Telecommunications Subscription Fraud Detection Using Naïve Bayesian Network

Ledisi G. Kabari
School of Applied Sciences,
Ken Saro-Wiwa Polytechnic, Bori,
Nigeria.
ledisigiokkabari@yahoo.com,

Domaka N. Nanwin, & Edikan Uduak Nquoh
Faculty of Natural and Applied Sciences,
Ignatius Ajuru of Education, Port Harcourt,
Nigeria.
cacusman@yahoo.com, nquedt@gmail.com

Abstract
The radical changes in the terrain of the telecommunications sector have made it difficult to control and detect fraudulent activities. Thus, to achieve positive results the problem of fraud requires to be handled with rapt and effective attention. The paper identifies the different subscription services provided by the telecommunications sector, identifies the different ways telecommunications fraud is perpetrated and proposes the use of Naïve Bayesian Network technology to detect subscription fraud in the telecommunications sector. The approach offers quick training, fast data analysis, straight forward interpretation of results and yet easy to use.

Keywords: *Telecommunications; Subscription; Naïve Bayesian classifier; Fraud detection; Subscription Fraud.*

INTRODUCTION

Experts agree that most telecom providers are losing 3 to 10 percent of their income to fraud. With a problem like that, you need intelligent fraud detection that works with your network. Telecommunications fraud continues to be a big problem in the industry today. Advancements in technology have made life easier and more convenient for most people today, but not without a price. These advancements not only bring innovation for good, but they also bring about increasingly sophisticated practices in which fraudsters can infiltrate a company. Communication Service providers are faced with enough challenges from competition, declining Average Revenue Per User(ARPU), lower margins and other growth-related challenges. While paying more attention to these other areas, it can leave them vulnerable to unsuspecting attacks. With fraud continuing to be a big problem, fraud management has evolved from a defensive and reactive strategy focused on prevention to a more proactive, revenue generating and innovative approach. Goals have shifted from simply detecting fraud to achieving higher customer satisfaction and creating new revenue streams.

Telecommunication fraud is defined as the theft of telecommunication services or the use of telecommunication service to commit other forms of fraud. This type of fraud happens on a daily basis, sometimes without anyone knowing until the damage has already been done.

Fraud primarily occurs to a company with a weak defence system. Billing systems and network vulnerabilities are easily exploited to gain access, when if proper procedures were put in place, could have easily been prevented. With new voice technologies becoming more attractive, improperly installed systems can be infiltrated easily and put a small company out of business in mere minutes.

The telecommunications sector is a wide sector with millions of users. This sector is broadly categorized into two categories based on its users; Domestic users and commercial users. The domestic users are provided with connections at an affordable rate while the commercial users are provided with connections at a comparatively higher rate as their usage scale is higher. But it has been discovered that there are cases where the subscription at the commercial level is fraudulently brought under the domestic level hence causing a significant loss to the sector. This kind of subscriptions if brought under the right category would have yielded a greater income to the sector. Fraud is defined as the deliberate and premeditated act perpetrated to achieve gain on false ground [1]. Fraud is also seen as any transmission of voice data across a telecommunications network, where the intent of the sender is to avoid or reduce legitimate call charges [2]. Telecommunication fraud is the theft of services or deliberate abuse of voice or data networks [3]. Telecommunications fraud can be broken down into several generic classes which describe the mode operators are defrauded but for the purpose of this paper the study is focused on **"Subscription Fraud"**.

SUBSCRIPTION FRAUD

Subscription fraud is a contractual fraud. In these kinds of fraud revenue is generated through the normal use of a service without having to pay. In this scenario, the fraudster operates at level of phone numbers where all transactions from this number is fraudulent and all activities in such cases are further abnormal throughout the active period of the account.

Subscription fraud can be divided into two categories. These are: Subscription fraud for the purpose of personal usage by the fraudster and Subscription fraud for profit. In the second category, the fraudster opens a small outfit where he starts up a call center. The fraudster has no intensions of paying his bills but he sells the airtime to people who intend to make cheap long distance calls for cash.

Fraud detection problems are found in many sectors of lives endeavor and the telecoms sector is not an exception. Hence fraud detection is referred to as the attempt engaged in discovering illegitimate usage of a communication network by identifying fraud as quickly as possible once it has been perpetrated[4].

The radical changes in the terrain of the telecommunications sector have made it difficult to control and detect fraudulent activities. Thus, to achieve positive results the problem of fraud requires to be handled with rapt and effective attention. The aim of this paper is to present an easy to use yet efficient subscription fraud detection system using Naïve Bayesian Network. To achieve it aim the paper is set out to:-
Identify the different subscription services provided by the telecommunications sector.
Identify the different ways telecommunications fraud is perpetrated.
Utilize Naïve Bayesian Network technology to detect subscription fraud in the telecommunications sector.

In terms of significance, the paper will benefit the private telecoms sector because they will be able to detect subscription fraud and hence reduce the losses incurred. The paper brought to limelight a novel means of combating telecommunications subscription fraud with the utilization of another machine learning agent like Naïve Bayesian Network technology. The paper will also benefit other researchers who intend to go into this area as this is the birthing of innovative ways of solving the problem of subscription fraud in the telecoms sector.

The existing system is the conventional Rules-based fraud detection schemes which had their place in fraud prevention efforts over the years, but on their own they are an inadequate solution to detecting fraud because the growth of the telecoms sector will force a reevaluation of the acceptance of a fraud status quo.

The Rules-based fraud detection pulls from knowledge of what tactics have been used to commit fraud in the past, and creates "if-then" rules to try to prevent the same tactics from working again. As an example, if a company like Globacom Nigeria has noticed that it's especially common for fraudulent data subscriptions to come from IPs based in a certain city or area, they can create a rule to ensure transactions from this area are denied or at least flagged for review. This is not scalable because there are billions of IP addresses.

TELECOMMUNICATION FRAUD

Telecommunication industry has expanded dramatically in the last few years with the development of affordable mobile phone technology[5]. With the increasing number of mobile phone subscribers, global

mobile phone fraud is also set to rise. It is a worldwide problem with substantial annual revenue losses of many companies. Telecommunication fraud which is the focus is appealing particularly to fraudsters as calling from the mobile terminal is not bound to a physical location and it is easy to get a subscription. This provides a means for illegal high profit business for fraudsters requiring minimal investment and relatively low risk of getting caught. Telecommunication fraud is defined as the unauthorized use, tampering or manipulation of a mobile phone or service.

Telecommunication fraud can be simply described as any activity by which telecommunications service is obtained without intention of paying. This kind of fraud has certain characteristics that make it particularly attractive to fraudsters. The main one is that the danger of localization is small. This is because all actions are performed from a distance which in conjunction with the mess topology and the size of network makes the process of localization time consuming and expensive. Additionally no particularly sophisticated equipment is needed if one is needed at all. The simple knowledge of an access code, which can be acquired even with methods of social engineering, makes the implementation of fraud feasible. Finally, in the product of telecommunication fraud, a phone call is directly convertible to money.

Types of Telecommunication Fraud

The telecom industry suffers major losses due to fraud [6] [7]. There are many different types of telecommunications fraud and these can occur at various levels. The two most common types of fraud are subscription fraud and superimposed fraud. Others are intrusion fraud, fraud based on loopholes in technology, social engineering, fraud based on new technology, fraud based on new regulation and masquerading as another user.

In subscription fraud, fraudsters obtain an account without intention to pay the bill. This is thus at the level of a phone number, all transactions from this number will be fraudulent. In such cases abnormal usage occurs throughout the active period of the account. The account is usually used for call selling or intensive self-usage.

In Superimposed fraud, fraudsters take over a legitimate account. In such cases the abnormal usage is superimposed upon the normal usage of the legitimate customers. There are several ways to carry out superimposed fraud, including mobile phone cloning and obtaining calling card authorization details. Examples of such cases include cellular cloning, calling card theft and cellular handset theft. Superimposed fraud will generally occur at the level of individual calls; the fraudulent calls will be mixed in with the justified ones.

Intrusion fraud occurs when an existing, otherwise legitimate account, typically a business, is compromised in some way by an intruder, who subsequently makes or sells calls on this account. In contrast to subscription calls, the legitimate calls may be interspersed with fraudulent calls, calling for an anomaly detection algorithm.

In fraud based on loopholes in technology, consider voice mail systems as an example. Voice mail can be configured in such a way that calls can be made out of the voice mail system (e.g., to return a call after listening to a message), as a convenience for the user. However, if inadequate passwords are used to secure the mailboxes, it creates vulnerability. The fraudster looks for a way into a corporate voice mail system, compromises a mailbox (perhaps by guessing a weak password), and then uses the system to make outgoing calls. Legally, the owner of the voice mail system is liable for the fraudulent calls; after all, it is the owner that sets the security policy for the voice mail system.

In social engineering, instead of exploiting technological loopholes, social engineering exploits human interaction with the system. In this case the fraudster pretends to be someone he or she is not, such as the account holder, or a phone repair person, to access a customer's account. Recently, this technique has been used by "pre-texters" in some high-profile cases of accessing phone records to spy on fellow board members and reporters[8].
For fraud based on new technology, new technology, such as Voice over Internet Protocol (VoIP), enables international telephony at very low cost and allows users to carry their US-based phone number to other

countries. Fraudsters realized that they could purchase the service at a low price and then resell it illegally at a higher price to consumers who were unaware of the new service, unable to get it themselves, or technologically unsophisticated. Detecting this requires monitoring and correlating telephony usage, IP traffic and ordering systems.

For fraud based on new regulation, occasionally, regulations intended to promote fairness end up spawning new types of fraud. In 1996, the Federal Communications Commission (FCC) modified payphone compensation rules, requiring payphone operators to be compensated by the telecommunication providers. This allowed these operators to help cover the cost of providing access to phone lines, such as toll-free numbers, which do not generate revenue for the payphone operator. This spawned a new type of fraud—payphone owners or their associates placing spurious calls from payphones to toll-free numbers simply to bring in compensation income from the carriers.

In masquerading as another user, credit card numbers can be stolen by various means (e.g., "shoulder surfing" looking over someone's shoulder at a bank of payphones) and used to place calls masquerading as the cardholder. There are many more fraud techniques, some of which are quite sophisticated and combine more than one known method.

Telecommunications fraud is not static; new techniques evolve as the telecom companies put up defenses against existing ones. The fraudsters are smart opponents, continually looking for exploitable weaknesses in the telecom infrastructure. Part of their motivation is accounted for by the fact that once an exploit is defined, there are thousands (or millions) of potential targets. New types of fraud appear regularly, and these schemes evolve and adapt to attempts to stop them.

Telecommunications Mobile Subscription Services

Telecommunication companies in Nigeria provide different services to win the heart of their subscribers because of the market competition. However, some of these mobile subscription services are common to all the mobile operators in Nigeria, these includes prepaid services and postpaid services.

Prepaid services are the most popular of the services provided by mobile operators. As the name implies 'Pre-paid', all transaction in this service is pay-as-you-go. This service is easy for the mobile operator to maintain in the event of fraud and it is less susceptible to fraud as compared to postpaid services.

Postpaid is the most conventional service offered by mobile operators all over the world. As the name implies 'Post-paid', credit facilities is given for services used for some period of time, usually between 1- 6 months. Though, this service is not common in Nigeria because there is no proper means of identification in case a subscriber defaults, yet, all mobile operators in Nigeria still render the service as a result of the stiff competition in the industry.

Fraud is a multi-billions problem around the globe. The problem with telecommunication fraud is the huge loss of revenue and it can affect the credibility and performance of telecommunication companies. The most difficult problem that faces the industry is the fact that fraud is dynamic. This means that whenever fraudster's feel that they will be detected they find other ways to circumvent security measures. Telecommunication fraud also involves the theft of services and deliberate abuse of voice and data networks. In such cases the perpetrator's intention is to completely avoid or at least reduce the charges for using the services. Over the years, fraud has increased to the extent that losses to telephone companies are measured in terms of billions of American dollars. Fraud negatively impacts on the telephone company in four ways such as financially, marketing, customer relations and shareholder perceptions.

Other services provided by Nigeria's mobile telecommunication companies that are susceptible to fraud includes: Roaming Services, Value Added Features and Service (VAS), Premium Rate Services (PRS).

Approaches to face fraud

Dealing with the fraud is a very complex task mainly due to its transversal nature to the operator's structure [9]. Traditional fraud techniques are evolving and adapting to the new network infrastructure. The fraud techniques are considered because basic ideas remain despite the underlying technology. Deceptions in telecommunications include subscription frauds where the cheater accesses the services without being subscribed. User can also suffer line or identity theft being charged for services used by others. Telecommunication operators can oversee users that exceed their download quote and rate performing illegal service redistribution, sometimes for an economic profit. Finally cloning or unauthorized access to services may lead to compromising privacy.

Anyway the most common types of fraud on telecommunications are subscription fraud and identity theft. After that voice mail fraud and calling card fraud prevail. The analysis of different fraud techniques points out that the tendency is a convergence of the fraud which increases the complexity of its detection.

Fraud management systems have proved to be a suitable tool to detect fraud in different networks with diverse techniques such as self-organizing maps (SOM), general data mining, rule based systems profiling through Artificial intelligence techniques like neural networks or decision trees based on the hierarchical regime switching models, Bayesian networks, fuzzy rules or other data mining techniques. There also exist works on how to discover new rules to detect fraud in telecommunications and on the privacy concerns of applying detection techniques to user's data.

Fraud detection can also be done at 2 levels call or behavior and with two different approaches user profile or signature based[10]. Most of the techniques use the CDR data to create a user profile and to detect anomalies based on these profiles. The mined large amounts of CDR have in order to find patterns and scenarios of normal usage and typical fraud situations. These scenarios were then used to configure monitors that observe the user behavior with relation to that type of fraud. These monitors are then combined in a neural network which raises an alarm when sufficient support of fraud exists. This type of system can be classified in a rule based approach since it relies in the triggering of certain rules due to abnormal usage. The rule based system has the drawback of requiring expensive management of rules. Rules need to be precise (avoid false positive alarms) and constantly evolving (detect new scenarios) which result in very time consuming programming.

The most common and best succeeded methods for fraud analysis are signature based. These methods detect the fraud based on deviation detection by comparing the recent activity with the user behavior data which is expressed through the user signature. The work can be adapted and extended by reformulating the notion of signature and by introducing the notion of statistical based distances to detect anomalies. Furthermore the Computational cost can be reduced by using simple statistical functions avoiding processing costly histograms. A clear problem with a histogram approach is that discretization intervals or bucket must be clean and what is right for one customer may be wrong for another. Other approaches have also been widely applied to fraud analysis like neural networks. Another applied technique is link analysis. Here the client links are updates over time establishing a graph of called communities of interest that can easily reveal networks of fraudster's. These methods are based on the observation that fraudsters seldom change their calling habits but are often linked to other fraudsters.

Telecommunication Companies and Fraud

Telecommunications companies have been studying fraud and fraud detection for many years and have probably spent more time and money on this than the research community [11]. However, most of their efforts do not reach beyond the limits of the companies and have not been available to the public research community. Still, a number of published papers on the subject are available. The author in [11] describes toll fraud, how it occurs and offers ways to secure systems from hackers. He also raises questions about who should be responsible for prevention of toll-fraud - subscribers, long-distance carrier, operators or manufacturer of telecommunication equipment. Shaffer in [13] discusses various aspects of digital transmission of wireless communication. It describes the vulnerability of wireless communication to a type of wireless fraud known as tumbling. Shaffer noted that this fraud could easily allow fraudsters to steal telephone

services and digital technology. The work also refers to clone fraud and attempts to foil cloners. Shaffer also describes a solution that creates a profile of normal use for subscribers and then track calling patterns, in terms of frequency, destination, length, origination, parties called, time of day and distance. The authors also describe a solution that can detect the unique signal characteristic of each individual cellular phone and compare it with a database of prints, each of which is assigned to a unique electronic serial number. The system denies call access once the calling telephone does not have the electronic fingerprint it is supposed to. Shaffer finally noted that none of these solutions however is foolproof and that their adoption is slow.

In [14] the authors report their first experiments detecting fraud in a database of simulated calls. They use a supervised feed forward neural network to detect anomalous use. Six different user types are simulated stochastically according to the users' calling patterns. Two types of features are derived from this data, one set describing the recent use and the other set describing the longer-term behavior. Both are accumulated statistics of call data over time windows of different lengths. This data is used as input to the neural network. The performance of their classifier is estimated to be 92.5 % on the test data, which has limited value in the light of simulated data and the need to give class-specific estimates on accuracy.

METHODOLOGY

This paper proposes the use of Naïve Bayesian Network technology in detecting telecommunication subscription fraud. The system takes care of the challenges encountered by the rule based system of detecting fraud.

The inputs are the data of a subscriber such as the name, gender, age, subscription duration, megabyte used and registration duration. The data is processed so as to ascertain if the subscriber is fraudulent or non-fraudulent. It outputs the state of a particular subscriber. That is "fraudulent" or "non-fraudulent".

Design of the Proposed System

We took five basic steps during the design process: Collecting data, Preprocessing data, Building the network, Train, and Test performance of model.

Collecting and preparing sample data is the first step in designing models. The data is gotten from historical data of fraudulent and non-fraudulent subscriptions.

After data collection, three data preprocessing procedures are conducted to train the system more efficiently. These procedures are: (1) solve the problem of missing data, (2) normalize data and (3) randomize data. The missing data are replaced by the average of neighboring values during the same week. Normalization procedure before presenting the input data to the network is generally a good practice, since mixing variables with large magnitudes and small magnitudes will confuse the learning algorithm on the importance of each variable and may force it to finally reject the variable with the smaller magnitude.

Training the System

The training algorithm used is the **Naïve Bayesian network** and for the purpose of fraud detection, we construct two Bayesian networks to describe the behavior of auto insurance. **First,** a Bayesian network is constructed to model behavior under the assumption that the driver is **fraudulent (F)** and another model under the assumption the driver is a **legitimate user (NF).** The 'fraud net' is set up by using expert knowledge. The 'user net' is set up by using data from non-fraudulent drivers. During operation user net is adapted to a specific user based on emerging data. By inserting evidence in these networks (the observed user behavior x derived from his toll tickets) and propagating it through the network, we can get the probability of the measurement x under two above mentioned hypotheses. This means, we obtain judgments to what degree an observed user behavior meets typical fraudulent or non-fraudulent behavior.
These quantities are referred to as: $P(X|NF)$ and $P(X|F)$. By postulating the probability of fraud P(F) and P(NF) = 1- P(F) in general and by applying Bayesian' rule, we get the probability of fraud, given the measurement X,

$$P(F|X) = \frac{P(F)P(X|F)}{P(X)} \dots\dots\dots (1)$$

Where, the denominator P(x) can be calculated as:

$$P(X) = P(F)P(X/F) + P(NF)P(X/NF) \dots\dots\dots (2)$$

Appling chain rule of probabilities, Suppose there are two classes C_1, C_2 for fraud and legal respectively. Given an instance, $X = (X_1, X_2, \dots, X_n)$ and each row is represented by an attribute vector $A = (A_1, A_2, \dots, A_n)$, The classification is to derive the maximum $P(C_i|X)$ which can be derived from Bayesian theorem as given in the following steps:

P(fraud|X) = [P(fraud | X) P(fraud)] / P(X)

P (non-fraudulent|X) = [P (non-fraudulent|X) P (non-fraudulent)] / P(X)

As P(*X*) is constant for all classes, only [*P* (*fraud* | *X*) *P* (*fraud*)] and [*P* (*non-fraudulent* | *X*) *P*(*non-fraudulent*)] need to be maximized.

The class prior probabilities may be estimated by: P(fraud) = S_i / S

$$P(fraud) = \frac{S_i}{S} \dots\dots\dots (3)$$

Here, *S* is the total number of training examples and s_i is the number of training examples of class *fraud.*

A simplified assumption of no dependence relation between attributes is made.

Thus,

$$P(X|fraud) = \prod_{k=1}^{n} (X_k|fraud) \dots\dots (4)$$

and

$$\text{P(X|fraud)} = \prod_{k=1}^{n} P(X_k|legal) \dots\dots (5)$$

The probabilities $P(X_1|fraud)$, $P(X_2|fraud)$ can be estimated from the training samples:

$P(X_k|\text{fraud}) = S_{ik} / S_i$

Here, S_i is the number of training examples for class *fraud* and S_{ik} is the number of training examples of class with value X_k for A_k.

RESULTS AND DISCUSSION

Bayesian learning algorithm is presented to predict occurrence of fraud. Using the "Output" classification results for Table1, there are 17 tuples classified as non-fraudulent, and 3 as fraudulent. To facilitate classification, the age attributes are divided into ranges as shown in Table2.

Table2 shows the counts and subsequent probabilities associated with the attributes. With these simulated training data, we estimate the prior probabilities.

The classifier has to predict the class of instance to be fraudulent or non-fraudulent.

P (fraud) = S_i / S = 3/20 = 0.15

P (legal) = S_i / S =17/20 = 0.85

Table1: Sample Training Set

S/N	Name	Age	Registration Duration	Subscription Bundle	Used	Gender	Output
1	David Okere	42	14.4	46.5	19.6	Male	Non-fraudulent
2	Benson Jackson	19	11.1	32.3	11.3	Male	Fraudulent
3	Adaobi Daniel	69	14.3	37.0	14.7	Female	Non-fraudulent
4	Rachael Okon	56	8.9	23.9	8.8	Female	Non-fraudulent
5	Katherine James	64	14.1	34.5	13.8	Female	Non-fraudulent
6	Amina Bala	53	8.1	21.6	7.7	Female	Non-fraudulent
7	Precious	13	10.5	31.6	13.1	Male	Non-fraudulent

	Okafor						
8	Chima Eze	17	9.9	27.1	9.8	Male	Fraudulent
9	Funmi Ajayi	73	12.2	31.9	11.5	Female	Non-fraudulent
10	Daniella Johnson	17	16.6	43.9	17.9	Female	Non-fraudulent
11	Mildred Bello	27	14.7	37.6	14.6	Female	Non-fraudulent
12	Joy Okpara	39	17.2	44.1	18.6	Female	Fraudulent
13	Boma West	28	17.9	46.4	17.9	Female	Non-fraudulent
14	Junior Adekunle	50	15.7	54.6	20	Male	Non-fraudulent
15	Hassan Audu	34	13	41.8	15.2	Male	Non-fraudulent
16	Susan Kuro	36	16.7	43.6	18.2	Female	Non-fraudulent
17	Favour Azubike	22	12.2	31.1	11.1	Female	Non-fraudulent
18	Obi Obi	24	11.5	32.2	13.1	Male	Non-fraudulent
19	Joshua Wilson	27	12.8	39.5	16.2	Male	Non-fraudulent
20	Mark Pepple	24	8.2	22.2	9.0	Male	Non-fraudulent
21	Aaron Azubike	37	12.6	40.3	14.5	Male	?

Table2: Probabilities Associated with Attributes

Attribute	Value	Count		Probabilities	
		Non-fraudulent	fraudulent	Non-fraudulent	Fraudulent
Gender	M	7	2	7/17	2/3
	F	10	1	10/17	1/3
Age	(13, 20)	2	2	2/18	2/3
	(21, 30)	7	0	7/18	0
	(31, 35)	1	0	1/18	0
	(36, 40)	1	1	1/18	1/3
	(41, 45)	1	0	1/18	0
	(46, 50)	1	0	1/18	0
	(51,75)	5	0	5/18	0

Probability calculation

This phase involves probability calculation. This phase uses Naïve Bayesian classification as a method for calculating probability. The reason why Naïve- Bayesian classification alone is not used for entire classification is that it requires building a trainer that will work on already classified data. To train the classifier to achieve the desired level of accuracy, huge number of records is needed. Since the occurrence of a fraudulent customer in this sector is sparse compared to other sector, building a classifier that requires utmost training may be arduous. To overcome this problem, the Naïve-Bayesian classification is combined with the idea of finding divergence.

By using the above values to classify a new tuple based on probabilities of gender and subscriber age, Let X = (Aaron Azubike, M, 37). Hence the following estimates are obtained:
P(X |Non Fraudulent) = 10/17 * 1/18 = 0.033
P(X |fraud) = 1/3 * 1/3 = 0.111
Thus, likelihood of being Non-fraudulent
= 0.033 *0.85=0.021
Likelihood of being fraudulent = 0.111 *0.15= 0.017
Let P(*X*) be the sum of these individual likelihood values since *X* will be either non-fraudulent or fraudulent:
P(X) = 0.021 + 0.017 = 0.038 = 0.04. Finally, to obtain the actual probabilities of each event the following calculations are performed:
P (non-fraudulent | X) = (0.03*0.85)/ 0.04 = 0.600
P (fraud |X) = (0.111 *0.15) /0.04 = 0.400
Therefore, based on these probabilities, the classification of the new tuple is non-fraudulent because it has the higher probability. Since attributes are treated as independent, the addition of redundant ones reduces its predictive power. To relax this conditional independence is to add derived attributes which are created from combinations of existing attributes. Missing data cause problems during classification process. Naïve Bayesian classifier can handle missing values in training datasets. To demonstrate this, seven missing values appear in dataset.

The naïve Bayesian approach is easy to use and only one scan of the training data is required. The approach can handle missing values by simply omitting that probability when calculating the likelihoods of membership in each class. The approach offers quick training and fast data analysis and consequently decision making. The approach is equally gives straight forward interpretation of result. The attributes usually are not independent. We could use subset of the attributes by ignoring any that are dependent on others. The technique does not handle continuous data. Dividing the continuous values into ranges could be used to solve this problem, but the division of the continuous values is a tedious task, and how this is done can impact the results.

CONCLUSION

Fraud is a multi-billions problem around the globe. The problem with telecommunication fraud is the huge loss of revenue and it can affect the credibility and performance of telecommunication companies. The most difficult problem that faces the industry is the fact that fraud is dynamic. This means that whenever fraudster's feel that they will be detected they find other ways to circumvent security measures. Telecommunication fraud also involves the theft of services and deliberate abuse of voice and data networks. In such cases the perpetrator's intention is to completely avoid or at least reduce the charges for using the services. Over the years, fraud has increased to the extent that losses to telephone companies are measured in terms of billions of American dollars.

Fraud detection problems are found in many sectors of lives endeavor and the telecoms sector is not an exception. Hence fraud detection is referred to as the attempt engaged in discovering illegitimate usage of a communication network by identifying fraud as quickly as possible once it has been perpetrated.

Telecommunications fraud is not static; new techniques evolve as the telecom companies put up defenses against existing ones. The fraudsters are smart opponents, continually looking for exploitable weaknesses in the telecom infrastructure. Part of their motivation is accounted for by the fact that once an exploit is defined, there are thousands (or millions) of potential targets. New types of fraud appear regularly, and these schemes evolve and adapt to attempts to stop them.

The paper thus identified different subscription services provided by the telecommunications sector, identify the different ways telecommunications fraud is perpetrated and utilized Naïve Bayesian Network model to train and implement a subscription fraud detection system for the telecommunications sector.
Naïve Bayesian classifier can handle missing values in training dataset. The naïve Bayesian approach is easy to use and only one scan of the training data is required.

REFERENCES

Alexopoulos, P. and Kafentzis, K. (2007): Towards a Generic Fraud Ontology in E Government, *ICE B*, pp. 269-276

Hollmen, J. (2000): *User Profiling and Classification for Fraud Detection in Mobile Communication Networks:* PhD thesis, Helsinki University of Technology, Department of Cognitive and Computer Science and Engineering. Espoo, Finland.

Hiyam, A. E. Tawashi (2010): *Detecting Fraud in Cellular Telephone Networks Jawwal Case Study*: MBA thesis Islamic University, Faculty of Commerce, Department of Business Administration, Gaza.

Bolton, R. J. and Hand, D. J. (2002): Statistical Fraud Detection. A Review, *Institute of Mathematical Statistics*, 17(3), 235–255.

Pieprzyk, J., Ghodosi, H. and Dawson, E. (2007): Information Security and Privacy: *12th Australasian Conference*, ACISP 2007, Townsville, Australia, July 2-4, 2007: Proceedings, Springer, Germany, pp 446-447.

Prasad, S. K., Routray, S. and Khurana, R. (2009): Information Systems, Technology and Management: *Third International Conference, ICISTM 2009, Ghaziabad, India, March 12 13, 2009*, Proceedings, Springer, Germany, pp 259-260.

Żytkow, J. M. and Rauch, J. (1999): Principles of Data Mining and Knowledge Discovery: *Third European Conference, PKDD'99, Prague, Czech Republic, September 15-18, 1999: Proceedings*, Springer, USA, pp 251.

Kaplan, D. A. (2006): Intrigue in High Places: To Catch a Leaker, Hewlett– Packard's Chairwoman Spied on the Home–Phone Records of Its Board of Directors, *Newsweek* (September).

Samarati, P. (2010): Information Security Theory and Practices: Security and Privacy of Pervasive Systems and Smart Devices: *4th IFIP WG 11.2 International Workshop, WISTP 2010*, Passau, Germany, April 12-14, 2010, Proceedings, Springer, USA, pp 201.

Perner, P. (2006) Advances in Data Mining: Applications in Medicine, Web Mining, Marketing, Image and Signal Mining: *6th Industrial Conference on Data Mining, ICDM 2006, Leipzig, Germany, July 14-15, 2006:* Proceedings, Springer, Germany, pp 535.

Kvarnstrom, H., Lundin, E. and Jonsson, E. (2000): Combining Fraud and Intrusion Detection Meeting New Requirements – *In Proceedings of the 5th Nordic Workshop on Secure IT systems (NordSec2000)*, Reykjavik, Iceland.

Michalecki, R. (1994): Toll Fraud: Multimillion-dollar Telecomm Problem. *Communication News*, 31(2), 34.

Shaffer, R. A. (1994): Good Guys, Bad Guys and Digital. Forbes, 154(4):122.

Barson, P., Field, S., Davey, N., McAskie, G. and Frank, R. (1996): The Detection of Fraud in Mobile Phone Networks. *Neural Network World*, 6(4), 477–484.