Data Mining Application for Cyber Credit-card Fraud Detection System

John Akhilomen

Abstract: Since the evolution of the internet, many small and large companies have moved their businesses to the internet to provide services to customers worldwide. Cyber credit-card fraud or no card present fraud is increasingly rampant in the recent years for the reason that the credit-card i s majorly used to request payments by these companies on the internet. Therefore the need to ensure secured transactions for credit-card owners when consuming their credit cards to make electronic payments for goods and services provided on the internet is a criterion. Data mining has popularly gained recognition in combating cyber credit-card fraud because of its effective artificial intelligence (AI) techniques and algorithms that can be implemented to detect or predict fraud through Knowledge Discovery from unusual patterns derived from gathered data. In this study, a system's model for cyber credit card fraud detection is discussed and designed. This system implements the supervised anomaly detection algorithm of Data mining to detect fraud in a real time transaction on the internet, and thereby classifying the transaction as legitimate, suspicious fraud and illegitimate transaction. The anomaly detection algorithm is designed on the Neural Networks which implements the working principal of the human brain (as we humans learns from past experience and then make our present day decisions on what we have learned from our past experience). To understand how cyber credit card fraud are being committed, in this study the different types of cyber fraudsters that commit cyber credit card fraud and the techniques used by these cyber fraudsters to commit fraud on the internet is discussed.

Keywords: Cyber credit card fraud, cyber credit-card fraudsters, black-hat hackers, neural networks, data mining.

I. INTRODUCTION

Imagine a scenario at the end of the month where you as a credit-card owner received your credit-card statement; you noticed on your credit-card statement that a purchase was made on your credit-card for a blackberry phone you never bought nor made an order for. You called your credit card company to explain to them that you never made this purchase but you were told that you did made that order since it was recorded on their system the purchase made with your legitimate information. Then they went ahead to tell you that from their logged file, you actually made that purchase for a blackberry on "www.ebay.com" and the monetary transaction was successfully made to EBay. Now afterwards your credit-card company decided to investigate further and they called the internet company EBay. EBay checking their logged file for the transaction told your credit-card company that the blackberry phone was delivered to a shipping address in Turkey while you the actual credit-card owner lived in the USA. Obviously in a case like this, you are a victim of an internet credit-card fraud or no card present fraud. When your credit card or credit card information is stolen and used to make unauthorized purchases on e-commercial systems on the internet, you become a victim of internet credit card fraud or no card present fraud. This is nothing new and there is nothing unusual about this because identity theft and credit-card fraud are present-day happenings affecting many people and involving substantial monetary losses. Fraud is a million dollar business and it's increasing every year. The PwC global economic crime survey of 2011 suggests that 34% of companies worldwide have reported being victim of fraud in the past year and increasing from 30% as reported in the year 2009 [9]. Fraud is as old as humanity itself and can take an unlimited variety of different forms. However, in recent years, the development of new technologies like the internet has provided further ways in which fraudsters can commit fraud.

Fraud is a very skilled crime; therefore a special method of intelligent data analysis to detect and prevent it is necessary [11]. These methods exist in the areas of Knowledge Discovery in Database, Data Mining, Machine Learning and Statistics. They offer applicable and successful solutions in different areas of fraud crime. The aim of this study is actually focused on modeling an applicable system for detecting fraud in a real-time transaction on the internet. This model implements the anomaly detection algorithm of Data Mining, using Neural Networks to learn patterns used by a particular credit-card owner and then match the patterns learned with the pattern of the current transaction to detect anomalies.

II. WHAT IS CYBER CREDIT-CARD FRAUD OR NO CARD PRESENT FRAUD?

Recent and current scholars investigating credit-card fraud have divided credit-card fraud into two types; the online credit card fraud (or no card present fraud) and the offline credit card fraud (card present fraud) [1]. The online credit-card fraud (in this paper is cyber credit card fraud) is committed with no presence of a credit-card but instead, the use of a credit-card information to make electronic purchase for goods and services on the internet. The offline credit-card fraud is committed with the presence of a credit-card which in most cases have been stolen or counterfeited and thereby used at a local store or a physical location for the purchase or some goods or services. However, to define cyber credit-card fraud, it is a scenario where

John Akhilomen is with the School of Computing and Mathematics, University Of Derby, Kedleston Rd, Derby DE22 1GB, United Kingdom.
(email: Johnblithe@gmail.com)

the credit-card information of a credit-card owner has been stolen, or in some cases valid credit-card information has been uniquely generated (just like credit-card companies or issuers do) and thereby used for electronic payment on the internet or via the telephone. In most cases, no I.T or computer skill may be required to commit online credit-card fraud because of the different techniques in which credit-card information can be stolen by cyber fraudsters.

III. WHO ARE THE CYBER CREDIT-CARD FRAUDSTERS?

I. ***Credit-card information buyers:*** They are fraudsters with little or no professional computer skills (e.g. Computer Programming, Networking, etc.) who buy hacked (or stolen) credit-card information on an illegal "credit-card sales" website. They buy this credit-card information with the intention of making electronic payment for some good and services on the internet.

II. ***Black hat hackers***: Recent research on Hackers in terms of Computer Security defined a "black hat hacker" (also known as a cracker) as a hacker who violates computer security with malicious intent or for personal gain [8]. They choose their targets using a two-pronged process known as the "pre-hacking stage"; Targeting, Research and Information Gathering, and Finishing the Attack. These types of hackers are highly skilled in Computer Programming and Computer Networking and with such skills can intrude a network of computers. The main purpose of their act of intrusion or hacking is to steal personal or private information (such as credit-card information, bank-account information, etc.) for their own personal gain (for instance creating a "credit-card sales" website where other cyber credit-card fraudsters with little or no computer skills can buy credit-card information).

III. ***Physical credit-card stealers***: They are the type of fraudsters who physically steal credit-cards and write out the information on them. They physically steal these plastic credit-cards (maybe by pick-pocketing in a crowded place) and write out the credit-card's information with the intention of using this credit-card information to make electronic payment for some good and services on the internet.

IV. DIFFERENT TECHNIQUES FOR CREDIT-CARD INFORMATION THEFT BY CYBER CREDIT-CARD FRAUDSTERS

In other to detect cyber credit-card fraud activities on the internet, a study conducted on how credit-card information is stolen is a good approach. Listed below are studied different techniques which are used for credit-card fraud information theft.

I. ***Credit-card fraud generator software***: These are software written to generate valid credit-card numbers and expiry dates. Some of these software are capable in generating valid credit-card numbers like credit-card companies or issuers because it uses the mathematical Luhn algorithm that credit-card companies or issuers use in generating credit-card numbers to their credit-card consumers or users. In other cases, this software is written by black-hat hackers with hacked credit-card information stored on a database file from which the software can display valid credit-card information to other type of cyber credit-card fraudsters who have bought the software to use. This technique is some cases used by black-hat hackers to sell their hacked credit-card information to other online credit-card fraudsters with little or no computer skills.

II. ***Key-logger and Sniffers***: Black-hat hackers with professional Programming or computer skills are able to infect a computer by installing and automatically running sniffers or key-logger computer programs to log all keyboard inputs made into the computer on a file with the intention of retrieving personal information (like credit-card information, etc.). These black-hat hackers or fraudsters are able to infect users' computers by sending spam emails to computer-users requesting them to download free software or games, or sometimes they create some porn-sites so that when these computer-users browse these porn sites or download those free software or games, the sniffers or key-loggers are automatically downloaded, installed and ran on the users' computers. While the sniffer or key-logger is running under the users' computer, they sniff and log all the keyboard-input made by the user over a connected network. Therefore, any computer-user can unknowingly share their private information (credit-card information, etc.) through viral-infecting software such as these. In some cases, no Programming or computer skill is required to sniff a computer-user's key-board input because this software are also being shared or sold to other cyber credit-card fraudsters with little or no computer skills.

III. ***Spyware, Site-cloning and False Merchant sites****:* They are software created by black-hat hackers, installed and ran on users' computer to keep track of all their website activities. From knowing the website activities of the victimized computer-user on the internet, electronic or banking websites regularly visited by the computer-user can be cloned and sent to the user for usage with the intention of retrieving personal or private information (like bank log-in's). In the case of false merchant sites, fake websites can be created to advertise and sell cheap products to computer-users, and thereby asking for payment via credit-card. If a credit-card payment is made on any of these fake merchant sites, the user's credit-card information is therefore stolen.

IV. ***CC/CVV2 shopping websites*****:** cyber credit-card fraudsters with no professional computer skills can buy hacked credit-card information on these websites to use for fraudulent electronic payment for some goods and services on the internet.

V. ***Physical stolen credit-card information***: Fraudsters can physically steal the credit-card of a user to write out the credit-card information and then use for fraudulent electronic payment on the internet.

V. METHODOLOGY

I. ***Implementing Data mining Techniques for Credit Card Fraud Detection System:*** Data mining is popularly used to effectively detect fraud because of its efficiency in discovering or recognizing unusual or unknown patterns in a collected dataset. Data mining is simply a technology that allows the discovery of knowledge in a dataset. In other words, with Data mining knowledge can be discovered in a dataset. Data is collected from different sources into a dataset and then with Data mining, we can discover patterns in the way all data in the dataset relates with another and then make predictions based on the patterns discovered. Data mining takes a dataset as an input and produces models or patterns as output. One of the popular effective Data mining techniques used in data security is the Neural Networks. The concept of the Neural Networks is designed on the functionality of the human brain. From kindergarten until college, we are developed from an infantry stage of life unto the adult stage through different experiences or a set of data through how we're schooled. And we use this past experience or training we have acquired to make present day decisions. This is the Neural Networks. The Neural Networks makes predictions and classifications from what it has learned. The Anomaly detection algorithm is an implementation of the Neural Networks. Anomaly detection (sometimes called deviation detection) is an algorithm implemented to detect patterns in a given data set that do not conform to an established normal behavior[10]. The patterns thus detected are called anomalies and often translate to critical and actionable information in several application domains. The Anomaly detection is categorized into three; Unsupervised anomaly, Semi-supervised and Supervised anomaly detection. Unsupervised anomaly detection techniques detect anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the data set. Supervised anomaly detection techniques require a data set that has been labeled as "normal" and "abnormal" and involves training a classifier (the key difference to many other statistical classification problems is the inherent unbalanced nature of outlier detection). Semi-supervised anomaly detection techniques construct a model representing normal behavior from a given normal training data set, and then testing the likelihood of a test instance to be generated by the learnt model [10]. As seen in the diagram on fig. 1, this data mining application uses Supervised Anomaly detection to detect credit card fraud in a transaction and thereby classifies a transaction as Ok, suspicious fraud or illegitimate transaction.

II. ***Credit Card Fraud detection model:*** This Data mining application applies the anomaly detection algorithm to detect cyber credit card fraud in an online credit-card transaction implementing Pattern recognition with Neural Networks. Anomaly detection algorithm is a technique used in Data mining applications to detect specific patterns or relations within the data provided for Fraud detection process. There is a fixed pattern to how credit-card owners consume their credit-card on the internet. This fixed pattern can be drawn from legitimate regular activities of the credit-card owner for the past one or two years on its credit-card; the regular merchant websites the credit-card owner regularly makes electronic payment for goods and services, the geographical location where past legitimate transactions have been made, the geographical location where goods have been shipped to by the credit-card owner, the email-address and phone number regularly used by the credit card owner for notification. Using the Neural Network technology, the computer-program or software can be trained with this fixed pattern to use it as knowledge in classifying a real-time transaction as fraudulent or legitimate transaction. In this Data mining application for credit-card fraud detection, the anomaly detection algorithm is implemented for cyber credit-card fraud detection process. Once the data to be analyzed is selected, the anomaly detection algorithms will be applied to perform a data mining process for matching the behavior of the current transaction if it differs in behavior with the owner's past transactions on its credit-card. If the behavioral pattern in the current transaction differs with the learned pattern of the original credit-card owner, the system will continue to match the pattern of the current transaction if it's similar with past cyber credit-card fraud transactions. If the system returns false (of mismatch patterns between the current transaction and past fraud transactions) then the system classifies the transaction as suspicious fraud but if true, then the system will classify the transaction as illegal fraud transaction.

III. ***Pattern Recognition To Train Neural Networks:***

A. ***Geolocation of real-time transaction:*** The geolocation technology provides the absolute geographical location of an internet-connected computer by its IP address. An IP address is a unique network identifier issued by an Internet Service Provider to a computer-user every time they are logged on to the Internet [12]. This Data mining application is trained with IP-addresses (City and Country location being formatted from the IP-addresses) of internet-connected-computers the credit-card owner has used in the past one or two years legitimate transaction on its credit-card. This is a good mechanism to train Neural Networks for cyber credit-card fraud detection because in training Neural Networks with the City and Country locations formatted from IP-addresses where the credit-card owner has regularly made legitimate transactions from for the past one or two years, Neural Networks can know if the internet-connected-computer of the current transaction behaves in pattern like the internet-connected computers the credit-card owner has regularly made his past one or two years legitimate credit-card transactions. While this is a very good anti-fraud mechanism and useful for tracking fraudsters, the IP addresses can also be changed using Proxy servers. Anonymous proxy servers allow Internet users to hide their actual IP address and run their computers behind a fake IP address of their desired region [13]. The main purpose using a proxy server is to remain anonymous or to avoid being detected. Fraudsters hide themselves behind anonymous proxy servers to commit credit-card fraud on the internet. This Data mining application automatically flags for suspicious fraud if a proxy-server is detected in a transaction.

B. ***Email address and Phone number:*** When a credit-card is issued to an individual by a credit-card issuer or company, an email address or phone number from the individual is registered with the credit card so that the individual can receive notification via telephone or email of any transaction that's been made on their credit-card. For this reason, fraudsters do use different email-addresses and phone numbers when committing cyber fraud on credit cards. Although, It is important to take note that the cyber fraudsters do not only use email-addresses registered with free domains (like Yahoo, Google or Hotmail), but also they do pay to get registered email-addresses with non-free domains. Therefore, in this data mining application, Neural Networks will be trained with the email addresses and phone number the credit-card owner has used in past one or two years internet credit-card transactions.

C. ***Shipping address:*** Although it is not uncommon for people sending gifts to others to request different shipping address. It is very difficult to retrieve goods or apprehend fraudsters once the goods have left the country of residence of the original credit-card owner. Fraudsters will possibly not send goods to the legitimate cardholder's billing-address. But it is possible that credit-card owners will send goods to legitimate shipping address different to their billing address. Therefore, in this data mining application, Neural Networks will be trained with Shipping addresses and oversea orders used by the credit-card owner in past one or two years transactions.

D. ***Merchants' websites, regular good and services purchased in past credit cardholder's transactions:*** Neural Networks will be trained with the merchant websites the credit-card owner has regularly visited and the type of goods and services they have regularly purchased on its credit-card for the past 1 or two years. Neural Networks will be trained with the cost range of goods and services purchased in the past one or two years transactions of the credit cardholder's credit card.

VI. CONCLUSIONS

In this paper, a data mining application has been modeled as a subsystem which can be used with software systems and applications in financial institutions to detect credit-fraud in a transaction on the internet. This Data mining application accepts input formatted on a pattern on which a transaction is being executed and matches it with the credit-card holder's patterns of its credit-card online consumptions it's been trained with to classify a real-time transaction as legit, suspicious fraud or illegitimate transaction. The data mining application modeled in this paper uses the anomaly detection algorithm of the Neural Networks to detect fraud in a real-time transactions and it not prone to errors because of its classification of Transactions (legitimate, Suspicious Fraud and illegitimate). In the case of the suspicious fraud classification, the financial institution using the system can investigate further by calling the credit-card owner regarding the suspicious fraudulent transaction.

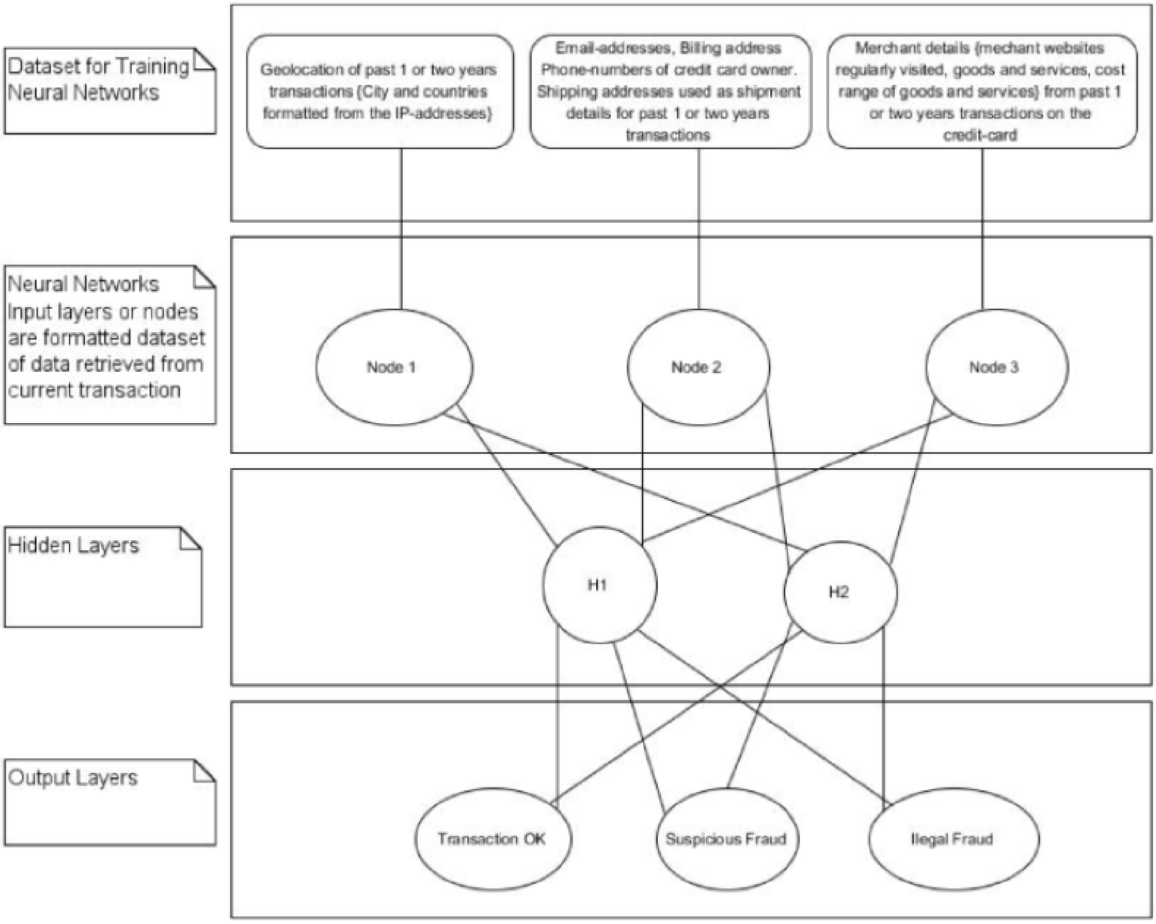

Dataset for Training Neural Networks
Geolocation of past 1 or two years transactions (City and countries formatted from the IP-addresses)
Email-addresses, Billing address Phone-numbers of credit card owner. Shipping addresses used as shipment details for past 1 or two years transactions
Merchant details (mechant websites regularly visited, goods and services, cost range of goods and services) from past 1 or two years transactions on the credit-card
Neural Networks Input layers or nodes are formatted dataset of data retrieved from current transaction
Node 1
Node 2
Node 3
Hidden Layers
H1
H2
Output Layers
Transaction OK
Suspicious Fraud
Ilegal Fraud

Fig. 1. The Learning and Classification of Neural Networks in the system

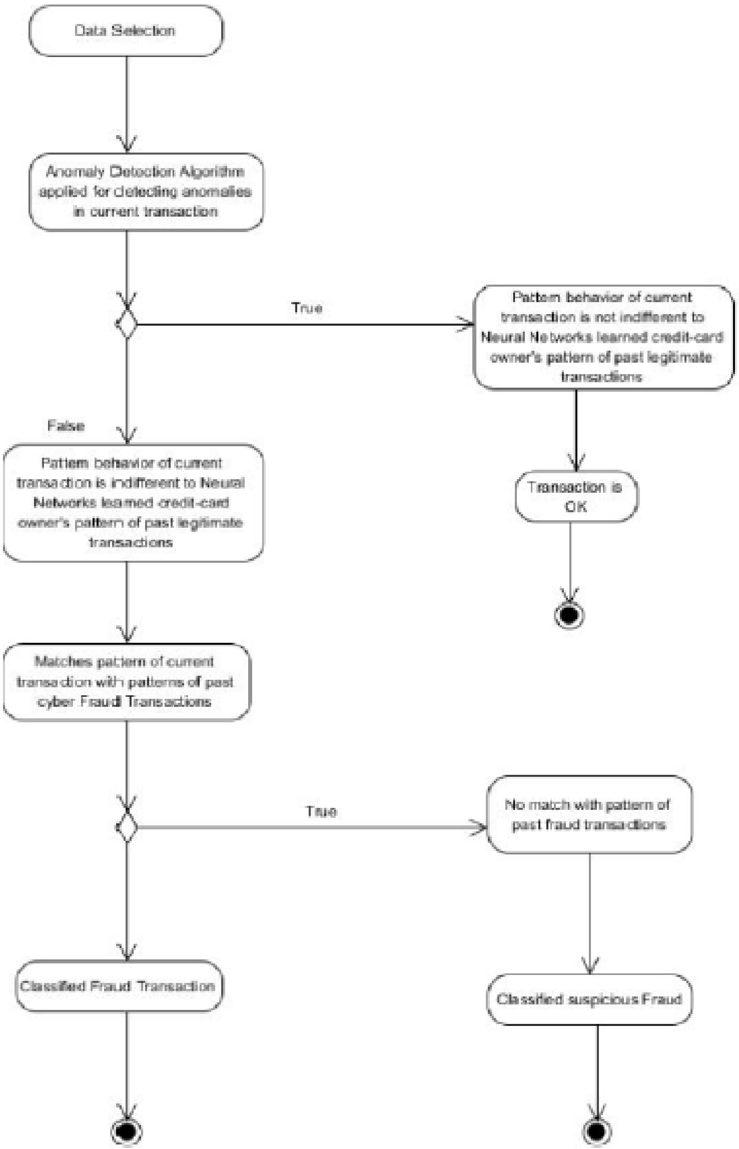
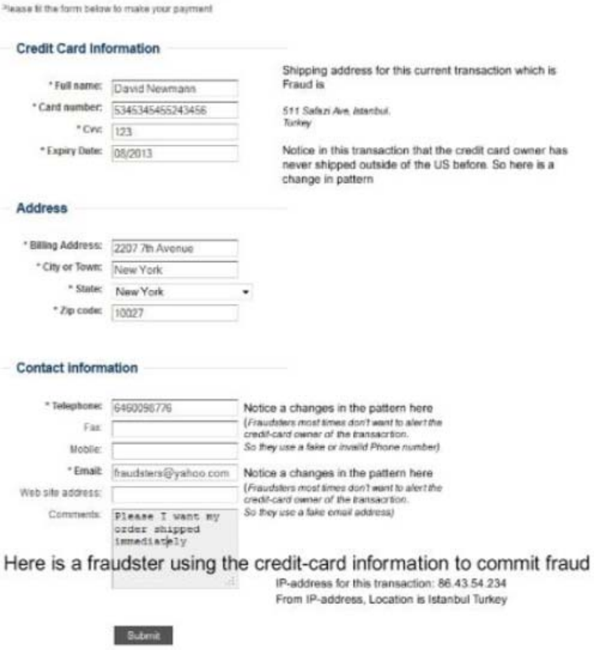

Data Selection
Anomaly Detection Algorithm applied for detecting anomalies in current transaction
True
Pattern behavior of current transaction is not indifferent to Neural Networks learned credit-card owner's pattern of past legitimate transactions
False
Pattern behavior of current transaction is indifferent to Neural Networks learned credit-card owner's pattern of past legitimate transactions
Transaction is OK
Matches pattern of current transaction with patterns of past cyber Fraud Transactions
True
No match with pattern of past fraud transactions
Classified Fraud Transaction
Classified suspicious Fraud

Fig. 2. The system's model for credit card fraud detection process in a transaction.


Please fill the form below to make your payment

Credit Card Information

* Full name: David Newmann
* Card number: 5345345455243456
* Cvv: 123
* Expiry Date: 08/2013

Shipping address for this current transaction which is Fraud is

511 Safazi Ave, Istanbul, Turkey

Notice in this transaction that the credit card owner has never shipped outside of the US before. So here is a change in pattern

Address

* Billing Address: 2207 7th Avenue
* City or Town: New York
* State: New York
* Zip code: 10027

Contact information

* Telephone: 6460098776
Fax:
Mobile:
* Email: fraudsters@yahoo.com
Web site address:
Comments: Please I want my order shipped immediately

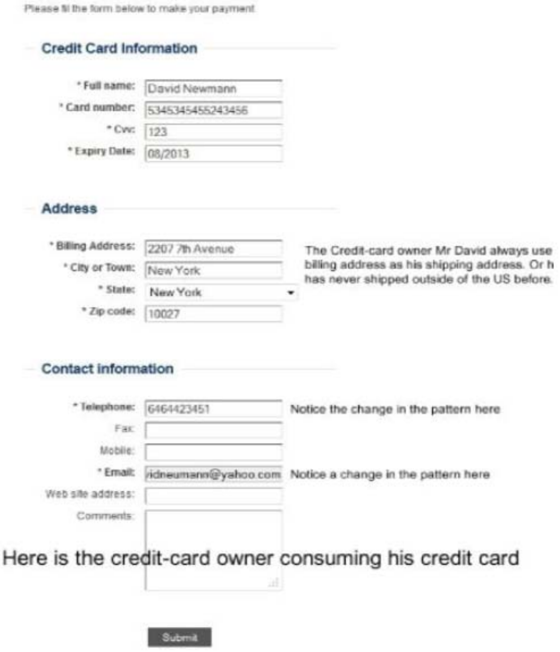
Notice a changes in the pattern here
(*Fraudsters most times don't want to alert the credit-card owner of the transacrtion. So they use a fake or invalid Phone number*)

Notice a changes in the pattern here
(*Fraudsters most times don't want to alert the credit-card owner of the transacrtion. So they use a fake email address*)

Here is a fraudster using the credit-card information to commit fraud

IP-address for this transaction: 86.43.54.234
From IP-address, Location is Istanbul Turkey

Fig. 3. A form used to purchase order by a cyber credit-card fraudster


Please fill the form below to make your payment

Credit Card Information

* Full name: David Newmann
* Card number: 5345345455243456
* Cvv: 123
* Expiry Date: 08/2013

Address

* Billing Address: 2207 7th Avenue
* City or Town: New York
* State: New York
* Zip code: 10027

The Credit-card owner Mr David always use billing address as his shipping address. Or h has never shipped outside of the US before.

Contact information

* Telephone: 6464423451
Fax:
Mobile:
* Email: idneumann@yahoo.com
Web site address:
Comments:

Notice the change in the pattern here

Notice a change in the pattern here

Here is the credit-card owner consuming his credit card

Fig. 4. A form used to purchase order by a credit card owner

References

[1]. Adnan M. Al-Khatib, Electronic payment fraud detection techniques, World of Computer Science and Information Technology Journal (2012), vol. 2, no. 4. pp. 137-141.

[2]. Francisca Nouyelum Ogwueleka, Data mining application in credit-card Fraud detection system, Journal of Engineering Science and Technology (2011), vol, 6, no. 3, pp. 311 - 322.

[3]. Dr. Yashpal Singh and Singh Chauhan, Neural networks in data mining. Journal of Theoretical and Applied Information Technology (2005-2009), vol, 5, no. 6. pp. 37-42.

[4]. Khyati Chaudhary and Bhawna Mallick. Exploration of data mining techniques in fraud detection: credit-card, International Journal of Electronics and Computer Science Engineering. vol. I, no. 3. pp. 1765-1771.

[5]. V.Dhecpa and Dr. RDhanapal, Analysis of credit-card fraud detection methods', International Journal of Recent Trends in Engineering (2009), vol, 2. no. 3, pp.126-128.

[6]. Khyati Chaudhary, Jyoti Yadav and Bhawna Mallick, A review of fraud detection techniques: credit-card, International Journal of Computer Applications (2012), vol. 45, no. I, pp.39-44

[7]. Sam Maes, Karl Tuyls and Bram Vanschoenwinkel, Credit-card Fraud Detection Using Bayesian and Neural Networks. [ONLINE] Available at: http://www.personeel.unimaas.nl/k-tuylslpublicationslpaperslmaenf02.pdf. [Accessed 12 December 2012].

[8]. Hacker (computer security) - Wikipedia, the free encyclopedia. [ONLINE] Available at: http://en.wikipedia.org/wiki/Hacker_(computer_security). [Accessed 12 December 2012].

[9]. Cybercrime: protecting against the growing threat Global Economic Crime Survey – PWC Global Economic. [ONLINE]. Available at: http://www.pwc.com/en_GX/gx/economic-crime-survey/assets/GECS_GLOBAL_REPORT.pdf. [Accessed 12 December 2012].

[10]. Anomaly Detection – Wikipedia, the free encyclopedia.[ONLINE] Available at http://en.wikipedia.org/wiki/Anomaly_detection. [Accessed 12 December 2012].

[11]. Data Analysis Techniques for Fraud Detection.[ONLINE] Available at http://en.wikipedia.org/wiki/Data_Analysis_Techniques_for_Fraud_Detection. [Accessed 12 December 2012].

[12]. Preventing Credit Card Abuse - Anti-Fraud Strategies. [ONLINE] Available at http://www.lawzilla.com/content/fed-bus-12301.shtml?&lang=en_us&output=json&session-id=3cd3dad0fc218a1ad59460ff032578fd. [Accessed 12 December 2012].

[13]. Precautions for internet traders to prevent fraudulent credit card. [ONLINE] Available at http://www.technade.com/2007/02/precautions-for-internet-traders-to_25.html?&lang=en_us&output=json&session-id=3cd3dad0fc218a12578fd. [Accessed 12 December 2012].