DETECTING FRAUD PATTERNS IN TELECOMMUNICATIONS USING CASE BASED REASONING

BALA MUSA SHUAIBU

UNIVERSITI UTARA MALAYSIA
2008

KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

BALA MUSA SHUAIBU

calon untuk Ijazah
(candidate for the degree of) **MSc. (Intelligent System)**

telah mengemukakan kertas projek yang bertajuk
(has presented his/her project paper of the following title)

DETECTING FRAUD PATTERN IN TELECOMMUNICATIONS USING CASE BASED REASONING

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory knowledge of the field is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **ASSOC. PROF. FADZILAH SIRAJ**

Tandatangan
(Signature) : _____ Tarikh (Date) : 29/04/2008

Nama Penyelia Kedua
(Name of 2nd Supervisor): **MS. NOORAINI YUSOFF**

Tandatangan
(Signature) : _____ Tarikh (Date) : 29/04/2008

ABSTRACT

All round the world, fraud situations are significantly causing huge revenue leakage in the telecommunication companies every year. The reuse of previous cases is an important issue in dealing with fraud pattern in a data with string features. Case Based Reasoning (CBR) systems have a set of cases inform of library used to facilitate the process of validation of new cases without the direct involvement of a domain expert. The proposed detection technique in this paper is based on Case Based Reasoning used to detect the occurrence of fraud with a meaningful confidence in telecommunication data. Experimental result on the fraud data indicates that the weight for all attribute used in this study needs to be set as 0.1 in order to get 98% similarity performance.

ACKNOWLEDGEMENTS

In the Name of Allah, the Most Gracious and the Most Merciful

It is my pleasure to acknowledge the immense contribution of some people who have assisted me one way or the other towards the successful completion of this project.

First of all, I give thanks to the Allah for his guidance and mercy throughout my life. Peace and Blessing to his last Prophet Muhammad (S.A.W.), his household and his companions. My sincere appreciation goes to my beloved parents and family members for their patience, prayers and understanding over the entire period of my study.

Secondly, my grateful thanks go to my supervisors, Prof. Madya Fadzilah Siraj and Nooraini Yusuf who had given their full support and contributed immensely towards the completion of this project. They have actually spent a lot of time patiently and painstakingly giving me the necessary advice, providing valuable information and correcting errors to ensure that the best effort has been given in the completion and achievement of this project.

I also wish to convey my appreciation to Associate Professor. Dr. Norita Md. Norwawi my evaluator who has giving me support and advise for the completion of this project.

Lastly, I recognize the efforts of all my friends, Staff of Faculty of Information Technology, Universiti Utara Malaysia and those who contributed directly or indirectly towards the completion of this project. Thanks to all.

Bala Musa Shuaibu
College of Arts and Sciences
Faculty of Information Technology
Universiti Utara Malaysia
June 2008

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This chapter gives a general overview of the project. It starts with a background review of the project which explains the motivation behind the project and the domain on which the project is based on. The chapter further describes the problem statement, the objective to be accomplished; significant to be derived from the project, the scope or coverage area and finally highlighted the way subsequent chapters will be organized.

1.1 Background

Telecommunications has brought tremendous achievement in life by providing a means to extend communication over a distance. This is done through telephony, computer networking, television, radio, and so on to transmit information in form of voice or data.. Telecommunications facilitates Internet service provision, networking of computers and telephones and cellular service for cars, fax or modem, or offers voice mail that meet specific business needs. Many telecommunications companies manufacture equipments ranging from simple pagers to mobile office packages for Internet service and high-speed connections. Therefore, telecommunications has brought revolution of new products and technologies to the market.

Examples of such products are:

a. ISDN

A high-capacity line that can transmit computer data such as large files, images, video, voice communications, and faxes much more quickly and at relatively low cost. It means more services from a single telephone line.

b. Video conferencing

A video phone call during which users see people and products from their computer screens, and will be able to speak back which in turn will cut down travel expenses.

c. Cellular modems and faxes

These transmit data from a wireless base rather than through regular telephone lines.

d. Voice mail and automated attendants

Answering services that can direct customers and organize messages. It provides for an exact time and day incoming calls tried to reach the recipient. Therefore it stores and secures office messages during business and non-business hours.

e. Internet services

Provide a connection from computer to the Internet, which can be a powerful tool for marketing and commerce. This facilitates global information and sales opportunities, also reduces promotional costs.

However, telecoms companies suffer enormous lose of their revenue through fraud. Fraudulent calls are those made with the intention of avoiding billing or reducing the tariff rate of billing. The Concise Oxford Dictionary defines fraud as "criminal deception; the use of false representations to gain an unjust advantage." It's an illegitimate use of telephone devices to initiate calls or terminate calls on a particular trunk. Perhaps, various types of fraud may be classified into two categories (Shawe-Taylor *et al*, 2000):

a. Subscription Fraud

This has to do with obtaining an account without the intention of paying the charges accrued to such account. Hence the fraudster makes extensive and abnormal call during the period. Also fall in this category, is the case of customer who do not necessarily have fraudulent intentions but never pays a single bill accrued to his account.

b. Superimposed Fraud

This is the case of fraudster taking over a legitimate line and superimposed on it more debt due to excessive calls. Examples of such cases include cellular cloning, calling card theft and cellular handset theft.

This project only identifies the fraud based on the usage of the line which could incorporate the various types of fraud.

According to Telecom and Network Security Review (1997), the estimated fraud losses in the U.S. telecommunications industry amount to between 4% and 6%of revenue. Internationally, the figures are generally worse, with several new service providers reporting losses over 20%. Also Business Daily from THE HINDU group of publications (2006) reported that the average revenue leakage across global telecom operators in 2006 has increased to 12.1 per cent of turnover ($176 billion), compared to 11.6 per cent in 2005. Fraud was reported as the single largest area of revenue leakage, almost 2.9 per cent of the turnover.

Furthermore, a press release by Communications Fraud Control Association (CFCA, 2007), of a comprehensive survey that estimates annual telecom fraud losses worldwide. In contrast to the organization's previous (1999) estimate of $12 billion, CFCA in 2003 considers annual worldwide telecom fraud losses to be in the range of $35 - $40 billion U.S. dollars. And now in 2007 the annual global loss due to fraud is $ 55-60 billion USD. According to Intec Telecom Systems PLC (2002) Malaysia is a prime target for fraud, as a leading high technology hub in Southeast Asia, with one of the most sophisticated

telecommunications facilities in the region, built on state-of-the art networks using fibre optics, ATM and ADSL. The increase of these new telecommunications services has also widened the opportunities for fraud to occur. Today, many of Malaysia's telecoms operators lose maybe 3 percent of their revenue to fraud each year, with mobile companies experiencing greater levels of crime. This adds up to millions of dollars worth of losses at a time when Malaysia's telecommunications industry can ill afford to lose money. With the economic slowdown of the region there has also been a steady drop in demand for highly invested next generation services and infrastructure projects are often now being financed in more costly foreign currencies. As a result, many companies are struggling to make a profit. The liberalization of the market has also aggravated the problem.

Hence, to detect these frauds in the telecommunication industry, several approaches are used. One of the approach been used is the Case Based Reasoning. Case-based reasoning is a prominent kind of analogy making. It has been argued that case-based reasoning is not only a powerful method for computer reasoning, but also a pervasive behavior in everyday human problem solving. Or, more radically, that all reasoning is based on past cases experienced or accepted by the being actively exercising choice – prototype theory most deeply explored in human cognitive science.

Case based reasoning is a methodological approach within the field of Artificial Intelligence which gives a good framework for case retrieval, reuse, solution testing and learning. It has argued that CBR is a generic methodology for building knowledge-based

systems, rather than an isolated technique that is capable of solving only very specific tasks (Kamp, 1998). Case-based reasoning will be a useful paradigm in the telecommunication fraud detection scenario since it differs with other AI techniques in many respects. It does not depend specifically on the general knowledge of a problem, but its ability to use knowledge of previously experienced (cases) in solving a similar case which is the intent of this project to discover the fraud pattern based on previous cases. Importantly, CBR is an incremental learning since a new knowledge is retained for future purpose when a problem is solved.

It has also been argued (Kamp, 1998) that CBR is more or less a generic methodology for building knowledge-based systems, rather than an isolated technique like in the case of Neural Network or Rule Based that is capable of solving only very specific tasks. Other advantages of CBR over other technique include:

- Reduces the knowledge acquisition effort
- Requires less maintenance effort
- Improve problem solving performance through reuse
- Makes use of existing data, e.g. in databases
- Improve over time and adapt to changes in the environment
- High user acceptance

1.2 Problem Statement

There is huge leakage in the telecommunications revenue due to the activities of fraudster which in turn result to poor services in terms of congestion and operators inability to withstand the competitive environment. Also the inconsistency of the fraud cases calls for a robust detection system.

Although initial works on fraud detection concentrate on using rule based system (Rupesh & Saroj, 2007: Jimmy, 2003), other methods such as Neural Networks has also been implemented to classify fraud cases (Grosser, 2005: Burge, 2001). In this study, a few examples of the attributes and their corresponding values are listed in Table 1.

Table 1.0: Sample data set.

Alarm_Code	Severity	User_Group	User_Cf	User_Usg	Ind_Case
038	0	R10	10	180.06667	C
033	1	DEF	6	30.16667	C
020	1	R10	10	101.15	C
029	3	R10	10	123.23336	C
038	4	R10	10	187.29998	C

The fact that the data is of type string rather than numbers, it is more appropriate to use case based reasoning since in CBR, Cases are often derived from legacy databases and need not be well structured (Wheeler & Aitken, 2000). Therefore, this study attempts to

explore the use of case based reasoning in fraud detection system. Its performance will be measured based on the similarities percentages produced by the prototype.

1.3 Research Questions

(i) Why do we need AI technique to guide human in detecting fraud pattern in a pool of Call Detail Rate for subscribers?

(ii) What is the appropriate AI technique to be used considering the fact that the Call Detail Rate is in string attribute?

(iii) How will Case Based Reasoning be used to achieve the desired objective?

1.4 Objectives

The primary objective of this study is to detect likely fraud patterns in Call Detail Record CDR made by users over a period of time in Telecommunication network using Case Base Reasoning technique.

Specifically, the objectives are:

(i) To identify AI technique for fraud

(ii) To develop and evaluate the prototype identified in (ii) for filtering fraud calls.

(iii) And to determine the appropriate weight for the attributes in order to obtain higher performance similarity.

1.5 Project Significance

This project aims to provide the fraud analyst with a system for detecting real fraud cases. It will also save the telecomm industry a considerable amount of money being lost due to activities of fraudsters by early detection of fraud calls. Perhaps, it will also provide an avenue for the company to fashion out measures to deal with culprits. Interestingly, it will go a long way to restore the growing lack of confidence among subscribers since it will guarantee security in their calls

1.6 Project Scope

Investigation begins when a trigger sends an alarm for a likely fraud call. To make a meaningful and absolute detection, call patterns of a subscriber are observed and a comparison is done between the historical data and the new data to determine any abnormality in the pattern.

This project will only analyze sample of usage pattern of CDR (Call Detail Records) from Malaysian Telecommunication which include:

- Alarm Code
- Severity
- User Group
- User Confidence
- User Usage
- Indication Case

It will uncover absurd usage based on established usage pattern of the subscriber. However, in this study only 6 attributes are used, example Alarm Code, Severity, User Group, User Confidence, User Usage and Indication Case.

1.7 Organization of the Report

The rest of the project is divided into four major chapters as follows: Chapter Two discusses about the literature review of several related projects and applications that uses AI technique of Rule Base, Neural Network, Case Base or Other Techniques. Chapter Three discusses the methodology used to achieve the design of this project. This methodology is the System Development Design Methodology comprising of steps as construct a conceptual frame work, develop system architecture, analyze and design the system, build prototype system and observe or evaluate the system, Nunamaker and Chen (1987). Chapter Four explains the steps taken in finding the most appropriate and suitable findings and results for this project. In this chapter, some explanation about the process of designing the system will be carried out. The results of this project will also be discussed under this chapter. Chapter Five concludes the project report based on the result and discussions achieved in the previous chapter.

1.8 Conclusion

This chapter gave an overall view of this project by first of all describing the benefit telecommunication has brought in improving the live condition of humans and it also pointed out the various kind of frauds and measure of the damages done due to activities of fraudster in the domain, thereby describing the problem statement as well as the primary objective of this project. Also not left out in this introduction is the significant benefit and scope this project will cover.

CHAPTER TWO

LITERATURE REVIEW

This chapter discusses related applications in the field of telecommunication which are developed using the AI techniques like Rule Base, Neural Network, Case Base and other techniques. More emphasis will be laid on the Case Base Reasoning Technique as it relates to the technique used in this project

2.1 TELECOMMUNICATIONS

Recently, a new category of emerging telecommunications applications offers other media in addition to voice (including data, video, animation, graphics, etc.), incorporates the computer as an interface to these media, and renders these dictionary definitions of telecommunications and computers archaic (David, 1998). Improvement in telecommunication leads to enhancement of our way of life. In earlier ages, the printing press and the telephone each had a dramatic impact on society, as did the automobile and the airplane for similar reasons. The emergence of the computer as a telecommunications tool will have an equally great impact (David, 1998). Significantly, telecommunications has provided an avenue for great changes in our individual life and collective. Obviously, mobile phones increase the ability to coordinate activities, especially across remote sites. This means that it is no longer necessary to conduct communications from an office desk

(John, 2004). Therefore the increasing globalization of commerce, largely enabled by modern telecommunications and transportation, dramatically increases the number of people with whom we may have occasion to communicate. Emerging technologies like personal communications are a double edged sword. While they make communications easier, they also impede personal effectiveness by generating constant interruptions, and in their extreme have the potential to decrease the quality of life (David, 1998).

Hence, appealing to fraudsters is the use of these telecommunication technologies as a back bone to gain profit illegally and to instantiate business in which minimal investment is required and at low risk.

2.2 FRAUD

Michael *et al.* (2000) described fraud as a big business. Calls, credit card numbers, and stolen accounts can be sold on the street for substantial profit. Fraudsters may subscribe to services without intending to pay, perhaps with the intention of reselling the services, or even the account itself, at a low cost until shut down. Call sell operations may extend their lives by subverting regulatory restrictions that are in place to protect debtors. Gaining access to a telephone or telephone line by physical intrusion still accounts for some fraud. Fraudsters also focus on the people who use and operate the network by applying "social engineering" to instruct an unsuspecting subscriber or operator to unknowingly agree to carry fraudulent traffic.

Similarly Richard *et al.* (2002) also noted that fraud is increasing dramatically with the expansion of modern technology and the global superhighways of communication, resulting in the loss of billions of dollars worldwide each year. Although prevention technologies are the best way to reduce fraud, fraudsters are adaptive and, given time, will usually find ways to circumvent such measures. Methodologies for the detection of fraud are essential if we are to catch fraudsters once fraud prevention has failed. Statistics and machine learning provide effective technologies for fraud detection and have been applied successfully to detect activities such as money laundering, e-commerce credit card fraud, telecommunications fraud and computer intrusion, to name but a few.

Finding telecommunications fraud in masses of call records is more difficult than finding a needle in a haystack. In the haystack problem, there is only one needle that does not look like hay, the pieces of hay all look similar, and neither the needle nor the hay changes much over time. Fraudulent calls may be rare like needles in haystacks, but they are much more challenging to find. Callers are dissimilar, so calls that look like fraud for one account look like expected behavior for another, while all needles look the same. Moreover, fraud has to be found repeatedly, as fast as fraud calls are placed, the nature of fraud changes over time, the extent of fraud is unknown in advance, and fraud may be spread over more than one type of service. For example, calls placed on a stolen wireless telephone may be charged to a stolen credit card. Finding fraud is like finding a needle in a haystack only in the sense of sifting through masses of data to find something rare.

In this review, systems based on the intelligent techniques including Rule Based, Neural Network, Case Based Reasoning and other Techniques are examined in the light of improvement or closing any gap available.

2.3 RULE-BASED

Tomoharu *et al.* (2000) proposed an ensembling method for pattern classification problems. The characteristic feature of the ensembling method is that two different types of fuzzy rule-based classification systems are used. One is a fuzzy rule-based classification that suggests a class of an input pattern. The other is a fuzzy rule-based ensembling system that assigns a weight to suggested class by each classification system. The assembling method system consist of one fuzzy ruled-based ensembling system, several fuzzy rule-based classification system and getting node that finally determine the final classification of the input pattern. Computer simulation shows the effectiveness of this ensembling method.

Similarly (Jimmy, 2003) examines the problem of managing fraud in emerging converged networks and presents work in progress on implementing rules based fraud detection system for deployment in a test bed environment "An Approach to Rules based Fraud Management in Emerging Converged Networks". This is done by examining the state of the art in telecoms fraud management and adapting this to emerging IP-based networks and services. Features of this fraud detection implementation include the use of

flexible data formats and spreadsheet/workbook base rules specification with the capability to apply arbitrarily complex rules.

Jian-Bo *et al.* (2003) used a generic Rule-base Inference Methodology using the Evidential Reasoning approach; he proposed a new knowledge representation scheme in a rule-base using a belief structure and fuzzy set theory. In this scheme, a rule-base is designed on the basis of the belief structure with belief degrees embedded in all possible consequents to capture vagueness, incompleteness and nonlinear causal relationships. Whilst traditional IF-THEN rules can be represented as a special case. In an established rule-base, an input to an antecedent attribute is transformed into a belief distribution. Subsequently, inference in such a rule base is implemented using the evidential reasoning approach. The scheme is further extended to inference in hierarchical rule bases.

Rupesh and Saroj (2007) presented a rule-based approach to detect anomalous telephone calls. The method described here uses subscriber usage CDR (call detail record) data sampled over two observation periods: study period and test period. The study period contains call records of customers' non-anomalous behavior. Customers are first grouped according to their similar usage behavior (like, average number of local calls per week, and so on). For customers in each group, probabilistic model to describe their usage is developed. Next, maximum likelihood estimation (MLE) is used to estimate the parameters of the calling behavior. Then a threshold by calculating acceptable change within a group is determined. MLE is used on the data in the test period to estimate the parameters of the calling behavior. These parameters are compared against thresholds.

Any deviation beyond the threshold is used to raise an alarm. This method has the advantage of identifying local anomalies as compared to techniques which identify global anomalies. The method is tested for 90 days of study data and 10 days of test data of telecom customers. For medium to large deviations in the data in test window, the method is able to identify 90% of anomalous usage with less than 1% false alarm rate

2.4 NEURAL NETWORKS

Neural Networks were inspired by biological findings relating to the behavior of the brain as a network of units called neurons. The human brain is estimated to have around 10 billion neurons each connected on average to 10,000 other neurons. Each neuron receives signals through synapses that control the effects of the signal on the neuron. These synaptic connections are believed to play a key role in the behavior of the brain. The fundamental building block in an Artificial Neural Network is the mathematical model of a neuron (Bishop, 1995). In depth study is carried out in the telecommunications domain using the neural network technique. Some of these studies are discussed as follows:

Bonchi *et al.* (1999) presented a case study, which illustrates how techniques based on classification can be used to support the task of planning audit strategies. The proposed approach is sensible to some conflicting issues of audit planning, e.g., the trade-off between maximizing audit benefits vs. minimizing audit costs. A methodological scenario, common to a whole class of similar applications, is then abstracted away from

the case study. The limitations of available systems to support the identified overall KDD process lead to point out the key aspects of a logic-based database language, integrated with mining mechanisms, which is used to provide a uniform, highly expressive environment for the various steps in the construction of the considered case-study

Frank *et al.* (2000) gave an overview of a project involving the application of neural networks to Telecommunications Systems. Five application areas are discussed, including cloned software identification and the detection of fraudulent use of cellular phones. The systems were summarized and the general results were presented. The conclusions highlighted the difficulties involved in using this technology as well as the potential benefits.

Also Burge (2001) uses a recurrent neural network technique, uniformly distribute prototypes over toll tickets, and sampled from the U.K. network operator Vodafone to build prototypes, which continue to adapt to cater for seasonal or long term trends, are used to classify incoming toll tickets to form statistical behavior profiles covering both the short- and the long-term past. This study introduce a new decaying technique, which maintains these profiles such that short-term information is updated on a per toll ticket basis whilst the update of the long-term behavior can be delayed and controlled by the user. The new technique ensures that the short-term history updates the long-term history applying an even weighting to each toll ticket. The behavior profiles, maintained as probability distributions, form the input to a differential analysis utilizing a measure known as the Hellinger distance between them as an alarm criterion. Fine tuning the system to minimize the number of false alarms poses a significant task due to the low

fraudulent/non fraudulent activity ratio. Hence the benefit from using unsupervised learning in that no fraudulent examples are required for training.

Similarly Isaac (2003) combines a state-of-the-art subscriber profiling and neural network with ad-hoc rules, case prioritization, and reporting capabilities. Link analysis can be provided as part of the solution as well. The streamlined case management process ensures that cases are prioritized according to risk in order to detect fraud at the earliest possible moment. Best-of-breed profiling and neural network technology are core to the complete solution that Fraud Analytics provides to the provider in order to deliver a superior, proven fraud management solution.

Munshi et *al.* (2003) presented the small signal and large signal models for an AlGaAs and a SiGe. Heterojunction Bipolar Transistor, using neural network techniques. The main advantage of this technique is the wide range of frequencies over which the small signal model is valid and the great accuracy of the large signal characteristics. Both the models have been verified by comparing the simulated values with the measured ones of the HBTs for both the material systems.

Olusola (2003) investigates the unsupervised learning potentials of two neural networks for the profiling of calls made by users over a period of time in a mobile telecommunication network. His study provides a comparative analysis and application of Self-Organizing Maps (SOM) and Long Short-Term Memory (LSTM) recurrent neural networks algorithms to user call data records in order to conduct a descriptive data

mining on users call patterns. His investigation shows the learning ability of both techniques to discriminate user call patterns; the LSTM recurrent neural network algorithm providing a better discrimination than the SOM algorithm in terms of long time series modelling. LSTM discriminates different types of temporal sequences and groups them according to a variety of features. The ordered features can later be interpreted and labeled according to specific requirements of the mobile service provider. Thus, suspicious call behaviours are isolated within the mobile telecommunication network and can be used to identify fraudulent call patterns. He gave results using masked call data from a real mobile telecommunication network.

Grosser (2005) focuses on the problem of detecting unusual changes of consumption in mobile phone users, the corresponding building of data structures which represent the recent and historic users' behavior bearing in mind the information included in a call, and the complexity of the construction of a function with so many variables where the parameterization is not always known. In his application of making a differential analysis, patterns of behavior of the mobile phone are monitored by comparing the most recent activities to the historic use of the phone; a change in the pattern of behavior is a suspicious characteristic of a fraudulent act. The results obtain in this experiment shows that, it is possible to obtain, with a high degree of certainty, a list of users who are using their mobile phone in a "not loyal" way. It is also proven, with the experiences carried out, that the differential analysis provides with much more information than the absolute analysis, which can only detect peaks of consumption and cannot describe the user behavior in question.

Furthermore Pattara and Peachavanish (2007) investigated an alternative method to estimate the degree of road traffic congestion based on a new measurement metric called Cell Dwell Time (CDT) using simple feed forward back propagation neural network. CDT is the duration that a cellular phone is registered to a base station before handing off to another base station. As a vehicle with cellular phone traverses along the road, cell handoffs occur and the values of CDT vary. Our assumption is that the values of CDT relate to the degree of traffic congestion and that high CDTs indicate congested traffic. They measured series of CDTs while driving along arterial roads in Bangkok metropolitan area. Human judgment of traffic condition was recorded into one of the three levels indicating congestion degree - free flow, moderate, or highly congested. Neural network was then trained and tested using the collected data against human perception. The results showed promising performance of congestion estimation with accuracy of 79.43%, precision ranging from 73.53% to 85.19%, and mean square error of 0.44.

Nenad *et al.* (2007) showed that the use of Hopfield-like neural network in solving routing and wavelength assignment (RWA) in all-optical networks is considered. Routing assignment was performed under the constraints of minimal wavelength conversions and minimal congestion. The wavelength assignment assumes four basic methods: first-fit, least-used, most-used, and maximal sum with relative capacity loss algorithm. Different RWA methods are simulated and results are compared.

2.5 CASE BASED REASONING

Case-based reasoning means using old experiences to understand and solve new problems. In case-based reasoning, a reasoner remembers a previous situation similar to the current one and uses that to solve the new problem. Case-based reasoning can mean adapting old solutions to meet new demands; using old cases to explain new situations; using old cases to critique new solutions; or reasoning from precedents to interpret a new situation much like lawyers do or create an equitable solution to a new problem much like labor mediators do (Janet, 2004).

Case-based reasoning (CBR), a general problem solving methodology for reusing previously stored solutions (Aamodt, 1994). Wheeler (2000) describes an application of Case Based Reasoning to the problem of reducing the number of final line fraud investigations in the credit approval process. The performance of a suite of algorithms which are applied in a combination to determine a set of diagnosis from a set of retrieved cases is reported. The result indicated that an adaptive solution can provide fraud filtering and case ordering function for reducing the number of final line fraud investigation necessary.

Zhi-Wei, *et al.* (2003) used case-based reasoning systems solve new problems by using previous problem solving experiences stored as cases in a case base. In recent years the integrated methods of case based reasoning have become an increasingly important research issue for the case-based reasoning community. In their study they carried out a

research on integrating case-based reasoning method based on rule-based reasoning, or induction learning technique so as to heighten the efficiency of problem-solving of case-based reasoning. They also utilize genetic algorithm in a integrated case-based reasoning approach. The experiments show that the new model of integrated case-based reasoning has got a better accuracy rate of classification.

Giulio and Alessandro (2004) also developed an Operational Knowledge Management System (OKMS), called Remoter. It has a tree tier web-based architecture which groups different technologies (Document Management, Forum, CBR). In particular, they used CBR methodology for Operational Knowledge Management Framework (OKMF). They show the synergies between OKMF and CBR techniques using a case study on the ADSL context of Telecom Italia, called Remoter.

Mingyang *et al.* (2004) highlighted that conversational Case-Based-Reasoning (CCBR) provides a mixed-initiative dialog for guiding users to construct their problem description incrementally through a question-answering sequence. Similarity calculation in CCBR, as in traditional CBR, plays an important role in the retrieval process since it decides the quality of the retrieved case. Hence, they analyzed the different characteristics of the query (new case) between CCBR and traditional CBR, and argue that the similarity calculation method that only takes the features appearing in the query into account, so called query-biased, is more suitable for CCBR. An experiment is designed and executed on 36 datasets. The results showed that on 31 datasets out of the total 36, the CCBR system using the query-biased similarity calculation method achieves more effective

performance than those using case-biased and equally-biased similarity calculation methods.

Similarly, Mingyang and Xin (2004) stated that one difficulty in component retrieval comes from users' incapability to well define their queries. They proposed a conversational component retrieval model (CCRM) to alleviate this difficulty. In CCRM, a knowledge-intensive conversational case based reasoning method is adopted to infer potential knowledge from current known knowledge, calculate the context-based semantic similarities between users' queries and stored components, and prompt users the most discriminative questions to extract more information to refine their component queries interactively and incrementally.

However, locating resources of interest in a large resource intensive environment is a challenging problem. Mehmet, *et al.* (2004) presented a research on addressing this problem through the development of a recommender system to aid in metadata discovery. Their recommender approach uses Conversational Case-Based Reasoning (CCBR), with semantic web markup languages providing a standard form for case representation. They presented their initial efforts in designing and developing ontologies for an Earthquake Simulation Grid, to use these to guide case retrieval, discuss how these are exploited in a prototype application, and identify future steps for this approach.

Marius and Anders (2004) described an approach to automatic situation assessment in a mobile environment within the AmbieSense research project. Even though the

AmbieSense system includes both mobile and fixed parts, this paper focuses on the mobile part of the system. The three major issues covered here are: the open-ended context model, the multi-agent system, and the reasoning mechanism.

Similarly Amani *et al.* (2005) developed a new using case-based reasoning (CBR) method for alarm correlation in telecommunication networks. The proposed method has been simulated by developing three main modules: a module for generating faults and alarms, defining network configuration, and alarm filtering and correlation by using case-based reasoning. One of the most important aspects of the obtained results was the speed of the system. Because of its simplicity, the case-based reasoning model is fast, requiring only a few floating-point calculations to produce the result. The accuracy of alarm correlation achieved in the simulation was higher than 90% in the case of unavailability of the required cases.

Joshua *et al.* (2005) also discusses how to extend Compensable Behavior Technology CBT to capitalize on previously stored and validated behavior models, and how to apply case based reasoning to this design problem. They propose an approach that incorporates text and graphical retrieval strategies and multiple levels of similarity analysis. Finally, they outline future research and development directions.

2.6 OTHER TECHNIQUES

Jaakko (2000) also identified a relevant user groups based on call data and then a user is assigned to a relevant group. The call data is subsequently used in describing behavioral patterns of users. Neural networks and probabilistic models are employed in learning these usage patterns from call data. These models are used either to detect abrupt changes in established usage patterns or to recognize typical usage patterns of fraud. The methods are shown to be effective in detecting fraudulent behavior by empirically testing the methods with data from real mobile communications networks

Ching and Chia (2000) presented an SVD-QR-based approach is proposed to design an appropriate fuzzy system directly from some gathered input-output data. A fuzzy system with fuzzy rule tables is defined to approach the input-output pairs of an identified system. In the rule base of the defined fuzzy system, each fuzzy rule table corresponds to a partition of an input space. In order to extract the most important fuzzy rules from the rule base of the defined fuzzy system, a firing strength matrix determined by the membership functions of the premise fuzzy sets is constructed. According to the firing strength matrix, the number of important fuzzy rules is determined by the singular value decomposition (SVD), and the most important fuzzy rules are selected by the SVDQR-based method. Consequently, a reconstructed fuzzy rule base composed of significant fuzzy rules is determined by the firing strength matrix.

Similarly Tomoharu, *et al.* (2003) examines credit assignment methods in fuzzy classifier ensembles. Multiple fuzzy rule-based classification systems are included in a fuzzy

classifier ensemble. Fuzzy rule based classification systems in a fuzzy classifier ensemble are constructed from given training patterns. In addition to fuzzy rule-based classification systems, we construct a credit assignment system that assigns a weight vector for each input pattern. For an input pattern, a class is suggested by each fuzzy rule-based classification system, and the final classification of the input pattern is determined by considering both the suggested class by the fuzzy rule based classification systems and the weight vector from the credit assignment system. Fuzzy and non-fuzzy rule-based system is used for the credit assignment systems. The credit assignment systems are constructed so that large weights can be assigned to the fuzzy rule based classification systems with high classification power.

Clifton *et al.* (2004) proposes an innovative fraud detection method, built upon existing fraud detection research and Minority Report, to deal with the data mining problem of skewed data distributions. This method uses back-propagation (BP), together with naïve Bayesian (NB) and C4.5 algorithms, on data partitions derived from minority over sampling with replacement. Its originality lies in the use of a single meta-classifier (stacking) to choose the best base classifiers, and then combine these base classifiers' predictions (bagging) to improve cost savings (stacking-bagging). Results from a publicly available automobile insurance fraud detection data set demonstrate that stacking-bagging performs slightly better than the best performing bagged algorithm, C4.5, and its best classifier, C4.5 (2), in terms of cost savings. Stacking-bagging also outperforms the common technique used in industry (BP without both sampling and partitioning). Subsequently, this paper compares the new fraud detection method (meta-learning

approach) against C4.5 trained using undersampling, oversampling, and SMOTEing without partitioning (sampling approach). Results show that, given a fixed decision threshold and cost matrix, the partitioning and multiple algorithms approach achieves marginally higher cost savings than varying the entire training data set with different class distributions. The most interesting find is confirming that the combination of classifiers to produce the best cost savings has its contributions from all three algorithms.

Pablo and Claudio (2006) also presented a system to prevent subscription fraud in fixed telecommunications with high impact on long-distance carriers. The system consists of a classification module and a prediction module. The classification module classifies subscribers according to their previous historical behavior into four different categories: subscription fraudulent, otherwise fraudulent, insolvent and normal. The prediction module allows for identifying potential fraudulent customers at the time of subscription. The classification module was implemented using fuzzy rules. It was applied to a database containing information of over 10,000 real subscribers of a major telecom company in Chile. In this database a subscription fraud prevalence of 2.2% was found. The prediction module was implemented as a multilayer perceptron neural network. It was able to identify 56.2% of the true fraudsters, screening only 3.5% of all the subscribers in the test set. This study shows the feasibility of significantly preventing subscription fraud in telecommunications by analyzing the application information and the customer antecedents at the time of application.

Adem (2007) uses various data mining techniques to obtain the best practical solution for the telecom fraud detection and offers the Adaptive Network based Fuzzy Inference System (ANFIS) method as a means of efficient fraud detection. ANFIS has provided sensitivity of 97% and specificity of 99%, where it classified 98.33% of the instances correctly.

Dehzangi *et al.* (2007) stated that classification Systems have been widely applied in different fields such as medical diagnosis. A fuzzy rule based classification system (FRBCS) is one of the most popular approaches used in pattern classification problems. One advantage of a fuzzy rule-based system is its interpretability. However, they were faced with some challenges when generating the rule-base. In high dimensional problems, they can not generate every possible rule with respect to all antecedent combinations. Hence, by making the use of some data mining concepts, they proposed a method for rule generation, which can result in a rule-base containing rules of different lengths. Then, their rule learning algorithm based on R.O.C analysis tunes the rule-base to have better classification ability. The goal was, to check if generating cooperative rule-bases containing rules of different dimensions, can lead to better generalization ability. To evaluate the performance of the proposed method, a number of UCIML data sets were used. The results show that considering cooperation in a rule-base tuned by rule weighting process can improve the classification accuracy. It is also shown that increasing the maximum length of rules in the initial rule-base, improves the classification accuracy.

Abhinav Srivastava *et al.* (2008) similarly modeled the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and showed how it can be used for the detection of frauds. An HMM is initially trained with the normal behavior of a cardholder. If an incoming credit card transaction is not accepted by the trained HMM with sufficiently high probability, it is considered to be fraudulent. At the same time, they tried to ensure that genuine transactions are not rejected. They presented detailed experimental results to show the effectiveness of the approach and compare it with other techniques available in the literature.

2.7 SUMMARY

Based on the review of the various techniques used in fraud detection CBR systems have a number of advantages over other AI techniques According to Wheeler (2000), the CBR has a number of advantages which is the fact that it provides meaningful confidence and system accuracy measures, it requires little or no direct expert knowledge acquisition and it's a technique which is easily updated and maintained. Therefore, this study tends to focus on the Case Based Reasoning approach to detect fraud patterns. The subsequent chapter will discuss the methodology to be followed in achieving the said objective.

CHAPTER THREE

METHODOLOGY

Based on the review of different Artificial Intelligent Technique in the previous chapter, Case Based Reasoning is chosen to be the most appropriate technique to deal with fraud pattern in the context of data used in this study. Hence this chapter will explain in details the methodology used in achieving the objective of this project.

3.1 INTRODUCTION

Classifying research according to its domains and purposes, as well as the processes and tools used will help in understanding where Information System research stands. Although these classification schemes overlap to some extent, they differ in their focus (Nunamaker & Minder, 1991).

Applied research is the application of knowledge to solve problems of immediate concern (Blake, 1978; Bailey, 19821). In this study knowledge and technique will be used to solve a problem of fraud in telecommunication domain and hence it is an applied research. Therefore the methodology that will best be appropriate to adopt is the System development Research Methodology proposed by Nunamaker in 1991.

The development of systems, especially the development of information systems, has to follow a certain research process and conform to some criteria to be qualified as academic research (Nunamaker & Minder, 1991). On this basis, this project is carried out observing the criteria and processes outlined in System Development Research Methodology.

3.2 System development Research Methodology

Methodology is the philosophy of the research process which "includes the assumptions and values that serve as a rationale for research and the standards or criteria the researcher uses for interpreting data and reaching conclusion (Bailey, 1982). Research process, the heart of research methodology, is the application of scientific method to the complex task of discovering answers (solutions) to questions (problems) (Blalock & Blalock, 1982).

The System Development Methodology consists of the following stages:

- Construct a conceptual framework.
- Develop a system architecture.
- Analyze and design the system.
- Build the system.
- Observe and evaluate the system.

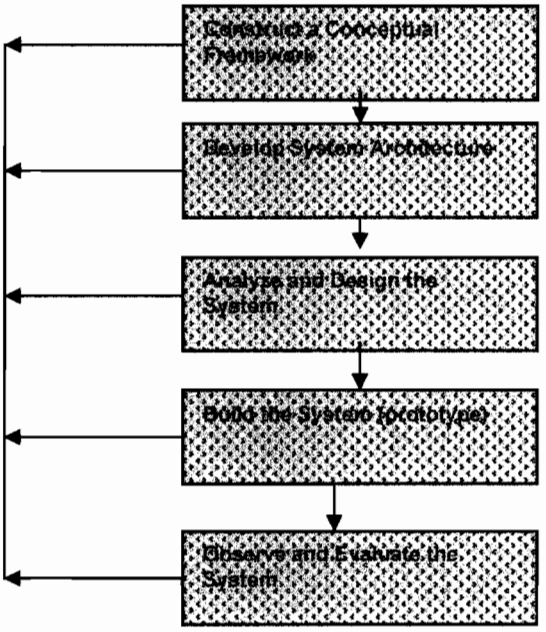


Figure 3.1: The Systems Development Research Methodology (Nunamaker , 1991)

3.2.1 Construct a conceptual frame work

A conceptual framework is to clearly define the research problem which provides a focus for the research throughout the development process. The research question is discussed at this stage based on the context of an appropriate conceptual framework. Various disciplines are also explored to gain additional knowledge on the approaches which will best be incorporated in the new system. In this project the research questions highlighted in the introductory chapter is followed by a clearly defined objective of the study and from the literature review in the subsequent chapter, an in-depth insight is gained on the

various techniques in relation to the domain of telecommunication. Furthermore a conclusion on the case based reasoning technique is arrived at.

3.2.2 Develop a System Architecture.

The system architecture is a road map for the systems building process. Various Modules are specified and their individual as well as collective functionalities in respect to the entire system is also identified. The modules are also put into correct perspective.

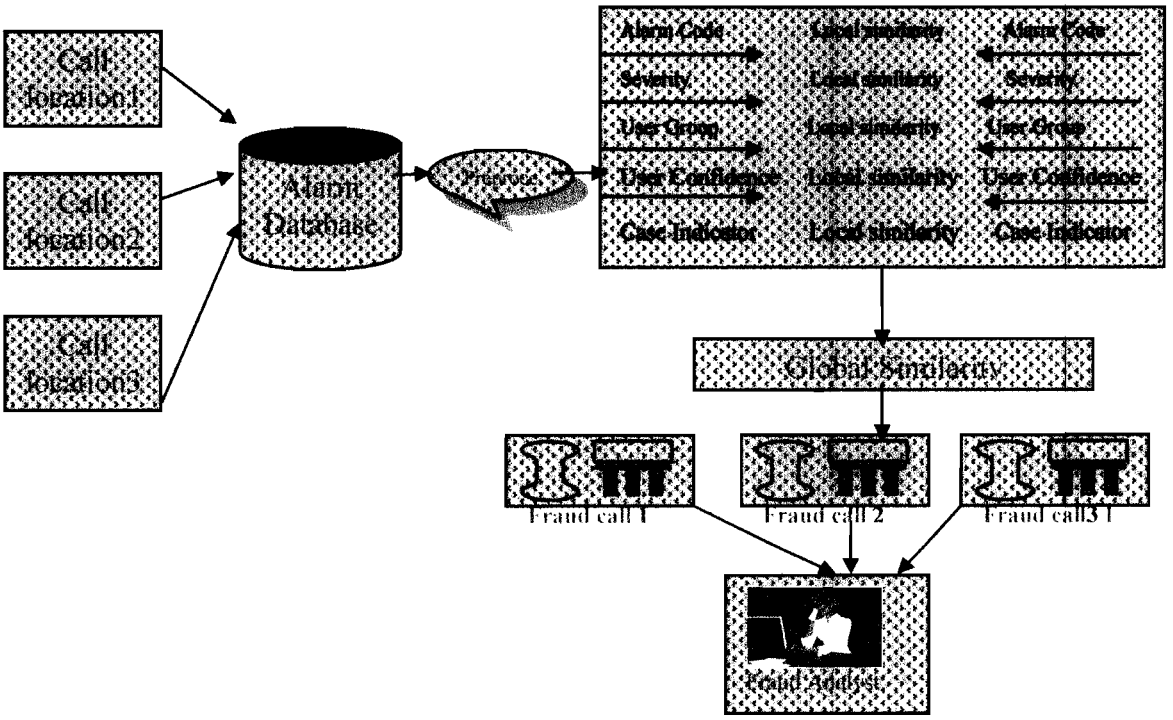


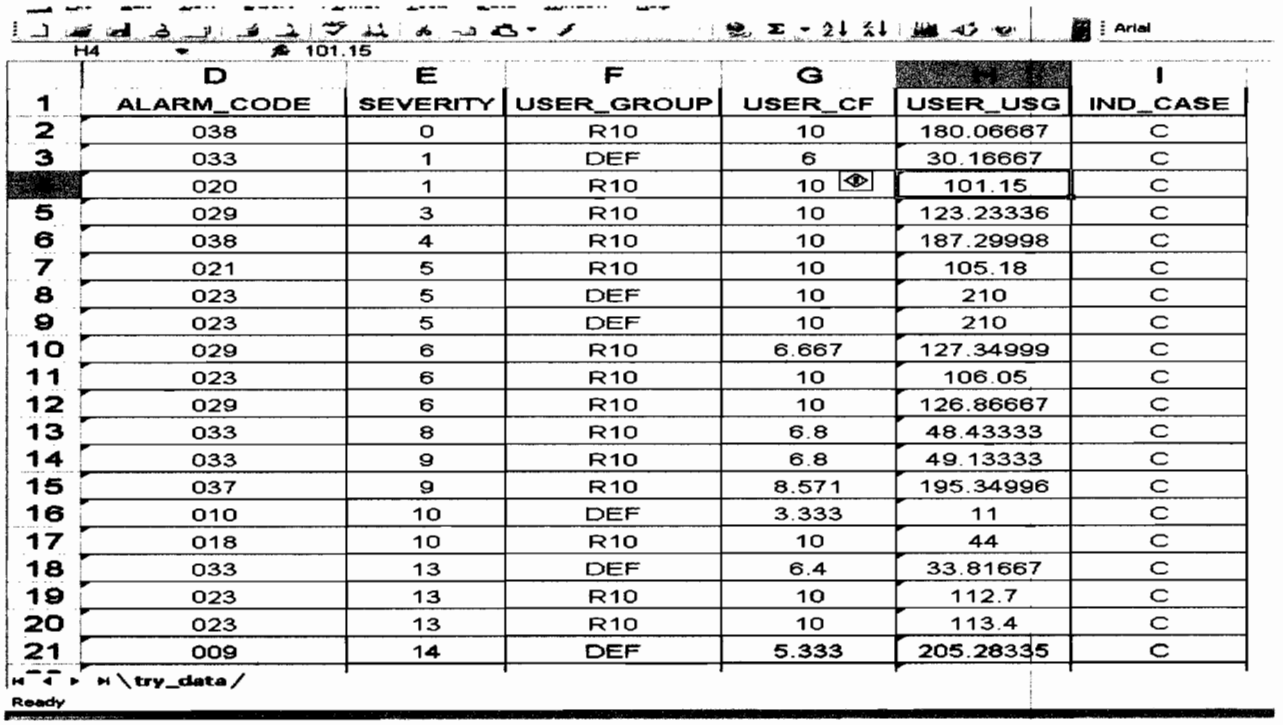
Figure 3.2: Fraud Detection System Architecture

3.2.3 Analyze and design the system

According to Nunamaker & Minder (1991) design is the most important part of a system development process. Design involves the understanding of the studied domain, the application of relevant scientific and technical knowledge, the creation of various alternatives, and the synthesis and evaluation of proposed alternative solutions. Design specifications will be used as a blueprint for the implementation of the system. For a software development project, design of data structures, databases, or knowledge bases should be determined at this phase. The program modules and functions also should be specified at this time after alternatives have been proposed and explored and final design decisions been made.

(a) Data Set:

This study uses a sample data set from Malaysian Telecommunications to develop the system. The data set is obtained from Consultant who has an ongoing research on Telecommunications. Figure 3.3 is the row sample data received.



H4 101.15

	D	E	F	G	H	I
1	ALARM_CODE	SEVERITY	USER_GROUP	USER_CF	USER_USG	IND_CASE
2	038	0	R10	10	180.06667	C
3	033	1	DEF	6	30.16667	C
4	020	1	R10	10	101.15	C
5	029	3	R10	10	123.23336	C
6	038	4	R10	10	187.29998	C
7	021	5	R10	10	105.18	C
8	023	5	DEF	10	210	C
9	023	5	DEF	10	210	C
10	029	6	R10	6.667	127.34999	C
11	023	6	R10	10	106.05	C
12	029	6	R10	10	126.86667	C
13	033	8	R10	6.8	48.43333	C
14	033	9	R10	6.8	49.13333	C
15	037	9	R10	8.571	195.34996	C
16	010	10	DEF	3.333	11	C
17	018	10	R10	10	44	C
18	033	13	DEF	6.4	33.81667	C
19	023	13	R10	10	112.7	C
20	023	13	R10	10	113.4	C
21	009	14	DEF	5.333	205.28335	C

try_data / Ready

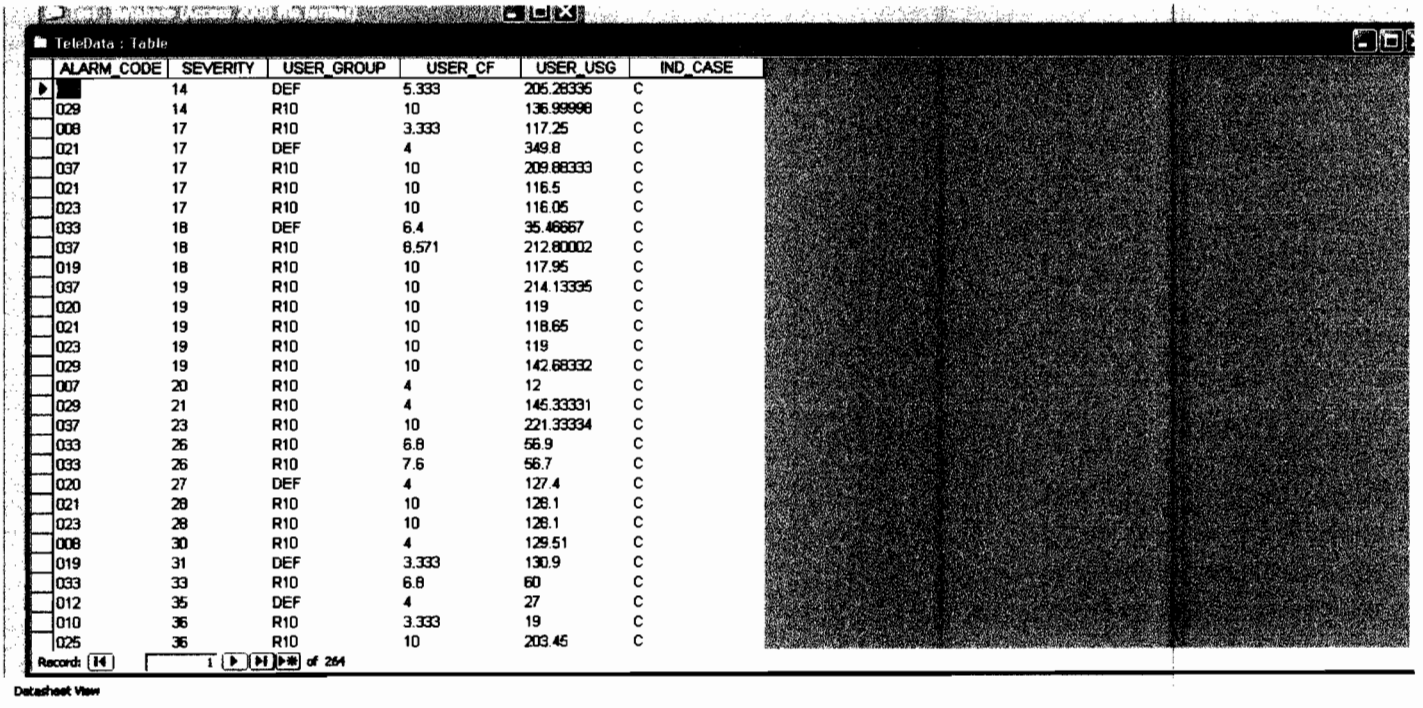
Figure 3.3: Sample data set in Microsoft Excel

The attribute of the data include Alarm_code, Severity, User_group,User_Confidence, User_Usage and Case_Indicator. All the attributes are in String.

Table 3.1: Data attributes

Attribute	Description
Alarm_code	The code responsible for trigger
Severity	Magnitude of the code
User_group	Subcribers group based on type of calls
User_Confidence	Text
User_Usage	Usage pattern of Subscriber
Case_Indicator	text

The data is then exported to Microsoft Access which will make it easier for connectivity to the programming language. Figure 3.4 shows the data in Microsoft access.



TeleData : Table

ALARM_CODE	SEVERITY	USER_GROUP	USER_CF	USER_USG	IND_CASE
	14	DEF	5.333	205.28335	C
029	14	R10	10	136.99998	C
008	17	R10	3.333	117.25	C
021	17	DEF	4	349.8	C
037	17	R10	10	209.88333	C
021	17	R10	10	116.5	C
023	17	R10	10	116.05	C
033	18	DEF	6.4	35.46667	C
037	18	R10	6.571	212.80002	C
019	18	R10	10	117.95	C
037	19	R10	10	214.13335	C
020	19	R10	10	119	C
021	19	R10	10	118.65	C
023	19	R10	10	119	C
029	19	R10	10	142.68332	C
007	20	R10	4	12	C
029	21	R10	4	145.33331	C
037	23	R10	10	221.33334	C
033	26	R10	6.8	56.9	C
033	26	R10	7.6	56.7	C
020	27	DEF	4	127.4	C
021	28	R10	10	128.1	C
023	28	R10	10	128.1	C
008	30	R10	4	129.51	C
019	31	DEF	3.333	130.9	C
033	33	R10	6.8	60	C
012	35	DEF	4	27	C
010	36	R10	3.333	19	C
025	36	R10	10	203.45	C

Record: 1 of 264

Datasheet View

Figure 3.4: Sample data set in Microsoft Access.

(b) Databases:

A database is created using Microsoft Access and the imported dataset from excel populate the table called Data table. The attributes of the table is shown in fig 3.5.

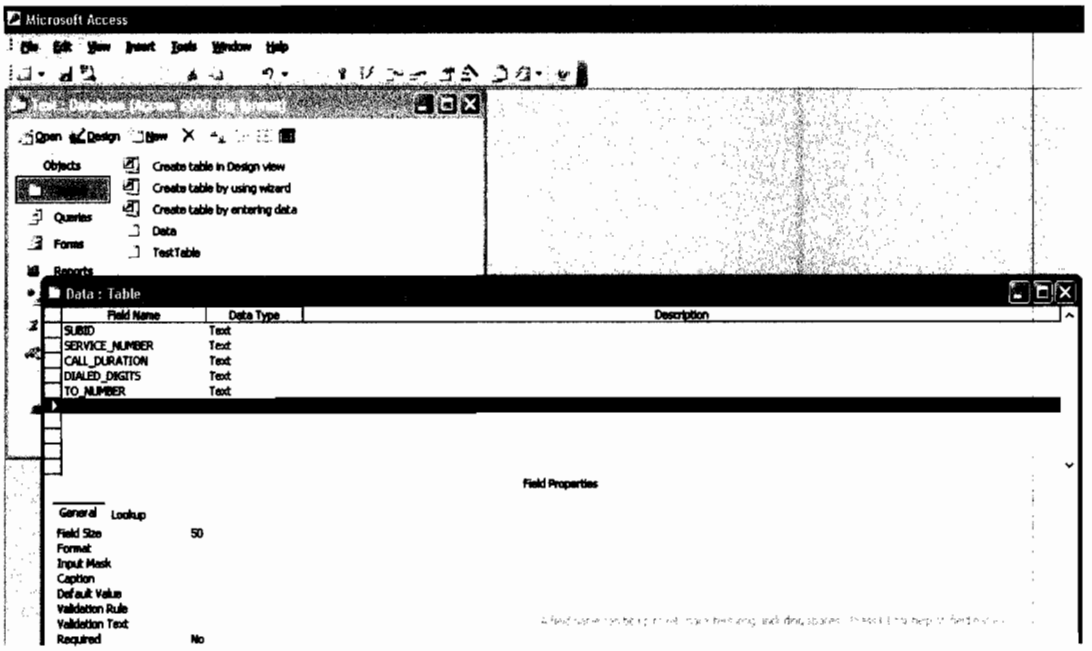


Figure 3.5: The attributes of Data table

(e) Case Based Reasoning

Expected Benefits of using CBR in this context are described as follows:

- In consideration of the fact that in practice problem areas are often not completely understood, even experts can't provide consistent rules for problem solving especially as it affects fraud pattern which is dynamically changing. In this regard, CBR systems often provide an acceptable quality of results.
- CBR systems guarantee a better utilization of the existing experience than other techniques by providing similarity-based searching.

- CBR systems problem solving approach typically uses the new specification of fraud to query and return and an appropriate result matching the specification considering the weight values specified by the expert.

(f) Case Retrieval Process

For retrieving cases that are most similar to the new created cases of attribute values, a matching component matches the new case against already existing best practice cases from the case base. The matching process consisting of identifying a basic set of similar cases using similarity measures; and, based on these identified most similar cases; the weight function is applied to find the Global Similarity. Therefore in order to achieve the basic set of matching best practice cases, the matching component is computing the similarity between attribute values.

Similarity is based on:

Non Numeric *sim* (a,b)= 1 if a=b .....(1)

$$Sim\ (a,b) = 0 \text{ if } a \neq b \quad \text{.....(2)}$$

$$\text{Global similarity (\%)} = 1/\sum w\ [/^{p}\sum w_i * sim(a_i, b_i)] * 100 \quad \text{.....(3)}$$

Table3.2.: Similarity Computation by Weighted Average

Profile Description (new case value)	Weight	Similarity [0-1]	Case X (from case base)
Alarm_Code (038)	1	1	038
Severity (1)	1	0	3
User_Group (R10)	1	0	DEF
User_Confidence (10)	1	1	10
Indication_Case (C)	1	0	V
User_Usergroup (210)	1 Total = 6	1	210
Global Similarity %	$1/\sum w\ [^{P}\sum w_i * sim(a_i, b_i)] * 100$ **[1/6 * [1*1+ 1*0+ 1*0+ 1*1+ 1*0+1*1]] * 100= 50%**		

Given the equation 1, 2, and 3, the sample data set with six attributes carrying uniform weight of 1 and similarity of three gives a global similarity of 50% after computation.

3.2.4 Build the System (Prototype)

Implementation of a system is used to demonstrate the feasibility of the design and the usability of the functionalities of a system development research project. It gives an insight into the advantages and disadvantages of the concepts, the frameworks, and the chosen design alternatives (Nunamaker & Minder, 1991).

For this project, a CBR technique is used as an AI technique to serve as basis for developing the proposed detection system through the recognition model. Therefore the approach is adopted from the (Mingyang & Agnar, 2005) pseudo code of the evaluation process for CBR as follows:

```
Procedure evaluation(CaseBase)
SuccessOnEqually,SuccessOnQuery,SuccessOnCase=false
TestCases,SessionsOnEqually,SessionsOnCase,SessionsOnQuery=0
GlobalWeights=weighting(CaseBase)
for each case X 2 CaseBase
        Xn=weighted1NNOnEqually(X,CaseBase-X)
        if Solution(Xn) = Solution(X) then
                TestCases=TestCases+1
                Xq=featureSelection(InitialFeatureNumber)
                do while not (SuccessOnEqually and SuccessOnCase
                        and SuccessOnQuery) and Xq 6= null
                        if not SuccessOnEqually then
                        ReturnedCasesOnEqually=weightedkNNOnEquall
                        (Xq,Ca  seBase-X)
                        SessionsOnEqually=SessionsOnEqually+1
                        if Xn 2 ReturnedCasesOnEqually then
                                SuccessOnEqually=true
                                End If
                        End If
                if not SuccessOnCase then
                        ReturnedCasesOnCase= weightedkNNOnCase(Xq,
                        CaseBase-X)
                        SessionsOnCase=SessionsOnCase+1
                        if Xn 2 ReturnedCasesOnCase then
                                SuccessOnCase=true
                        End If
                End If
                if not SuccessOnQuery then
                        ReturnedCasesOnQuery=
                        weightedkNNOnQuery(Xq,CaseBase-X)
                        SessionsOnQuery=SessionsOnQuery+1
                        if Xn 2 ReturnedCasesOnQuery then
                                SuccessOnQuery=true
                        End If
                End If
                Xq=Xq+featureSelection(1)
                End Loop
        End IF
```

End Loop
Return SessionsOnEqually
TestCases , SessionsOnCase
TestCases , SessionsOnQuery

For timely attainment of this system, Rapid Application Development (RAD) methodology is also invoked at this stage.

i) Rapid Application Development

RAD is a repetitive process in which analysts and users build a rudimentary version of an information system based on user feedback. It involves building a small portion and allowing the end users to constructively criticise then a modification is done and such iteration is repeated until a stable system is obtained. The processes of specification, design and implementation are concurrent. There is no detailed Specification and design documentation is minimised. The system is developed in a series of increments. End users evaluate each increment and make proposals for later increments. System user interfaces are usually developed using an interactive development system.In the case of this project, the consultant is serving this purpose. Fig 3.6 shows the rapid application life cycle architecture.

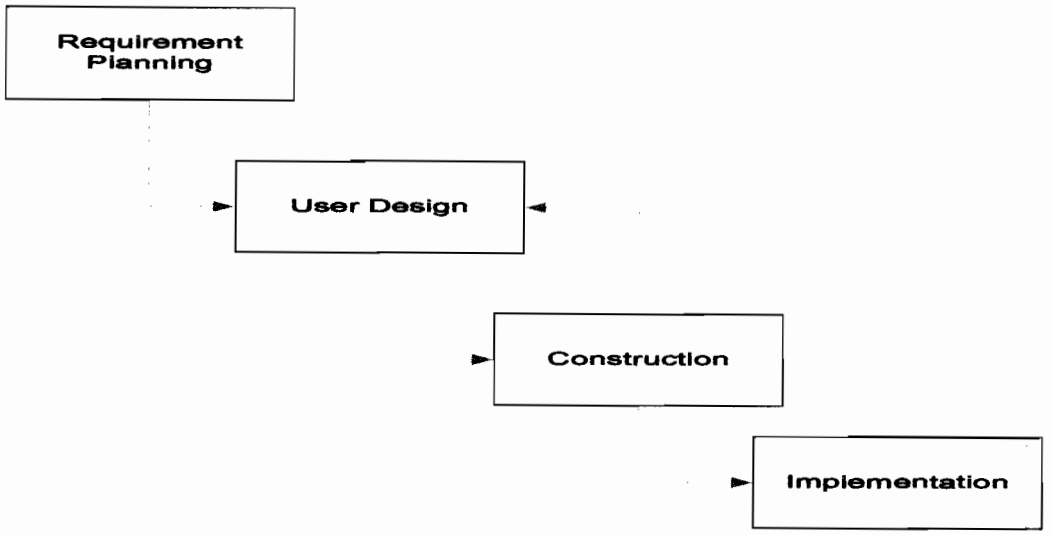


Figure 3.6: Rapid Application Development Architecture

Although some of the phases in the RAD has already being accomplished in the earlier stages, elaborate explanation will be given on the phases not yet attained.

a) Planning Requirement

Also known as the Concept Definition Stage, this stage defines the business functions and data subject areas that the system will support and determines the System's scope. This phase of planning requirement is done at the earlier stage of the project and Gantt chart is used to outline clearly the time schedule allocated for each stage to its logical conclusion.

b) User Design

Also known as the Functional Design Stage, this stage uses workshops to model the system's data and processes and to build a working prototype of critical system components. At this stage the design specification from the user perspective of a CBR is as depicted from the fig 3.7.

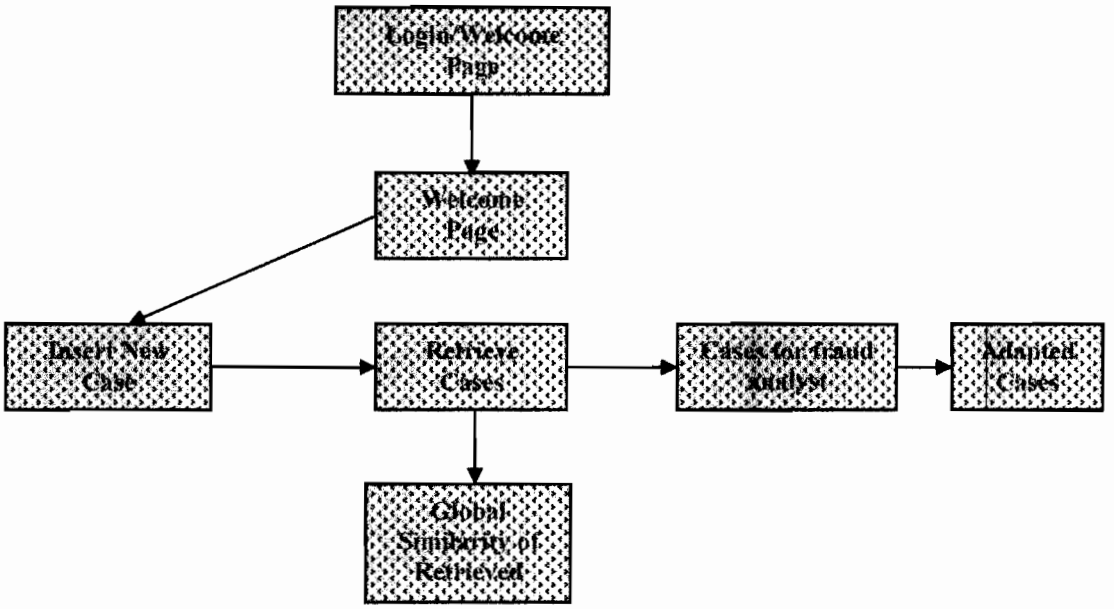


Figure 3.7: User Design Architecture

- **Use Case Diagram**

Use cases describe basic functions of the information system, they describes activities that are performed by the users of the systems. In this project the basic activity are as shown in the fig 3.8.

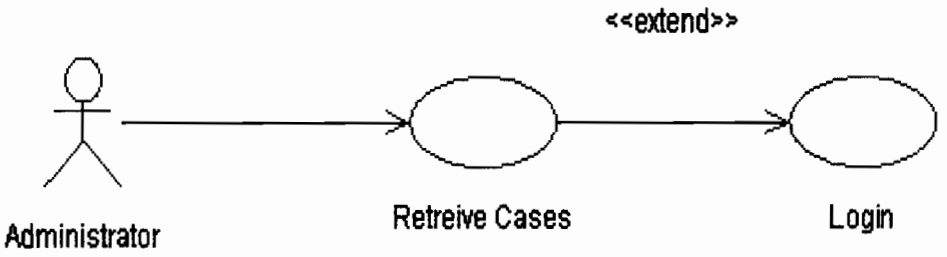


Figure 3.8: Use case Diagram

Fig 3.8 shows the major actor which is the system administrator and has the only two functions which is the login and retrieves function. The use case begin when the administrator login to the system and then can use the retrieve function.

- **Sequence diagram**

According to Hoffer *et al.,* (2002), sequence diagram depicts the interaction among objects during a certain period of time. Each sequence diagram shows the interaction related to use case diagram since the pattern of interaction varies from one use to another. In this project sequence diagram is illustrated as shown in Fig 3.6:

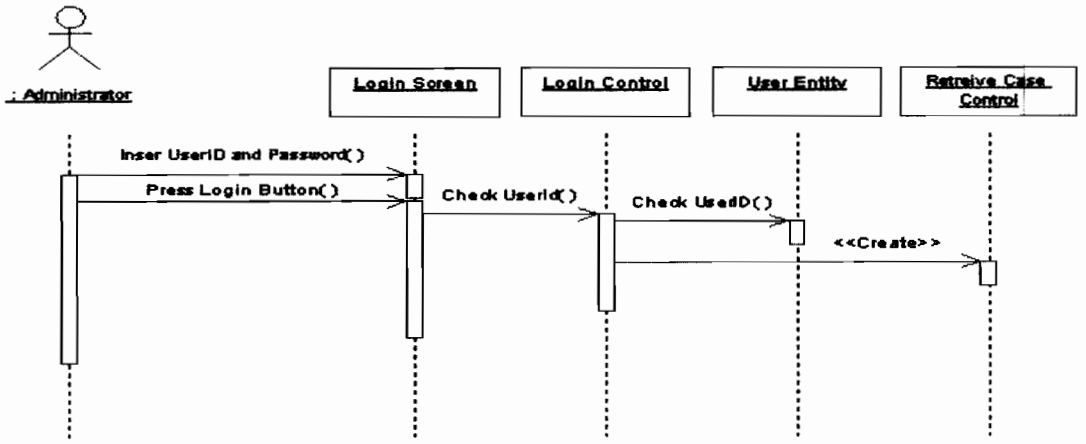


Figure 3.9: the sequence diagram for login

When the admin supply a valid username and password the login control checks with the user entity for the validity of the username and password, then the system for retrieved

control case.

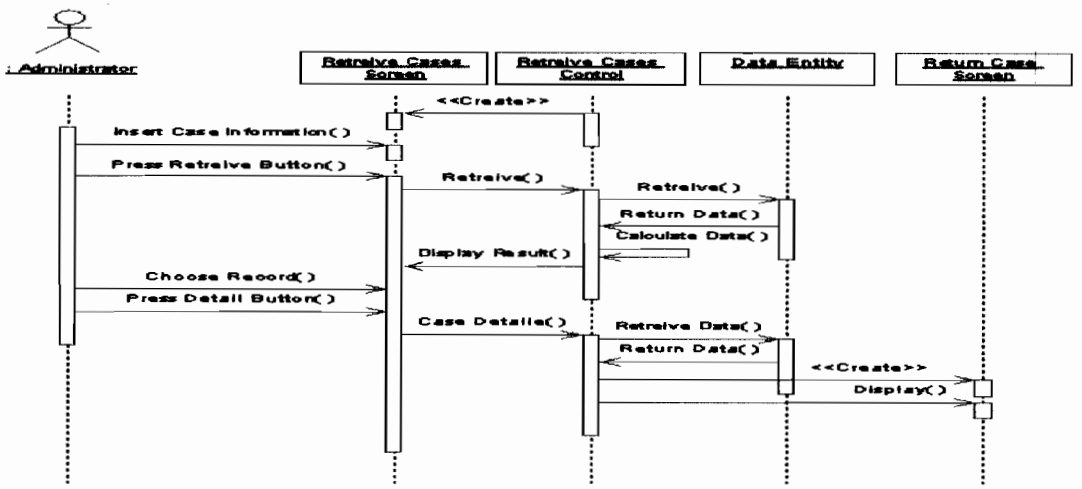


Figure 3.10: Retrieved case sequence diagram

Also when the administrator clicks on the retrieve button, the retrieve control send the data to the data entity for retrieval and return the display retrieved result to screen for administrative action. The corresponding collaboration diagrams are shown in fig 3.9.2 and 3.9.3.

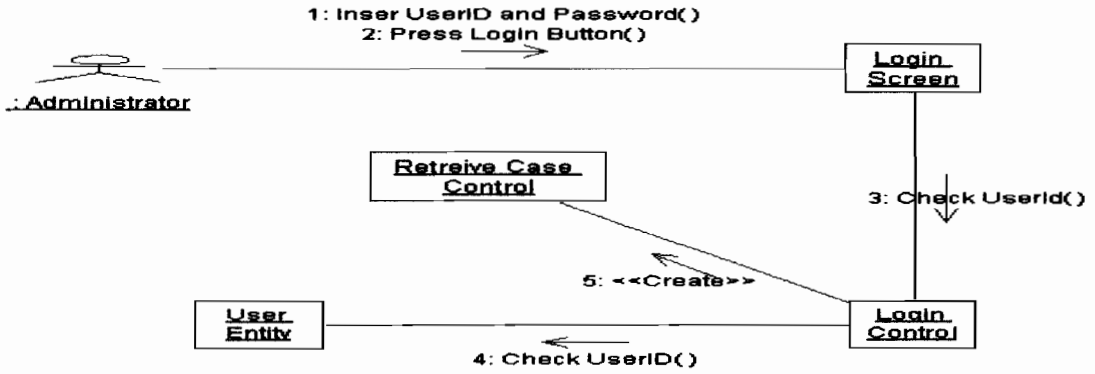


Figure 3.11: collaborative diagram

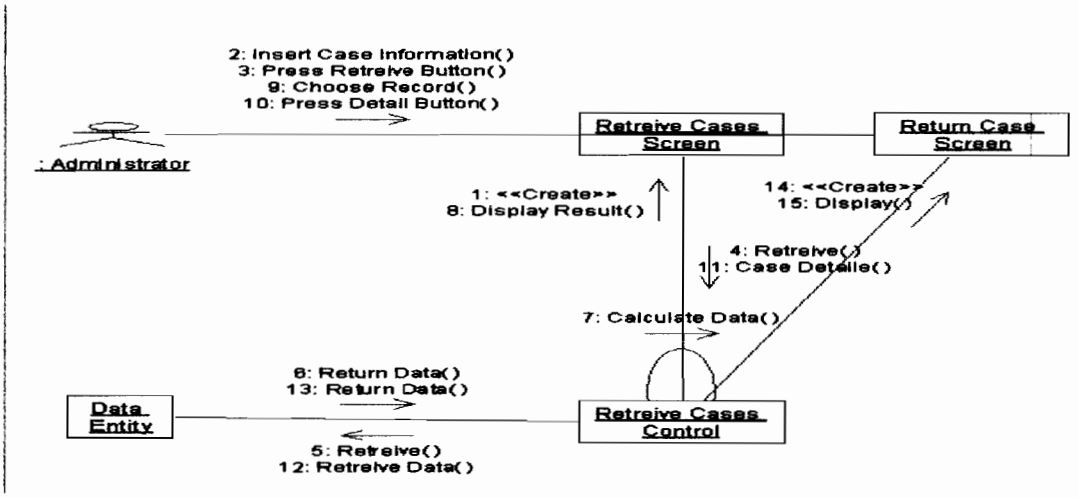


Figure 3.12: collaborative diagram

The detail for the use cases description is attached as Appendix A with the chapters of this project.

c) Construction

Also known as the Development Stage, this stage completes the construction of the physical application system, builds the conversion system, and develops user aids and implementation work plans. This study uses the ASP.Net to design the web page where the CBR engine is developed. Features of the Telecommunications and some walkthrough instructions are also inserted in the application for proper guide to users. The figure 3.9 below shows the design of the Welcome page using ASP.Net

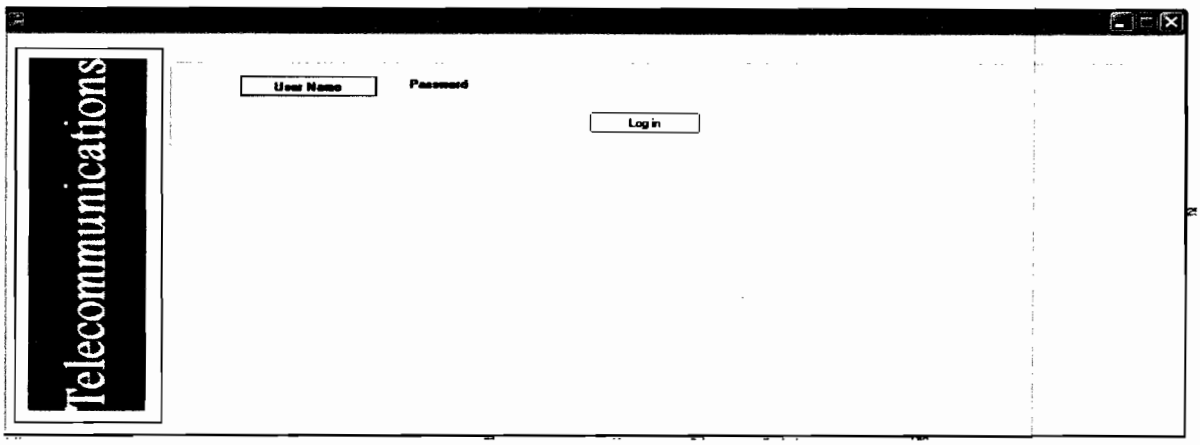


Figure 3.13: CBR Engine Welcome Page

d) Implementation

This is also known as the Deployment Stage, this stage includes final user testing and training, data conversion, and the implementation of the application system. The usability testing for this project will be discussed in detail in chapter 4 under result and findings.

3.2.5 Observe and Evaluate the System

Once the system is built, researchers can test it performance and usability as stated in the requirement definition phase, as well as observe its impacts on individuals, groups, or organizations. The test results should be interpreted and evaluated based on the conceptual framework and the requirements of the system defined at the earlier stages. Therefore test cases for various features in the CBR engine are formulated and the results or output from these cases were documented. Also included in the cases is the interaction between the new cases a user enters and the old cases for validity of what is happening in

the back end. The results of the evaluation are discussed in details in the subsequent chapter.

3.3 SUMMARY

Design is the key to Information System (IS) and the emphasis is on rigorous IS development (Konsynski and Nunamaker, 1982). Therefore System Development Research Methodology is used in this project due to its complimentary approach with other methodologies which in turn generates a fruitful research results in IS research. The next chapter will detail the findings of this project and the system evaluation.

CHAPTER FOUR

FINDINGS AND RESULTS

This chapter will explain the result of the use of CBR in detecting the fraud pattern in the telecom data. Already the system pictorial representation is done in previous chapters; more focus will be done on the functionalities, interface, and the system evaluation.

4.1 Functionalities

The functionality of this project followed the requirements set out by the users. The architecture diagram shown in previous chapter explains the routine of the system. Nevertheless the CBR approach fetches or retrieves data from the old cases in the case base that has similarity to the new case by applying the similarity computation with the given weights on the data.

4.2 Interface design

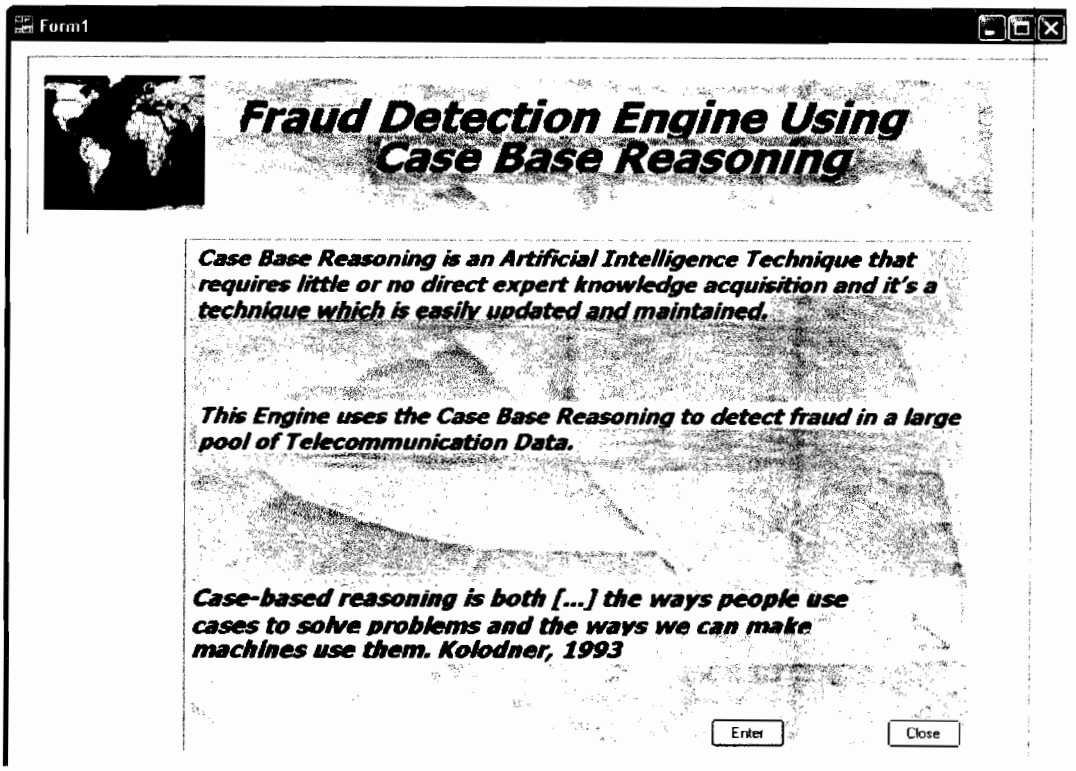


Figure 4.1: Welcome Page

Fig 4.1 shows the welcome page carrying the title of the engine and the Artificial Intelligence technique used to achieve the requirements of the system. It also contains a brief introduction on the technique as well as advantage of using this technique. Furthermore a highlight is given on the purpose and what is meant to be achieved from the CBR engine and the domain for the system is equally mentioned which is the telecommunications domain.

Enter button is clearly shown for the user to proceed to the next function in the system. Similarly a Close button is also provided for a user to exit at this point if he so wishes.

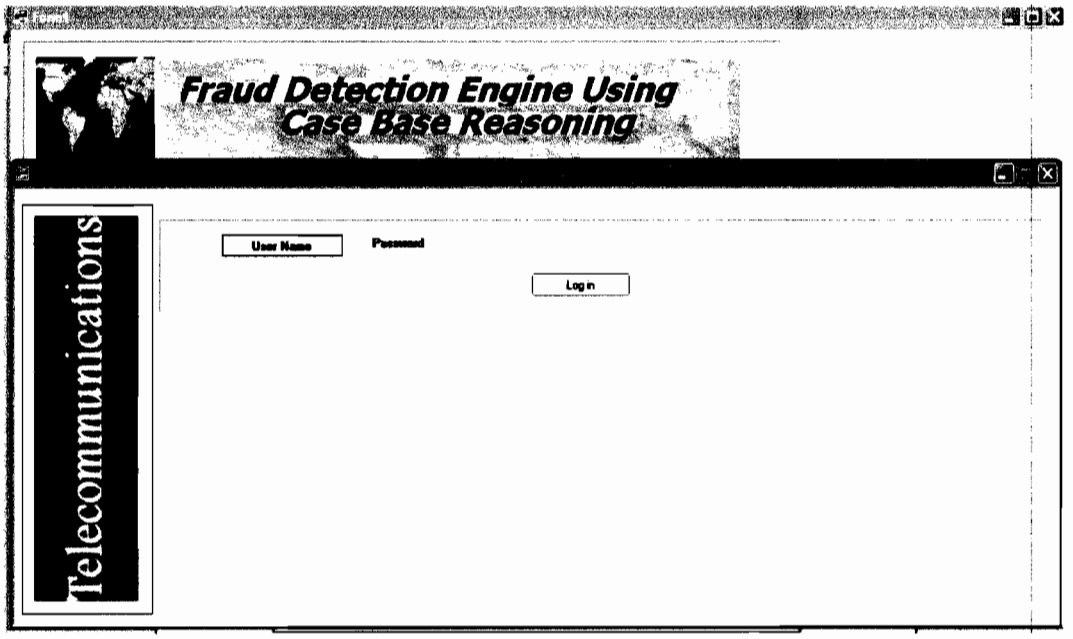


Figure 4.2: Login Page

Fig 4.2 shows the login page of the system. Here, the user must insert the right username and password in order to login into the system. At this stage only valid and authorized user is allowed into the system. This checks any unauthorized use of the system. Any invalid username or password caused the system to display an error message.

Figure 4.3 Error Message

Hence the page is loaded again for correction of user name or password.

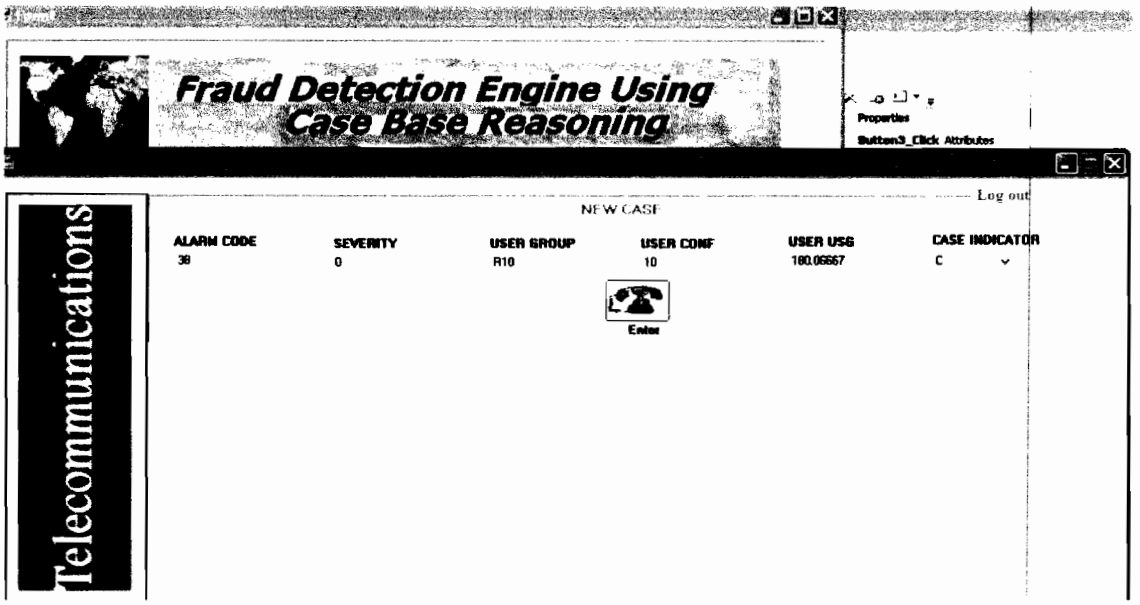


Figure 4.4: System Main Page

The fig 4.4 is the system's main engine page. It carries clearly the data attributes used during this study. The attributes include the Alarm Code, Severity, User Group, User Confidence. Others are Case Indicator and User Usage. The user is expected to insert the new cases based on these attributes. This new case when the enter button is pressed, is retrieved from the case base that is of similar pattern with the new case. Although the result that is displayed based on the CBR similarity computation earlier on discussed in Chapter 3, the retrieved cases are ranged based on the number of attributes that are similarly starting with only one attribute to all the six attributes.

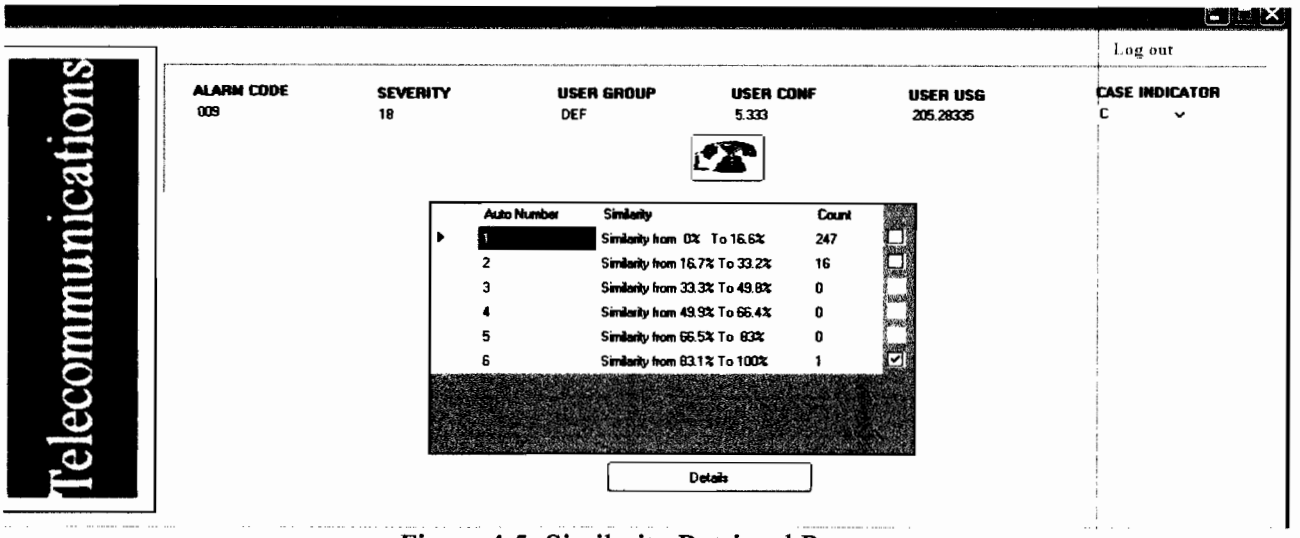


Figure 4.5: Similarity Retrieval Page

As can be seen in Fig 4.5, the new case is compared to those cases in the case base and retrieval based on 16.66% similarity, 33.33% similarity, 50% similarity until 100% similarity is displayed for the user to view and possible adopt in the case base. Details of the interface is attached in Appendix C including the user manual of the system as well as the system navigation for users understanding.

A prototype of the system is developed using Microsoft Visual Basic 2005 version. It also uses Microsoft access as the back end of the system. The process of CBR had been discussed in the previous stage.

4.3 System Test

The correctness of the retrieval process is one of major concerns in CBR systems. The CBR retrieval test is designed to evaluate the correctness of the retrieval function. The indexing system used, although not evaluated independently, is clearly part of the retrieval evaluation test. Therefore in order to achieve this test and to ascertain whether the system had been fully functioning and meet the user's requirement, the detail retrieval test algorithms is described as follows.

1) Copy a case in the case library to a test-case list

a) Use the first test case.

b) Present test case to the CBR system as the current problem (without the solution attribute), and obtain its solution.

c) Inspect the retrievals made by the CBR system.

- If the top retrieved case is the same case in the library which was used as the test case, the test is designated as "Successful," and the case is added to a successful-case list.

- Else, test is designated as "Failed,"

d) Delete test case from test-case list.

e) Go to step (a).is done.

2) Until the test-case list is empty:

a) Point to the first test case.

b) Remove from the case library the historical case corresponding to the current test case.

c) Present test case to the CBR system as the current problem to be solved (without the solution attribute), and obtain its solution.

d) Inspect the solution generated by the CBR and compare it to the solution of the test case.

- if so, test is designated as "Successful;" add it to a successful-case list.
- Else, test is designated as "Failed;" add it to a failed-case list.

f) Return the removed historical case back into the case library.

g) Delete test case from test-case list.

h) Go to step a.

4.3.1 Percentage Similarity

The test cases are presented in terms of the similarities obtained from each experiment to demonstrate the evaluation performance by the proposed prototype. The next cases are described in the following subsection.

(i) Test Cases 1

For this experiment, all weights attributes are assigned to 1. Therefore all attribute correspond to the cases in the case base attribute. The result is shown in Fig 4.6.

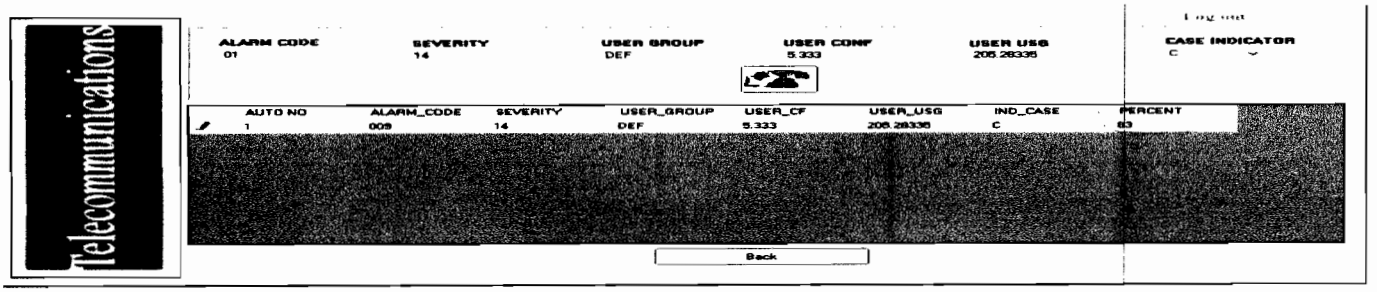


Figure 4.6: CBR output when attribute 1 match

(ii) Test Case 2

In this case all weight attributes are assigned to 1; however one of the attribute is not equal to the case based. Hence the result is 84%

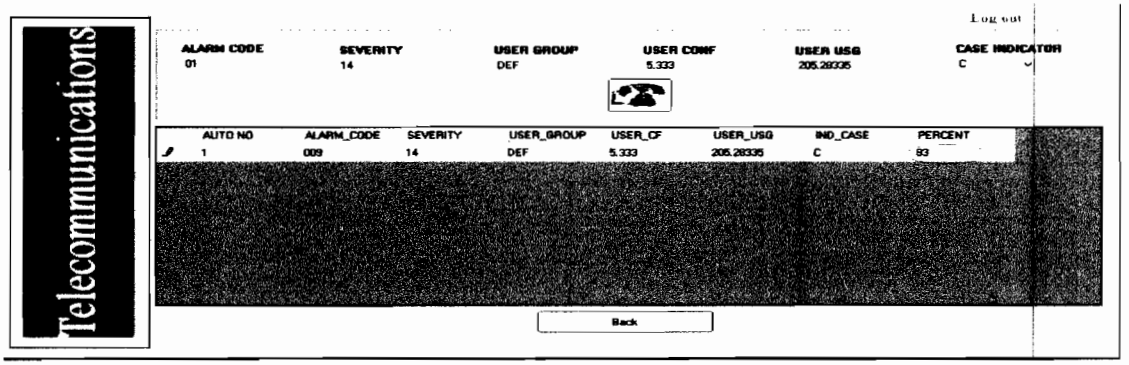


Figure 4.7: CBR output when attribute 1 is not matched

(iii) Test Case 3

In this case all weight attributes are assigned to 1 but attribute 1 is assigned to 0.9, However one of the attribute is not equal to the case based. Hence the result is 84.74% as shown in Fig 4.7

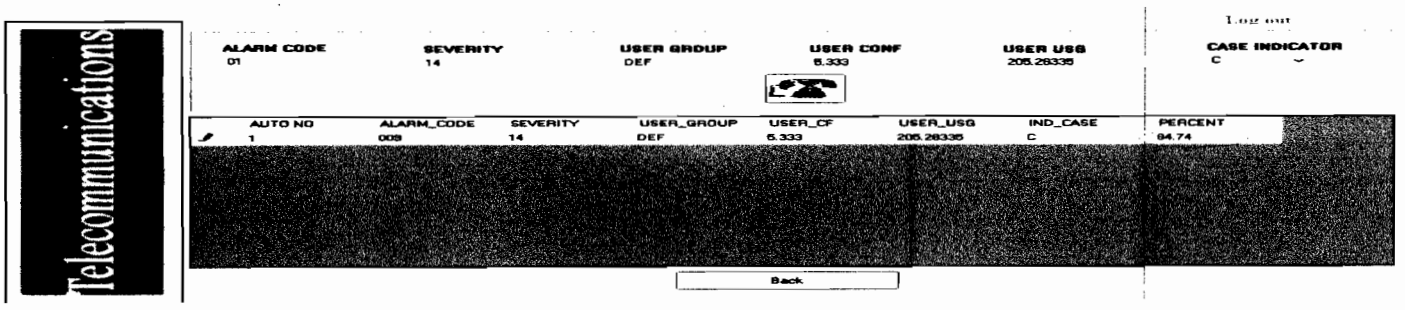


Figure 4.7: CBR output when attribute 1 is not match

(iv) Test Case 4

In this case all weight attributes are assigned to 1 but attribute 1 is assigned to 0.8, However one of the attribute 3.05 is not equal to the case based. Hence the result is 83.05 % as shown in Fig 4.8

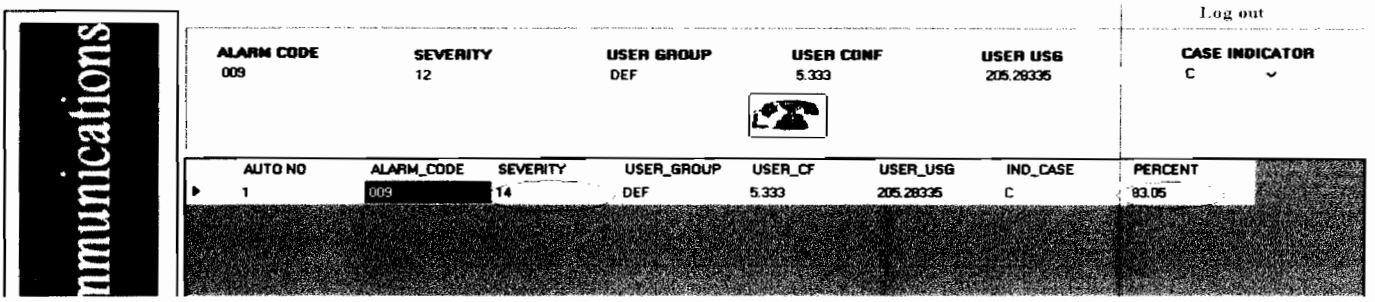


Figure 4.8: CBR output when attribute 1 is not matches

Therefore it can be observed that the percentage does not change with other attributes when the weight value remains the same.

Hence investigation is further carried out to determine the appropriate weight of the attributes in order to fix the suitable weight. This gives confidence to the detection in the CBR engine and guarantee minimal involvement of the fraud analyst rather than the data itself. Detail of these test cases is shown in Appendix D and Table 1.1 shows the summary result as follows:

Table 4.1 the summary of result for test 1,2 various attributes weight

Case 1		Case 2	
Attribute 1 weight	Percentage Similarity Not Matched	Attribute 2 weight	Percentage Similarity Not Matched
1.0	83	1.0	90.41
0.9	84.74	0.9	92.23
0.8	86.21	0.8	92.56
0.7	87.7	0.7	93.18
0.6	89.2	0.6	97.77
0.5	90.90	0.5	97.82
0.4	91.25	0.4	97.87
0.3	96.03	0.3	97.9
0.2	96.15	0.2	98.00
0.1	98.03	0.1	99.12

Table 4.2: the summary of result for test case 3,4 various attributes weight

Case 3		Case 4	
Attribute 1 weight	Percentage Similarity Not Matched	Attribute 2 weight	Percentage Similarity Not Matched
1.0	76.19	1.0	69.70
0.9	75.61	0.9	71.88
0.8	80.00	0.8	74.19
0.7	82.05	0.7	76.67
0.6	84.21	0.6	79.31
0.5	86.49	0.5	82.14
0.4	88.89	0.4	85.19
0.3	91.42	0.3	88.46
0.2	94.12	0.2	92.00
0.1	96.96	0.1	95.83

Table 4.3 the summary of result for test case 5,6 various attributes weight

Case 5		Case 6	
Attribute 1 weight	Percentage Similarity Not Matched	Attribute 2 weight	Percentage Similarity Not Matched
1.0	58.33	1.0	33.33
0.9	60.89	0.9	35.71
0.8	63.63	0.8	38.46
0.7	66.67	0.7	41.67
0.6	70.00	0.6	45.46
0.5	73.68	0.5	50.00
0.4	77.79	0.4	55.55
0.3	82.35	0.3	62.50
0.2	87.5	0.2	71.43
0.1	93.33	0.1	83.33

One interesting point is that the smaller the weight value is the higher similarity percentage to this end; the study proposed the weight for all attributes are set to 0.1 in order to get higher similarity values. Furthermore attribute 2 has recorded 99.12% similarity more than the rest of the attributes which shows its significant compared to the

rest of the attributes. On the other hand attribute 6 recorded a lowest similarity compared to the rest of attribute.

4.4 CONCLUSION

This chapter explains the careful implementation and testing of the prototype system. The main goal was to use the case based reasoning technique in AI to find or detect fraud in the telecom data. The next chapter will draw a conclusion from the project and to suggest a way forward in subsequent enhancement.

CHAPTER FIVE

CONCLUSION

This chapter concludes the entire project by stating in summary the achievement and findings. Suggestion for future enhancement or improvement is also made, since the system has not reached a complete perfection. Therefore based on the previous chapters the following summary is made:

5.1 Project Summary

The objective as stated in the introductory chapter of this project is to develop a prototype system using a case based reasoning technique to find a fraud pattern in the telecommunication data. Hence the implementation has been done specifically to conform to the objective in its entire ramification.

As mentioned earlier use of Microsoft Visual Basic 2005 has facilitated the development by creating a user friendly interface where an initial login page guarantee some level of security into the system such that access or privilege is given to an authorized user. Furthermore the system has made available a legible page for to insert new case which is required to be detected whether it is a fraud case or not. On inserting a new case, the

system checks in the case based of similar fraud pattern and return the result with computation of how much similar is the new case to the existing case in the case base. The system also provide for the user the flexibility of retaining cases below a certain threshold in the case base in an event that the new case do not match the existing cases. However due to time constrain for the project accomplishment, the following aspects were not accomplished.

5.2 Project Limitations

i) There have been very few functions in the system. Specifically the retrieval and adoption of cases are the main functions with respect to the objective set out.

The system does not follow a real time design for a fraud engine where a new case of fraud comes into the system on real time and not inserted by the user.

5.3 Recommendation for Future Works

It is recommended that the limitations stated earlier should be followed to its logical conclusion and specifically more functionalities in term of fraud case revision and real time application should be included. Future work can be focus also on the use of hybrid techniques together with the CBR in terms of string attribute nature of data. This will guarantee precision in saying that a call made is actually a fraud call or not.

Reference:

Adem, K., Senay Y., Cengiz, K., & Mert, S. (2007). Fraud Detection Using an Adaptive Neuro-Fuzzy Inference System in Mobile Telecommunication. Networks Information Sciences 2007, pp1440-1446.

Amani, N., Fathi, M., Dehghan, M., (2005). A case-based reasoning method for alarm filtering and correlation in telecommunication networks Electrical and Computer Engineering, 2005. in Canadian Conference, 2005, pp2182 - 2186

Burge, P., & Shawe-Taylor, J., (2001). An Unsupervised Neural Network Approach to Profiling the Behaviour of Mobile Phone Users for Use in Fraud Detection. Journal of Parallel and Distributed Computing 61, pp915-925.

THE HINDU group of publications. Business Daily (n.d.). Retrieved January 23, 2008, from http://www.thehindubusinessline.com/2006/09/08/stories/2006090803460400.htm

Clifton, P., Vincent, L., Kate, S. & Ross, G., (2005). Comprehensive Survey of Data Mining-based Fraud Detection Research. School of Business Systems, Faculty of Information Technology, Monash University, Clayton campus Wellington Road, Clayton, Victoria 3800, Australia

CFCA press release (2007). Retrieved on 15th April 2008 from http://www.cfca.org/pdf/press/3-28-6PR.pdf,

Fair, I., (2003). Prepaid Telecommunications Fraud Techniques and Detection Retrieved on 15th April 2008 from: www.fairisaac.com/telecom.

Jaakko, H., (2000). User profiling and classification for fraud detection in mobile communications networks. Dissertation for the degree of Doctor of Science in Technology, Helsinki University of Technology Department of Computer Science and Engineering Laboratory of Computer and Information Science.

Jimmy, M., & Seán, H., (2003). Approach to Rules based Fraud Management in Emerging Converged Networks. Telecommunications Software & Systems Group, Waterford Institute of Technology, Ireland

Joshua, D., Bruce, M,. and David, W., (2004). Towards Applying Case-Based Reasoning to Composable Behavior Modeling. In the Proceedings of the 2004 Conference on Behavior Representation (BRIMS), Arlington, Virginia.

Jian-Bo Yang; Jun Liu; Jin Wang; Sii, H.S.(2003). The Evidential Reasoning approach for Inference in rule-based systems. In Systems, Man and Cybernetics, 2003. IEEE International Conference, 3, pp2461 - 2468

Michael, H., Diane, L., Jos´e C., & Don, X. (2000). Detecting Fraud in the Real world. Retrieved from http://cm.bell-labs.com/cm/ms/departments/sia/jcp/HMDS.ps

Pablo, A., Claudio, M., & Claudio, A. (2006). Subscription Fraud Prevention in Telecommunications using Fuzzy Rules and Neural Networks. Expert Systems with Applications Journal, August 2006, pp337-344.

Rupesh, K. G., and Saroj, K., (2007). Rule-based Approach for Anomaly Detection in Subscriber Usage Pattern. World Academy Of Science, Engineering And Technology Journal , 25, pp1307-6884.

Wheeler, R., Aitken, S (2000). Multiple Algorithms for Fraud Detection. Knowledge Based Systems Journal, 13 pp2-3 & 93-99.

Wikipedia, the free encyclopedia (2008). Case Based Reasoning retrieved on 24th January 2008 from http://en.wikipedia.org/wiki/Case-based_reasoning

Giulio, V., and Alessandro, R. (2004). Using Case-based Reasoning to support Operational Knowledge In Proceedings of 14th International Conference, EKAW 2004 Whittlebury Hall, UK,

Tomoharu, N., Gaku, N., and Hisao I., (2003). Credit Assignment by Fuzzy Rule-Based Systems in Fuzzy Classifier Ensembles. In a Proceedings 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation, Kobe, Japan.

Ching-Chang, W., and Chia-Chong, C., (2000). An SVD-QR-based approach to fuzzy modeling. In Systems, Man and Cybernetics, 2000. IEEE International Conference.

Pattara, W., and Peachavanish, R.,(2007). Estimating Road Traffic Congestion from Cell Dwell Time using Neural Network. In Telecommunications, 2007. ITST '07. 7th International Conference on ITS.

Munshi, K. Vempada, P., Sheila P., Sonmez, E., and Schumacher, H., (2003). Small signal and large signal modeling of HBT's using neural networks. In Telecommunications in Modern Satellite, Cable and Broadcasting Service, 2003. TELSIKS 2003. 6th International Conference.

Zhi-Wei, N., Shan-Lin, Y., Long-Shu, L., and Rui-Yu J., (2003). Integrated case-based reasoning. In machine Learning and Cybernetics, 2003 International Conference

G. Kamp, S. Lange and C. Globig, (1998) Case-based reasoning technology: related areas. In: M. Lenz Editor, *Case-based Reasoning Technology: from Foundations to ApplicationLNAI no. 1400* Springer, Berlin, pp. 325

Intec Telecom Systems PLC (2002). The Boom of New Technologies and the Rise of Telecom Fraud in Malaysia. Retrieved on 15th April 2008 from http://www.cybersecurity.org.my/bahasa/knowledge_bank/news/2002/main/detail/937/index.html

Mingyang Gu, Xin Tong, and Agnar Aamodt, (2005). Comparing Similarity Calculation Methods in Conversational CBR. Department of Computer and Information Science, Norwegian University of Science

Mingyang Gu1 and Xin Tong2(2004) An Intelligent Component Retrieval System Using Conversational CBR. a Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC'04) 0730-3157/04

Mehmet S. Aktas, Marlon Pierce, Geoffrey C. Fox, David Leake (2004). A Web based Conversational Case-Based Recommender System for Ontology aided Metadata Discovery. In proceedings of the Fifth IEEE/ACM International Workshop on Grid Computing (GRID'04) 1550-5510/04

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7, pp39-59.

Nenad S. Kojić, Irini S. Reljin, Branimir D. Reljin (2007). Different Wavelength Assignment Techniques in Al l-Optical Networks Controlled by Neural Network.

O. Dehzangi, M. J. Zolghadri, S. Taheri and S.M. Fakhrahmad (2007).Efficient fuzzy rule generation: A new approach using data mining principles and rule weighting. IN Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007).

J. Shawe-Taylor, K. Howker, P. Gosset, M. Hyland, H. Verrelst, Y. Moreau, C. Stoermann, and P. Burge (2000). In Business Applications of Neural Networks, chapter Novel techniques for prolong and fraud detection in mobile telecommunications, pp113-139.

Wheeler, R. and Aitken, S.(2000) Multiple Algorithms for Fraud Detection. In Journal Knowledge-Based Systems, 13,pp2-39 & 3-99.

David G. Messerschmitt (1996).Convergence of telecommunications with computing Department of Electrical Engineering and Computer Sciences University of California. In Proceeding of the IEEE, 84,8, pp1167-1186

John Beacon (2004). The impact of Mobile telephoning in Australia. In a conference Australian Mobile telecommunication Association conference.

Bolton R. J. And Hand D. J. (2002). Statistical Fraud Detection: A Review. In journal Statistical Science 17, 3, pp235–255.

Clifton P., Damminda A., and Vincent L. (2004). Minority Report in Fraud Detection:Classification of Skewed Data. In ACM SIGKDD Explorations Newsletter, 6 , 1. pp50-59.

Abhinav S., Amlan K., Shamik S., and Arun K. M. (2008) Credit Card Fraud Detection Using Hidden Markov Model. In proceeding Ieee Transactions On Dependable And Secure Computing 5, 1, pp37-48

Marius M., and Anders K. (2004). Representing and Reasoning about Context in a Mobile Environment Proceedings Workshop on Modelling and Retrieval of Context, 2004

Janet L. K. (2004). An introduction to case-based reasoning, Netherlands: Springer.

Olusola A. (2005). Data Mining, Fraud Detection and Mobile Telecommunications: Call Pattern Analysis with Unsupervised Neural Networks. Master of Science thesis, Computer Science, University of the Western Cape..

Frank R. J., Hunt S. P., and Davey N. (2000). Applications of Neural Networks to Telecommunications Systems.

Bonchi F., Giannotti F., Mainetto G., Pedreschi D. (1999). A classification-based methodology for planning audit strategies in fraud detection. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp175 – 184.

Bailey, K. D. (1982), ***Methods of Social Research,*** The Free Press,.

Blake, S. P.,(1978) ***Managing for Responsive Research and Development,*** W. H. Freeman and Company,

Nunamaker J.F., Chen M. and Purdin T.D.M. (1991). Systems development in information systems research. In Journal of Management Information Systems. 7, 3, pp. 89-106, 1991.

Blalock, A. **B.** and Blalock, H. M., Jr (1982). Introduction to Social Research, Prentice-Hall, second edition,.

Konsynski, B. R. and Nunamaker, J. F., Jr., (1982). Plexsys: A System Development System," in Couger, Colter, and Knapp (eds.), Advanced Systems Development/Feasibility Techniques, John Wiley & Sons, Inc.,.

APPENDIX A

1 USE CASE : LOGIN

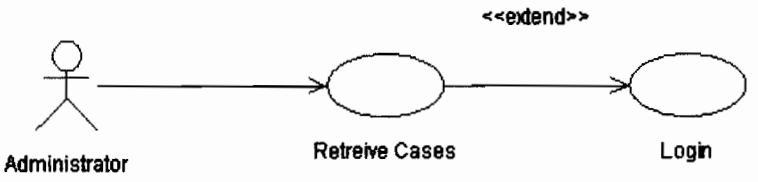


Figure 1.0 Login use case

1.1 BRIEF DESCRIPTION

This use case is initiated by the administrator when he clicks on the login button. This will enable the administrator to enter to the system.

1.2 PRE-CONDITIONS

The administrator must have user name and password.

1.3 CHARACTERISTIC OF ACTIVATION

Event Driven (on administrator's demand)

1.4 FLOW OF EVENTS

1.4.1 Basic Flow (VMS_01_01)

- This use case begins when the administrator enter to the system.
- The system shall display main screen.
- The administrator will insert his user name and password and press login.
- The system shall check the user name and password with the existing user name and password in the database.
- The system will open.

1.4.2 Alternative Flow

Not Applicable.

1.4.3 Exceptional Flow

E-1: invalid user name and password

The system shall display the message "Invalid user name and password". The system waits until the voter insert new user name and password.

1.5 POST-CONDITIONS

- The admin can Retrieve new case

1.6 RULE(S)

Not applicable.

1.7 CONSTRAINT(S)

The admin can only insert valid user name and password otherwise will be identified as hacker.

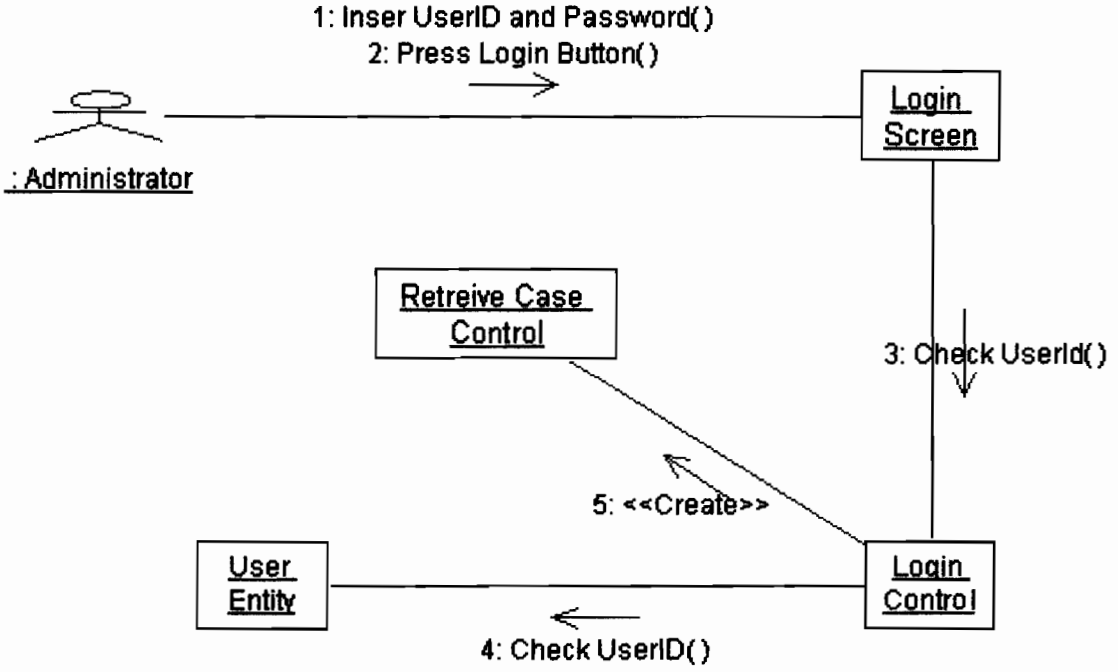


Figure 1.1: Collaboration Diagram for Login

1.0 APPENDIX B

1.1 USE CASE: Retrieve case

1.2 BRIEF DESCRIPTION

This use case is initiated by the administrator. This use case will enable the admin to insert new case for retrieval by the system.

1.3 PRE-CONDITIONS

Not applicable.

1.4 CHARACTERISTIC OF ACTIVATION

Event Driven (on admin's demand)

1.5 FLOW OF EVENTS

1.5.1 Basic Flow (VMS_01_01)

- This use case begins when the admin selects login from main page.
- The system shall display field for new cases to be entered as new case.
- The admin will insert the details of the new case
- The system will search the comparison of the new case with the old case in the database
- The system will return the similarity range for further action.

1.5.2 Alternative Flow

Logout will clear the window and return to previous position

1.5.3 Exceptional Flow

Not applicable.

1.6 POST-CONDITIONS

- The admin can select the range of similarity to display.

1.7 RULE(S)

Not applicable

1.8 CONSTRAINT(S)

Not applicable

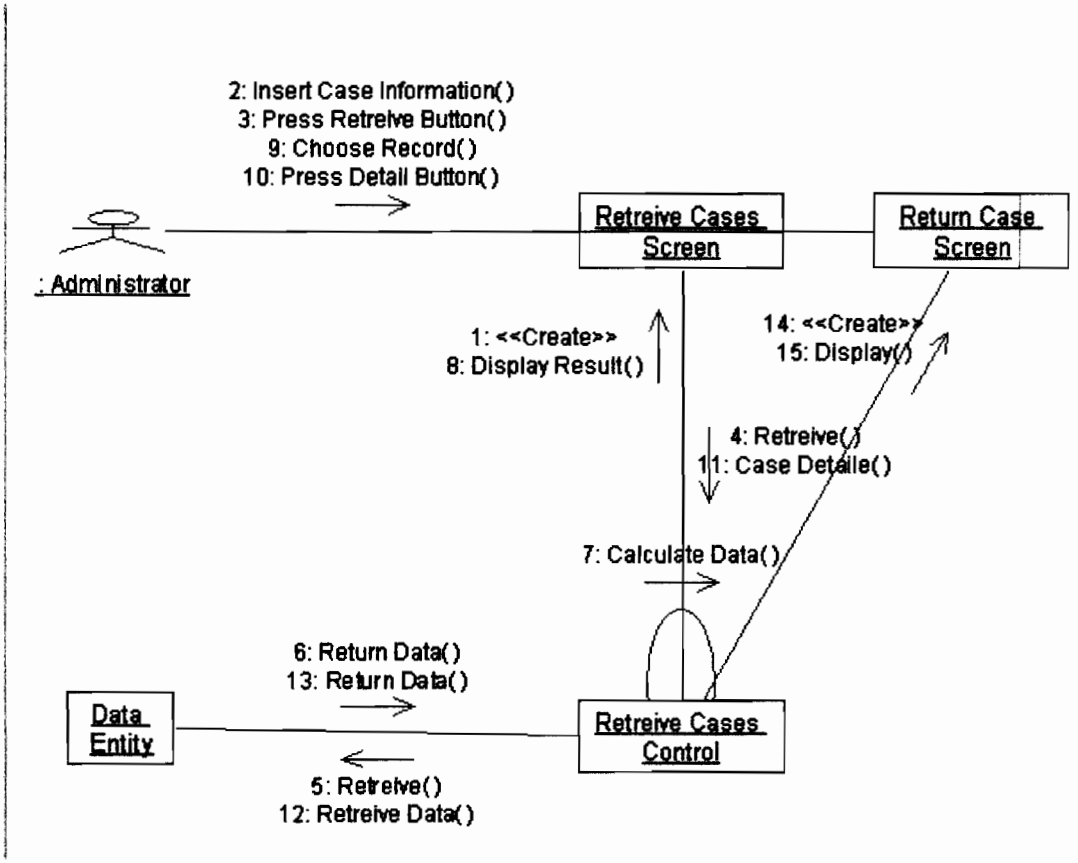


Figure 1.2: Collaboration Diagram for Case Retrieval

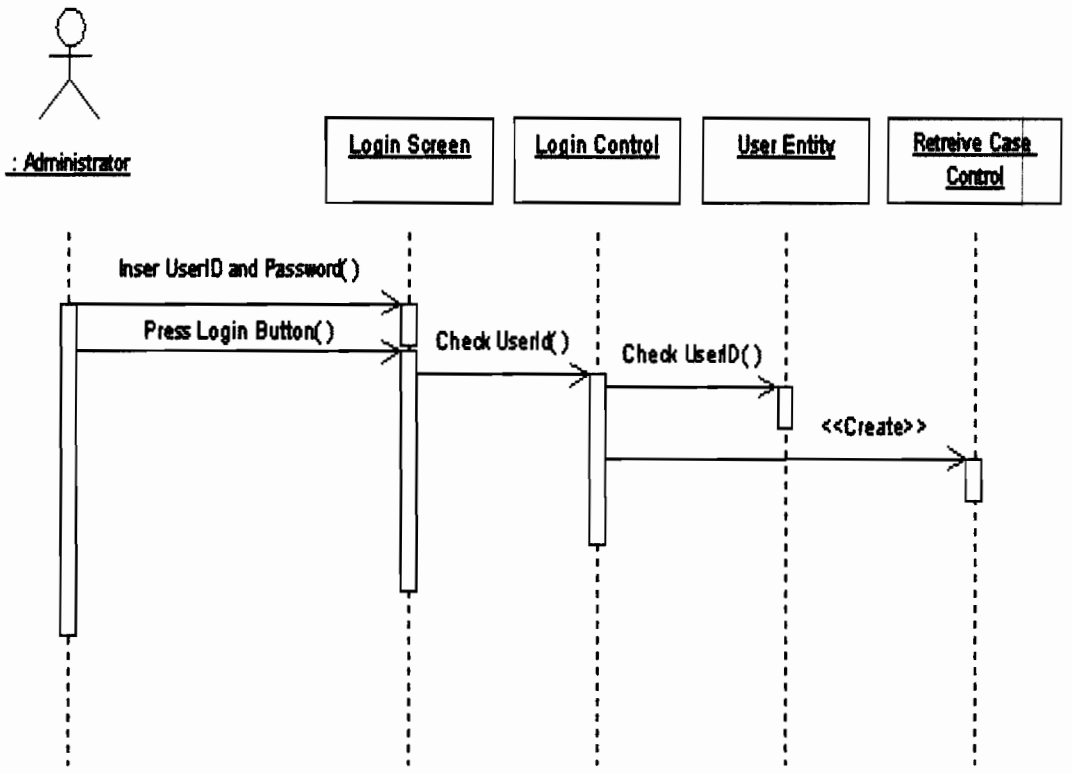


Figure 1.3: Sequence Diagram for Login

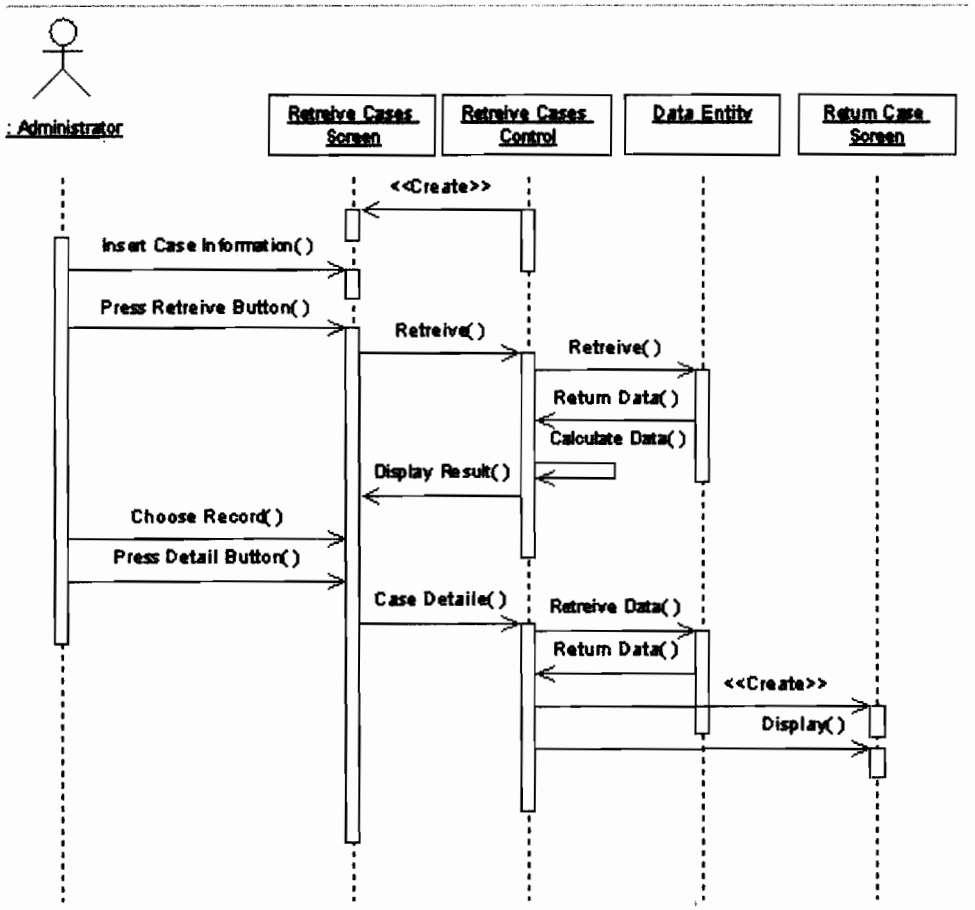


Figure 1.4: Sequence Diagram for Case Retrieval

APPENDIX C:

1.1 User manual for CBR engine for fraud detection;

Welcome page

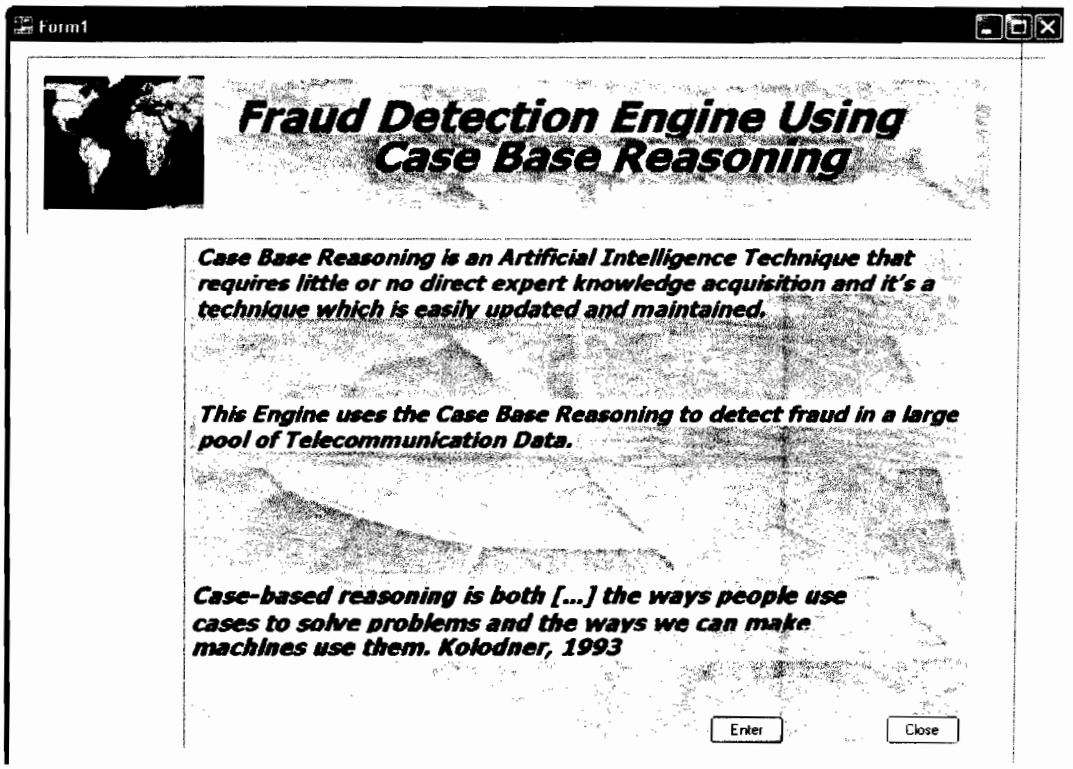


Figure 1.5: Welcome Page

1. The welcome is the first page when the system is started. The user is expected to click the enter button to proceed or the close button to exit the application
2. The title of the application is boldly shown and a brief description is given on the CBR.

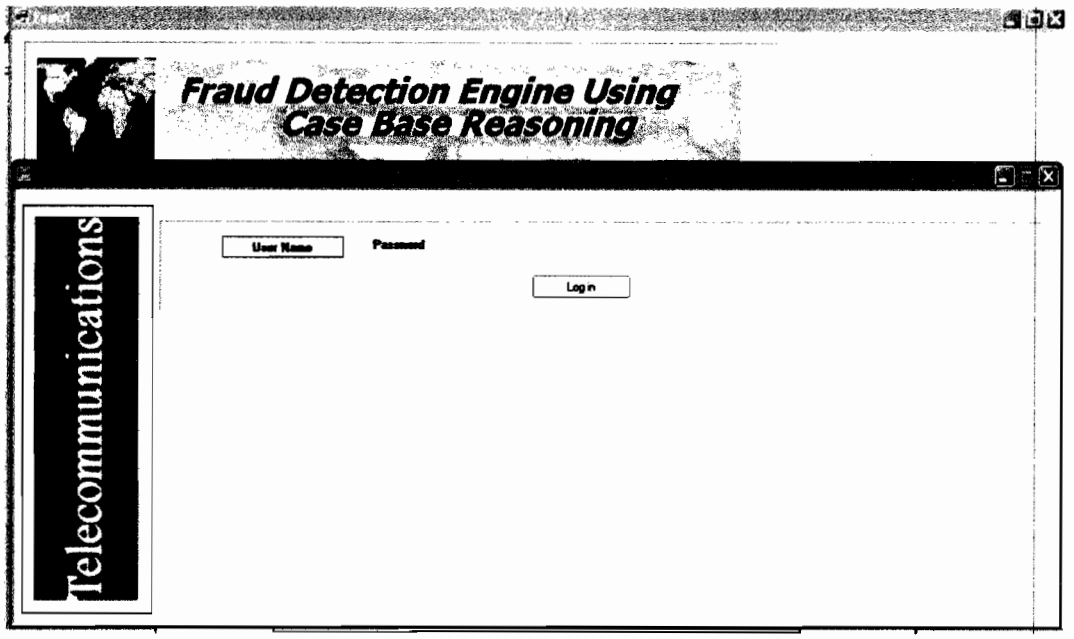


Figure 1.6: Login Page

3. When the user clicks on the enter button in the welcome page, the system navigates to this welcome page
4. The user is expected to insert a valid user name and password in the space provided from the page.
5. A correct user name and password will open the next page while wrong user name or password or both will result to an error message as shown in below

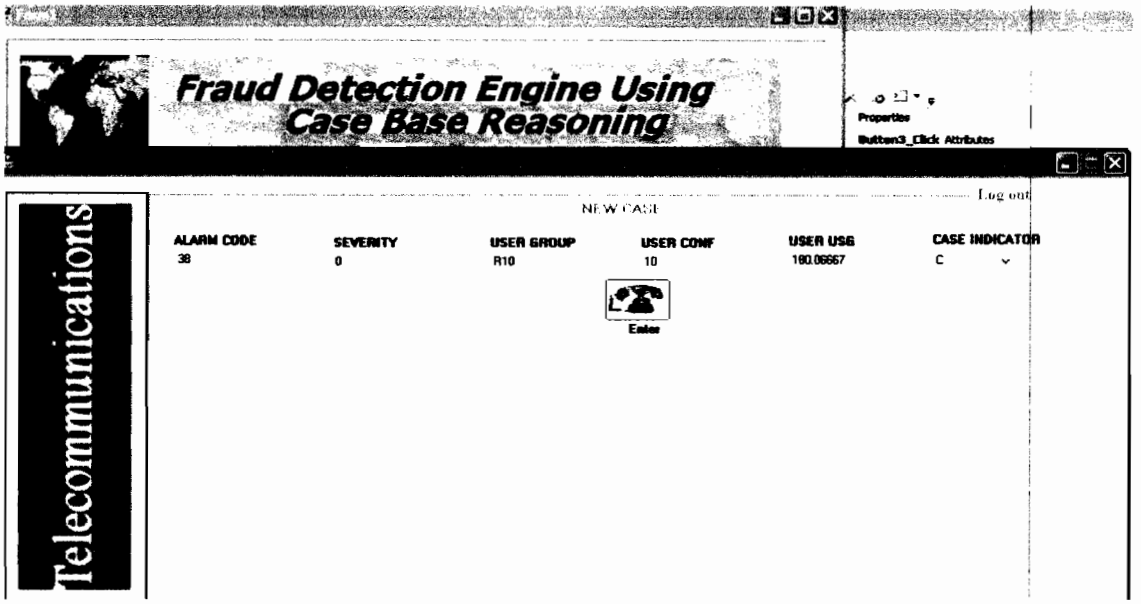


Figure 1.7: Main page

6. The main page provide for new fraud case to be inserted
7. the user is expected to click on the enter after inserting the new case so that the retrieval process will be effected.

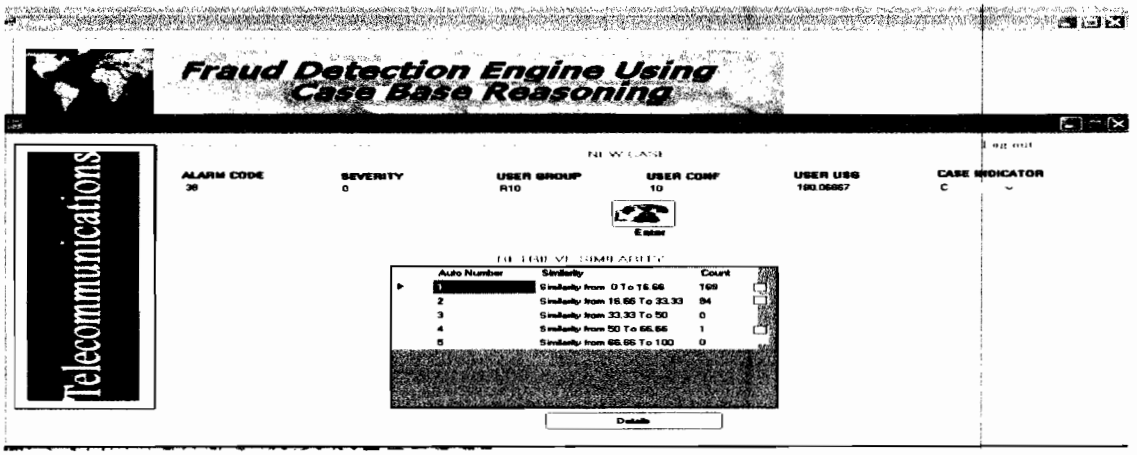


Figure 1.8: Similarity Retrieval Page

8. the case retrieval page will be displayed when the user in the previous page insert the new case and presses enter as shown above.

9. the similarly for the retrieved cases are also shown for the user to select which of the case similarity he wants to view the details.
10. checking on any of the check box for the retrieved similarity and pressing the detail button will cause the system to display the details of the particular item selected.
11. A user can logout also using the logout button.

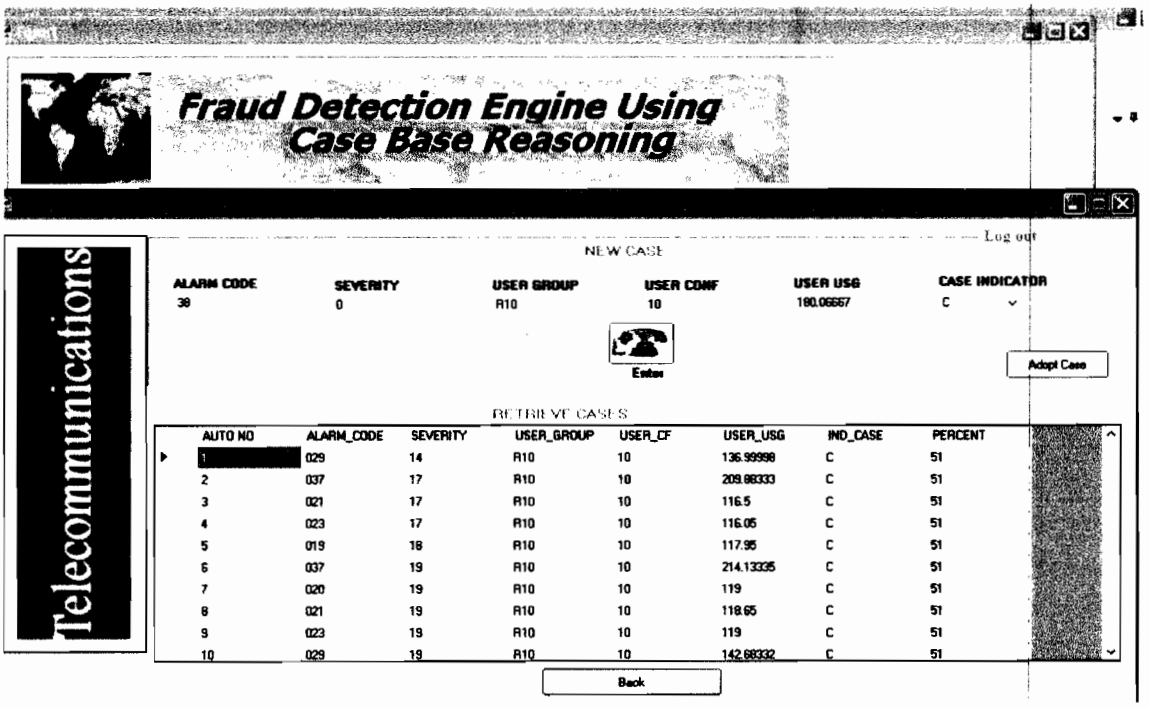


Figure 1.9: Case Detail Page

12. The retrieved detail is shown after the user clicks on detail in the previous page.
13. Also a user can return back to the previous page using the back button.
14. User can also logout using the logout at the top right corner of this page.
15. Finally a user can also decide to retain the new case in the case base using the Adopt case button.

APPENDIX D:

Test Cases 1

Attribute 1 not match with case base, weight for attribute1 = 1.0

Weights for rest attributes =1.

We obtain the result as:

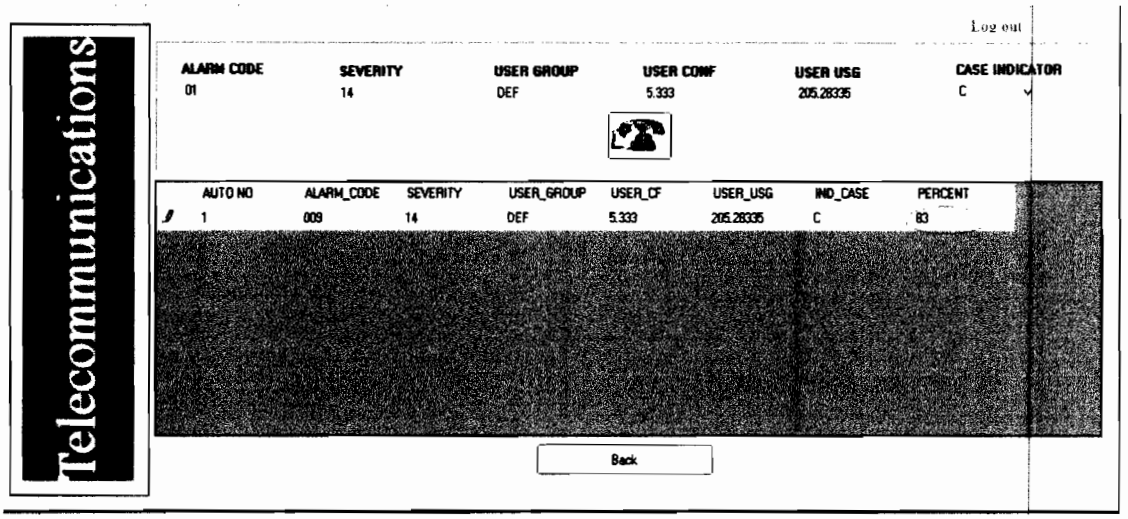


Figure 1.10: Test case 1

Attribute 1 not match with case base, weight for attribute1 = 0.9

Weights for rest attribute also =1.

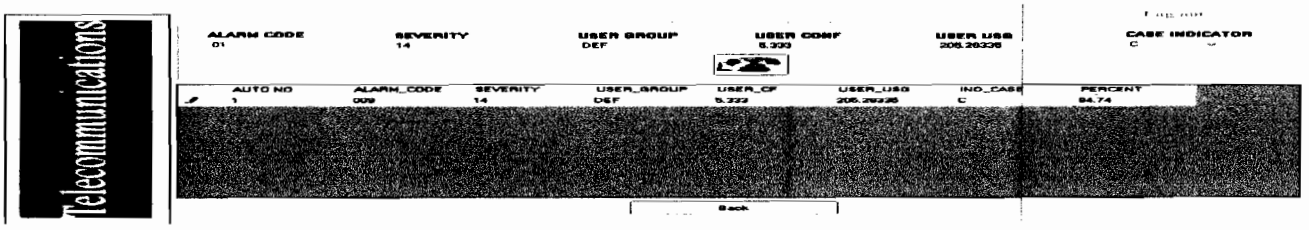


Figure 1.11: Test case 1

Attribute 2 not match with case base, weight for attribute1 = 0.9

Weights for rest attribute also =1.

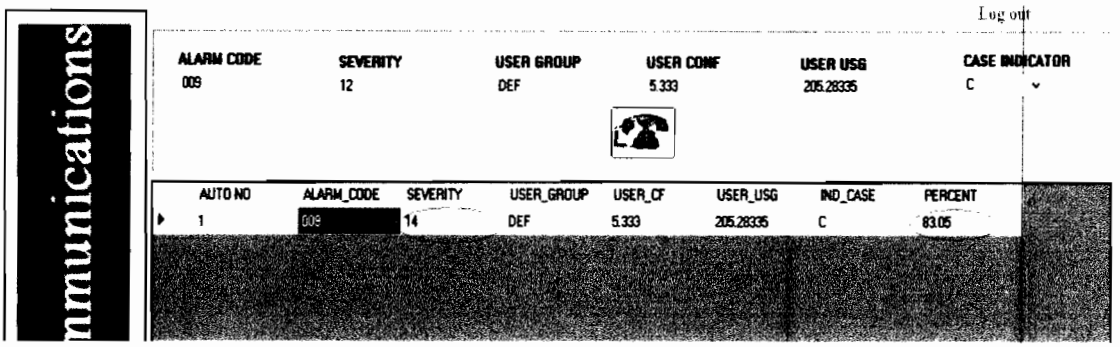


Figure 1.12: Test case 1

Attribute 3 not match

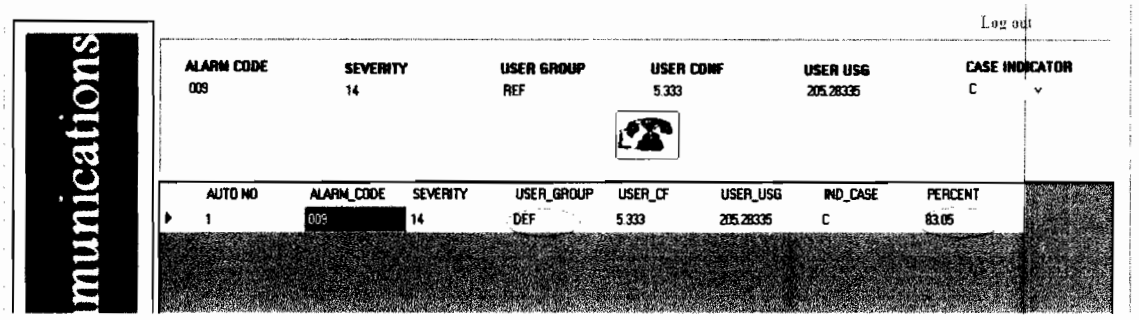


Figure 1.13: Test case 1

Attribute 4 not match

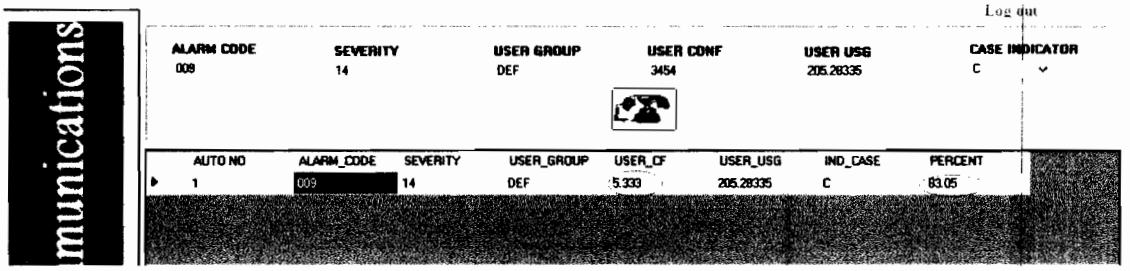


Figure 1.14: Test case 1

Attribute 5 not match:

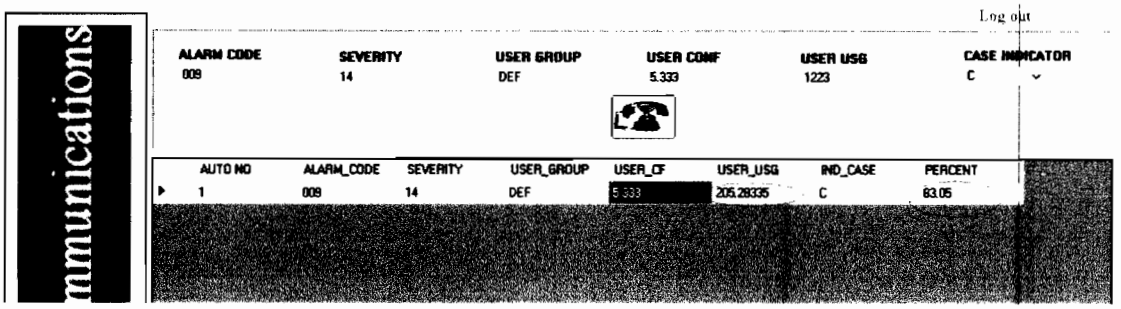


Figure 1.15: Test case 1

Attribute 6 not match:

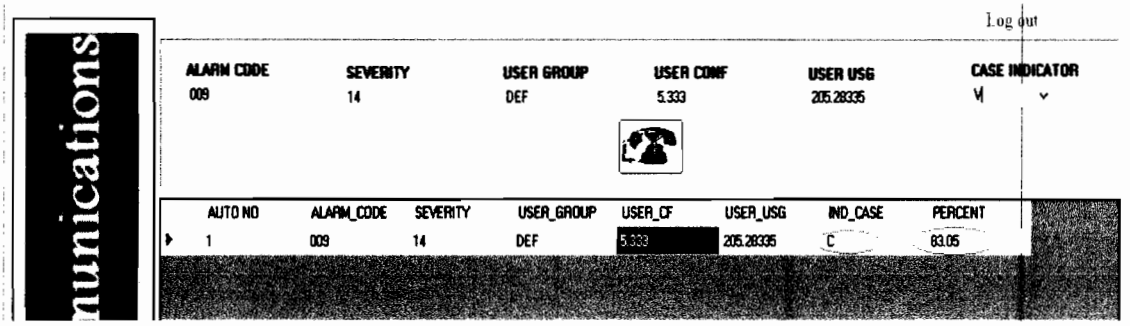


Figure 1.16: Test case 1

Therefore it can be observed that the percentage does not change with other attributes when the weight value remains the same. Hence further investigation is done on 2 attributes as follows:

Test Case 2:

Attribute 1 not match with case base, weight for attribute1 = **0.8**

Weights for the remaining attributes =1.

We obtain the result as

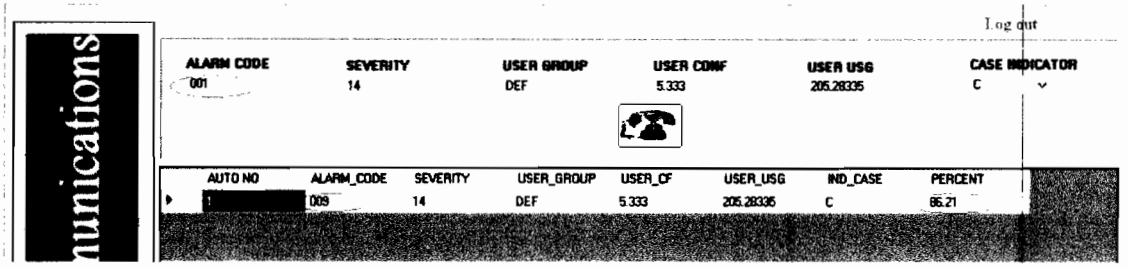


Figure 1.17: Test case 2

Attribute 2 not match

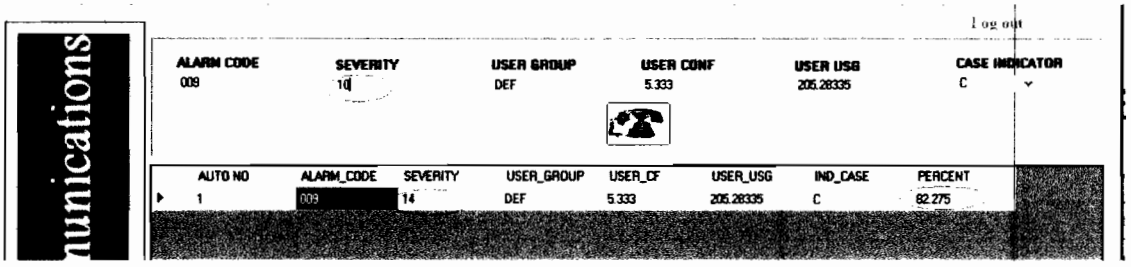


Figure 1.18: Test case 2

Test Case 3

Attribute 1 not match with case base, weight for attribute1 = **0.7**

Weights for the remaining attributes =1.

We obtain the result as

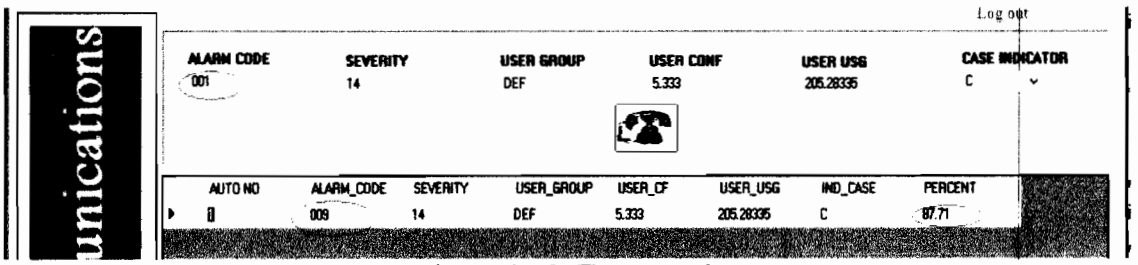


Figure 1.19: Test case 2

Attribute 2 not match

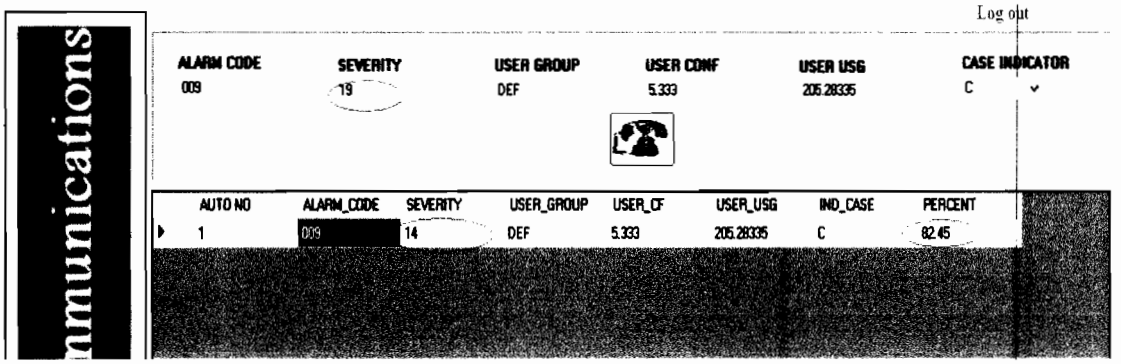


Figure 1.20: Test case 2

Test Case 3

Attribute 1 not match with case base, weight for attribute1 = **0.6**

Weights for the remaining attributes =1.

We obtain the result as

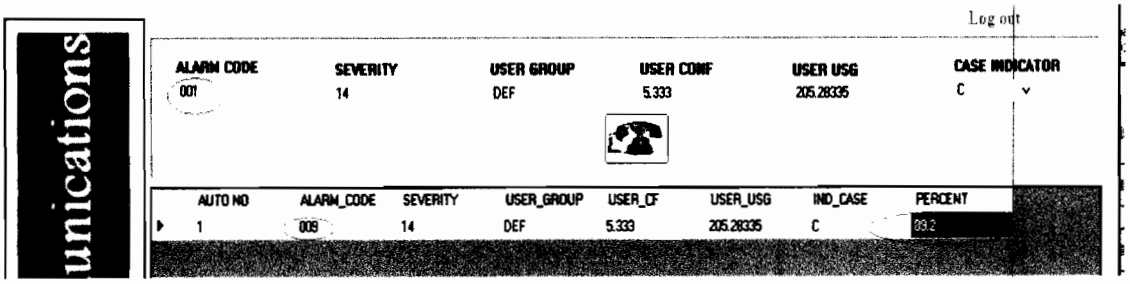


Figure 1.21: Test case 3

Attribute 2 not match

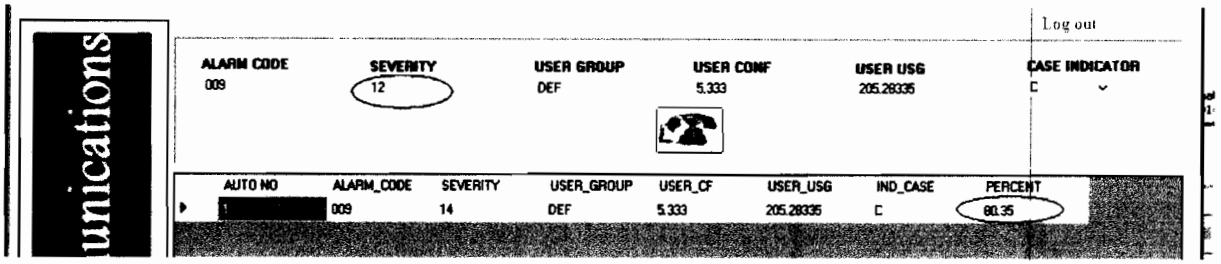


Figure 1.22: Test case 3

Test Case 4

Attribute 1 not match with case base, weight for attribute1 = **0.5**

Weights for the remaining attributes =1.

We obtain the result as

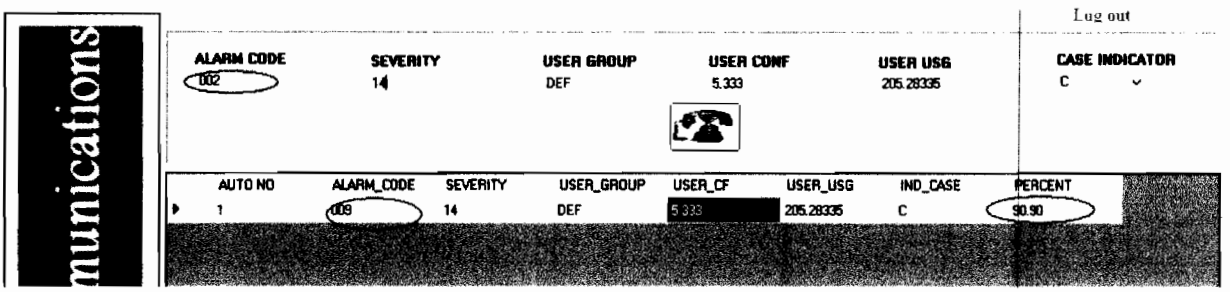


Figure 1.23: Test case 4

Attribute 2 not match

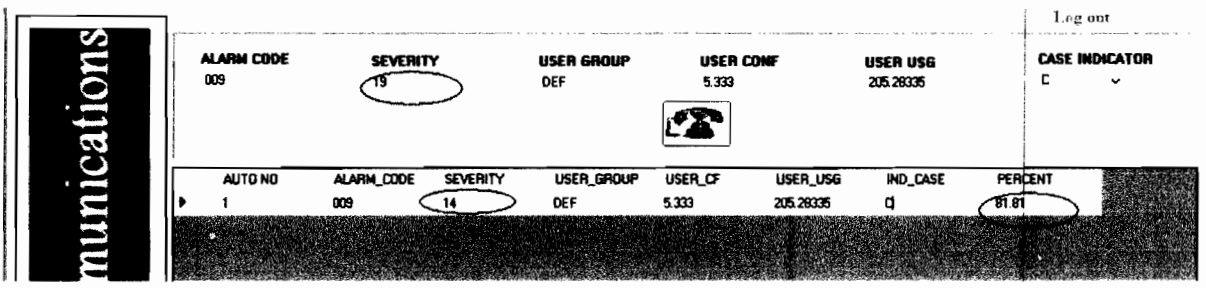


Figure 1.24: Test case 4

Test Case 5

Attribute 1 not match with case base, weight for attribute1 = **0.4**

Weights for the remaining attributes =1.

We obtain the result as

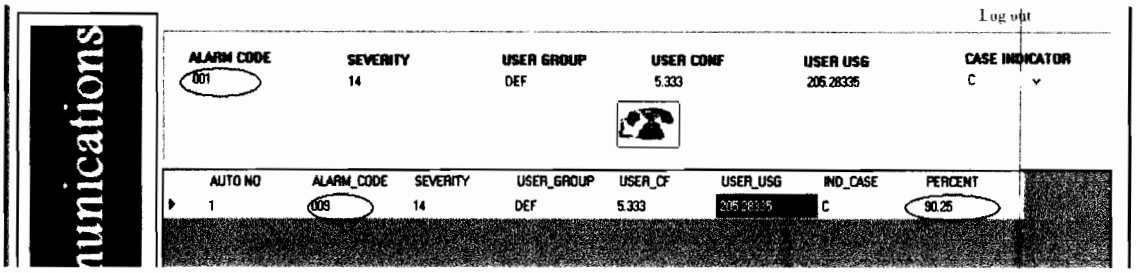


Figure 1.25: Test case 5

Attribute 2 not match

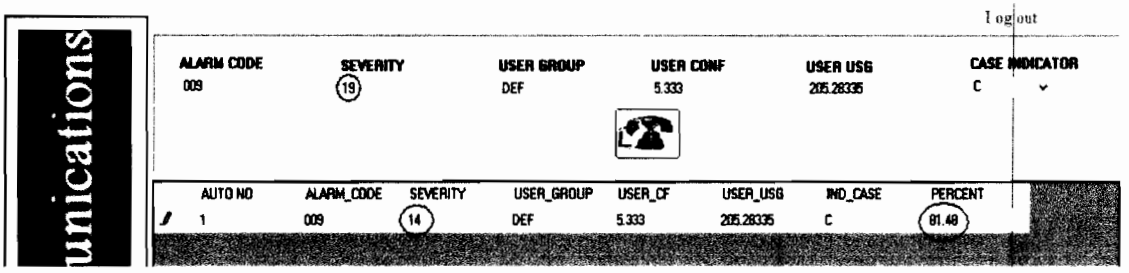


Figure 1.26: Test case 5

Test Case 6

Attribute 1 not match with case base, weight for attribute1 = **0.3**

Weights for the remaining attributes =1.

We obtain the result as

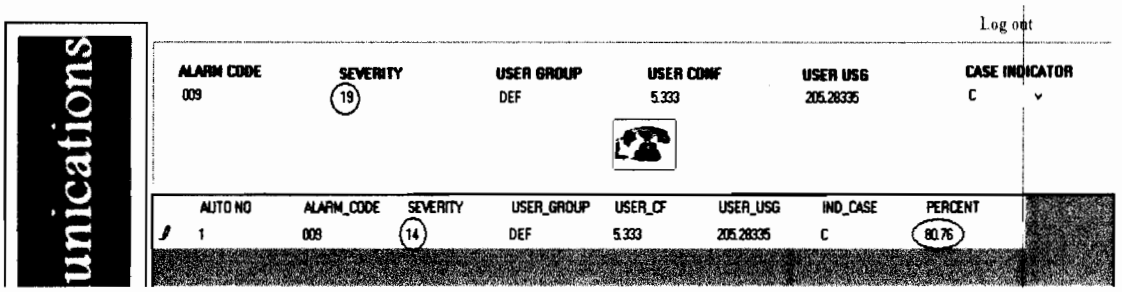


Figure 1.27: Test case 6

Attribute 1 not match

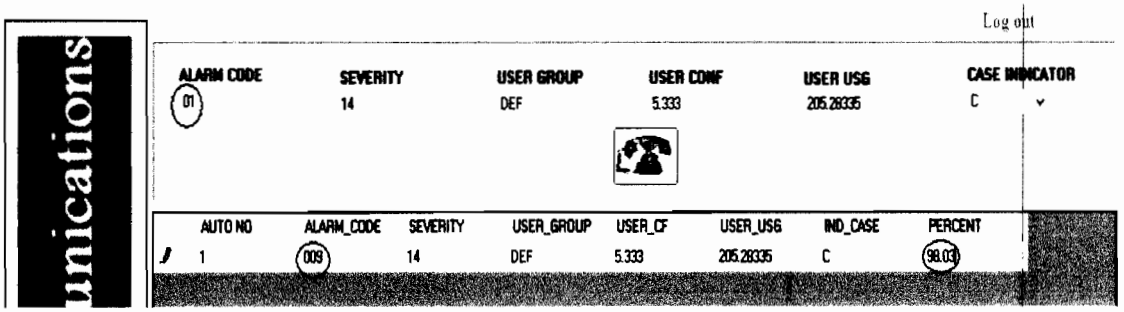


Figure 1.28: Test case 6

Test Case 7

Attribute 1 not match with case base, weight for attribute1 = **0.2**

Weights for the remaining attributes =1.

We obtain the result as

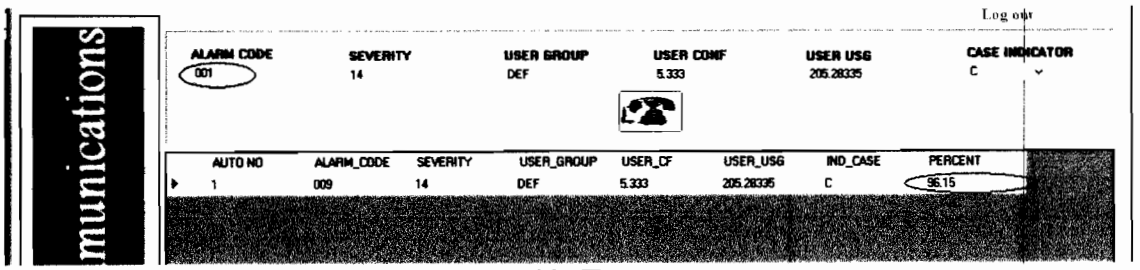


Figure 1.29: Test case 7

Attribute 2 not match

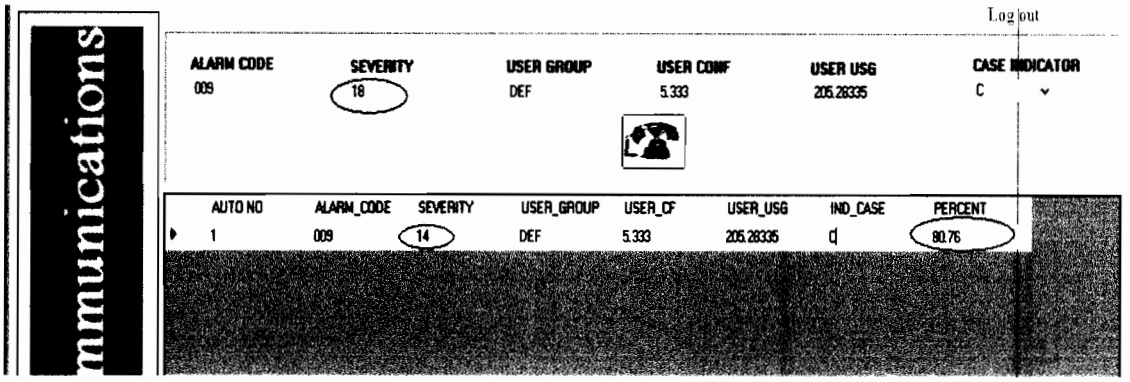


Figure 1.30: Test case 7

Test Case 8

Attribute 1 not match with case base, weight for attribute1 = **0.1**

Weights for the remaining attributes =1.

We obtain the result as

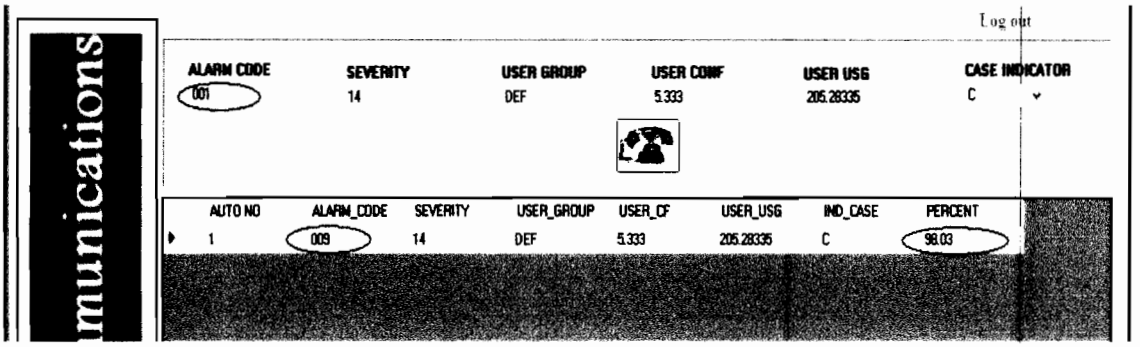


Figure 1.31: Test case 8

Attribute 2 not match

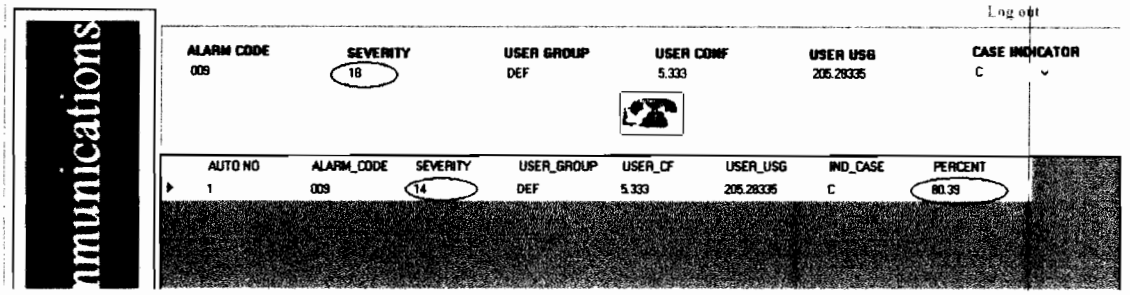


Figure 1.32: Test case 8

The highest percentage is recorded for case when the percentage similarity is 0.1. Therefore the weight for attribute 1 is taken as 0.1 and further investigation is carried out for case attribute 2 as follows:

Test Case 9

Attribute 1 not match with case base, weight for attribute1 = **0.1**

Weights for Attribute 2 = **0.9** and the remaining attributes =1.

We obtain the result as

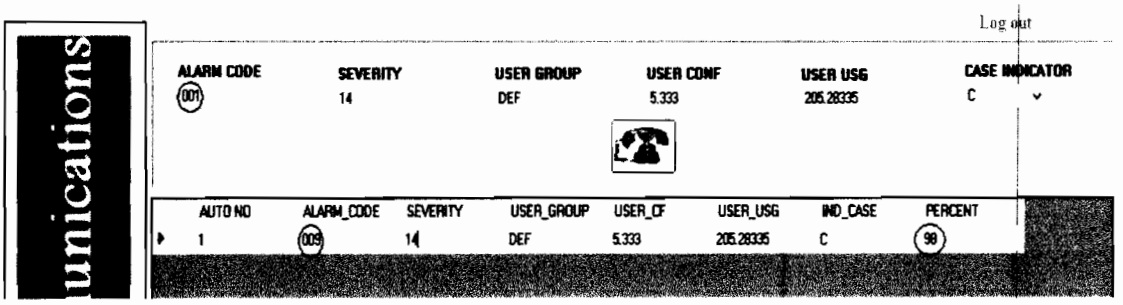


Figure 1.33: Test case 9

Attribute 2 not match

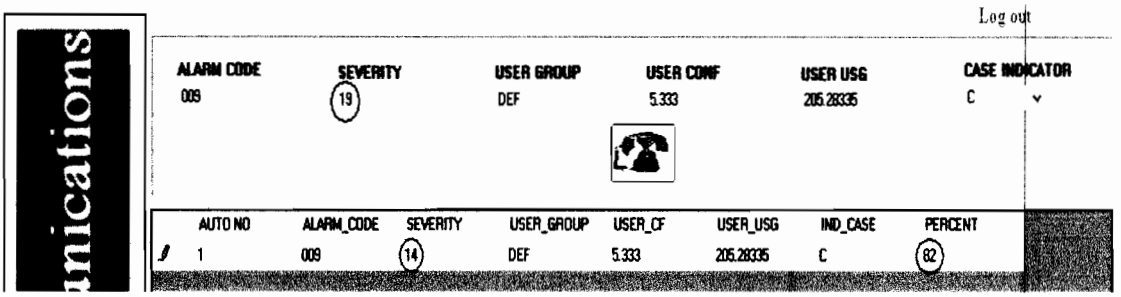


Figure 1.34: Test case 9

Test Case 10

Attribute 1 not match with case base, weight for attribute1 = **0.1**

Weights for Attribute 2 = **0.8** and the remaining attributes =1.

We obtain the result as

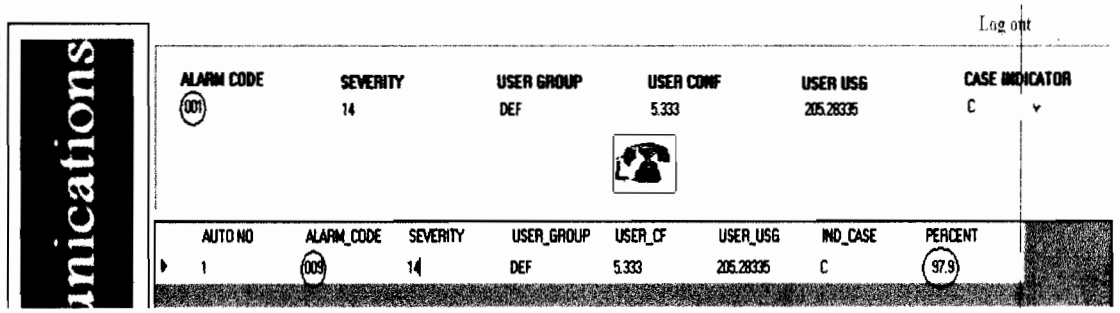


Figure 1.35: Test case 10

Attribute2 not match

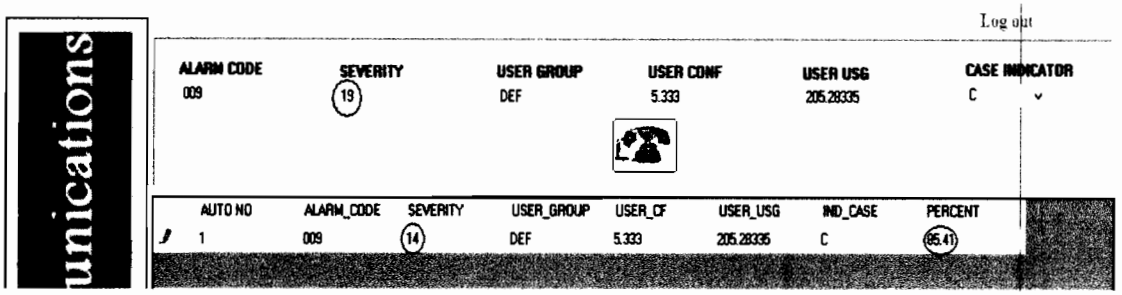


Figure 1.36: Test case 10

Test Case 11

Attribute 1 not match with case base, weight for attribute1 = **0.1**

Weights for Attribute 2 = **0.7** and the remaining attributes =1.

We obtain the result as

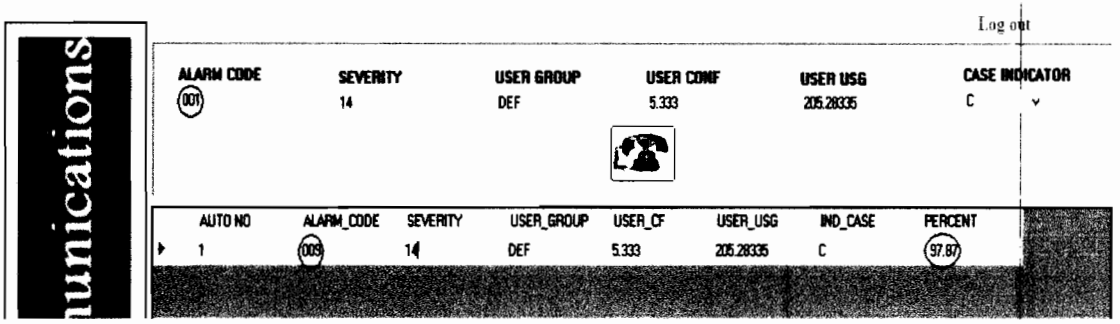


Figure 1.37: Test case 11

Attribute 2 not match

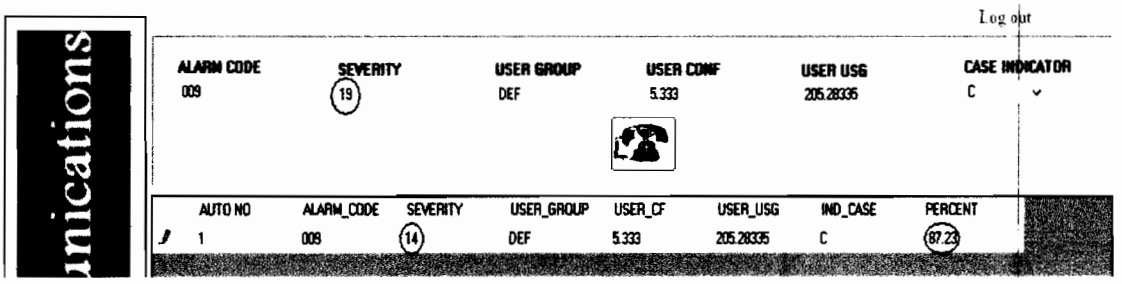


Figure 1.38: Test case 11

Test Case 12

Attribute 1 not match with case base, weight for attribute1 = **0.1**

Weights for Attribute 2 = **0.6** and the remaining attributes =1.

We obtain the result as

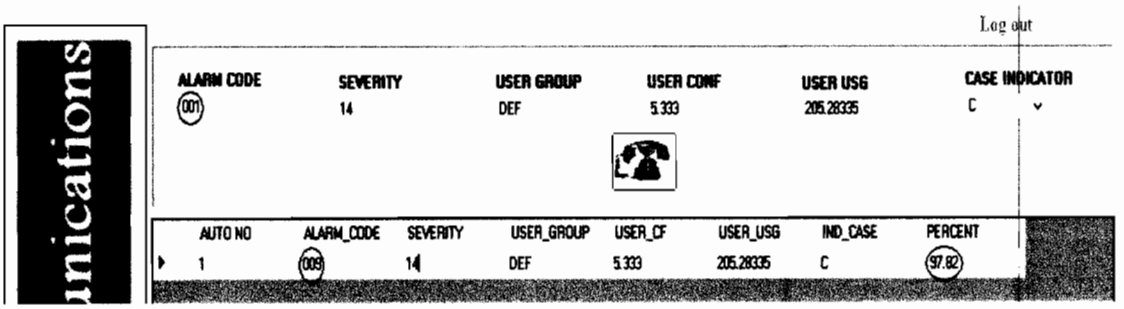


Figure 1.39: Test case 12

Attribute 1 not match

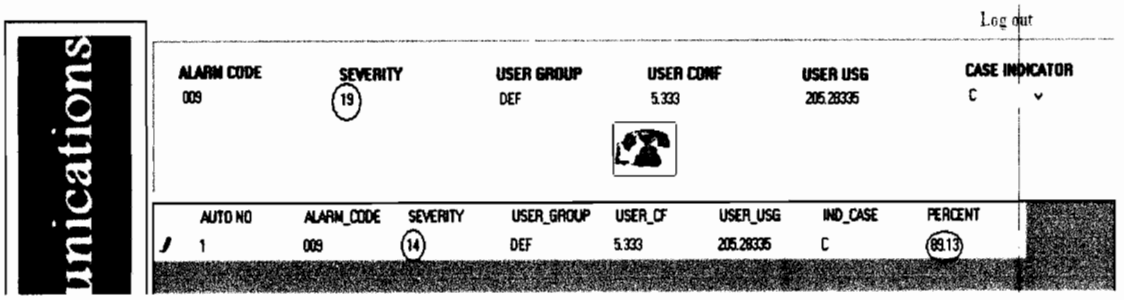


Figure 1.40: Test case 12

Test Case 13

Attribute 1 not match with case base, weight for attribute1 = **0.1**

Weights for Attribute 2 = **0.5** and the remaining attributes =1.

We obtain the result as

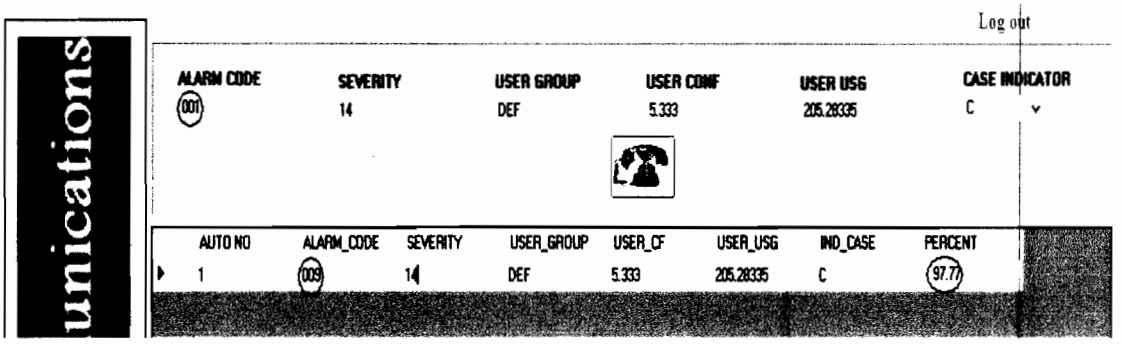


Figure 1.41: Test case 13

Attribute 1 not match

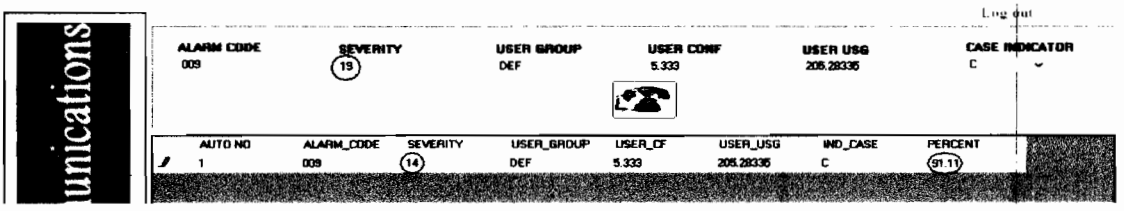


Figure 1.42: Test case 13

Test Case 14

Attribute 1 not match with case base, weight for attribute1 = **0.1**

Weights for Attribute 2 = **0.4** and the remaining attributes =1.

We obtain the result as

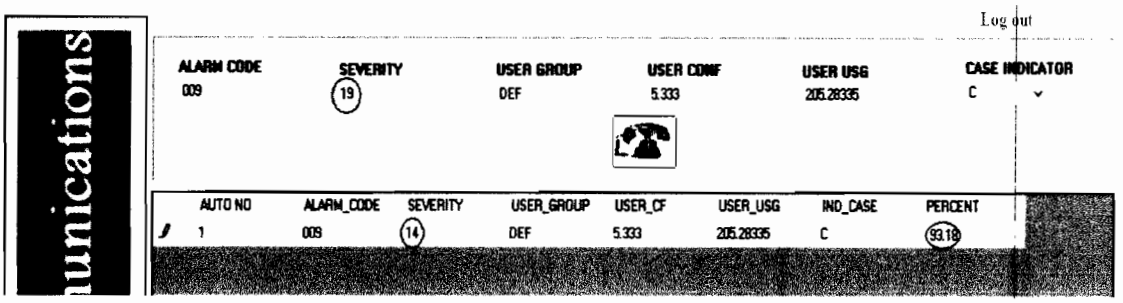


Figure 1.43: Test case 14

Test Case 14

Attribute 1 not match with case base, weight for attribute1 = **0.1**

Weights for Attribute 2 = **0.3** and the remaining attributes =1.

We obtain the result as

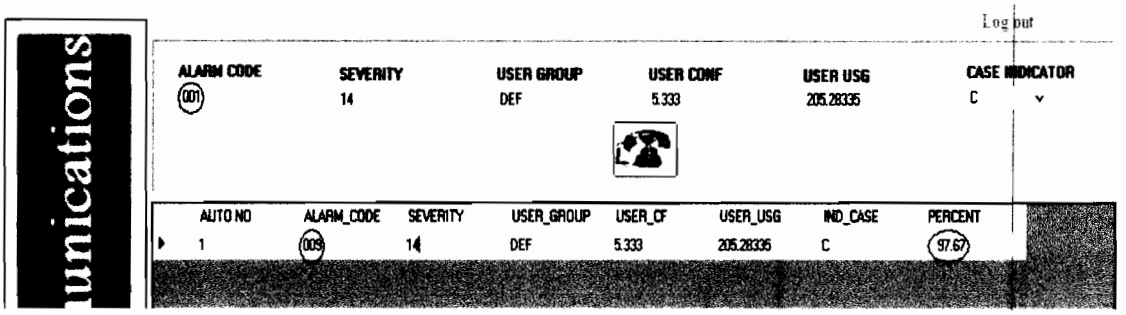


Figure 1.44: Test case 14

Attribute 1 not match

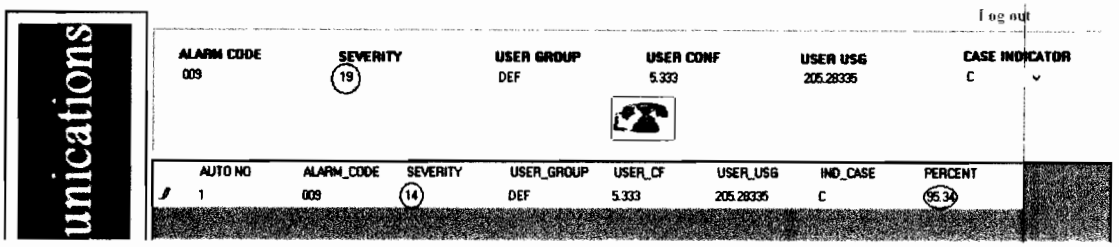


Figure 1.45: Test case 14

Test Case 15

Attribute 1 not match with case base, weight for attribute1 = **0.1**

Weights for Attribute 2 = **0.2** and the remaining attributes =1.

We obtain the result as

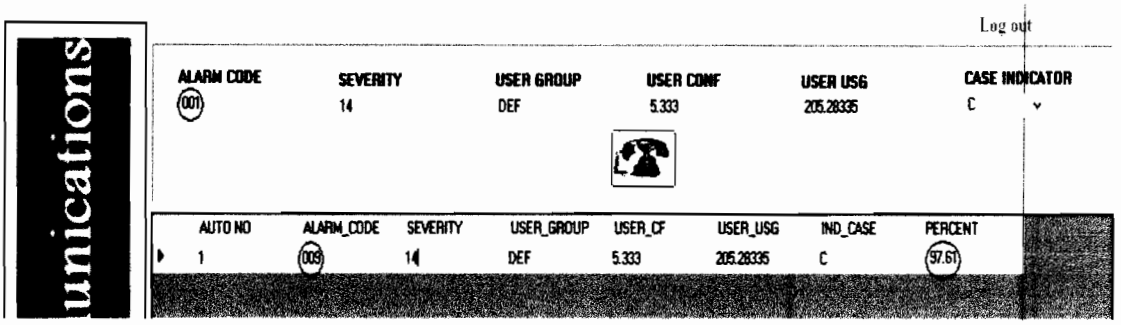


Figure 1.46: Test case 15

Attribute 1 not match

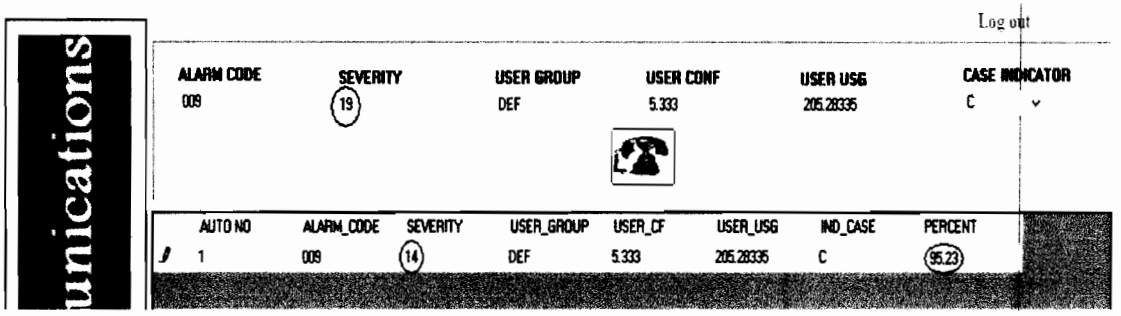


Figure 1.47: Test case 15