Credit Card Fraud Detection Using Hidden Markov Model (HMM) and Artificial Neural Network (ANN)

Amanze, B. C. & Nwoke, B. C.  
Department of Computer Science  
Imo State University, Owerri

Abstract

The growth of electronic commerce ('e-commerce') in the world has been dramatic over the past few years, with forecasts suggesting that this explosive trend will continue. However, this growth has given rise to electronic payments for goods which presently is posed with challenges of fraudulent activities worldwide. E-commerce customers most times run into hackers where their credit card details are hacked and funds in it used. Another problem associated with e-commerce systems is their inability to automatically deactivate any victim's account and send real-time alert to the victim's financial company. The aim of this paper is to develop an online fraud detection, monitoring and real-time alerting system to reduce fraudulent activities associated with e-commerce using the joint approach of Hidden Markov Model (HMM) and Artificial Neural Network (ANN). The system developed is an online fraud detection and monitoring system which uses the Hidden Markov Model (HMM) to generate One Time Password (OTP) and email Code. These codes (input layers) are integrated using Artificial Neural Network (ANN) to determine okay and fraudulent activities. This system will help ecommerce industries to detect and monitor online fraud.

Keywords: Electronic commerce, Automatically deactivate, Customers, ANN, HMM.

Corresponding Author: Amanze, B. C.
Background to the Study

Globally, information technology has become a key element in economic development of many countries in the world. Over the years, many innovations have taken place in the world, the most striking and most celebrated is the aspect of information technology. Many sectors have adopted Information Technology to better and facilitate operations. Some have gone electronically to reach to consumers all over the world with the use of Information Technology innovation. Some of these sectors are banking, education, communication, commerce (e-commerce), media, etc.

The growth of electronic commerce ('e-commerce') in the world has been dramatic over the past few years, with forecasts suggesting that this explosive trend will continue (Pastore, 2012). The birth of the dot-coms who have capitalized on the advantages of e-commerce provides, such as lower barriers to market entry, as well as the extensive integration of e-commerce systems into incumbent organizations is testament to this. This growth has arisen as the benefits of e-commerce have been realized by businesses and consumers alike. In the 'digital economy', e-commerce has facilitated efficient information exchange, enabled cost reductions, provided new revenue opportunities and increased process efficiencies. Customers also reap these benefits through the reception of better customer service and the greater convenience of paperless online transactions (Turban, 2012). E-commerce is ever more becoming intertwined with how organizations do business and are operated.

However, this tight integration with e-commerce has increased the exposure of businesses to a broader range of risks such as security, privacy and reliability concerns (Daigle & Lampe 2000). Actual and perceived security concerns, in particular, are large barriers preventing a more rapid uptake and growth of ecommerce (Elliot & Fowell, 2010). Managing these risks becomes of great importance to companies partaking in any e-commerce operation – both in protecting company e-commerce revenue flows from security-related mishaps (such as fraud, theft and systems failure), and in assuring hesitant customers of the safety of engaging in e-commerce.

One of the chief security risks is fraud (Cerpa & Jamieson, 2011). In any transaction, participants want to ensure proper receipt of payment in exchange for goods or services. Failure by one party to receive what they expect may indicate the occurrence of fraud. As a consequence of the advent of ecommerce, new methods of carrying out financial transactions mean that new methods by which fraud is perpetrated also arise (such as shill bidding in online auctions (Wang et al., 2011). The presence of fraud or even the threat of it is a deterrent to businesses and customers alike, which may choose to resort to more traditional means of performing transactions (Elliot & Fowell, 2010).

Currently, a variety of tools using a myriad of approaches to detect fraud do exist, but their use is limited, fragmentes and their effectiveness is untested. The fraud detection solutions that do exist for businesses engaging in e-commerce tend to be proprietary in nature and how they work is unpublicized. The electronic environment therefore has need for effective controls, built around a generalized, tested framework that will mitigate the risk of fraud that e-commerce poses. It is clear that businesses stand to benefit from the ability to reduce fraud, but the development of these controls is also important for auditors. An essential responsibility of auditors is to plan and conduct audits for irregularities induced by fraud, other illegal acts and errors that impact upon the financial reports of an entity (AICPA, 2012). Therefore, enabling better detection of fraud would facilitate an auditor’s job.
Fraud prevention is difficult in the faceless world of the Internet, and any measure designed to respond to it must be able to do so in a timely manner. Continuous assurance (CICA, 2009) offers a timely method of assurance where, by monitoring transactions (flows of information, especially payment and order details) in real-time, irregularities that point to illicit behavior may be promptly detected and dealt with. Continuous assurance systems capitalize on the infrastructure and real-time nature of e-commerce systems. In fact, continuous assurance systems rely on the system being assured to be a quick and reliable source of relevant data, because the assurance system must, in turn, provide its own service of delivering timely assurance and reporting information. Such a system will be able to detect fraudulent activity in an e-commerce system in an unobtrusive manner. There is a need to develop an assurance system that can be easily integrated into existing systems, be flexible to adapt to different organizations and organizational change, and provide control over the assurance process.

This paper focuses primarily on improving our understanding of detecting and preventing fraud in e-commerce systems by some other fraud detection methods. It went ahead to provide a real-time alerting system as soon as potential fraud is detected or noticed. Meanwhile, a conceptual model relating the aspects and concepts associated with the real-time monitoring of e-commerce transactions for fraud will be developed.

**Objective of the Study**
The general objective of the study is to design and develop a fraud detection, monitoring and real-time alerting system to reduce fraudulent activities associated with e-commerce.

**The Specific Objectives are to:**
(1) Analyze some existing fraud detection and monitoring systems in order to have wider understanding fraud detection in e-commerce transactions.
(2) Generate one time password and email code (input layers) using the Hidden Markov Model and integrate these input layers using Artificial Neural Network to determine okay and fraudulent transaction.

**Significance of the Study**
This work would be of great significance to the following:

**Business Organizations:** they will gain a better understanding of fraud and be able to address it in a more systematic and informed manner than is currently done in industry. For business organizations, the problem of electronic fraud is a large one. If fraud detection methods can be improved, then fraud can be reduced, which produces a few benefits – apart from reducing financial losses sustained from fraud, reducing it will also have the effect of increasing the confidence of e-commerce users and increasing adoption and usage of e-commerce systems.

**Auditors:** This work will enable auditors (auditing firm) carry out e-commerce transactions audit effectively. This allows them to better meet their responsibilities in detecting irregularities induced by fraud that impact upon businesses by providing them with a tool that can manage fraud detection.

**Academic:** In the academic sphere, this research will provide a better understanding of fraud, of fraud detection methods that have been evaluated by experts in the field of auditing, and will produce a continuous assurance system which can be used as a tool in future research.
Literature Review
The term 'fraud' commonly includes activities such as theft, corruption, conspiracy, embezzlement, money laundering, bribery and extortion. The legal definition varies from country to country, and it is only since the introduction of the Fraud Act in 2006, that there has been a legal definition of fraud in England and Wales. Fraud essentially involves using deception to dishonestly make a personal gain for oneself and/or create a loss for another. Although definitions vary, most are based around these general themes. Meanwhile, there are two principal methods of getting something from others illegally. They can either be physically forced, or they can be deceived into giving up their assets. The first type is called robbery and the second is fraud. Albrecht et al. (2009) defines fraud as a deception made for personal gain. The most common definition of fraud according to Webster's Dictionary (2001) is:

"Fraud is a generic term that embraces all the multifarious means which human ingenuity can devise, which are resorted to by one individual, to get an advantage over another by false representations. No definite and invariable rule can be laid down as a general proposition in defining fraud, as it includes surprise, trickery, cunning and unfair ways by which another is cheated. The only boundaries defining it are those which limit human knavery."

Australian Government Fraud Control Guidelines define fraud as (ComLaw 2011):

"theft; accounting fraud (false invoices, misappropriation etc); unlawful use of, or obtaining property, equipment, material or services; causing a loss, or avoiding and/or creating a liability; providing false or misleading information, or failing to provide it when there is an obligation to do so; misuse of assets, equipment or facilities; making, or using false, forged or falsified documents; and wrongfully using information or intellectual property."

Furthermore, the Government identifies fraud as targeting revenue, benefits, property, information and intelligence, funding and grants, entitlements, facilities, and money or property. Benefits obtained fraudulently are not restricted to monetary or material benefits, and may be tangible or intangible, including unauthorized provision of access to or disclosure of information. Benefits may also be obtained by third parties in addition to the fraud perpetrator.

Roles Involved in Credit Card Processing
Credit card processing is the process where it happens with many parties, it also referred to as roles involved in the processing of a credit card transaction. Namely, the issuer, the cardholder, the merchant, the acquirer, the card association, and the settlement bank (Keith Lamond, 2013). The card issuer is a licensed financial institution or its agent that issues the credit card to the cardholder and is responsible for the provision of responses to authorization requests. Those financial institutions can be a bank, also referred to as the issuing bank that is member of a card association and adopts a payment card product promoted by the card association. The issuer here keeps the cardholder accounts to which he charge the bills. The issuer guarantees payment for authorized transactions, processing the payment card in accordance with the payment card product regulations and local legislation. The issuer supports the clearing and settlement functions between the cardholder and the acquirer. The issuer host is the computing system that accesses the cardholder accounts database and represents the issuer during the authorization, clearing, and settlement.
The cardholder is a customer of the issuer that uses a payment card in a business to consumer (B2C) payment transaction. The card acceptor is the party that accepts a payment card at the point of service, formats the data of the transaction in a payment message, and forwards the payment message to the acquirer.

Figure 1: Credit Card Transactions: Real World and online (Lamond, 2013)

Online Banking Fraud Detection Based on Local and Global Behavior

Kovach and Ruggiero (2011) stated that fraud prevention describes the security measures to avoid unauthorized individuals from initiating transactions on an account for which they are not authorized. In spite of many advanced mechanisms available for fraud prevention for online banking applications, it can fail. Fraud detection consists in identifying such unauthorized activity once the fraud prevention has failed. In practice, fraud detection must be used continuously, since the system is unaware that fraud prevention has failed (Bolton, and Hand, 2010). Among the approaches used by fraudsters, phishing is one of the most common forms for stealing account details for authentication from the customers. Social engineering is the most common method used in phishing. Social engineering usually comes in the form of e-mails trying to convince users to open attachments or by directing them to some fraudulent site, and most of the time it is so well designed that many customers are led to informing their account details. This paper presents a framework, and the corresponding system, for online banking fraud detection in real time. It uses two complementary approaches for fraud detection. In the differential analysis approach, the account usage patterns are monitored and compared with the history of its usage, which represent the user’s normal behavior. Any significant deviation from the normal behavior indicates a potential fraud (Murad and Pinkas, 2009).

Kovach and Ruggiero (2011) presented a fraud detection system proposed for online banking that is based on local and global observations of users' behavior. Differential analysis was used to obtain local evidence of fraud where a significant deviation from normal behavior indicates a potential fraud. This evidence is strengthened or weakened by the user's global behavior. In this
In this architecture, each access device from which transactions are performed is supposed to have an identity. These identities are used along with a set of counters to monitor the number of different accounts accessed by each device. The system uses two independent approaches for detecting frauds: a differential analysis approach that detects significant changes in transaction patterns in individual accounts, and a global analysis approach that uses the set of counters to detect unusual number of accounts accessed by a single device. The fraud evidences determined by the two approaches are then combined in order to determine an overall score that may trigger an alarm depending on a prefixed threshold. Meanwhile, their main contribution is a fraud detection method based on effective identification of devices used to access the accounts and assessing the likelihood of being a fraud by tracking the number of different accounts accessed by each device.

**Fraud Detection within Mobile Money: A Mathematical Statistics Approach**

Kappelin and Rudvall (2015) stated that today it is easy to do banking transaction digitally, both on a computer or by using a mobile phone. As the banking-services increases and gets implemented to multi-platforms it makes it easier for a fraudster to commit financial fraud. In their research, the discovered the need to focus on investigating log-files from a Mobile Money system that makes it possible to do banking transactions with a mobile phone. They developed a system whose main objective is to evaluate if it is possible to combine two statistical methods, Benford’s law together with statistical quantiles, to find a statistical way to find fraudsters within a Mobile Money system. To achieve this, rules were extracted from a case study with focus on a Mobile Money system and limits were calculated by using quantiles. A fraud detector was implemented that use these rules together with limits and Benford’s law in order to detect fraud. The fraud detector used the methods both independently and combined.
Finally, the results obtained showed that it is possible to use the Benford's law and statistical quantiles within the studied Mobile Money system. It is also shown that there is only a very small difference when the two methods are combined or not both in detection rate and accuracy/precision. Meanwhile, Kappelin and Rudvall (2015) concluded that by combining the chosen methods it is possible to get medium-high true positive rates and very low false positive rates. The most effective method to find fraudsters is by only using quantiles.

Credit Card Fraud Detection System through Observation Probability Using Hidden Markov Model

Khan, et al. (2013) proposed credit card fraud detection model using Hidden Markov Model. Hidden Markov Models (HMMs) which is a statistical tool and extremely powerful method used for modeling generative sequences characterized by a set of observable sequences. Hidden Markov Model is probably the simplest and easiest models which can be used to model sequential data, i.e. data samples which are dependent from each other. An HMM is a double embedded random process with two different levels, one is hidden and other is open to all. The Hidden Markov Model is a finite set of states, each of which is associated with a probability distribution. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated, according to the associated probability distribution. It is only the outcome, not the state visible to an external observer and therefore states are “hidden” to the outside; hence the name Hidden Markov Model (Ghosh and Reilly, 2014). HMM has been successfully applied to many applications such as speech recognition, robotics, bio-informatics, data mining etc.

Khan, et al. (2013) achieved their aim by storing all the information about credit card (Like Credit card number, credit card CVV number, credit card Expiry month and year, name on credit card etc.) in the credit card database. If details entered by User into the database are correct then it will ask for Personal Identity number (PIN). After matching of Personal Identity number (PIN) with database and account balance of user's credit card is more than the purchase amount, the fraud checking module will be activated. The verification of all data will be checked before the first page load of credit card fraud detection system. If user credit card has less than 10 transactions then it will directly ask to provide personal information to do the transaction. Once database of 10 transactions is developed, then fraud detection system will start to work. Observation probabilistic in an HMM Based system is initially studied, spending profile of the cardholder and followed by checking an incoming transaction against spending behavior of the cardholder we can show clustering model is used to classify the legal and fraudulent transaction using data conglomeration of regions of parameter, HMM based credit card fraud detection during credit card transaction. Khan, et al. (2013) presented experimental result to show the effectiveness of our approach.

From the above, the work of Kovach and Ruggiero (2011) summarizes the fact that: Kovach and Ruggiero (2011) developed a work on fraud prevention by describing the security measures to avoid unauthorized individuals from initiating transactions on an account for which they are not authorized. The key point of this work was to identify unauthorized activity once the fraud prevention has failed. Kappelin and Rudvall (2015) had a work whose main objective is to evaluate if it is possible to combine two statistical methods, Benford's law together with statistical quantiles, to find a statistical way to find fraudsters within a Mobile Money system. Khan, et al. (2013) developed credit card fraud detection model using Hidden Markov Model (HMM) which is a statistical tool and extremely powerful method used for
modeling generative sequences characterized by a set of observable sequences. However, all the above mentioned works concentrated on fraud detection. They lack fraud monitoring and real-time alert. The idea of generating one time password (OTP) and email code using a standard statistical approach like HMM, Support Vector Machine (SVM) was not captured. These codes should be generated and sent to the customer through different media (email and phone) before transactions are completed.

Analysis of Existing System

Data Mining and Detection Techniques

This uses data mining approach and techniques in detecting credit card fraud. Data mining refers to a family of machine learning techniques capable to analyze and extract non-trivial patterns from data (Chen, et al., 2012). The detection techniques using data mining are:

a) Artificial Neural Network: This imitates the way human brain works. It makes use of nodes which are called neurons. The neurons are computational units which are used to process some input information and produce some output (Ngai, et al., 2011).

b) Support Vector Machines (SVMs): This is suitable for credit card fraud detection because only two classes are needed; namely the “legitimate” and “fraudulent” class. SVM tries to calculate an optimal hyper plane which will separate the samples of the two classes (Meyer, 2012).

Fraud Detection Using Mathematical Statistics Approach: This technique as described in chapter two uses Benford's law and statistical quantiles to find a statistical way to find fraudsters within a mobile money system. To achieve this, the following algorithm is used.

**Fraud detector Algorithm**

1. Procedure Detect Fraud(quantile, time Span, logs Path)
2. If quantile = 95 then
3. Limits = limits for 95%-quantile*time Span
4. Else
5. Limits = limits for 99%-quantile*time Span
6. Records = read Records(logs Path)
7. For all record in records do
8. If records = transaction Record then
9. Is Fraud = checkIfFraud (record, rule, limits)
10. If  is Fraud = true then
11. Alarm(record)

Disadvantages of the Existing System

Implementing a fraud detection tool using data mining techniques involves a number of challenges which needs to carefully be considered (Maes, et al., 2012).

1. **Noise:** This is simply the presence of errors in the data, for example incorrect dates”. Missing values are also considered as noise. Noise can result in an erroneous model construction with bad predictive accuracy. The process of removing noise is called data cleansing. Depending on the concerned data set; data cleansing can be a very complex task.

2. **Supplying Labeled Training Samples:** Finding training samples and providing the right class labels for model construction can also be a very complex task. This is one of the biggest challenges of supervised learning techniques since labeled training samples may not always be available.
3. **Overlapping Data:** Overlapping occurs when a fraudulent transaction looks very similar to a legitimate one or when a legitimate transaction looks very similar to a fraudulent one. This is also a problem because it can lead to an erroneous model construction.

4. **Choosing Parameters:** Most of data mining techniques require a number of parameters including thresholds to pre-set by the user. Different parameters can lead to completely different model performance. This increases the complexity of model construction.

5. **Feature Selection:** Selecting the features – also known as attributes or columns – of the data set that should be used to construct the detection model can also be a challenge. Many articles in the literature suggest the features that should be used to achieve better results.

6. **Over-fitting:** Generally the training data set always contains few errors or random values even after data cleansing. These are known as “small fluctuations” in the data. Over-fitting occurs when the algorithm used in model construction, tries to learn as many information as possible from the training data set including this small fluctuations which do not represent the real situation. This can lead to a very complex model with poor predictive accuracy.

**Analysis of the Proposed System**

The new system is a joint approach of Hidden Markov Model (HMM) and Artificial Neural Network (ANN) in online fraud detection and monitoring. The HMM will be used to generate PIN, One Time Password (OTP) and e-mail Code. These authentication codes will be manipulated using ANN before online transaction will be completed. This system will be embedded in the present e-commerce system to enhance security and reduce external fraud. This approach will enable the system detect fraudsters and send alert to the admin if imposture (intruder) tries to complete financial transaction with customer’s e-commerce account. As soon as this alert is sent to the main e-commerce system, the account is temporary blocked and the Internet Protocol (IP) address of the intruder is automatically sent and stored in the fraud database as an evidence for forensic investigation.
Data Flow Diagram of the Proposed System
The Data Flow Diagram of the Proposed System is shown in Figure 1.3

Figure 3  Data flow of the new system

Conclusion
The Internet has become a major resource in modern business, thus electronic shopping has gained significance not only from the entrepreneur's but also from the customer's point of view. For the entrepreneur, electronic shopping generates new business opportunities and for the customer, it makes comparative shopping possible. As per a survey, most consumers of online stores are impulsive and usually make a decision to stay on a site within the first few seconds. E-commerce store is like a conventional shop interior. If the shop looks poor or like hundreds of other shops the customer is most likely to skip to the other site. Meanwhile, the issue of security and fraud has discouraged many not patronize and make effective use of e-commerce. Hence we have designed the paper to help e-commerce companies and payment companies to detect fraud. In this paper, the user is provided with a scalable security platform that is to be attached to the existing e-commerce platform that can be used to buy products and authenticate payments online. It can be agreed that e-commerce is not a new nomenclature within the modern day economic parlance but yet has not yet been fully embraced amongst many people because of fear of fraud.
References


.IJARPPSDES | 186 of 235


