Development of android-based sms scam detector

Technology, Mobile Phone

This chapter reviews the origin of scam, when and where it all started, what motivates people to embark on advanced fee fraud (419 scam), how advanced fee fraud has affected individuals, other business corporation and countries, some of the various types of scams today, some approaches to detecting scam, a review of related works, and a summary of the reviewed works.

SHORT MESSAGE SERVICE (SMS)

The short messaging service (SMS) is a bi-directional service to transfer text over wireless communication systems. It is a text messaging service of many mobile phones today. SMS is one of the most widely used data application with an estimate of about 3. 5billion users that is 80% of mobile subscribers in 2010 (Ahonen, Tomi T 2011). It consists of a message that can be up to 160 alphanumeric characters in text and 70 alphanumeric in Unicode characters. SMS has been existence from the second generation (2G) until present of fourth generation (4G) GSM mobile (Pereira & Sousa, 2004). This GSM data service has established the simplest one-to-one communication by exchanging short text messages. Now SMS has been the most popular messaging service due to the low cost of SMS, network reliability has made sending SMS messages an economic option for GSM subscribers (Yoon, Kim, & Huh, 2010).

SMS SCAMA scam short message service is an effort to defraud an individual by gaining their confidence through a text message.

TYPES AND FEATURES OF SMS SCAM

There are a many types of scams occurring today, some of which are described below with their identified features.

SMSIHING

“ Smishing is a word that means SMS phishing or phishing that occurs through text messaging. Cyber criminals will either obtain phone numbers from the dark web following a data breach or a a random number generator. They will send messages asking users to call a number or click a link. The messages usually involve bank accounts, in some cases credit card details or bank verification number (BVN). A report from NBC news reports a smishing scam that tried to get a victim activate a new credit card. The messages prompted individuals to call a private number to enter their private information over the phone. At times smishing may lead to one installing a virus on their device without them knowing. In such cases the results are worse because the devices are been monitored especially during bank transactions. Overall, what smishers are looking for is a missing piece in a puzzle which could be a pin, password or other private details that will help them access your accounts.

LOTTERY SCAM

This type of scam usually involves a scammer sending a fake notice to the victim about a lottery won although the victim has not entered the lottery. The “ winner” is asked to send sensitive information such as his residential address, phone contact, and occupation/position lottery number and so on to an email address given. In addition to harvesting this information, the scammer then notifies the victim that releasing the funds involves some small fee (shipping, insurance, and registration).

EMPLOYMENT SCAM

In this type of scam, the scammer sends a letter with a falsified company logo on it to the victims who posted their resumes on some job websites. The job requires a work permit and a fake government official contact is given to the victims in order to obtain the work permit. The government official then proceeds to extract money from the victim for the work permit and then in the end nothing is given.

DATING AND ROMANCE SCAM

The scammer makes contact with the victim through social media, instant messaging online dating sites and so on. The scammer post pictures of someone very attractive, uses his communication to gain confidence of the victim and then ask for money to meet with them. They may also gain access to the victim’s bank accounts, credit cards, passport, and email accounts, to commit fraud with their details. In 2014 dating and romance scams occupied the top position in terms of financial losses, with $27, 904, 562 reported lost which accounts for 34 per cent of all losses that were reported.

ONLINE SALES AND AUCTION SCAMS

In this type of scam, a scammer sells a product online and then sends an inferior or low quality good or nothing at all to the victim. The scammer sometimes pretends to sell a product just to gather the victim’s bank details. In an Online auction scam, the victim is been told he has a second chance to buy an item in which he placed a bid on because the winner pulled out. The scammer will request that the money be paid outside of the auction site. By doing so, his money is lost already.

CHARITY AND MEDICAL SCAMS

Charity scam involves scammers collecting money from people pretending to work for a legitimate cause or charity organization. They exploit recent natural disaster or crisis news to play on the emotions of people of collect funds. Medical scam deals with sales of drugs and product that appear to be legitimate based on false testimonies of people.

SMALL BUSINESS SCAMS

If you own a small business, you can be targeted by scammers such as issuing of fake bills for unwanted or unauthorized products and services.

SOCIAL ENGINEERING

The act of manipulating people into performing actions or disclosing confidential information.

PHISHING

It is an attempt or effort to obtain classified information such as password, usernames, and credit card details for malicious purpose, disguising as a trustworthy party in an electronic communication. It is carried out by spoofing, email or instant messaging and it often directs users to enter their personal details on a fake website. FraudulentCreate fake websitesSend thousands of phishing emails link having links to fake websitesVictims click on this emails links and enter their personal information. Fraudulent stole the data from users. Figure 2. 3 the process of phishing a website.

CLONE PHISHING

This type of phishing attack occurs when a legitimate delivered text with an attachment or link inside the mail is cloned. The attachment in the text is then replaced with a malware or malicious version and then sent to people. Usually the original address is been spoofed to carry out this operation. It is usually claimed to be an updated version of the original text sent.

SPEAR PHISHING

In this type of phishing, a specific individual or company is been targeted and so all details about the individual or company is been obtained first to increase the probability of success. Spear phishers focus on a selected group of people who have some similarities, for example, employees who work at firm, in phone phishing the phisher makes a telephone call to the user and asks the user to dial a number. The purpose is to acquire personal information of the bank account over the phone as opposed to sending thousands of emails randomly. Phone phishing is frequently done with a forged caller identification. This kind of phishing denotes the messages that claim to be from a bank. Most times customers' bank, asking clients to dial a phone number regarding problems with customer's bank accounts. (Rathore 2014).

EFFECT OF SCAM

Fraud is a serious, complex and complicated offense that affects all facets of a society. Fraud always have a negative impact on individuals, an organization or corporation and the nation.

INDIVIDUALS

The advanced fee fraud cost individuals millions of dollars every year. In the United States of America, it is currently estimated that losses range from tens to hundreds to thousands of dollars per day (419 coalition website). Also victims are been threatened with violence unless they corporate with scammers. Smith et al (1993) reported that victims have also been held hostages until the ransom had been paid. He also reported that 17 people died in 1991 in attempt to recover their funds. Another victim ran from the United States of America to London because he had lost money that he owed farmers. The victim was the director of Ashburton Stock firm and he had siphoned $4million from the company’s fund to scam. He died of heart attack in London as the money was never recovered (Van Beynen 2002).

COPORATIONS AND BUSINESS

In as much as scams affect individuals, large firms such as banks have also become victims of fraudsters. It is reported that sometimes these fraudsters approach the banks through a middleman, perhaps a trusted customer whom they have managed to convince and who is willing to invest in their scheme. This trusted customer may wish the bank to loan money to the scheme, hold documents against receipts, issue guarantees or any number of other actions which all add up to the involvement by the bank and possible financial exposure. An incidence occurred when a bank was filed for bankruptcy when a corrupted bank official invested $250million in a scheme which was meant to be used to extend the airport in Abuja, Nigeria (SAPA-AFP 2005).

NATIONS

It is a fact that advanced fee fraud scammers do not only affects individuals and businesses, but they also have a huge impact on countries in which they operate. Countries are affected because advanced fee fraud leads to corruption and bribery, forgery, money laundering and loss of business confidence in countries and regions where such offenders are located (The 419 Coalition 2003).

MOTIVATION OF SCAMMERS

It is acknowledge that the basic motivation for fraud is greed. Another element present in advanced fee fraud is pride, the desire to possess what one cannot afford even when true financial deprivation may not exist. It stems from a desire to match a standard in terms of lifestyle, comfort and material possession of others who are better off (Grabosky 2002). It is also stated that ego/power is another motivation that applies to all fraudsters. According to Duffield et al. (2001: 3) scam also arise from imprudence, misfortune or a combination of both. The financial reward of advanced fee fraud was found to be another motivating factor for this crime. Bium (1972) said “ because the financial rewards are enormous. Maybe you’ve been selling sweets in the streets or maybe you’ve been working in a tuck-shop, and just because a letter you’ve sent, somebody pays you a million dollars. I mean obviously for a rationally thinking person, everybody would be interested in something like that would give you a financial reward for that kind of money”. Also in Nigeria the northerners have ruled for 43years out of the 58years of independence be it military or democratic regime. As a result thousands of people who have worked under the northern regime are sitting on billions of naira stolen from public coffers. It is therefore suggested that the exclusivity of the south in advanced fee fraud business can be explained as there way of getting back at the northerners who have also stolen the countries money (Adaora Reports 2002).

REVIEW OF RELATED WORKS

Huang Wen-Liang, Liu Yong, Zhong Zhi-Qiang, and Shen Zhong-Ming (2008) proposed a complex-network based SMS filtering algorithm which compares an SMS network with a phone call communication network. This comparison provides additional features, which makes SMS networks and obtaining well-aligned phone-calling networks difficult. Khemapatapan (2010) proposed two filtering methods for SMS spam message and then applying Support Vector Machine (SVM) and Naïve Bayesian (NB) algorithms for filtering. The two filtering methods performed were used to classify the words in SMS message. The support vector machine are supervised learning models with associated learning algorithm that analyze data used for classification. The binary classification of the data is created by using a separating hyper plane to maximize the space of the margin base on kernel functions, and extracting data and storing it in the vector, to reach the best solution of the problem and finding the suitable classification. This technique is beneficial for finding solutions of problems with unfamiliar history. Naive Bayes classifiers are widely used for text classification in machine learning are based on the conditional probability of features belonging to a class, which the features are selected by feature selection methods. It determines features by an existing feature selection method, and selects an auxiliary feature which can reclassify the text space aimed at the chosen features

(Zhang & Wang, 2009). Liu Jun, Ke Haifeng and Zhang Gaoyan (2010) proposed pattern-matching algorithm called BM algorithm.

The pattern matching technique filters messages based on patterns specified, such as words, text strings, and character sets mentioned in the message content or subject. The filter searches through the message for these specified patterns to classify it as either spam or ham messages. They evaluated the system and filtering algorithms by using the actual SMS data. The experimental data was 100, 000 short messages randomly extracted from the actual system of operators, and tested the BM algorithm and proved that BM algorithm is suitable to run under the condition of high concurrency and real-time environment. (Liu et al. , 2010)Yuanchun Zhu and Ying Tan (2011) proposed a Local concentration approach for extracting local-concentration- features for messages. Two implementation strategies of the approach, namely the LC-FL strategy and the LC-VL strategy, have been designed. Extensive experiments have shown that the proposed LC strategies have quite promising performance. Sarah Jane Delany, Mark Buckley. Derek Greene (2012) proposed different approaches to spam detection. It also discusses work on filtering SMS spam and reviews recent developments in SMS spam filtering. Delany, Buckley and Greene (2012) presented a state of the art SMS scam detection and filtering techniques and they reviewed some of different approaches to the SMS scam. They also discussed important issues with data collection and availability for further research. They analyzed a large dataset of SMS scam.

They collected instances of SMS dataset by collecting messages from two public consumer complaints websites: GrumbleText and WhoCallsMe, which have assembled a corpus of 1, 353 unique SMS spam messagesUysal, Gunal, Ergin and Sora Gunal (2012) investigated the impact of several feature extraction and feature selection approaches on filtering of SMS scam. This study extensively analyses the effects of several feature extraction and feature selection methods together on filtering SMS scam messages in two different languages, Turkish and English. The selected features are then combined with the structural features and fed into two distinct pattern classification algorithms, namely K-nearest neighbor and SVM to classify SMS messages as either spam or legitimate. The filtering framework is evaluated on two separate SMS message datasets consisting of Turkish and English messages. Experimental work indicated that the combinations of bag-of-words (BoW) and structural features, rather than BoW features alone, offer better classification performance most of the time. Efficacy of the utilized feature selection strategies was not significantly superior to each other for both languages. Chao Gao and Jiming Liu (2013) proposed Two-layer network model for simulating and analyzing the propagation dynamics of SMS based scam and viruses. S. -E. Kim, Jo and Choi (2015) proposed a light and fast algorithm for SMS filtering which can be performed within mobile phones independently. It employs techniques for remove unneeded data. These techniques include data filtering, feature selection, data clustering, etc.

They select important features using relative volume of feature values. (S. -E. Kim, Jo, & Choi, 2015)Akbari and Sajedi (2015) proposed an algorithm called (GentleBoost algorithm) for SMS scam detection. They tried to reduce the number of word attributes significantly without reducing accuracy in comparison to other successful methods. They used content of messages and tried to extract the words which are more repeated in scam messages. They applied tokenization by removing stop words or symbols such as “ the”. For classification, they applied GentleBoost algorithm and finally by optimizing the features and applying GentleBoost algorithm which combines features of AdaBoostM1 and LogitBoost algorithms. They obtained only 32 word attributes and 98. 30% accuracy. (Akbari & Sajedi, 2015)Suraj J. Warade, Pritish A. Tijare, Swapnil N. Sawalkar (2016) proposed a mobile service provider level SMS scam detection system. The system will first look up in log data and call log data base and check a direct or the mutual relation between sender and receiver if system found no relation and if the message content are found spamming then it will treat message as a spam message and forward message with spam tag or directly reject it. Table 2. 1 shows a summary of related works to sms scam detection.