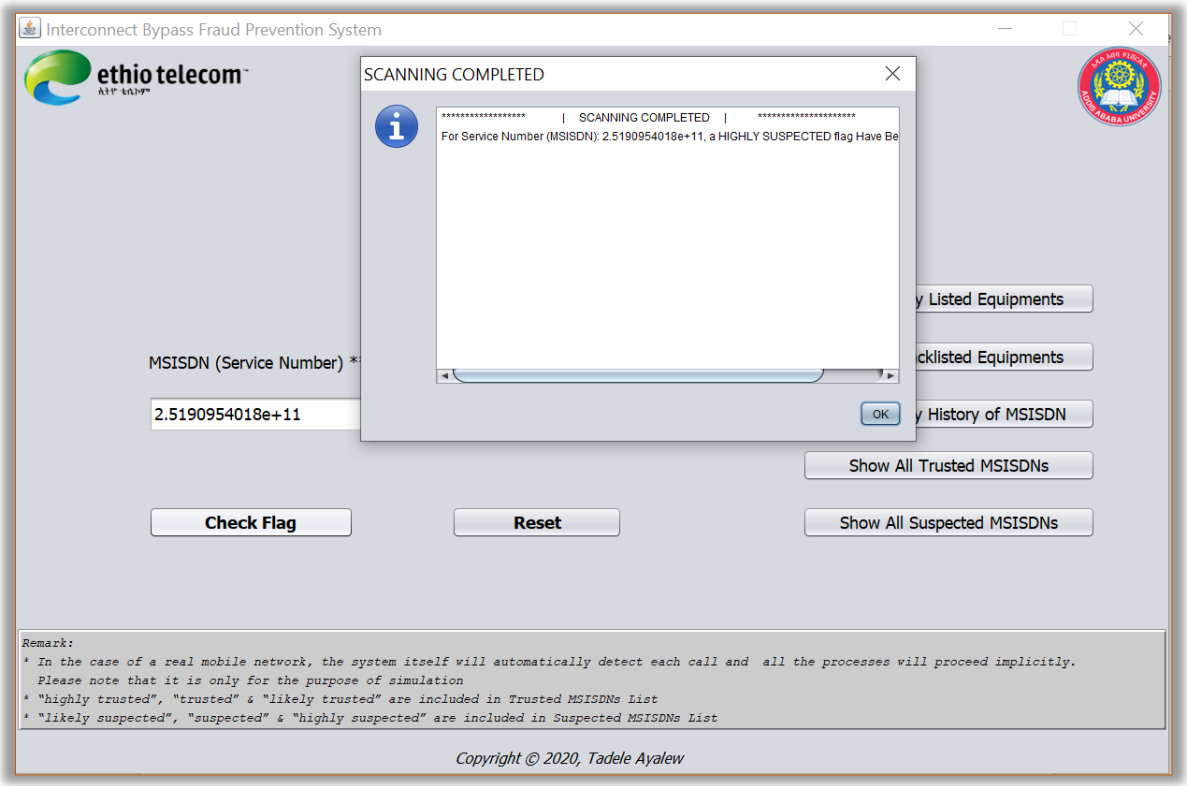
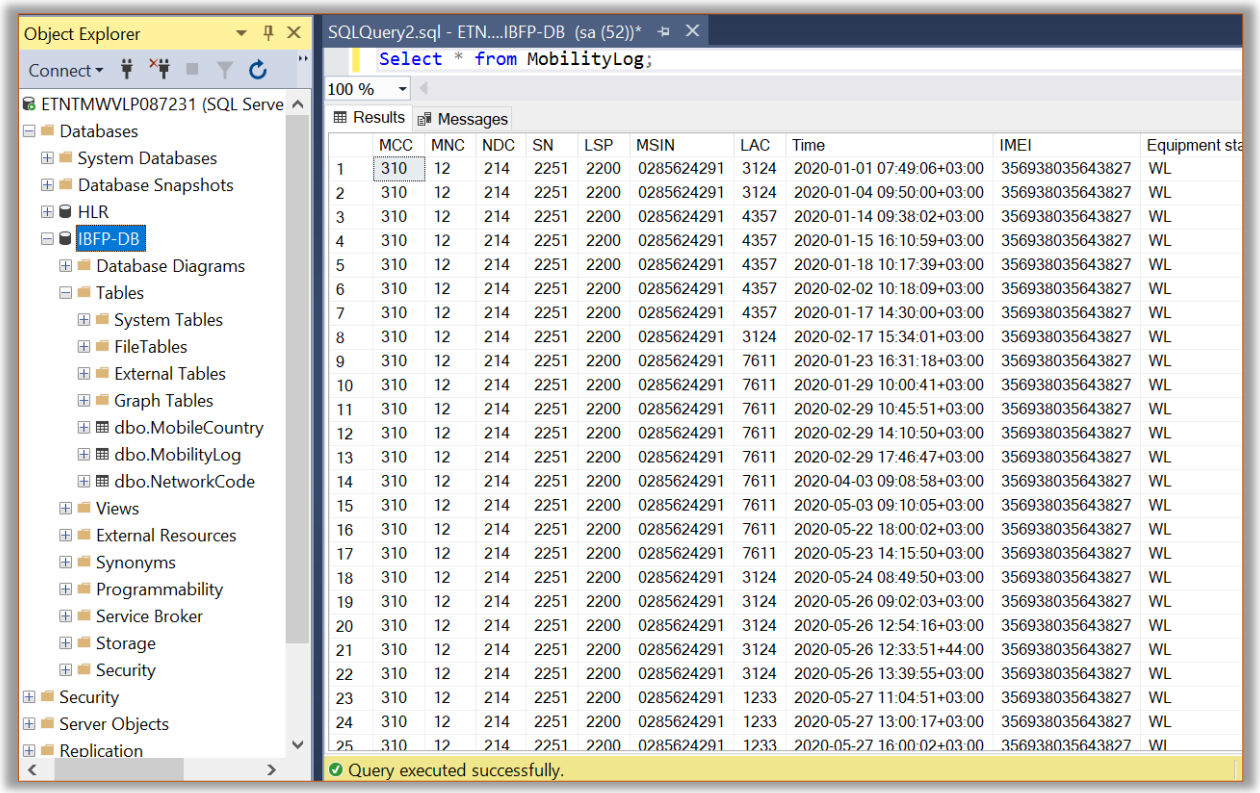


Figure 5.4: *Monitoring-end Side User Interface*

The screenshot which is displayed in Figure 5.4 shows the monitoring end (client) side graphical user interface. Normally, this figure shows how to check the legitimacy of a single user manually. However, in the case of a real mobile network, the system itself will automatically detect each call and all the processes will proceed implicitly. During the evaluation, however, it is required to monitor the detection hence the user interface is created.



	MCC	MNC	NDC	SN	LSP	MSIN	LAC	Time	IMEI	Equipment sta
1	310	12	214	2251	2200	0285624291	3124	2020-01-01 07:49:06+03:00	356938035643827	WL
2	310	12	214	2251	2200	0285624291	3124	2020-01-04 09:50:00+03:00	356938035643827	WL
3	310	12	214	2251	2200	0285624291	4357	2020-01-14 09:38:02+03:00	356938035643827	WL
4	310	12	214	2251	2200	0285624291	4357	2020-01-15 16:10:59+03:00	356938035643827	WL
5	310	12	214	2251	2200	0285624291	4357	2020-01-18 10:17:39+03:00	356938035643827	WL
6	310	12	214	2251	2200	0285624291	4357	2020-02-02 10:18:09+03:00	356938035643827	WL
7	310	12	214	2251	2200	0285624291	4357	2020-01-17 14:30:00+03:00	356938035643827	WL
8	310	12	214	2251	2200	0285624291	3124	2020-02-17 15:34:01+03:00	356938035643827	WL
9	310	12	214	2251	2200	0285624291	7611	2020-01-23 16:31:18+03:00	356938035643827	WL
10	310	12	214	2251	2200	0285624291	7611	2020-01-29 10:00:41+03:00	356938035643827	WL
11	310	12	214	2251	2200	0285624291	7611	2020-02-29 10:45:51+03:00	356938035643827	WL
12	310	12	214	2251	2200	0285624291	7611	2020-02-29 14:10:50+03:00	356938035643827	WL
13	310	12	214	2251	2200	0285624291	7611	2020-02-29 17:46:47+03:00	356938035643827	WL
14	310	12	214	2251	2200	0285624291	7611	2020-04-03 09:08:58+03:00	356938035643827	WL
15	310	12	214	2251	2200	0285624291	7611	2020-05-03 09:10:05+03:00	356938035643827	WL
16	310	12	214	2251	2200	0285624291	7611	2020-05-22 18:00:02+03:00	356938035643827	WL
17	310	12	214	2251	2200	0285624291	7611	2020-05-23 14:15:50+03:00	356938035643827	WL
18	310	12	214	2251	2200	0285624291	3124	2020-05-24 08:49:50+03:00	356938035643827	WL
19	310	12	214	2251	2200	0285624291	3124	2020-05-26 09:02:03+03:00	356938035643827	WL
20	310	12	214	2251	2200	0285624291	3124	2020-05-26 12:54:16+03:00	356938035643827	WL
21	310	12	214	2251	2200	0285624291	3124	2020-05-26 12:33:51+44:00	356938035643827	WL
22	310	12	214	2251	2200	0285624291	3124	2020-05-26 13:39:55+03:00	356938035643827	WL
23	310	12	214	2251	2200	0285624291	1233	2020-05-27 11:04:51+03:00	356938035643827	WL
24	310	12	214	2251	2200	0285624291	1233	2020-05-27 13:00:17+03:00	356938035643827	WL
25	310	12	214	2251	2200	0285624291	1233	2020-05-27 16:00:02+03:00	356938035643827	WL

Figure 5.5: *Mobility History Log Samples of IBFD-DB*

As per our database design and our data extraction component of our main system, we have populated IBFD-DB. The database has three tables and in Figure 5.5, the records of mobility log have been shown.

5.3 Results and Discussion

In this section, the results of our experiment are discussed with respect to the research objective. The experiment conducted after implementing the fraud detection system by applying fuzzy logic to conclude the mobility of a subscriber is evaluated by comparing it with mobility trends. Also the accuracy has been measured to determine the false positive/negative rate.

5.3.1 Assessing Mobility Trends

To learn the mobility trends, we considered the mobility history of sampled mobile subscribers with respect to the five Representative Functions as shown in Table 5.2.

Table 5.2: *Mobility Trend of Sampled Mobile Subscribers*

Representational Function	Number of Mobile Users	Share (%)
Immobile	8	0.77
Less Mobile	26	2.51
Moderately Mobile	105	10.13
Highly Mobile	349	33.65
Extremely Mobile	549	52.94

As a result, only 0.77% of the samples were identified to be immobile users (stationed in one place throughout their call-history) while 2.51%, 10.13%, 33.65% and 52.36% were identified to be less, moderately, highly and extremely mobile, respectively.

Table 5.3: *Association of Fraudster Distribution with Representational Functions*

Rules	***IM***	***LM***	***MM***	***HM***	***EM***	***Total***
Highly Suspected	7	5	1	0	0	**13**
Suspected	1	15	2	5	0	**23**
Likely Suspected	0	4	13	9	2	**28**
Likely Trusted	0	1	72	22	23	**118**
Trusted	0	1	16	188	86	**291**
Highly Trusted	0	0	1	125	438	**564**
Total	8	32	105	349	543	**1037**

For the IBFD method, we have implemented six rules (i.e., highly suspected, suspected, likely suspected, likely trusted, trusted and highly trusted). These rules are used to evaluate each of the representational functions (i.e., IM, LM, MM, HM and EM). As shown in Table 5.3, a "highly suspected" flag have been observed for 87.5% of the immobile (IM) users while 79.78% of the extremely mobile (EM) users were not fraudsters ("highly trusted"). In the same way,

- From those which have a minimum variation between their location histories (LM): 19.23%, 57.69%, 17.38%, 3.85% and 3.85% were flagged as “highly suspected”, “suspected”, “likely suspected”, “likely trusted” and “trusted” respectively.
- From moderately mobile (MM) users: 0.95%, 1.9%, 12.38%, 68.57%, 15.24% and 0.95% were flagged “highly suspected”, “suspected”, “likely suspected”, “likely trusted”, “trusted” and “highly trusted” respectively.
- From highly mobile (HM) users: 0%, 1.43%, 2.58%, 6.3%, 53.9% and 35.8% were flagged as “fraud”, “suspected”, “likely suspected”, “likely trusted”, “trusted” and “highly trusted” respectively.

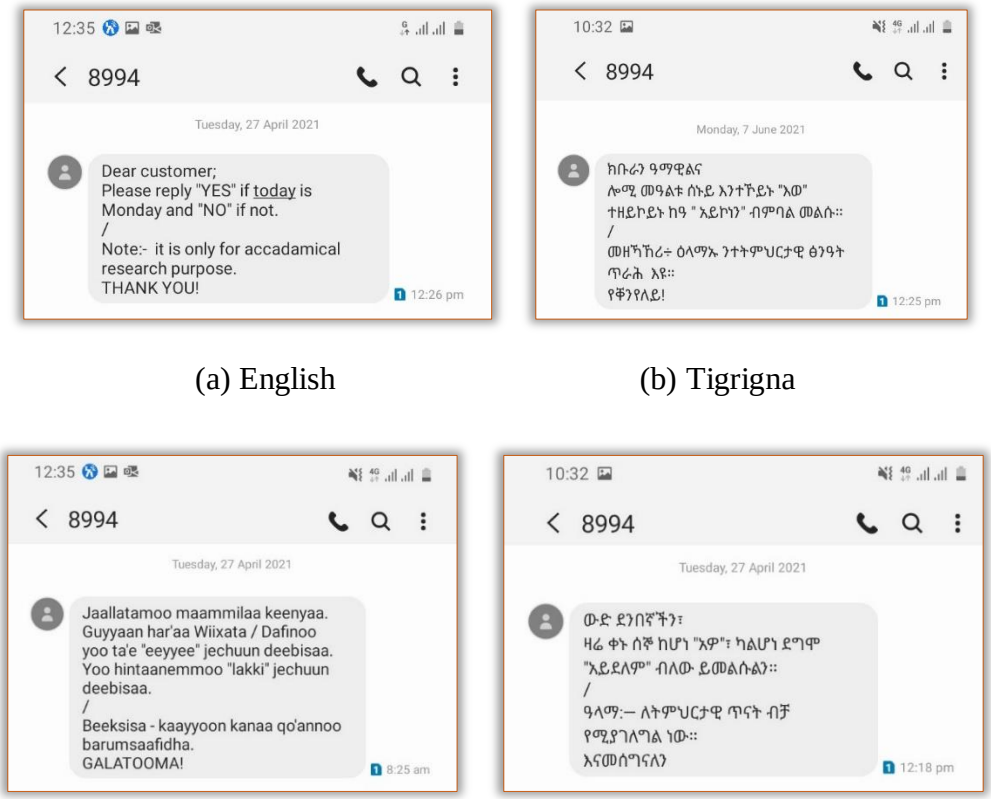
5.3.2 Rate of False Positive/Negative

The rate of false positive and false negative is major problem that we had sought to address. For this reason, we have evaluated if our method (fuzzy logic based mobility management) had reduced the rate significantly or not.

Considering the fact that interconnect bypass orchestrated by the use of mobile subscription in a fixed device, the immobility of a subscriber and the fuzzy logic leads to determines the subscription at different level of suspicion based on the predefined rules. To verify the correctness of the suspicion, we test if the sample subscriber is interactive or not. The assumption here is that if a subscription is used in a device it will not be interactive otherwise if the subscription is being used by real mobile user, he/she will be interactive.

Therefore, we test the interactive-ness of the subscriber by imitated execution of reCaptcha technique as first phase of the verification. By comparing the result of the method with the result of the interactive-ness the false positive/negative rate has been calculated. Then the result of interactive-ness test has been compared with the false positive/negative results of ethio telecoms’ FMS.

During the first phase, where we test the interactive-ness of the subscriber, we began the process by sending a bulk short message to all sampled mobile subscribers by intending to assess their legitimacy from the interaction behavior.



(a) English (b) Tigrigna

(c) Afan Oromo (d) Amharic

Figure 5.6: *Verification Short Messages*

As shown in Figure 5.6, the message was contented by a simple request which was written in the users' language preference (referred from CRM profile). Because, by simply taking the model of reCaptcha verification technique [29], it is assumed that robots (i.e., SIMBOX machines in our case) are not intellectual to understand human questions. In addition, for those who didn't reply to our first request, we gave additional two chances by resending the short message.

Table 5.4: *Phase I Result Versus Representational Functions*

	Total Excluded	Human-like Interaction	IVR-Like Interaction	Total Included	Included + Excluded
Highly Suspected	5	0	8	8	**13**
Suspected	2	0	21	21	**23**
Likely Suspected	4	1	23	24	**28**
Likely Trusted	5	110	3	113	**118**
Trusted	22	267	2	269	**291**
Highly Trusted	37	525	2	527	**564**
Total	**75**	**903**	**59**	**962**	**1037**

As shown in Table 5.4, only 7.3% (75/1037) were kept irresponsive while 92.77% (962 of 1037) have replied to our request. Since we didn't get any means to know why those 75 users didn't respond to our verification request, we excluded them and considered only those who had replied. So, we have considered only 962 of the sampled users for Phase-I verification, and then, we have categorized those 962 respondents into two. The first ones are humans (real users) which had interacted us like humans by responding "YES/NO" while others have responded what an IVR (SIMBOX machine) might do by responding with text out of context (e.g., "in God, we trust", "how can I help you", "good days for you", etc.).

Then, it was important to verify how many of the suspects become responsive or not. As a result, since five of highly suspected users were completely irresponsive, we have excluded lately from Phase-I verification process. Thus, only 8 of 13 have responded and considered. Again from those 8 interactive users, we have counted no users who had responded like human and eight users had interacted like IVR. Likewise, we got 21, 23, 3, 2, and 2 users interacted like IVR/SIMBOX from highly suspected, suspected, likely suspected, likely trusted, trusted and highly trusted users respectively.

Using the result of interactive-ness test, we calculated the rate of false positives and false negatives as follows.

a) False Positive Rate: the rate in which legitimate mobile users considered as fraudster.

$$\text{False Positive Rate} = \frac{\text{number of false positives}}{(\text{number of false positives}) + (\text{number of True negative})}$$

While we were experimenting our method; out of the sample data, 13 MSISDNs were flagged as "highly suspected", 23 were flagged as "suspected" and 28 were flagged as "likely suspected". So, when we associate this by looking the interactivity of users (manual verification) from the experiment result, we got the following results.

- **Highly Suspected:** based on our experimentation, 13 MSISDNs were highly suspected to a fraudulent (i.e., 7 from IM, 5 from LM & 1 from MM). On the other hand, in the Phase-I verification process also, we have noticed that all of the "highly suspected" users were not legitimate (interacted as IVR). This implies, the false positive rate is ***0%***.
- **Suspected:** we got no wrongly suspected user (false positive rate ***0%***).
- **Likely Suspected:** 1 out of 24 (likely suspected users) has interacted us like human and which results the false positive rate to ***4.17%*** (i.e., 1 ÷ 24).

This implies, besides to this meaningful result of each, when we look the average false positive rate of anyhow suspected (highly suspected, suspected and likely suspected) users, we got it **1.92%** (i.e., (0+0+1) ÷ (8+21+24)). But since we were not able to confirm for 75 of the sample subscribers, as they did not respond, we cannot conclude on the false positive rate to be only 1.92%. As there will be a room to detect the non-responded subscribers wrongly, the maximum rate could be 7.3%.

b) False Negative Rate: it is also known as "the miss rate" which is the proportion of actual fraudsters to be accepted as legitimate mobile user.

$$\text{False Negative Rate} = \frac{\text{number of False Negatives}}{(\text{number of False Negatives}) + (\text{number of True Positive})}$$

The number of "likely trusted", "trusted" and "highly trusted" MSISDNs were 118, 291, and 564 respectively. When we verify this result with the interactive-ness test, we got the following confirmations.

- **Likely Trusted:** the total number of "likely trusted" MSISDNs found in our experiment was 118. However, during the verification process, we got a response

from 113 users. So, when we compare the two results, we have noticed only three of them were missed.

➲ False negative rate (likely trusted) = 2.65% (i.e., 3÷113).

- **Trusted:** from the 291 trusted MSISDNs, 22 of them will not be considered since they have not responded (interacted) to the interactive-ness test. From the experimentally trusted and manually considered list, there were 2 users missed from detection out of 269.

  ➲ False negative rate (trusted) = 0.74% (i.e., 2÷269).

- **Highly Trusted:** Out of the 564 highly trusted list, 37 were excluded (since there was no response from them). Then, we have verified 525 as legitimate and 2 fraudulent users.

  ➲ False negative rate (highly trusted) = 0.38% (i.e., 2÷527).

This implies, the average false negative rate of anyhow trusted (highly trusted, trusted and likely trusted) users is **0.77%** (i.e., (3+2+2) ÷ (113+269+527)). Here as well, since we cannot conclude having those subscribers who did not respond to our message, there will be a room to not detect the non-responded subscribers while they are fraudulent. Hence the maximum rate of false negative could be 7.3%.

Therefore, from this phase, we got the two extreme ends reduced to 0% of highly suspected as false positive and 0.38% of highly trusted as false negative rates. Even if we consider the average rates of "anyhow suspected" and "anyhow trusted" users, we got **1.92%** of false positive and **0.77%** of false negative rates respectively.

However, even from those 75 users which were excluded since they didn't respond to our short messages, we are still uncertain. There is still a probability that some of them might be missed wrongly from detection (i.e., false negative), some of them detected wrongly (i.e., false positive) and/or some of them detected correctly (i.e., perfect). In order to furtherly identify the accuracy of these results, including the uncertain probabilities, we have gone through phase-II verification.

As we have discussed in Section 2.2, telecom service providers already have deployed their own FMS. Since our samples were randomly taken from the subscribers of ethio telecom, we thought we could compare at least the list that our system detects and ethio telecoms' FMS false positive result. However, the trend of ethio telecoms false positive identification basically depends on the users compliant. For example, if mobile subscriber X was detected and blocked from the cellular network though it was legitimate, then, subscriber X will complain and when the number gets verified in anyway, it will be resumed.

Table 5.5: *Phase-II Verification Result*

IBFD / ethio telecom	Phase-I Result (962)			Phase-I Excluded (75)
	False Positive	*False Negative*	*Detected Correctly*	
False Positive & Resumed	0	0	1	1
False Negative & re-blocked	0	1	1	0
Detected Correctly	1	4	203	13
Record not Found	0	2	749	**61**
Total	**1**	**7**	**954**	**75**

Due to the ethio telecoms' "complain-then-resumed" trend, we could not find a complete collection of falsely detected users. So, as shown in Table 5.5, only 21.7% of those 1037 mobile users were found in the output archives of ethio telecoms' FMS.

Since we are in an extra phase, the insufficiency of the FMS archive would not have that much impact to our conclusion. However, we don't know about the status of the lately excluded samples (61 or 5.88%). Some or all of them might be detected correctly. There is also a probability to fall either in false negative or false positive detections. Since we didn't get any way to confirm this, we have considered the maximum probability and obliged to put the results in range. Yet, after we assess the correctness of Phase-I with respect to Phase-II results, we got:

- ☑ False positive rate of **anyhow** suspected = **1.92% up to 5.88%** (i.e., average of Phase-I & Phase-II results. Because, as indicated on Table 5.5, the user that we thought we've falsely detected in Phase-I is found in the correctly detected list of Phase-II).

- ☑ False negative rate of **anyhow** trusted = **<u>0.88% up to 5.88%</u>** (Because, in Phase-I, we thought we have missed a total of 7 fraudulent users. However, in Phase-II, we have found another 1 missed user).

Anyhow suspected includes the suspected, likely suspected and highly suspected users while anyhow trusted includes trusted, likely trusted and highly trusted users of the experimental (our systems') result.

5.3.3 Our Method Versus Previous Approaches

There are medias, and official and unofficial reports that estimates the actual loss caused due to an interconnect bypass fraud is worst [7, 9]. Compared to those estimations, our method could be taken effective enough.

Table 5.6: *Comparison Between Fuzzy Logic Based Method and Others*

Metric / Approach	Implementation	Basic Input	FP & FN Rates
Speaker Recognition	Edge, Business	Voice Recording	Unidentified
Boxed Out	Edge, Operational	Transmitting Voice	"Small" **&** 13%
Machine Learning	Edge, Operational	Attributes of CDR	Unidentified
Fuzzy Logic Based	Core, Operational	LAI from HLR	[1.92%, 5.88%] **&** [0.88%, 5.88%]

As shown on table 5.6, we got a mutual metric to compare our method with "speaker recognition bas "machine learning based" [25] and "boxed out" [26]. As we have discussed on Se 2.4, the access and the core parts are major units of a cellular network. In addition, cellular network providers might have two types of systems; the one that supports to run the business (i.e., business support systems) and those which are essential for the technical operation (i.e., operational support systems). For example, the speaker recognition based approach [19] was implemented on the business support system of edge part while "machine learning based" [25] and "boxed out" [26] were implemented on the operational support system of the cellular networks edge part. On the other hand, our method (fuzzy logic based detection) is implemented on the core part (i.e., HLR) while others were proposed to be on the edge part (in between the access and the core).

CHAPTER SIX

CONCLUSION AND FUTURE WORK

In this Chapter, we have indicated how our objectives and research questions were addressed in a summarized way. We have also included the major contributions of our thesis. Finally, we have indicated our future activities.

6.1 Conclusion

In order to overcome the limitations of FMS, we have proposed to detect interconnect bypass frauds by monitoring the mobility of cellular network users by taking the visible signature differences of an interconnect bypass fraudster and legitimate mobile user. In addition, the existing cellular networks already have a two-tier database known as HLR and VLR for holding actual profile (including the current location) of mobile users.

However, since HLR and VLR would communicate only if the user enters from one wide location area to another, it doesn't show the specific location which might let us to make not the right decision. For this reason, we took the logs of locations in a cell level update in addition to the identified statuses found from EIR. Then, in order not to make a Boolean decision (i.e., simply "fraud" or "legitimate") we have applied a fuzzy logic which lets us to consider even the uncertainties.

In this thesis, the following achievements have been attained.

- An interconnect bypass fraud detection method has been developed with minimum complexities.
    - ☑ *Easy to implement:* it is checked by integrating on a cellular network testbed.
    - ☑ *Minimum cost:* since it stores a minimum amount of data and it will be integrated with existing databases and algorithms (i.e., HLR, fuzzy logic), the computational and implementation cost will be reduced.
    - ☑ *Better performance:* since it processes a comparatively simple data, it performs better with a reasonable throughput.

- We have detected an interconnect bypass fraudulent with a false positive rate of 1.92% and the false negative rate of our detection was also 0.88%. However, since we could not identify the actual legitimacy of 61 sampled users, the probability could be up to 5.88% for both cases.
- Since we proposed our method to be implemented on the core unit of a cellular network, it would not be as easy as CDR based methods to be discovered by the fraudsters.

6.2 Future Work

At the time of experimenting this thesis, we were limited to base on a single telecom service provider (i.e., ethio telecom). In addition, due to the security policies which were predefined by ethio telecom, we didn't test our method by directly connecting it to a live network. Rather, we took a logged data from the live cellular network. Therefore, experimenting it in a diversified cellular network (i.e., various service providers), a live network and a larger sample data is left for future.

Moreover, for keeping the false positive and false negative rates closer to zero, the method needs to be extended by including the following elements and features.

- During our analysis, even though we were successful on identifying highly trusted and highly suspected users, the false positive rates of suspected/likely suspected and the false negative rates of trusted/likely trusted users are observed weighty. So, in addition to managing the mobility of users, more signatures should be investigated in the future. Finding all potential signatures will make the method away from fraudsters and keep it undisclosed.
- The approach of prevention (reaction) is left for future.

References

[1] "World-Newspapers.com", [Online]. [Accessed 12 01 2021].

[2] Tamal ChakrabortyIti, Saha Misra and Ramjee Prasad, "VoIP Technology: Applications and Challenges", Vol. 10, New Jersey: Springer International Publishing AG, part of Springer Nature, 2019.

[3] Frehiwot Mola, "Analysis and Detection Mechanisms of SIM Box Fraud in The Case of Ethio Telecom", Unpublished Master Thesis, Department of Electrical and Computer Engineering at Addis Ababa University, Addis Ababa, 2017.

[4] Nazish Yaqoob, Seemab Latif, Rabia Latif, Haider Abbas and Asif Yaseen, "An Adaptive Rule-Based Approach to Resolving Real-Time VoIP Wholesale Billing Disputes", *Journal of information science and engineering,* Vol. 33, pp. 1433-1446, 2017.

[5] "Teletopix.org", Telecom Techniques Guide, [Online]. Available: http://teletopix.org/gsm/what-is-hlr-and-vlr-and-its-function-in-gsm/. [Accessed 5 December 2019].

[6] Ilsun You, Yuh-Shyan Chen, Sherali Zeadally and Fei Song, "A Brief Overview of Intelligent Mobility Management for Future Wireless Mobile Networks", *EURASIP Journal on Wireless Communications and Networkin,* No. 188, 2017.

[7] ethio telecom, "Three Years Strategy (2020 - 2023) of ethio telecom", ethio telecom, Addis Ababa, 2020.

[8] Godfred Yaw Koi-Akrofi, Joyce Koi-Akrofi, Daniel Adjei Odai and Eric Okyere Twum, "Global Telecommunications Fraud Trend Analysis", *Innovative Space of Scientific Research Journals,* Vol. 25, pp. 940-947, 2019.

[9] C. Gibson, "Cyber-Telecom Crime Report", Trend Micro Research, 2019.

[10] Kala N., "A Study on Internet Bypass Fraud: National Security Threat", *Forensic Res Criminol International Journal,* Vol. 7, No. 1, p. 31–35, 2019.

[11] "Fast money on SIMBOX fraud in Ghana: high rates of international calls and tax avoidance", ANTRAX, 19 March 2019. [Online]. Available: https://en.antrax.mobi. [Accessed 18 October 2019].

[12] Tewodros Hailu, "Network Traffic Classification Using Machine Learning: A Step Towards Over-the-Top Bypass Fraud Detection", Unpublished Master Thesis, Addis Ababa University, Addis Ababa, 2018.

[13] J. M. Kizza, Guide to Computer Network Security, Fourth Edition, Springer International Publishing, 2017, pp. 1-6.

[14] James F. Kurose and Keith W. Ross, Computer Networking: A Top Down Approach - 7th edition, New Jersey: Pearson Education, Inc., publishing, 2017.

[15] James Graham, Richard Howard and Ryan Olson, Cyber Security Essentials, Boca Raton: Taylor & Francis Group, LLC, 2012.

[16] Radim Belohlavek, Rudolf Kruse and Christian Moewes, "Fuzzy Logic in Computer Science", in *Computer Science: the Hardware, the software and the heart of it*, New York, Springer, 2011, pp. 385-419.

[17] "Population Proportion – Sample Size Calculators", Select Statistical Services Limited, [Online]. Available: https://select-statistics.co.uk. [Accessed 03 01 2020].

[18] ethio telecom, "Telecom Fraud Prevention Best Practices", Telecom Excellency Accadamy (TExA), Addis Ababa, Ethiopia, 2018.

[19] Osama Mohamed Elrajubi, Ali Mustafa Elshawesh and Mustafa Ali Abuzaraida, "Detection of Bypass Fraud based on Speaker Recognition", *International Conference on Information Technology (ICIT),* Vol. 8, pp. 50-54, 2017.

[20] "Fraud Managment”, SIGOS, 9 October 2019. [Online]. Available: https://www.sigos.com/fraud-management. [Accessed 29 February 2020].

[21] "GSM Gateway”, SysMaster, [Online]. Available: http://www.sysmaster.com/products/gsm_termination.php. [Accessed 27 December 2020].

[22] "Viber Out”, Viber, [Online]. Available: https://www.viber.com. [Accessed 27 December 2020].

[23] "FraudBuster”, [Online]. Available: https://www.fraudbuster.mobi/solutions/simbuster. [Accessed 22 February 2020].

[24] M. Sharma, "Fuzzy Logic System Architecture in Artificial Intelligence”, 15 June 2019. [Online]. Available: https://www.includehelp.com. [Accessed 25 April 2020].

[25] Mhair Kashir and Sajid Bashir, "Machine Learning Techniques for SIM Box Fraud Detection”, *International Conference on Communication Technologies,* pp. 4-8, 2019.

[26] Bradley Reaves, Ethan Shernan, Adam Bates, Henry Carter and Patrick Traynor, "Boxed Out: Blocking Cellular Interconnect Bypass Fraud at the Network Edge”, *USENIX Security Symposium,* Vol. 24, pp. 833-848, 2015.

[27] Sharmila Subudhi and Suvasini Panigrahi, "Use of Fuzzy Clustering and Support Vector Machine for Detecting Fraud in Mobile Telecommunication Networks”, *Int. J. Security and Networks,* Vol. 12, No. 1/2, 2016.

[28] Debasis Samanta, "Defuzzification Methods”, in *Fuzzy Logic: the Logic and its Application*, Indian Institute of Technology Kharagpur, 2018, pp. 200-207.

[29] "Captcha: Codes and Images for Spam Protection”, IONOS, 28 08 2020. [Online]. Available: https://www.ionos.com/digitalguide. [Accessed 06 10 2021].

DECLARATION

I, the undersigned, declare that this thesis is my original work and has not been presented for a degree in any other university, and that all source of materials used for the thesis have been duly acknowledged.

Declared by:

Name: Tadele Ayalew Degu

Signature: _____

Date: _____

Confirmed by Advisor:

Name: Dagmawi Lemma (PhD)

Signature: _____

Date: _____