Journal of Intelligent Learning Systems and Applications, 2024, 16, 274-315
https://www.scirp.org/journal/jilsa
ISSN Online: 2150-8410
ISSN Print: 2150-8402

Emoti-Shing: Detecting Vishing Attacks by Learning Emotion Dynamics through Hidden Markov Models

Virgile Simé Nyassi[1*], Franklin Tchakounté[1], Blaise Omer Yenké[2], Duplex Elvis Houpa Danga[1], Magnuss Dufe Ngoran[1], Jean Louis Kedieng Ebongue Fendji[2]

[1]Department of Mathematics and Computer Science, Faculty of Science, University of Ngaoundéré, Ngaoundéré, Cameroon
[2]Department of Computer Engineering, University Institute of Technology, University of Ngaoundéré, Ngaoundéré, Cameroon
Email: *simevirgilesnv@gmail.com

How to cite this paper: Nyassi, V.S., Tchakounté, F., Yenké, B.O., Danga, D.E.H., Ngoran, M.D. and Fendji, J.L.K.E. (2024) Emoti-Shing: Detecting Vishing Attacks by Learning Emotion Dynamics through Hidden Markov Models. *Journal of Intelligent Learning Systems and Applications*, **16**, 274-315.
https://doi.org/10.4236/jilsa.2024.163015

Received: June 13, 2024
Accepted: August 27, 2024
Published: August 30, 2024

Abstract

This study examines vishing, a form of social engineering scam using voice communication to deceive individuals into revealing sensitive information or losing money. With the rise of smartphone usage, people are more susceptible to vishing attacks. The proposed Emoti-Shing model analyzes potential victims' emotions using Hidden Markov Models to track vishing scams by examining the emotional content of phone call audio conversations. This approach aims to detect vishing scams using biological features of humans, specifically emotions, which cannot be easily masked or spoofed. Experimental results on 30 generated emotions indicate the potential for increased vishing scam detection through this approach.

Keywords

Social Engineering, Hidden Markov Model, Vishing, Voice Mining

1. Introduction

Advancement in technology is improving the quality of services in almost all aspects of life. Thanks to these technologies, people meet their needs comfortably and in less time. We depend on Information and Communication Technologies for business and leisure, store and use valuable information across numerous platforms, as we quest for a comfortable life. Thanks to mobile Smartphone, people can perform several operations like banking transactions anytime, anywhere at a go. They can withdraw cash, deposit and transfer money, make payments to utility services, and shop via the mobile device.

Malicious agents, hackers and criminals are looking for security weaknesses and gaps in the services we use. A vast range of threats exploit vulnerabilities in the cyberspace, creating more and more victims in our society. A lot of work is being done to secure the user's data and information by securing hardware, software (platforms), and procedures. Criminals are becoming more and more interested in passing through the human components of the information system (people) to penetrate the system [1]. They use social engineering (SE) to manipulate human emotions and exploit the human nature of trust to steal users' data and lure them into financial lose. These attacks are the most dangerous and successful attacks as they involve human interactions [2].

The process of involving the user to give out confidential data or get into fraudulent transactions via phone call is known as vishing [3]. This is phone call phishing where the scammer impersonates and lures the victims into dishing out sensitive information and money so fast that the victim has no time to verify the veracity of the caller. Receiving a call from a scammer who intends to manipulate you into dishing out sensitive information or losing money from mobile money accounts is the order of the day in our society. The National Agency for Information and communication Technologies in Cameroon reports that banks lose a lot of CFA due to phone call fraud [4]. Therefore, there is a need for a system that can intelligently detect a scam during a phone call conversation (vishing) and save individuals and institutions from losing money via scams.

A number of solutions have been proposed to track this kind of fraud. Some solutions are based on features such as caller ID identification ("who is calling") [5], content analysis ("what is said") [6]. Others focus on education to help reduce social engineering attack [7]. However, technologies like caller ID spoofing weakens caller ID based strategies. Also, scammers use different information in various scenarios to lure their victim into dishing out sensitive information; this weakens solutions based on content analysis. Education based solutions may not save an individual who is not informed about the strategies or manipulations of the scammers. They are therefore likely to fall victim.

We propose Emoti-Shing, a model that offers a means through which vishing scam can be tracked by analysing the manner in which a phone call audio conversation is done, which is independent of the caller, the content of the call and the education a potential victim may have as far as vishing is concerned. This approach works well to reinforce others, because vishers use social engineering, a well-developed tool in psychology, to manipulate their victims. Enabling machines to intelligently identify scam calls based on "how the conversation is made" can go a long way to reduce vishing attack. Here, we use Hidden Markov Models (HMM) of artificial intelligence to model the vulnerability states of a potential victim to a scam call, observed via sequences of emotions that can be extracted from a scam conversation. We are therefore interested in the sequences of emotional changes that can lead to a social engineering attack during a phone call conversation, as these observed emotions are a reflection of the vul-

nerability state of the victim at a given time. We propose an approach that brings together knowledge from psychology and sociology, affective computing and artificial intelligence to predict a conversation that is likely a scam conversation via emotion analysis. A successful manipulation leaves the victim in a state where the victim is highly vulnerable and will definitely do what the social engineer wants the victim to do. This approach has a potential of increasing vishing detection since it relies on the intrinsic and biological characteristics of humans that cannot be easily masked or spoofed as the case with other solutions.

The main contribution of our work is that it reveals formally that the emotional changes in a potential victim of a scam can be used to indicate the level of vulnerability of the potential victim to a scam call during a vishing scam. Our work therefore incites research in the use of intrinsic biological features such as emotions of a potential victim to detect vishing scam. This work, to the best of our knowledge, has not been used by authors. It therefore proposes an approach that can go along way to reinforce other approaches that are based on the properties of the scammer and the call. More specifically:

- We propose Emoti-Shing, an approach that models vulnerability states of victim expressed through observable emotions by using Hidden Markov Models;
- We implement and simulate our approach on 30 variable sequences of emotions, to explore and observe potentials of emotion dynamics towards scam detection.

The rest of the paper is structured as follows: Section 2 describes literature about vishing detection, the problem statement and the research hypotheses. Section 3 presents key concepts about social engineering in general, vishing in particular, emotion related psychology and Hidden Markov Models. Section 4 is about the proposal based HMM to curb vishing activities. Section 5 concerns results analysis after the implementation and simulation of what we propose, on generated sequences of emotions as well as the limitations. The paper ends with a conclusion and future works to improve results.

2. Related Works

Some works have been put in place to curb vishing scams. Authors in [5] propose a fraud detection system based on genetic programming. They use features like the caller ID, the chargeable duration of the call, the called party ID, the date and the time of the call etc., from historical call records of each user to construct five normal calling profiles and uses genetic programming classifier to identify illegitimate calls. Scammers bypass this by using techniques that spoof the caller ID. They may make calls of lengths that are not tracked by the historical records, leading to false positives and negatives.

Olszewski [8] demonstrated the significance of subscriber account visualization in the context of mobile phone fraud detection by employing a self organizing map in the context of mobile phone fraud detection, to model the user activities from historical data. A user may change his/her activities based on legitimate changes in

his/her environment. This might be misinterpreted by their approach.

The paper [9] presents an approach that identifies the fraudulent calls by initially forming groups of mobile phone users based on their calling instances present in the training set. A behaviour pattern matching algorithm is then used for matching a new call record with the normal user groups. The call is marked as normal if maximum similarity is found; otherwise, it is labeled as malicious.

The paper [10] demonstrates the usefulness of two clustering methods, namely, hierarchical agglomerative and K-means for identifying illicit actions in the calling profiles by constructing five subscriber profiles from their respective call records. Any sign of illegitimate activities found in the incoming call is analyzed by visualizing the clustering output generated from those profiles.

An approach proposed in [11] used Fuzzy C-Mean (FCM) and Support Vector Machine (SVM) on the past call records of each user for detecting fraudulent calls. The FCM clustering technique has been applied to certain calling features for user profile construction. The clustering outputs are then fed to SVM as input for building a trained SVM model, which then identifies a recent call record as a malicious one for not complying with the model.

Subudhi and Panigrahi [12] proposed a two-stage fraud detection system in mobile telephone networks to identify malicious calls amongst normal ones. They use a genetic algorithm based optimised fuzzy c-means clustering on user's historical call records to construct a calling profile. In the first stage of detection, the incoming call is passed to the clustering module that identifies the call as genuine, malicious or suspicious by comparing the distance value of the new calling instance from the cluster centers against a two predefined threshold values. Genuine and malicious calls are classified once while suspicious calls are further scrutinised in the second stage by previously trained groups of data handling models for final decision making.

In the work of [13], authors have developed a method that detects social engineering attacks based on natural language processing and artificial neural networks. This method can be applied in offline texts or online environments and flag a conversation as a social engineering attack or not.

Also, the authors in [7] suggest that the best way of detecting social engineering attack is to build behavioral biometrics into fraud prevention systems. These metrics detect abnormalities in the user's behavior caused by social engineering in real time. For example, the manner in which the user inputs data during a call can be a metric.

The works of Michael [14] considers the problem of detecting social engineering over telephone lines and propose Social Engineering Defense Architecture (SEDA) that generates attack signatures in real time. The authors designed SEDA to detect attacks based on intent and deception from the attacker. The authors however present a challenge and the need to detect social engineering via the target.

Lansley and his collaborators [6] propose and demonstrate a two-stage ap-

proach that detects social engineering attack based on natural language processing, case based reasoning and deep learning. They check on the URL of the site for anomalies, check for spellings in the chat text to get the number of misspelled words, check for intent verbs and adjectives and use this to track a social engineering attack. Overall, existing works on social engineering and vishing attacks can be summarized in Table 1.

Table 1. Existing works on social engineering/vishing attacks.

Author	Year	Phishing	Vishing	Call prop.	Princ. psy.	Biol. prop.
[5]	2014	Yes	Yes	Yes	No	No
[10]	2015	Yes	Yes	Yes	No	No
[8]	2014	Yes	Yes	Yes	No	No
[9]	2015	Yes	Yes	Yes	No	No
[11]	2016	Yes	Yes	Yes	No	No
[12]	2018	Yes	Yes	Yes	No	No
[13]	2020	Yes	No	Yes	No	No
[7]	2020	Yes	Yes	Yes	Yes	No
[14]	2005	Yes	Yes	Yes	No	No
[6]	2020	Yes	No	No	Yes	No

From the works presented above, we can say that most of the works to detect vishing scams involve building a profile from the attributes of the calls made. The existing works that deal with social engineering detection in general involve the use of natural language processing which requires text mining. However, the problem of confidentiality in security is bridged. Since the system is based on text mining. Again, conversation in a language that is not yet formal may fail to fit here. Hence there is a need for an approach that uses intrinsic and biological properties of the parties involved in a scam conversation to suggest the vulnerability of the victim towards a vishing scam. We therefore propose an approach that reinforces the existing approaches, by looking at those factors that are intrinsic to the parties involved in the conversation: emotions, that can't be masked by technology and can be detected in any conversation, independent of the natural language used. This approach also guarantees confidentiality, where the features extracted from the voice do not require any transcription into text.

2.1. Motivation

Our motivation stem from the fact that a lot of people in our society are becoming victims of scam calls nowadays. This is because these scammers use techniques that bypass hardware and software security mechanisms, and even security by education as they change their strategies to get the victims into doing what the scammers want. Boosting research in a direction that tracks biological

elements in the victim that can contribute to scam detection may go a long way to reduce the number of scam victims in our society. Given that a scammer knows beforehand where he/she wants to lead the victim, and cannot mask the vulnerability state of the victim, a system that monitors the changes in state of the vulnerabilities of the victim to reveal the probability that the victim is being scammed, leaves the scammer in the dark. Research from affective computing and other related fields has proposed various means by which emotions can be extracted from the human voice [15]. Authors in [16] have worked on the extraction of sequences of emotions from a conversation between two parties. In addition, a lot of work has been done in psychology that identifies the important role played by emotions to influence people in decision making [16]. Psychologists have exposed the "dark arts" of social engineering, and how social engineers manipulate human emotions to achieve their goals in business, politics, religion and other social spheres [17]. Inspired by this, and given that vishing is a form of social engineering attack in computer security, we propose a system that models the vulnerability states of the victim as a function of the observed emotional changes in the victim from the sequences of the utterances uttered by this victim during a scam conversation as the victim is manipulated by the social engineer.

2.1.1. Relevance of Emotions in Scam Detection

The choice of emotion in our approach towards scam detection stems from the nature of emotions and the power of emotions to influence behaviour and decision making. Considered to be an affective state of a person, created by the perception of the environment to trigger a reaction, the effects of emotions can't be masked by an individual [18]. Emotions can trigger a set of physiological, behavioural, communication and experience response that causes individuals to quickly deal with situations or opportunities [19]. Again, cognitive processes triggered by emotions interrupt current cognitive processes and directs the attention, memory and judgment to handle the emotion eliciting event [20]. This makes emotions a powerful ingredient in a social engineers' recipe as he/she seeks to alter the belief system of the victim and get the victim do what the social engineer wants. Social engineering is about manipulating the emotions of the victim. So keeping track of the emotions dynamics related to scam calls (social engineering attack) can give an insight towards detecting a vishing scam call.

2.1.2. Problem Statement

From the existing works presented, we can say that vishing scam detection using the attributes of the call and the caller leaves some gaps for scammers to exploit. This is due to the fact that the scammer can manipulate these attributes as they change their strategies to carry out this malicious act. This raises the fundamental question of how possible is it to track vishing scam by using the intrinsic and biological properties of the potential victim: emotional dynamics that can not be masked during a vishing scam conversation?

2.2. Research Hypothesis

This work is designed to assess the hypothesis that emotional dynamics observed during a vishing scam can be used to trace the vulnerability state of a potential victim to the scam.

3. Background

3.1. Social Engineering Attack

Social engineering is the act of manipulating a person to take an action that may or may not be in the target's best interest [21]. This may include obtaining information, gaining access, money transfer or getting the target to take some other actions that he would normally not take. Social engineering can be used in many areas of life. However, not all of these uses are malicious. Medical doctors use social engineering to obtain useful information from patients during diagnosis; detectives use social engineering to squeeze out useful information from people during investigations. In computer security, the focus is on the malicious use of social engineering. As software companies and organisations are leaving no stone unturned to strengthen their information systems, hackers and agents are redirecting their attention to the weakest part of the infrastructure: *i.e.* people [22]. They exploit the human nature of trust to pass through the weakest link in the security chain. By using well crafted means, the hackers get their victims into a vulnerable position, and then hit. Social engineering attacks aim at manipulating people to divulge valuable and sensitive data in the interest of cyber criminals. Security mechanisms [23] put in place to protect networks such as firewalls, intrusion detection systems, honey pots, cryptographic methods, and malware detection systems can be bypassed with a well planned and executed social engineering attack [24]. This is because social engineering attackers manipulate human emotions to develop an environment of trust in the human victim, which exposes the victim to attacks [25]. Malicious agents can influence people psychologically during interactions to give out sensitive information or break security procedures [26]. This makes social engineering attacks a permanent threat to all systems and networks, hence the best alternative for cyber criminals to attack a system with no technical vulnerabilities.

3.1.1. Attack Stages

Taking advantage of a victim by a social engineer to get sensitive information for malicious intentions remains one of the biggest threats in network security. Social engineering attacks are done in various ways, by various social engineers, to various categories of victims. However, these social engineering attacks share a common pattern with four phases [2]. The process begins with information collection about the target, where the attacker uses some criteria to select a victim. The second stage is the development of a relationship with the target, where the attacker communicates with the victim in order to get trust from the victim. In this stage the attacker uses psychology techniques that create strong emotional

responses to prepare the grounds for attack. Then, follows the exploitation of the available information. Here, the attacker manipulates the emotions of the victim and influence the victim to give out more sensitive information or perform an act that may harm the victim (the attack). The last stage is that of exiting with no traces. Once the social engineer succeeds or fails in the execution of his attack, he may simple leave, cleaning his traces as much as possible and prepares for the next attack. In summary, the stages of social engineering attacks are illustrated in Figure 1 below.

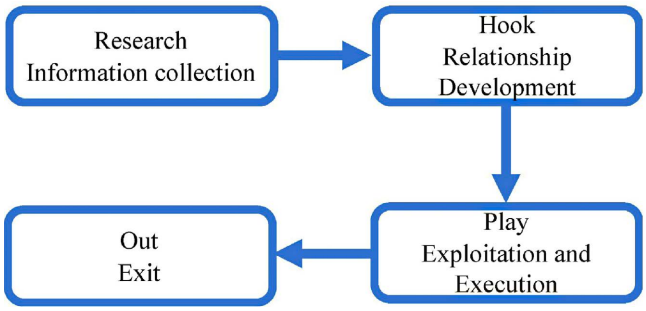

Figure 1. Social Engineering attack stages [2].

3.1.2. Attack Classification

Social engineering attacks can be grouped into two main categories: direct attacks and indirect attacks [2]. With the direct attacks, the attacker comes onto direct contact with the victim as they interact. Indirect attacks involve the use of information and communication technology tools like computers to lure a great number of people into falling victim. Several techniques including human, technical, social, computer and physical-based aspects can be put together to perform successful social engineering attacks. Common social engineering attacks include phishing, impersonation on help desk calls, shoulder surfing, dumpster diving, stealing important documents, diversion theft, fake software, baiting, quid pro quo, tailgating, Pop-Up windows, Robocalls, ransomware, online social engineering, reverse social engineering, and phone social engineering as shown in Figure 2.

Phishing attacks. These are attacks in which the attackers mislead victims to fraudulently obtain private and confidential information via phone calls, email, and Short Message Service (SMS). Phishing attacks are the most common attacks conducted by social engineers [27] [28]. They use fake websites, ads, emails, anti-virus, awards, PayPal websites, and free offers to lure their victims. Phishing attacks can be classified into five categories: spear phishing, whaling phishing, vishing, interactive voice response phishing, and business email compromise phishing as illustrated in Figure 3.

Phishing attacks in which the attacker targets specific individuals or selected groups using their personal information to make claims or communications is referred to as Spear Phishing. They collect information about the victim using available data online. Spear phishing that targets high profiles in companies is known as Whaling phishing. Interactive voice response phishing is performed by

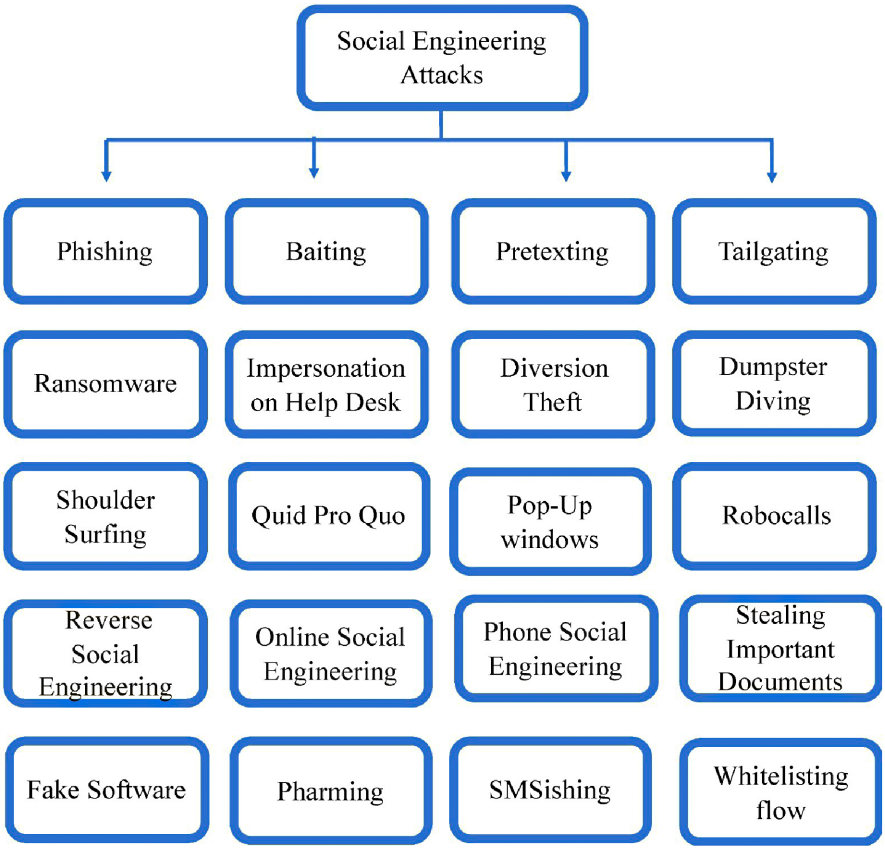

Figure 2. Social Engineering attacks [2].

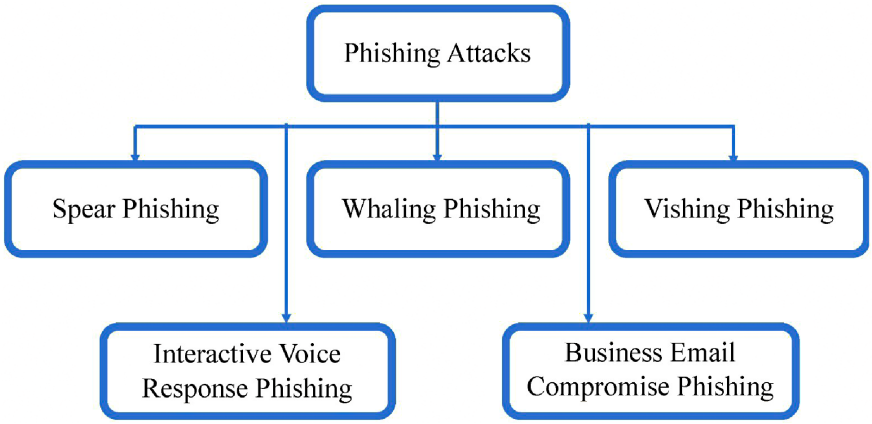

Figure 3. Phishing attacks [2].

using an interactive voice response system to make the target enter the private information as if it is from a legitimate business or bank [2]. Business email compromise phishing mimics the whaling by targeting high profiles in corporate businesses in order to get access to their business emails, calendar, payments, accounting, or other private information. The social engineer uses this data to send emails by mutating past emails, change meeting schedules, read professional information about the enterprise, and contact clients or service providers. Vishing attacks are performed when Phishers manipulate people to get their

sensitive information via phone call [29]. Phone Scam attacks involve the attacker contacting the victim via phone seeking specific information or promising a prize or free merchandise. They aim at influencing the victim to break the security rules or to provide personal information that can lead to harm. SMishing attacks consist of sending fraudulent messages and texts via cell phones to victims to influence them [30]. Robocall attacks have recently emerged as massive calls coming from computers to targeted persons with known phone numbers. They target cell phones, residential, and work phones. A robocall is a device or computer program that automatically dials a list of phone numbers to deliver pre-recorded messages. It is mainly based on voice over the internet protocol (VoIP) to ensure several VoIP functions such as interactive voice response and text to speech [2]. These calls can be about offering or selling services or solving problems.

3.2. Emotions

An emotion is considered to be an affective state of a person, created by the perception of the environment, which triggers a reaction [31]. Emotions influence our daily lives in a number of ways. The perceptions we have, the choices we make and the actions we take are all influenced by the emotions we are experiencing at a given time. Our lifestyle and interactions with others are greatly influenced by emotions. Theories in Psychology have identified the different types of emotions that people experience. According to Paul Eckman [29], there are six basic emotions universally experienced in all human cultures. These emotions are happiness, sadness, disgust, fear, surprise, and anger. Other emotions like pride, shame, excitement, and embarrassment were later added to these six emotions.

1) **Happiness:** This is a pleasant emotional state that is characterized by feelings of contentment, joy, gratification, satisfaction, and well-being. Happiness is sometime expressed via facial expressions: such as smiling, body language: such as a relaxed stance, tone of voice: an upbeat, pleasant way of speaking.

2) **Fear:** Fear is the emotional response to an immediate threat, anticipated threats or even our thoughts about potential dangers. It takes you through the flight or fight response. This response prepares you to deal with the threats in your environment that triggered the fear. Fear can be expressed via facial expressions such as widening of the eyes and pulling back chin, body language such as attempting to hide or flea from the threat, physiological reactions such as breathing and heartbeat.

3) **Anger:** This is an emotion characterised by a feeling of agitation, hostility, frustration and antagonism towards others or situations. It equally takes the body through the flight response. Humans express anger through facial expressions: such as frowning or glaring, body language: such as taking a strong stance or turning away, tone of voice: such as speaking gruffly or yelling, physiological responses: such as sweating or turning red, aggressive behaviours: such as hit-

ting, kicking, or throwing objects

4) **Disgust:** This is an emotion that manifest a sense of revulsion due to unpleasant situations, sight or smell. People experience moral disgust when they observe others engaging in behaviours that they find distasteful, immoral, or evil. This emotion can be manifested via body language: turning away from the object of disgust, Physical reactions: such as vomiting or retching, Facial expressions: such as wrinkling the nose and curling the upper lip.

5) **Sadness:** Sadness can be seen as a transient emotional state characterized by feelings of disappointment, grief, hopelessness, disinterest, and dampened mood. This emotion is experience via crying, dampened mood, lethargy, quietness, withdrawal from others.

6) **Surprise:** Surprise is an emotion characterised by a physiological startle response following something unexpected. Surprise is often characterized by facial expressions: such as raising the brows, widening the eyes, and opening the mouth. Physical responses: such as jumping back, verbal reactions: such as yelling, screaming, or gasping.

These six basic emotions can be combined to form other feelings as seen in the work of [32], who put forth a "wheel of emotions" that shows the emotion combinations. This is illustrated in Figure 4.

Figure 4. The wheel of emotion [29].

3.2.1. Role of Emotions in Decision Making

Human reasoning and decision making are highly linked to emotions. A scammer seeks to alter the belief system of the victim by playing with his reasoning and decision making power. Emotion dynamics during a scam conversation can be studied to trace scam prints in the conversation. Emotions play a decisive role in human behavior in many situations. They play an important role in human interactions where humans in stressful situations keep aside their cognitive processes and react according to emotion [33]. This is a great tool exploited by scammers as they aim at taking their victim into a position of acting even beyond reason. Keeping track of stressful moments in a conversation can reveal elements of scam possibilities and used to track a scam call. In [33], authors affirm the use of emotions in the field of social simulation. They emphasize on the use of psychological theories developed to understand the way people dynamically make decisions to make social simulations real. Emotions can therefore be used in the race to track scams, since emotions are an important psychological component in human decisions, in order to improve their realism. Emotions must be integrated when modeling humans who need to make decisions. Studies supported by Zeelenberg *et al.* [34] hold that emotions modify motivations during a decision making process, making them a key component of human cognition. Scammers seek to create different motives as they incite their victims during the scam conversation. The role of emotions in cognition has been studied at a neurological level by Bechara *et al.* [35] Their results support the idea that emotions have to be taken into account to correctly reproduce the human decision making process. Frijda *et al.* [5] links emotions to a specific action. They hold that the actions taken by a person are highly correlated to the emotional state of that person. Emotions are relevant in any simulation involving emotionally-impacted human decisions. Social engineering involves high manipulations of emotions to cause the victim act or behave in a particular way. Hence keeping track of the emotional dynamics can give insights of scams. Hatfield *et al.* [36] reveal that emotional contagion occurs during social interactions. Emotional contagion is the process where the emotions of a person are influenced by the emotions of other people nearby. Studying this contagion process in a scam call can reveal malicious intensions of the scammer. Emotions greatly influence our mode which has a direct effect on our decision making process. For example, depressed mood is characterized by feelings of guilt and sadness which have a significant effect on decision making. According to associative network models, this process explains why people in good moods make optimistic judgements and people in bad moods make pessimistic judgements [37]. Finally, emotions have a social role, as described by Frijda and Mesquita [38], meaning that people can communicate through emotions and social relations and that these factors produce a behaviour. Hence, people in a social group will communicate their emotions, which leads people around them to react to these emotions. For example, a person may express fear when looking at something hidden to another

person. This other person will anticipate a reaction and change his/her behaviour depending not only on the perceived fear, but also on the social link with the person. This reaction to others' emotions gives the scammer the possibility to mimic the emotions that he wants to activate in the victim. As argued by Tähtinen and Blois [38]: "human decision making and actions are embedded in emotions and therefore cannot be meaningfully separated". Therefore, emotions should be part and parcel of a social engineering attack, since social engineering is all about influencing the victim to take a decision or act in a particular way.

3.2.2. Emotion Detection from the Human Voice

The human voice can be characterized by several attributes such as pitch, timbre, loudness, and vocal tone. Speech carries linguistic content, *i.e.*, sentences and words, and paralinguistic content [39], such as mood, affect, speaker states such as intoxication and sleepiness, and speaker traits such as age, gender, and personality [39]. It has often been observed that humans express their emotions by varying different vocal attributes during speech generation. Hence, deduction of human emotions through voice and speech analysis has a practical plausibility and could potentially be beneficial for improving human conversational and persuasion skills [15]. Human voices are highly personal, hard to fake, and contain surprising information about our mental health and behaviours [40]. The key to voice analysis research is not what someone says, but how they say it: the tones, the speed, the emphases, the pauses. Technically, speech and music are acoustic signals, represented in the physical world by micro variations of pressure, mostly air pressure, in the range from approx. 50 - 8000 Hz. Emotion detection from speech offers means for estimating with considerable accuracy human emotion states. Research on emotion detection from speech has witnessed great advances in recent years. Authors in [40] found correlation between emotion and facial cues. In [41], authors fused acoustic information with visual cues for emotion recognition. Recently, [16] successfully used RNN-based deep networks for multimodal emotion recognition. Reproducing human interaction requires deep understanding of conversation [42] used memory networks for emotion recognition in dyadic conversations, where two distinct memory networks enabled inter-speaker interaction, Recent works [16] describe a new method based on recurrent neural networks that keeps track of the individual party states throughout the conversation and uses this information for emotion classification. We can therefore note here that emotion detection from speech in a conversation between two parties consist of extracting the emotion associated to each utterance in the conversation speech. Here the emotion of an utterance affects the emotion of the next utterance uttered by the parties in the conversation. This is illustrated in **Figure 5**.

At the utterance level, each utterance undergoes the steps summarized in **Figure 6** for its emotions to be extracted.

3.2.3. Emotional Arousal as a Scam Tool

The vulnerability to fraud in a victim can be increased through emotional arousal

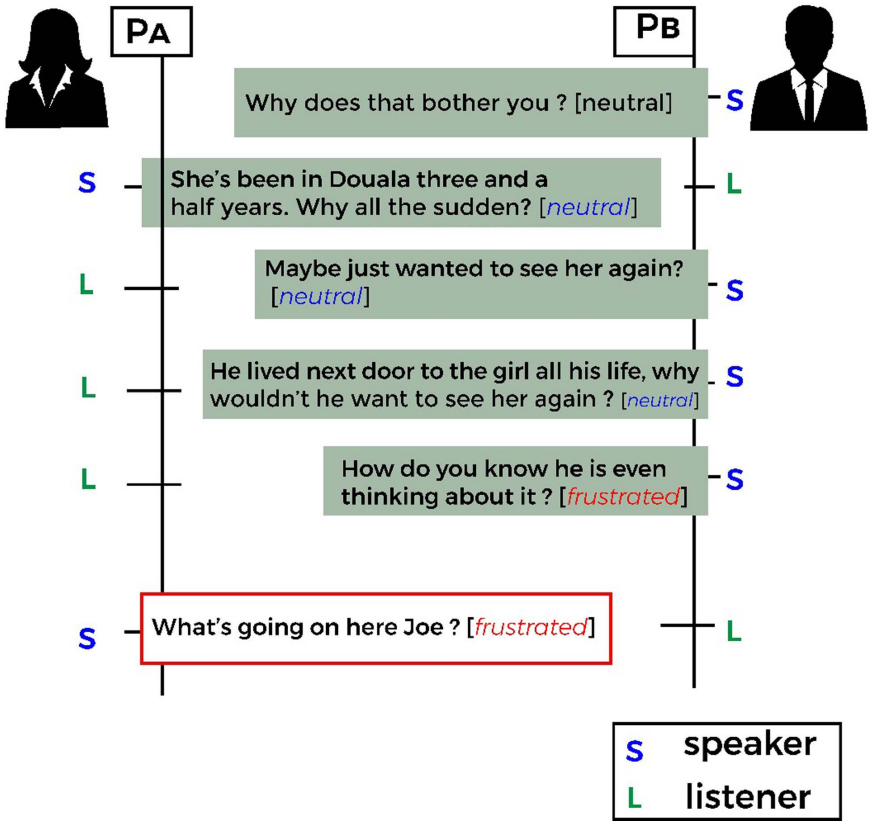

Figure 5. An illustration of a dialogue as an emotional sequence between two parties.

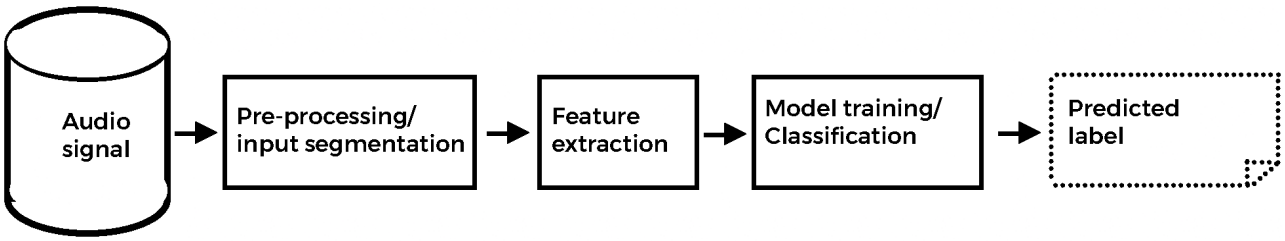

Figure 6. Processing steps for signal processing of an utterance.

[43]. Researchers have proven that emotion elicitation is a powerful technique used by scammers to influence their victims into doing what they want [44]. As a subjective feeling of state, neural and psychological activities work together to shape different emotional states. These emotions can be highly aroused in both the positive domain as seen in excitement or in the negative domain as seen with anger and fear [45]. Scammers frequently invoke emotional arousal to persuade their targets to comply. Activating highly positive emotions to influence decision making can be effective because they may promote heuristic or biased information processing rather than the effortful, higher-order cognitive processing needed for complex decision-making tasks For instance, high emotional arousal makes the victim to focus attention on reward cues associated, promised by a scammer, and to decrease attention to indicators of deception that may mitigate the likelihood of responding [46] analysed more than 100 Nigerian scam letters

and found consistent appeals to greed, guilt, lust, and charity. [47] found that sales agents were specifically trained to put targets into an emotional state when persuading them to buy bogus annuity products. In another study, researchers analyzed undercover audiotapes of phone calls in which fraudsters induced a sense of urgency by claiming the product was in short supply and induced excitement by dangling the prospect of wealth before the target. In [44], the authors found that consumers who were excited were more inclined to buy falsely advertised items than those in the neutral group. We can say that highly positive emotions like excitement manifested via greed, lust, and charity, as well as highly negative emotions like anger and fear are excellent ingredients in a social engineers recipe. Although neutral emotions may not directly influence the victim to take a decision, scammers may activate a neutral emotion to bring the victim's emotion from negative to positive and vice versa when need be.

3.3. Hidden Markov Model

The concept of observation is the building blocks of HMM. This concept is characterised by three main elements:

- **Symbols:** A symbol is an entity that can be observed. That is, it can be seen, identified, touched, felt, heard, etc. It can be an object, an activity, living thing, or an abstraction.
- **States**: A state is an observation point from which symbols can be observed. It could be a place, time of the day, seasons, physiological state etc.
- **State transitions**: Let S be a set of state and s_1 and s_2 be states in S. A state transition from s_1 to s_2 moves the system from s_1 to s_2 when an observation is made.

3.3.1. Markov Chains

A Markov chain of length T is an ordered sequence of T consecutive observations implicitly produced by state transitions, when we move from one observation to another. That is, given that $S=\{s_1,\cdots,s_N\}$ is a set of states and $O=\{o_1,\cdots,o_N\}$ O is a set of symbols observed, a Markov chain δ of length $|\delta|=T$ is given by Figure 7 with $S_{it}\in O_{it}$.

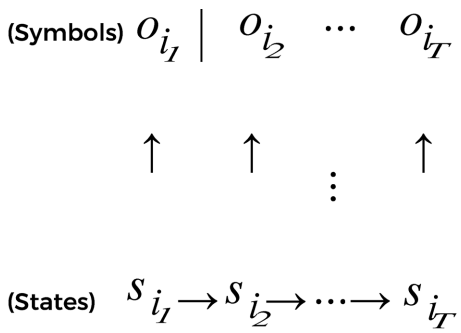

Figure 7. Illustration of a Markov Chain [48].

Here, the state S_{i1} is called the initial state. The symbol $O_{i1}O_{i2}\cdots O_{iT}$ is

called a sequence and $S_{i1}S_{i2}\cdots S_{iT}$ is called a path. Therefore a Markov chain is built by going through a path made up of states, observing a symbol in a particular state. A Markov chain is useful when we need to compute a probability for a sequence of observable events. In many cases, however, the events we are interested in are hidden: we do not observe the states directly. Hence the need for a model that considers observations emitted by hidden states.

3.3.2. Hidden Markov Model

A HMM λ [28] can be formally defined as a 5-upplet made up of:

1) A set of N states $S=\{s_1,\cdots,s_N\}$.

2) A set of M observable symbols $O=\{o_1,\cdots,o_N\}$.

3) A $N\times N$ state transition probability matrix, $A=\{a_{ij}\}$ where $\{a_{ij}\}=p\{S_i\mid S_j\}$.

4) A $N\times M$ observation matrix, $B=\{b_j(k)\}$ where $\{b_j\}=p\{S_j\mid o_k\}$.

5) An N dimension vector of initial states, denoted $\pi=\{\pi(i)\}$ where $\pi(i)=P(S_i \text{ is an initial state})$.

From these definitions, we can say that the components A, B and π of a HMM are probability distributions. Therefore the elements of each of their rows sum up to 1. Given appropriate values of N, M, A, B and π, we can use the HMM as a generator to generate an observation sequence

$$O=\{o_1,\cdots,o_T\} \tag{1}$$

as specified by the algorithm below:

1) Choose an initial state $q_1=S_i$ according to the initial state distribution π.

2) Set $t=1$.

3) Choose $O_t=v_k$ according to the symbol probability distribution in state S_i, *i.e.*, $b_i(k)$.

4) Transit to a new state $q_{t+1}=S_j$ according to the state transition probability distribution for states S_i, *i.e.*, a_{ij}.

5) Set $t=t+1$; return to step (3) if $t<T$; otherwise, terminate the procedure.

The above procedure can be used as both a generator of observations, and as a model for how a given observation sequence was generated by an appropriate HMM. A HMM can therefore be specified by a set of model parameters: the set states and observations and three probability measures A, B and π. For convenience, we shall use the compact notation

$$\lambda=(A,B,\pi) \tag{2}$$

to indicate the complete parameter set of the model. A HMM can be represented graphically by using a directed graph in which the states of the HMM are the nodes and there exist an edge from the node s_i to the node s_j on the graph, if the value of a_{ij} of the HMM is different from zero. The weight of this edge is therefore the value of a_{ij}. The observation probabilities are linked to each state s_i by an arrow that points out from the hidden state to the i^{th} row of the matrix B. The initial probability vector is not represented on the graph **Figure 8**.

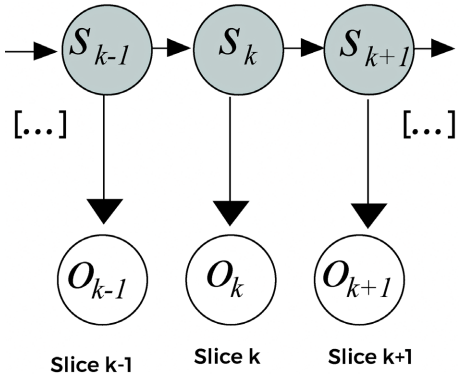

Figure 8. Diagrammatic representation of a HMM [49].

3.3.3. Hidden Markov Model Related Problems

HMM are generally used to solve problems linked to the following:

1) Evaluation:

Finding the probability of an observed sequence given a HMM. This evaluation must be done independent of the path taken, since the states that constitute these paths are hidden. We therefore need to calculate $P(O|\lambda)$. This problem can be solved using the Forward Backward algorithm.

2) Decoding:

This consists of finding the sequence of hidden states that most probably generated an observed sequence. The idea is to find the ideal sequence of states that maximises the value of $P(O|\lambda)$. The solution to this problem is obtained from the Viterbi algorithm.

3) Learning:

This involves generating a HMM λ given a sequence of observations. The idea here is to optimise the parameters (A, B, π) of λ to obtain a new model such that $P(O|\lambda)$ is optimal. This problem is addressed by the Baum-Welch algorithm.

3.3.4. Solutions to Hidden Markov Problems

1) Solution to the evaluation problem: the Forward Backward Algorithm

Basic solution:

Given a set of states $S=\{s_1,\cdots,s_M\}$ and a set of observable symbols $O=\{o_1,\cdots,o_M\}$.

Let q_t denote the state at time t and o_t denote the symbol observed by the model in the state q_t.

Consider a sequence of observations $O=\{o_1,\cdots,o_t\}$ and a HMM $\lambda=(A,B,\pi)$.

The goal is to calculate the probability of observing the sequence O, no matter the path taken, given that we use the model λ.

i.e. $P(O|\lambda)$.

This therefore consist of evaluating for each possible path Q, the probability of observing O, taking the path Q and given that we use the model λ. *i.e.* $P(O, Q|\lambda)$.

Let $Q=q_1,q_2,\cdots,q_t$ be one of the possible paths, calculating $P(O, Q|\lambda)$ consist of evaluating the probability of:

- Taking the path Q given that we use the model λ.
  That is $P(O \mid \lambda)$.
- Observing O given that we take the path Q and we use the model λ.
  That is $P(O/Q,\lambda)$.
  We therefore have

$$P(O,Q \mid \lambda) = P(Q \mid \lambda)P(O/Q,\lambda). \tag{3}$$

But

$$a - P(O/Q,\lambda) = b_{q1}(o_1) * b_{q2}(o_2) * \cdots * b_{qT}(o_T) = \prod_{t=1}^{T} b_{qt}(o_t) \tag{4}$$

$$b - P(Q/\lambda) = \pi_{q1} a_{q1,q2} * a_{q2,q3} * \cdots * a_{qT-1,qT} = \pi_{q1} \prod_{t=2}^{T-1} a_{qt-1,qt}$$

Now when we consider all the possible paths of $Q = q_1 q_2 \cdots q_T$ we have:

$$\begin{aligned} P(O \mid \lambda) &= \sum_{\forall Q} P(O/Q,\lambda) \\ &= \sum_{\forall Q} P(Q \mid \lambda)P(O/Q,\lambda) \\ &= \sum_{\forall\, q_1 q_2 q_3 \cdots q_T} \left(\pi_{q1} \prod_{t=2}^{T-1} q_{t-1}, q_t\right)\left(\prod_{t=1}^{T} b_{q_t}(o_t)\right) \end{aligned}$$

This computation can be interpreted as follows: Initially (at time $t = 1$) we are in state q_1 with probability π_{q_1}, and generating a symbol o_1 (in this state) with probability $b_{q_1}(o_1)$.

The clock changes from time t to $t + 1$ ($t = 2$) and we make a transition to state q_2 from state q_1 with probability $a_{q_1 q_2}$, and generate symbol o_2 with the probability $b_{q_2}(o_2)$.

This process continuous in this manner until we make the list transition (at time T) from state q_{T-1} to state q_T with probability $a_{q_{T-1} q_T}$ generating the symbol o_T with probability $b_{q_T}(o_T)$.

This method is so complex, with a complexity of $2T * N^T$. Since for every t from 1 to T, there are N possible states which can be reached. And each state sequence has about $2T$ operations for each term in the sum.

Forward Backward Algorithm:

Given an HMM $\lambda = (A, B, \pi)$ with a set of states $S = s_1, \cdots, s_N$ and a set of observable symbols $O = o_1, \cdots, o_M$. Let $O = o_1, \cdots, o_T$ be a set of observable sequence of symbols. The forward variable with index j at time t, denoted by $\alpha_t(j)$ is the probability of observing the sub-sequence $o_1, \cdots, o_t$, what ever path taken, being in the state s_j in λ at the instant t. That is $\alpha_t(j) = P(o_1, \cdots, o_t, q_t = s_j \mid \lambda)$ as shown in **Figure 9**.

$\alpha_t(j)$ can be calculated recursively for all $t = 1, 2, \cdots, T$. Firstly, $\alpha_1(j)$ is calculated and then $\alpha_{t+1}(j)$ is calculated from $\alpha_t(j)$.

1)

$$\begin{aligned} \alpha_1(j) &= p\left(o_1, \left(q_1 = s_j \mid \lambda\right)\right) \\ &= p\left(q_1 = s_j \mid \lambda\right) * p\left(o_1 \mid q_1 = s_j, \lambda\right) \\ &= \pi_j * b_j(o_1) \end{aligned}$$

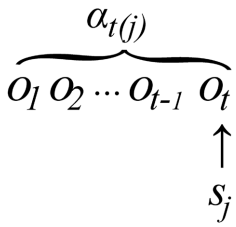

Figure 9. Illustration of the forward variable.

2)

$$\alpha_{t+1}(j) = p\left[o_1 o_2 \cdots o_t o_{t+1}, q_{t+1} = s_j \mid \lambda\right]$$

This is seen as N different Markov chains with i leaving from 1 to N we therefore have:

$$\begin{aligned}\alpha_{t+1}(j) &= \sum_{i=1}^{N} \alpha_t(i) * a_{ij} * b_j(o_{t+1}) \\ &= b_j(o_{t+1}) * \left(\sum_{i=1}^{N} \alpha_t(i) * a_{ij}\right)\end{aligned}$$

The forward variables are therefore given by the equation below: $\alpha_t(j)$

$$\begin{cases}\alpha_1(j) = \pi_j * b_j(o_1) \\ \alpha_{t+1}(j) = b_j(o_{t+1}) * \left(\sum_{i=1}^{N} \alpha_t(i) * a_{ij}\right)\end{cases}$$

The backward variable with index i at time t, denoted by $\beta_t(i)$ is the probability of observing the sub-sequence $o_{t+1}o_{t+2}\cdots o_T$, what ever path taken, being in the state in λ at the instant t. That is $\beta_t(i) = P\left(o_{t+1}o_{t+2}\cdots o_T \mid q_t = s_i, \lambda\right)$ as shown in Figure 10.

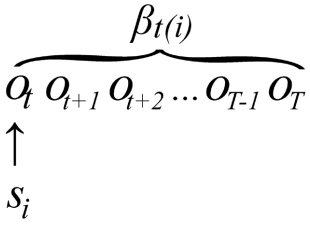

Figure 10. Illustration of the backward variable.

In a similar way with the forward variable calculations, we shall recursively calculate the $\beta_t(i)$ by calculating the first $\beta_T(i)$, then $\beta_T(i)$ from $\beta_t(i)$. which makes not sense!!! 1) $\beta_T(i) = P\left[o_{T+1}\cdots o_T \mid q_T = s_i\lambda\right]$ We therefore attribute $\beta_T(i) = 1$ by convention.

2) $\beta_{T-1}(i) = P\left[o_{T+1}\cdots o_T \mid q_{T-1} = s_i\lambda\right]$.

This is seen as N different Markov chains with j leaving from one to N we therefore have $\beta_t(i)$ $\begin{cases}\beta_T(i) = 1 \\ \beta_{t-1}(i) = \sum_{i=1}^{N} a_{ij} * b_j(o_t) * \beta_t(j)\end{cases}$.

Given the observable sequence $0 = o_1, \cdots, o_T$ and a HMM λ. When we combine the Forward and Backward variables, we have a global situation for any ar-

bitrary value of t^- taken between 1 and T in Figure 11.

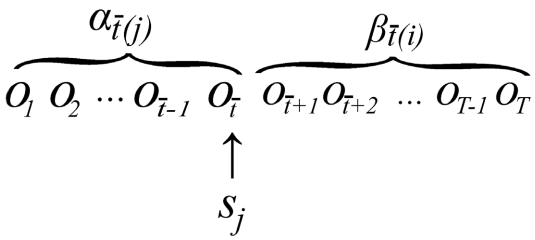

Figure 11. Combining the forward and backward variables.

From the above analysis, we can deduce that

$$P(O|\lambda) = sum_{j=1}^{N} \alpha_{\bar{t}}(j)\beta_{\bar{t}}(j) \tag{5}$$

Some algorithm can be used to find the best solution. Their investigation is out of this work's scope. We therefore just describe them in the next.

2) Solution to the evaluation problem: the Viterbi Algorithm

The Viterbi algorithm [28] is a formal technique for finding the single best state sequence that exist for a given observation of the HMM based on dynamic programming method. To find the single best state sequence, $Q = \{q_1, q_2, \cdots, q_T\}$, for a given observation sequence $O = \{o_1, o_2, \cdots, o_T\}$, we need to define the quantity

$$\delta_t(i) = q_1, q_2, \cdots, q_{t-1}{}^{max} P[q_1, q_2, \cdots, q_t = s_i, o_1, o_2, \cdots, o_t | \lambda] \tag{6}$$

That is, $\delta_t(i)$ is the best score along a single path, at time *t*, which accounts for the first *t* observations and ends in state S_i. By induction we have

$$\delta_t(i) = \left[i^{max} \delta_t(i) a_{ij} \right] \cdot b_j o_{t+1} \tag{7}$$

To actually retrieve the state sequence, we need to keep track of the argument which maximised for each *t* and j. This is done via the array $\psi_t(j)$. The complete procedure for finding the best sequence can be stated as follows:

1) Initialisation:

$$\delta_1(j) = \pi_i b_i(o_1) \quad 1 \le i \le N \psi_1(i) = 0 \tag{8}$$

2) Recursion:

$$\delta_t(j) = \max_{1 \le i \le N} \delta_{t-1}(i) a_{ij} b_j o_t \qquad 2 \le i \le T \quad 1 \le j \le N \tag{9}$$

$$\psi_t(j) = \arg\max_{1 \le i \le N} \delta_{t-1}(i) a_{ij} \qquad 2 \le i \le T \quad 1 \le j \le N \tag{10}$$

3) Termination:

$$P^* = \max_{1 \le i \le N} \left[\delta_T(i) \right] q_t^* = \arg\max_{1 \le i \le N} \left[\delta_T(i) \right] \tag{11}$$

4) Path (state sequence) backtracking:

$$q_t^* = \psi_{t+1}\left(q_{t+1}^*\right), t = T-2, \cdots, 1. \tag{12}$$

3) Solution to the evaluation problem: the Baum-Welch algorithm

Given a HMM λ, and an observation sequence $O = (o_1, o_2, \cdots, o_t)$, we aim at

finding a new HMM λ' that explains the observations better, *i.e.*, such that $P[O \mid \lambda'] \geq P[O \mid \lambda]$. The solution to the third problem of HMM takes observed data as input, and uses heuristic methods to find locally optimal parameters of the HMM that generated that data. This solution assumes that data come from some random process that we can fit to a HMM. With the observations and number of states fixed, transition, emission and initial distribution probabilities all not known. Also, the data are a set of observed sequences, each of which has a hidden state sequence. All parameters and probabilities can be set to some initial values.

4. Model Development

In this section, we develop Emoti-Shing, a model for emotion dynamics towards scam detection. We shall give a clear problem statement addressed by the model, formulate the vulnerability states of a potential victim as well as the emotions that are manipulated by a scammer. This shall enable us to propose the HMM that has the vulnerability states as hidden states observed via emitted emotions. We end the section by implementing our model.

4.1. Privacy and Ethical Considerations

The Emoti-Shing model has been designed with a strong focus on protecting the privacy of participants. Importantly, the model does not store any personal information, such as telephone conversations, telephone number or the names of individuals involved, thereby preserving the anonymity of all parties.

Responsible data usage is a key element of this project. The emotional sequences used as inputs to the Emoti-Shing model have been handled with the utmost care and respect for the participants' privacy. For example, the model analyzes the emotional content of phone call audio conversations without storing or transcribing the actual conversations. Additionally, the emotional data is aggregated in an anonymous manner, preventing the identification of individual participants. The model's outputs, such as the predicted vulnerability state, will be used solely for the intended research purposes and not for any unintended or harmful applications.

Furthermore, any software utilizing this model will be required to obtain the explicit consent of the phone owner before execution, ensuring the respect for individual privacy and autonomy. The research team has also taken proactive measures to protect participants' privacy, such as limiting access to the data to only authorized personnel involved in the study.

Moreover, the research findings is reported in a manner that protects individual privacy and avoids any potential misuse or misinterpretation of the developed techniques. For instance, the predicted vulnerability state is reported without identifying individual information (names, gender, etc.).

By placing data protection and ethical considerations at the heart of the development and deployment of the Emoti-Shing model, this research aims to ad-

vance the field of vishing detection while upholding the standards of participant protection and responsible technology development.

4.2. Problem Statement

Given two parties, $P1$ and $P2$ in a conversation, consisting of a sequence of utterances $u_1u_2\cdots u_n$ with constituent emotions $E_1E_2\cdots E_n$. Given the vulnerability states of a potential victim to a scam call, $V_1V_2\cdots V_k$. $n,k\in\mathbf{N}$. We aim at:

1) Proposing a Hidden Markov Model, λ that models the vulnerability states $V_1V_2\cdots V_k$ as hidden states with emotions sequence $e_1e_2\cdots e_n$ as observations or outcomes.

2) Validating the model λ by generating a sequence of emotions, $e_1e_2\cdots e_n$ randomly via the model.

3) Decoding the most likely sequence of states, $V_1V_2\cdots V_p$ the model λ can produce, given a sequence of observations $e_1e_2\cdots e_n$.

4.3. Model Formulation

A social engineering attack involves four main stages: research and information collection, relationship development, attack execution and exit [2]. However, only the relationship development and the attack execution stages require that the attacker should be actively connected with the victim in a conversation. Scam detection of vishing can really be vital during the conversation between the attacker and his potential victim. Our model therefore comes in at the level of the conversation between the attacker and the victim. So we shall consider the second and the third stages of a social engineering attack to define three vulnerability states of a potential victim to a vishing attack. Once the attacker gathers enough information about the victim, the second stage of social engineering consists of establishing a relationship of trust with the victim, a state the victim believes that the attacker is the person the attacker claims to be. From social engineering literature [1] [2] [29], and from a history of recorded scam calls, we can say that prior to this state of trust, as the potential victim picks the call, the victim is in an initial state, say V_1. In this state, the attacker prepares his ground by inciting very neutral emotions in the victim [1], and then he activates other emotion like excitement and content to move the victim to the next state of vulnerability, a state of trust which we shall consider as V_2. If the attacker activates negative emotions in V_1, there are high chances that he does not move to the V_2. This can occur when the victim gets angry and hangs up the call. In V_2 the attacker has succeeded to make the victim belief that the attacker is the person he claims to be. The victim is therefore ready to listen to the well crafted and meaningful pretext (story/scenario) being told by the attacker. As the attacker convinces the victim with his attack scenario, he gradually moves the victim into the highest level of vulnerability in our model, V_3. In state V_2, if the attacker fails to activate highly positive emotions in the victim, there are high chances that the victim gets back to the first state, V_1, and the attack process starts over or fails. If

the attacker maintains the victim in the vulnerability state V_3, the attacker has high chances of getting the victim do anything the attacker wants at anytime, making the attack to succeed. We also model the transition from V_1 to V_3 where the attacker convinces the potential victim to believe that he is the person he claims to be, and request the victim to act (give out sensitive information) directly without formulating a pretext. For example, the victim takes a work leave, the attacker presents himself as the victim's boss, and requests for the password to his computer. The victim has the tendency of just obeying. The victim therefore has the three states of vulnerabilities as outlined in Table 2.

Table 2. The three vulnerability states of a potential victim.

State	Symbol
Indifferent to attacker	V_1
Trust the attacker	V_2
Ready to act	V_3

William Sargant, a controversial psychiatrist and author of the book entitled "Battle for the Mind" [50], talks about the methods by which people are manipulated. According to Sargant, various types of beliefs can be implanted in people after the target has been disturbed by fear, anger, or excitement. These feelings cause heightened suggestibility and impaired judgment. A social engineer can use this device to their advantage by offering the target a suggestion that causes fear or excitement and then offering a solution that turns into a suggestion. We consider the following biological universal emotions that are generally manipulated by scammers during a scam conversation at the different stages of the scam: anger, fear, neutral, excitement.

- **Anger:** Anger is a strong negative emotion that social engineers try to avoid as much as possible during a scam. When the attacker meets the potential victim in angry emotion, he begins by neutralising this anger. He highly activates the neutral emotion in order to maintain the conversation. An attacker would avoid getting the potential victim angry at all cost throughout the attack, except for very few cases of targeted anger. For this reason, we shall assign a relatively low probability to anger in our model, for we aim at observing the situation that likely leads to a scam.
- **Fear:** Fear is one of the emotions manipulated by social engineers in their victims throughout their attacks. At the early stage of a social engineering attack, the social engineer avoids creating fear in the target as much as possible. For a feeling of fear at this stage may cause the victim to hang the call. So the emotion of fear is given a very low probability at the first level of vulnerability in our model. As the attack process progresses, the social engineer may design a pretext that aims at creating an environment of fear, and then proposes solutions that require the victim to act in a way beneficial to the attacker. So a high emotional response of fear at the third level of vulnerability

in our model is an indication for high chances of a successful scam.

- **Excitement:** Excitement is an emotion that falls under the category of happy emotion. Social engineers seek to get their victim excited at every stage of the attack. When the victim is excited about the imposter the social engineer claims to be, the victim can easily be moved from V_1 to V_2 in our model. While in V_2, getting an exciting pretext pulls the victim to V_3 as his excitement take over his cognition, preparing the victim to do whatever the attacker wants. The victim can be excited over financial gains promised by the scammer, or over a better/dream life promised by the attacker. In our model we place this emotion together with greed, one of the emotions targeted by the attackers. So to initialise our model, we need to give a relatively high probability to observing excited emotions in order to succeed in getting the victim vulnerable.
- **Neutral:** A neutral emotion, is considered to be the absence of positive and negative emotions by some literature [51]. Social engineers apply the concept of emotional neutrality to remove greed, fear and other human emotions that may prevent the attack progress from progressing. In the early stage of our model, the social engineer prepares the victim by neutralizing any emotion the victim is in, especially negative emotions that may cause the victim to hang up the call. So for the attack to succeed, our model should have high chances to emit a neutral emotion when emotions of anger are observed at the very early stages of the attack. This emotion gradually reduces as the attack progresses as other emotions are gradually activated by the attacker.

The four emotions presented so far have been repeatedly used by researches to highlight the role of emotions as a powerful fraud tool. However, other emotions that contribute to fraud may have more or less effect as the above four. We shall use the following symbols shown in Table 3 to denote the observation emotions used in the model.

Table 3. The four observation emotions in the model.

Emotion	Symbol
neutral	e_n
anger	e_a
fear	e_f
Excitement	e_e

The vulnerability states of the victim can be observed through the emotions the victim emits as the victim is manipulated by the attacker. We therefore propose the following Hidden Markov Model, that considers the vulnerability states of the victim as the hidden states and the emitted emotions as observations. The labels on the arrows indicate the transition and emission probabilities as show in Figure 12 with:

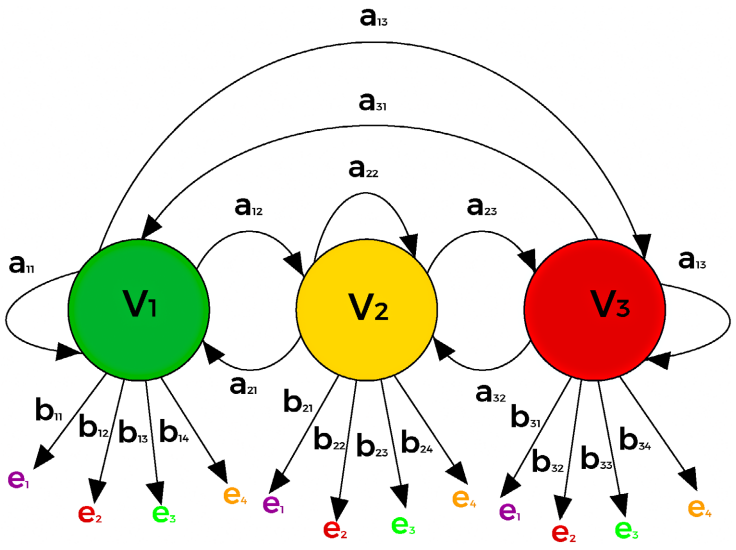

Figure 12. The proposed model.

a_{ij} being the probability of transiting from the hidden state V_i to V_j.

b_{lm} being the probability that we are at the hidden state V_l, observing the emotion e_m.

The model can therefore be defined by:

$$\lambda = (A, B, \pi) \tag{13}$$

where:

$$A = \left[a_{ij}\right] \quad \text{for } 1 \le i \le 3, 1 \le j \le 3$$

is the transition probability matrix

$$B = \left[b_{lm}\right] \quad \text{for } 1 \le l \le 3, 1 \le m \le 4$$

is the emission probability matrix

$$\pi = \pi_1$$

is the initial probability over the state V_1.

4.4. Determining the Values of the Transition

From the works in psychology, we observe the following for the success of a vishing scam. At the first level of our model, V_1 the victim is at the lowest level of vulnerability with regards to the attack. The attacker is considered to have initiated the call, with the potential victim picking the call in an emotional state that is independent of the attack. That is, the victim's mood before the call. Here, the attacker's goal is to move the victim from this state, V_1 to state V_2. However, there are some situations that require the victim's state to transit from V_1 to state V_3. The transition from the state V_1 to state V_1 keeps the victim in the same state as before. So the initial probability distributions of our model at this state should satisfy the expression:

$$a_{11} < a_{13} < a_{12} \tag{14}$$

Applying Binmore's method [52] in the interval [1; 2], Where a_{ij} is the probability that the victim leaves from the vulnerability state V_i to the vulnerability state V_j, with $1 \le i \le 3$, $1 \le j \le 3$.

$$a_{11} < a_{13} < a_{12} \Rightarrow 1 < \frac{1+2}{2} < 2$$

$$= 1 < \frac{3}{2} < 2$$

This gives us weighted values $\beta\left(V_{1j}\right)$ that reflect a stable arrangement:

$$\begin{cases} \beta\left(V_{11}\right) = 1 \\ \beta\left(V_{13}\right) = 1.5 \\ \beta\left(V_{12}\right) = 2 \end{cases} \tag{15}$$

To scale the $\beta\left(V_{1j}\right)$ to probability distributions, we apply the formula:

$$p\left(V_{1j_k}\right) = \frac{\beta\left(V_{1j_k}\right)}{\sum_{j=1}^{3} \beta_{V_{1j}}} \qquad k \in 1,2,3 \tag{16}$$

his gives us the following results:

$$\begin{cases} a_{11} = p\left(V_{11}\right) = \dfrac{2}{9} \\ a_{12} = p\left(V_{12}\right) = \dfrac{1}{3} \\ a_{13} = p\left(V_{13}\right) = \dfrac{4}{9} \end{cases} \qquad \text{where } \sum_{k=1}^{3} p\left(V_{1k}\right) = 1 \tag{17}$$

Once in V_2, the attacker aims at moving the victim from V_2 to V_3. He first of all has to maintain the victim at the state V_2 as he develops his pretext; V_2 to V_2. Should he fail in his pretexting, he has the high chances of moving the victim from V_2 back to V_1. So for the attack to progress from state V_2 to V_3 the following probability equation should be satisfied.

$$a_{21} < a_{22} < a_{23} \tag{18}$$

where a_{ij} is the probability that the victim leaves from the vulnerability state V_i to the vulnerability state V_j, with $1 \le i \le 3$, $1 \le j \le 3$.

Similarly as with the transitions from V_1 above, we obtain the following probabilities after applying Binmore's method and the formula (16) above:

$$\begin{cases} a_{21} = p\left(V_{21}\right) = \dfrac{2}{9} \\ a_{22} = p\left(V_{22}\right) = \dfrac{1}{3} \\ a_{23} = p\left(V_{23}\right) = \dfrac{4}{9} \end{cases} \qquad \text{where } \sum_{k=1}^{3} p\left(V_{1k}\right) = 1 \tag{19}$$

At V_3, the attacker wants to maintain the victim at V_3 until the victim does what he wants. So we have:

$$a_{31} < a_{32} < a_{33} \tag{20}$$

In a similar way, we have

$$\begin{cases} a_{31} = p(V_{31}) = \dfrac{2}{9} \\ a_{32} = p(V_{32}) = \dfrac{1}{3} \\ a_{33} = p(V_{33}) = \dfrac{4}{9} \end{cases} \quad \text{where } \sum_{k=1}^{3} p(V_{1k}) = 1 \tag{21}$$

The transition probability matrix is therefore given by Table 4.

Table 4. Model initial transition probabilities.

	V_1	V_2	V_3
V_1	0.22	0.44	0.33
V_2	0.22	0.33	0.44
V_3	0.22	0.33	0.44

4.5. Determining the Values of the Emission Probabilities

Let E_{lm} be the event that we are at the hidden state V_l, observing the emotion e_m. b_{1m} be the probability that we are at the hidden state V_l, observing the emotion e_m. With and $l \in \{1,2,3\}$ and $m \in \{\text{neutral}, \text{anger}, \text{fear}, \text{excitement}\}$ having n = neutral, a = anger, f = fear, e = excitement.

For example, E_{1n} is the event that we are at the hidden state V_1, observing the emotion neutral. At each state of vulnerability, the attacker seeks to activate specific emotions to guarantee the success of his attack. Also, activating some emotions at each state might play negatively on the attack.

At V_1, the attacker avoids getting the potential victim angry or afraid. For these emotions can make the victim hang-up the call. His goal is to activate the neutral emotion and as he prepares his grounds to activate the emotions he needs for the success of attack. Getting the victim excited contributes positively to get the attacker succeed at this state. This analysis enables us to formulate the following well ordered relationship.

$$E_{1a} < E_{1f} < E_{1n} < E_{1e} \tag{22}$$

Applying Binmore's method in the interval [1; 2],

$$E_{1a} < E_{1f} < E_{1n} < E_{1e} \tag{23}$$

$$\Rightarrow 1 < \frac{1+2}{2} < \frac{\frac{1+2}{2}+2}{2} < 2 = 1 < \frac{3}{2} < \frac{7}{4} < 2 \tag{24}$$

This gives us weighted values $\beta(E_{1j})$ that reflect a stable arrangement:

$$\begin{cases} \beta(E_{1a}) = 1 \\ \beta(E_{1f}) = 1.5 \\ \beta(E_{1n}) = 1.75 \\ \beta(E_{1e}) = 2 \end{cases} \tag{25}$$

To scale the $\beta(b_{1m})$ to probability distributions, we apply the formula: (16) for $k \in 1,2,3$. This gives us the following results:

$$\begin{cases} b_{1a} = p(E_{1a}) = \dfrac{4}{25} \\ b_{1f} = p(E_{1f}) = \dfrac{6}{25} \\ b_{1n} = p(E_{1n}) = \dfrac{7}{25} \\ b_{1e} = p(E_{1e}) = \dfrac{8}{25} \end{cases} \tag{26}$$

At V_2, the attacker seeks to get the victim excited by creating an environment of confidence that can enable the victim to listen to his pretext. Getting the victim angry at this state is not favourable. A neutral emotion is activated to counter any event of anger that may arise, as the emotion of fear may be an indicator that the attacker's—pretext seeks to create an atmosphere of fear to get the attack successful. This gives us the well ordered relationship below:

$$E_{2a} < E_{2n} < E_{21f} < E_{2e} \tag{27}$$

In a similar way, the application of equations (24) and (16) above gives us

$$\begin{cases} b_{2a} = p(E_{2a}) = \dfrac{4}{25} \\ b_{2f} = p(E_{2f}) = \dfrac{6}{25} \\ b_{2n} = p(E_{2n}) = \dfrac{7}{25} \\ b_{2e} = p(E_{2e}) = \dfrac{8}{25} \end{cases} \tag{28}$$

At V_3, the attacker avoids getting the potential victim angry as much as possible, neutralising any attempt that leads to anger. The attacker seeks to activate emotions that can maintain the potential victim in this state and get the victim do what the attacker wants. There are two possible scenarios:

◦ The attacker aims at creating an environment of fear in the victim to get the victim do what the attacker wants. In this case, we have:

$$E_{3a} < E_{3f} < E_{3n} < E_{3e} \tag{29}$$

◦ The attacker promises some gains or benefits and gets the potential victim excited.

$$E_{3a} < E_{3f} < E_{3n} < E_{3e} \tag{30}$$

For the model to keep track of the nature of the scam pretext, we apply Binmore's method once more, considering the emotions of fear and excitement to have the same probability values.

$$E_{3a} < E_{3f} < E_{3n} < E_{3e} \Rightarrow 1 < \frac{1+2}{2} < 2 < 2 = 1 < \frac{3}{2} < 2 < 2 \tag{31}$$

This gives us weighted values $\beta\left(E_{1j}\right)$ that reflect a stable arrangement:

$$\begin{cases} \beta\left(E_{1a}\right)=1 \\ \beta\left(E_{1f}\right)=1.5 \\ \beta\left(E_{1n}\right)=2 \\ \beta\left(E_{1e}\right)=2 \end{cases} \tag{32}$$

To scale the $\beta\left(b_{1m}\right)$ to probability distributions, we apply (28) $k \in 1,2,3$ having Table 5:

Table 5. Emission probability matrix for the model.

	a	*f*	*n*	*e*
V_1	0.28	0.16	0.24	0.32
V_2	0.28	0.16	0.24	0.32
V_3	0.23	0.15	0.31	0.31

Determining the values of the initial probability π

In the context of this model, we suppose that the victim shall always be at the vulnerable state V1. This can be achieved with the following initial probability vector:

$$\pi = (1.0, 0.0, 0.0) \tag{33}$$

5. Implementation and Simulation

5.1. Implementation Model

To implement this model, we use the R programming language. This is a language and a free software environment for statistical computing and graphics supported by the R Foundations for Statistical Computing. It is an integrated suite of software facilities for data manipulation, calculation and graphical display. R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS. We used the IDE for R, the RStudio in our work. It is available in two formats: RStudio Desktop is a regular desktop application while RStudio Server runs on a remote server and allows accessing RStudio using a web browser. We implement the transition probability matrix A, as shown in **Algorithm 1**.

Furthermore, we implement the emission probability matrix, *B*, as shown in **Algorithm 2**.

In Section 5, we have used the psychological principles of social engineering to propose a HMM that models the vulnerability states of a potential victim, observed via emitted emotions at each state of vulnerability. This transition and emission probabilities have been obtained by applying mathematical formulae on situations that contribute the success or failure of scams, as far as emotion

Algorithm 1 Pseudo code for Transition probability [2]

1: Start
2: Define the vulnerability states to V1, V2 and V3
3: **For** $i = 1$ **to** 3 **do**
4: Viprobs = the transition probabilities from Vi to other states
5: **End For**
6: Create a 3 x 3 transition probability matrix using the transition probability from each state.
7: Rename the rows of the matrix to states
8: Rename the columns of the matrix to states.
9: **End**

Algorithm 2 Pseudocode for the Implementation of the emission probability matrix for the model [2]

1: Start
2: Define the vulnerability states to neutral, n; angry, a, fear, f, excitement, e
3: **For** $i = 1$ **to** 3 **do**
4: Viemotions = the emission emotion probabilities from Vi to other states
5: **End For**
6: Create a 3 x 4 emission probability matrix using the emission probability from each state.
7: Rename the rows of the matrix to states
8: Rename the columns of the matrix to emotions.
9: **End**

dynamics are concerned. The initial probability is chosen such that the model shall always start at the vulnerability state V_1. We implemented our model in R programming language using the IDE Rstudio. Next, we simulate the model and analyse the results obtained.

5.2. Simulation and Results Analysis

The simulation concerns the HMM $\lambda = (A, B, \pi)$. We begin with the model validation, by simulating the generating power of our HMM in generating random sequences of emotions. We shall then use different sequences of emotions, favourable to the success of the scam (inspired from psychology and social engineering), and see the changes in the vulnerability states of the victim. In the same way, we shall use emotion sequences that are likely to fail a scam call and equally observe the changes in the hidden states on our model. We shall end this section with a discussion on the results obtained.

5.3. Generating a Random Sequence of Emotions with Our Model

We use **Algorithm 1** and **Algorithm 2** presented above to write a function, *sequence generator* that receives the model parameters, initial probability and the length of emotion sequence, to generate a random sequence of emotions based on the transition and emission probabilities upon an observation. Generating an emotion sequence of 30 emotions respectively a 45 emotions in one of the runs

gives the results in Table 6 and Table 7.

From Table 6 and Table 7, we can clearly notice the changes in the vulnerability states of the victim and the associated emotional dynamics observed due to these changes. The psychologic based principles used to develop the model are

Table 6. Sample output of a sequence of 30 emotions generated by our model.

Step/Position	Vulnerability State	Observed Emotion
	V_1	NEUTRAL
2	V_2	ANGRY
3	V_2	NEUTRAL
4	V_2	FEAR
5	V_3	EXCITEMENT
6	V_1	EXCITEMENT
7	V_1	NEUTRAL
8	V_2	NEUTRAL
9	V_3	EXCITEMENT
10	V_3	EXCITEMENT
11	V_1	EXCITEMENT
12	V_2	NEUTRAL
13	V_3	FEAR
14	V_1	EXCITEMENT
15	V_3	ANGRY
16	V_1	EXCITEMENT
17	V_3	EXCITEMENT
18	V_1	EXCITEMENT
19	V_2	EXCITEMENT
20	V_1	NEUTRAL
21	V_2	ANGRY
22	V_2	FEAR
23	V_3	EXCITEMENT
24	V_3	EXCITEMENT
25	V_3	EXCITEMENT
26	V_3	EXCITEMENT
27	V_3	FEAR
28	V_2	ANGRY
29	V_2	FEAR
30	V_3	EXCITEMENT

Table 7. Sample output of a sequence of 45 emotions generated by our model.

Step/Position V_1	Vulnerability State FEAR	Observed Emotion 1
	V_3	NEUTRAL
3	V_3	EXCITEMENT
4	V_3	EXCITEMENT
5	V_3	ANGER
6	V_2	NEUTRAL
7	V_3	EXCITEMENT
8	V_1	ANGER
9	V_2	NEUTRAL
10	V_1	EXCITEMENT
11	V_2	ANGER
12	V_3	EXCITEMENT
13	V_3	ANGER
14	V_1	EXCITEMENT
15	V_1	EXCITEMENT
16	V_2	FEAR
17	V_3	EXCITEMENT
18	V_1	FEAR
19	V_1	EXCITEMENT
20	V_3	FEAR
21	V_2	ANGER
22	V_3	FEAR
23	V_2	ANGER
24	V_3	FEAR
25	V_1	FEAR
26	V_2	EXCITEMENT
27	V_3	NEUTRAL
28	V_2	NEUTRAL
29	V_2	ANGER
30	V_2	FEAR
31	V_3	ANGER
32	V_3	FEAR
33	V_3	FEAR
34	V_3	NEUTRAL
35	V_3	FEAR

Continued

36	V_3	FEAR
37	V_3	ANGER
38	V_3	FEAR
39	V_3	FEAR
40	V_3	NEUTRAL
41	V_2	FEAR
42	V_2	FEAR
43	V_3	EXCITEMENT
44	V_3	EXCITEMENT
45	V_3	EXCITEMENT

reflected in these changes. For instance, in Table 6, when the scammer observes that the potential victim is angry (at position 2) the scammer neutralizes this emotion in the same state, V_2 and creates an atmosphere that moves the victim to V_3 (excitement). Generally, the emotions that highly favour the success of vishing scams are seen to have a high occurrence at the highest level of vulnerability, V_3. The emotion of anger, that plays negatively to the success of a scam moves the victim from a higher level of vulnerability to the lower level (e.g. position 15 and 16) or keeps the victim at the same level. We can therefore affirm that the model generate sequences of emotions that reflect the internal changes in the hidden vulnerability states of the potential victim. From the above analysis, we can see clearly that the model is a reflection of the psychological principle of social engineering, specifically vishing scam.

5.4. Determining the Vulnerability State of the Victim That Most Probably Emits a Sequence of Emotions

Here, we implement the Viterbi algorithm described in 3.3.4.2).

5.5. Testing with a Random Sequence of Emotions Generated

We begin by invoking this function with the emotion sequences generated in section 5.3 by the model. The most probable states, that produced the various emotion sequences in the sequence of 30 emotions generated, are illustrated in Figure 13.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Emotion Sequence	n	n	n	n	e	n	n	e	e	e	e	e	e	e	e	e	f	f	f	f	f	f	f	f	e	e	e	e	e	e
Most Probable State	V1					V2		V3																						

Figure 13. The most probable states that produced the generated random sequence of emotions.

From the results, we note that the potential victim to call (represented by these sequences of emotions observed in a potential victim) gets into the conversation with neutral emotions, observed in the first four utterances he or she makes. This keeps the victim in the vulnerability state, V_1. The potential victim suddenly becomes excited, and moves into the next vulnerability state, V_2. Here, the victim become neutral and later gets highly excited and moves to the highest level of vulnerability, V_3, where he stays till the end of the conversation. The results reveal that the potential victim is most probably in the highest level of vulnerability. This can be explained from the discussions above, as we saw that most of the emotions generated here are emotions that favour the success of the vishing scam.

5.6. Testing with Emotion Sequences Based on Principles of Psychology

We consider a situation where the victim is angry throughout the conversation. When we run this with 30 sequences of emotions, we obtain the results in Figure 14.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Emotion Sequence	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
Most Probable State	V1																													

Figure 14. The most probable states that produced a sequence of 30 emotions of anger.

From the results, we note that the potential victim to the simulated sequence of emotions corresponding to a scam call gets into the conversation with an angry emotion and stays in this emotional state throughout the conversation. This is to simulate the principle of psychology and social engineering that holds that the anger emotion should be avoided as much as possible to succeed in social engineering. Here, the victim remains in an indifferent vulnerability state, V_1 throughout the conversation. This shows that no vishing scam can be successful if the victim remains angry throughout the conversation. This is true in real life, since the social engineer avoids getting the potential victim angry as much as possible. Next we consider a sequence of emotions that reflects the principles of social engineering, with the emotion sequence consisting of emotions that a social engineer should activate at each stage of a social engineering attack to guarantee the success of the attack. Here the potential victim is not angry at all. He begins with a series of neutral emotions, gets excited, develops fear and later on gets excited as shown in Figure 15.

We notice that the model takes the victim through the vulnerability states V_1 V_2 V_3, back to V_2 and terminates in V. It is possible that the scammer generates an atmosphere of confidence (where the victim accepts the scammer to be the

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Emotion Sequence	n	a	n	f	e	e	n	n	e	e	e	n	f	e	a	e	e	e	e	n	a	f	e	e	e	e	f	a	f	e
Most Probable State	V1		V2		V3																									

Figure 15. The most probable states that produced a sequence of emotions that reflects a scam scenario.

person he pretends to be). Then the scammer presents a pretext that creates an atmosphere of fear in the victim, and reassures the victim that he got the solution to the presented pretext. This scenario, observed in real life has been confirmed by the simulation to be have high chances of leading to a successful scam as the victim is mostly in the highest level of vulnerability. The next simulation concerns a scene where the victim displays a sequence of neutral emotions. The intention is to investigate the situation where the potential victim is aware of the fact that this is possibly a scam call and may only want to play around with the scammer. The potential victim is therefore in a neutral emotional state since his emotions are not influenced by the manipulations of the scammer shown in Figure 16.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Emotion Sequence	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
Most Probable State	V1																													

Figure 16. The most probable states that produced a sequence of neutral emotions.

We notice here that the victim moves to the second level of vulnerability and remains there throughout the conversation. This confirms our situation under investigation, where the potential victim is aware of the fact that it is a scam conversation, and remains indifferent throughout the conversation. From the above scenario, we see clearly that the model reflects the reality of social engineering scams in general and vishing scams in particular.

5.7. Case Study

The caller is a scammer. He basically presents a neutral emotion throughout the conversation up to the point where he actually gets angry and insults the scammer off. Table 8 shows the extract of their conversation. We listened to the conversation and manually attributed corresponding emotions to utterances.

We further isolated the utterances of the potential victim and simulated their associated emotions as a sequence in the model. The results obtained are shown in Figure 17.

We notice that the potential victim begins in at V_1, and as he pretends to

Table 8. Case study of the real scam conversation between a scammer and a potential victim.

SN	Scammer	Potential Victim
1	Allo	
2		Allo (n)
3	Yeah Mtn customer service	
4	Good afternoon Sir	
5		Yeah good afternoon (n)
6	You are a loyal customer to all our mtn services, we wish to enquire from you Sir, do you use our MTN mobile money service?	
7		Yeah I do (n)
8	Do you face any difficulties using your MTN mobile service Sir?	
9		No (n)
10	Ok Sir I wish to inform you that we have updated the mtn mobile service, ok	
11		Hmmm (n)
12	Instead of the 5 digit code you were using before, we have updated it into 8 digits for security reasons OK	
13		Yes (n)
14	And once your account has been updated Sir, any deposit you make into the account, you will have an interest rate of 5 percent directly into your account and we have also reduced the withdrawal fee down?	
15		Ok (n)
16	Yeah, I'm called Daniel, I am calling from the main service Akwa Douala. I will be the one to give you directives on how to update your account.	
17		I will be waiting (n)
18	So do you have any question about our mobile money service before we proceed?	
19		I don't. there have been no problems there (n)
20	So I am just going to send a 4 digit confirmation code to you, through an SMS, then from there?	
21		I hear you, I hear you (n)
22	Do you prefer to create a new code or do you wish to add a digit to the one you were using before?	
23		Yeah just to add a digit to the code I was using before (n)
24	Ok Sir, and which digit do you wish to add?	
25		Naut (n)
26	Zero (excited)	
27		Naut Just naut (n)

Continued

28	Ok Sir and what is the former code you were using before?		
29			Why do you want to know my code? (a)
30		Just to	
31			Get out (a)
32			Get out (a)

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Emotion Sequence	n	n	n	n	n	n	n	n	n	n	n	n	n	a	a	a
Most Probable State	V1	V2												V3	V1	

Figure 17. Case study.

acknowledge the scammer to be a trust worthy person that he can listen to, the model moves him to the state V_2. He stays in this state in the greater part of the conversation (from his 2nd to 13th utterance). This is because he is conscious of the fact that the caller is a scammer. So he does not activate any other emotion at this state. He keeps his calm (neutral) perhaps to see to what extend the scammer may want to go. And when the scammer finally asks for his secret code, he becomes very angry and ends the conversation in this state as the scammer goes on to insult him. This case study reflects reality as we see clearly that the victims remains in the state V_2 throughout the conversation and did not activate any emotion that has a high probability of luring the potential victim to act.

5.8. Limitations

The main challenge in this work is to get a sequence of emotion that reflects a true scam. We based our analysis on results from psychology and emotional theories to define the transition probabilities and emission probabilities in this model. Obtaining these probabilities from real world data could give more credit to the model. Also, in the experimentation, the emotional sequences used were theoretical. We based the analysis on psychology principles and a few recorded scam calls (where the victim is already conscious that it is a scam), to come out with the emotions that are likely to contribute to the success of a vishing scam and those that are likely to contribute to a failure. Further studies may consider real life unbiased scam records, and extract the emotion sequences of the victim that intervene in the calls for our experimentation.

6. Conclusion and Perspectives

Mastering the activities of scammers remains a challenge in a society characterised by scamming. These scammers use different pretexts and strategies to scam

different victims. As the victims become aware of one strategy, the scammers device another strategy to get the victim do what the scammers want. Inciting research in the direction that explores aspects of a scam that cannot be masked by the scammer can go a long way to reinforce existing measures to track scams in our society. Scamming is a social engineering attack that involves the manipulation of the victim to do what the scammer wants. Researchers in psychology have proven that emotional arousal is a powerful fraud tool. In this work, we explored the possibilities of tracking scam through human emotional changes observed in the victim. We have used the stages of a social engineering attack in general to define three vulnerability states of a potential victim to a scam call. These vulnerabilities are observed through emotions emitted by the victim at each stage during a conversation. With the increase attention given to research in emotion detection from speech, the emotions associated to each utterance of a conversation speech can be tracked and the dynamics in these emotions used to predict a scam. The proposed model shows the changes in the vulnerability states of a potential victim to a scam call observed through emitted emotions. We simulate this model with real life psychological scam principles and the results obtained reflects reality. The main contribution is to incite research on emotions dynamics towards scam detection. The proposed model shows that it is possible to track the changes in vulnerability states of a potential victim, and say if the conversation he/she is involved in is likely to be a scam.

As future work, it could be interesting to obtain the emission probabilities and the transition probabilities from real world scam data. Moreover, we intend to continue research in this area as follows:

- **Optimization of the Model:** Consider a hybrid approach by combining the Hidden Markov Model with other machine learning techniques (neural networks, random forests, etc.) to leverage the strengths of each model.
- **Expansion of the Data Set:** To enhance the robustness and generalizability of the model, researchers could explore the expansion of the data set by including a larger and more diverse set of emotion sequences from various social engineering scenarios during phone conversations.
- **Integration of Multimodal Data:** Combining emotional data with other modalities (such as facial expressions, pitch variations, rhythm and timing, speaking rate and fluency) could provide a more comprehensive and accurate approach to vishing scam detection. Fusing multiple biological characteristics through multimodal fusion may lead to improved performance compared to relying on a single modality.

Therefore, exploring these avenues for model optimization, data set expansion, and multimodal integration, the Emoti-Shing model can be further enhanced to deliver more robust and reliable vishing scam detection capabilities. Leveraging the strengths of various machine learning techniques and incorporating a diverse range of biological signals can contribute to the development of a more comprehensive and effective solution for protecting individuals against

vishing attacks.

Acknowledgements

The authors thank the anonymous reviewers for their valuable suggestions.

Disclosure Statement

This study was conducted with the sole purpose of advancing knowledge in the field of victim-offender overlap and to contribute to the development of effective interventions for victims and offenders.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Mitnick, K.D. and Simon, W.L. (2003) The Art of Deception: Controlling the Human Element of Security. John Wiley & Sons.

[2] Salahdine, F. and Kaabouch, N. (2019) Social Engineering Attacks: A Survey. *Future Internet*, **11**, Article 89. https://doi.org/10.3390/fi11040089

[3] Yeboah-Boateng, E.O. and Amanor, P.M. (2014) Phishing, SMiShing & Vishing: An Assessment of Threats against Mobile Devices. *Journal of Emerging Trends in Computing and Information Sciences*, **5**, 297-307.

[4] ANTIC (2020) Cameroon Lost XAF6 Bln to Bank Frauds in 2019 (ANTIC).

[5] Hilas, C.S., Kazarlis, S.A., Rekanos, I.T. and Mastorocostas, P.A. (2014) A Genetic Programming Approach to Telecommunications Fraud Detection and Classification. *International Conference on Circuits*, *Systems*, *Signal Processing*, *Communications and Computers*, Venice, 29 September-1 October 2014, 77-83.

[6] Lansley, M., Polatidis, N., Kapetanakis, S., Amin, K., Samakovitis, G. and Petridis, M. (2019) Seen the Villains: Detecting Social Engineering Attacks Using Case-Based Reasoning and Deep Learning. *Proceedings of the ICCBR Workshops*, Otzenhausen, 8-12 September 2019, 39-48.

[7] Nicholas, J. (2020) Types of Social Engineering Attacks: Detecting the Latest Scams. BioCatch.

[8] Olszewski, D. (2014) Fraud Detection Using Self-Organizing Map Visualizing the User Profiles. *Knowledge-Based Systems*, **70**, 324-334. https://doi.org/10.1016/j.knosys.2014.07.008

[9] Ko, M.M. and Su Thwin, M.M. (2015) Anomalous Behavior Detection in Mobile Network. *Genetic and Evolutionary Computing*, Yangon, 26-28 August 2015, 147-155. https://doi.org/10.1007/978-3-319-23207-2_15

[10] Hilas, C.S., Mastorocostas, P.A. and Rekanos, I.T. (2015) Clustering of Telecommunications User Profiles for Fraud Detection and Security Enhancement in Large Corporate Networks: A Case Study. *Applied Mathematics & Information Sciences*, **9**, 1709-1718.

[11] Subudhi, S. and Panigrahi, S. (2016) Use of Fuzzy Clustering and Support Vector Machine for Detecting Fraud in Mobile Telecommunication Networks. *International Journal of Security and Networks*, **11**, 3-11.

https://doi.org/10.1504/ijsn.2016.075069

[12] Subudhi, S. and Panigrahi, S. (2018) A Hybrid Mobile Call Fraud Detection Model Using Optimized Fuzzy C-Means Clustering and Group Method of Data Handling-Based Network. *Vietnam Journal of Computer Science*, **5**, 205-217. https://doi.org/10.1007/s40595-018-0116-x

[13] Lansley, M., Mouton, F., Kapetanakis, S. and Polatidis, N. (2020) SEADer++: Social Engineering Attack Detection in Online Environments Using Machine Learning. *Journal of Information and Telecommunication*, **4**, 346-362. https://doi.org/10.1080/24751839.2020.1747001

[14] Hoelscher, D.M., Ranjit, N. and Pérez, A. (2017) Surveillance Systems to Track and Evaluate Obesity Prevention Efforts. *Annual Review of Public Health*, **38**, 187-214. https://doi.org/10.1146/annurev-publhealth-031816-044537

[15] Dasgupta, P.B. (2017) Detection and Analysis of Human Emotions through Voice and Speech Pattern Processing. arXiv preprint arXiv:1710.10198.

[16] Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A. and Cambria, E. (2019) Dialoguernn: An Attentive RNN for Emotion Detection in Conversations. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 6818-6825. https://doi.org/10.1609/aaai.v33i01.33016818

[17] Liu, F. and Maitlis, S. (2013) Emotional Dynamics and Strategizing Processes: A Study of Strategic Conversations in Top Team Meetings. *Journal of Management Studies*, **51**, 202-234. https://doi.org/10.1111/j.1467-6486.2012.01087.x

[18] Lee, S., Lovelace, K.J. and Manz, C.C. (2014) Serving with Spirit: An Integrative Model of Workplace Spirituality within Service Organizations. *Journal of Management, Spirituality & Religion*, **11**, 45-64. https://doi.org/10.1080/14766086.2013.801023

[19] Lerner, J.S. and Keltner, D. (2001) Fear, Anger, and Risk. *Journal of Personality and Social Psychology*, **81**, 146-159. https://doi.org/10.1037//0022-3514.81.1.146

[20] Levenson, R.W. (1999) The Intrapersonal Functions of Emotion. *Cognition & Emotion*, **13**, 481-504. https://doi.org/10.1080/026999399379159

[21] Johnson-Laird, P.N. and Oatley, K. (1992) Basic Emotions, Rationality, and Folk Theory. *Cognition and Emotion*, **6**, 201-223. https://doi.org/10.1080/02699939208411069

[22] Hadnagy, C. (2010) Social Engineering: The Art of Human Hacking. John Wiley & Sons.

[23] Edwards, M., Larson, R., Green, B., Rashid, A. and Baron, A. (2017) Panning for Gold: Automatically Analysing Online Social Engineering Attack Surfaces. *Computers & Security*, **69**, 18-34. https://doi.org/10.1016/j.cose.2016.12.013

[24] Kalniņš, R., Puriņš, J. and Alksnis, G. (2017) Security Evaluation of Wireless Network Access Points. *Applied Computer Systems*, **21**, 38-45. https://doi.org/10.1515/acss-2017-0005

[25] Lunceford, B.L. (2006) Democracy and the Hacker Movement: Information Technologies and Political Action. The Pennsylvania State University.

[26] Myers, C.D. and Tingley, D. (2016) The Influence of Emotion on Trust. *Political Analysis*, **24**, 492-500. https://doi.org/10.1093/pan/mpw026

[27] Gupta, S., Singhal, A. and Kapoor, A. (2016) A Literature Survey on Social Engineering Attacks: Phishing Attack. 2016 *International Conference on Computing, Communication and Automation* (*ICCCA*), Greater Noida, 29-30 April 2016, 537-540. https://doi.org/10.1109/ccaa.2016.7813778

[28] Rabiner, L.R. (1989) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, **77**, 257-286. https://doi.org/10.1109/5.18626

[29] Thornburgh, T. (2004) Social Engineering. *Proceedings of the* 1*st Annual Conference on Information Security Curriculum Development*, Kennesaw, 8 October 2004, 133-135. https://doi.org/10.1145/1059524.1059554

[30] Choi, K., Lee, J. and Chun, Y. (2017) Voice Phishing Fraud and Its Modus Operandi. *Security Journal*, **30**, 454-466. https://doi.org/10.1057/sj.2014.49

[31] van Gorp, T. and Adams, E. (2012) Why Design for Emotion? In: van Gorp, T. and Adams, E., Eds., *Design for Emotion*, Elsevier, 1-18. https://doi.org/10.1016/b978-0-12-386531-1.00001-6

[32] Plutchik (2019) Motivation and Emotion/Book/2014/Plutchik's Wheel of Emotions. https://en.wikiversity.org/wiki/Motivation_and_emotion/Book/2014/Plutchik%27s_wheel_of_emotions

[33] Lee, K., Caverlee, J. and Webb, S. (2010) Uncovering Social Spammers: Social Honeypots + Machine Learning. *Proceedings of the* 33*rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Geneva, 2010 July 19-23, 435-442.

[34] Bourgais, M., Taillandier, P., Vercouter, L. and Adam, C. (2018) Emotion Modeling in Social Simulation: A Survey. *Journal of Artificial Societies and Social Simulation*, **21**, Article 5. https://doi.org/10.18564/jasss.3681

[35] Zeelenberg, M., Nelissen, R.M.A., Breugelmans, S.M. and Pieters, R. (2008) On Emotion Specificity in Decision Making: Why Feeling Is for Doing. *Judgment and Decision Making*, **3**, 18-27. https://doi.org/10.1017/s1930297500000139

[36] Hatfield, E., Cacioppo, J.T. and Rapson, R.L. (1993) Emotional Contagion. *Current Directions in Psychological Science*, **2**, 96-100. https://doi.org/10.1111/1467-8721.ep10770953

[37] Kavanagh, D.J. and Bower, G.H. (1985) Mood and Self-Efficacy: Impact of Joy and Sadness on Perceived Capabilities. *Cognitive Therapy and Research*, **9**, 507-525. https://doi.org/10.1007/bf01173005

[38] Tähtinen, J. and Blois, K. (2011) The Involvement and Influence of Emotions in Problematic Business Relationships. *Industrial Marketing Management*, **40**, 907-918. https://doi.org/10.1016/j.indmarman.2011.06.030

[39] Schuller, B. and Batliner, A. (2013) Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. John Wiley & Sons.

[40] Chen, A. (2019) Why Companies Want to Mine the Secrets in Your Voice: Voices Are Highly Personal, Hard to Fake, and Contain Surprising Information about Our Mental Health and Behaviors. https://www.theverge.com/2019/3/14/18264458/voice-technology-speech-analysis-mental-health-risk-privacy

[41] Lindquist, K.A., Barrett, L.F., Bliss-Moreau, E. and Russell, J.A. (2006) Language and the Perception of Emotion. *Emotion*, **6**, 125-138. https://doi.org/10.1037/1528-3542.6.1.125

[42] Hazarika, D., Poria, S., Zadeh, A., Cambria, E., Morency, L. and Zimmermann, R. (2018) Conversational Memory Network for Emotion Recognition in Dyadic Dialogue Videos. *Proceedings of the* 2018 *Conference of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies*, *Volume* 1 (*Long Papers*), New Orleans, 1-6 June 2018, 2122-2132. https://doi.org/10.18653/v1/n18-1193

[43] Kircanski, K., Notthoff, N., DeLiema, M., Samanez-Larkin, G.R., Shadel, D., Mottola, G., *et al.* (2018) Emotional Arousal May Increase Susceptibility to Fraud in Older and Younger Adults. *Psychology and Aging*, **33**, 325-337. https://doi.org/10.1037/pag0000228

[44] Allen, K. (2017) Scammers Play on Older Adult Emotions, Study Finds. https://www.mcablab.science/news/2017/12/21/scammers-play-on-older-adults-emotions-study-finds

[45] Russel, J. (1980) Three Dimensions of Emotion. *Journal of Personality and Social Psychology*, **9**, 1161-1178.

[46] Cukier, W., Nesselroth, E. and Cody, S. (2007) Genre, Narrative and the "Nigerian Letter" in Electronic Mail. 2007 40*th Annual Hawaii International Conference on System Sciences* (*HICSS'07*), Waikoloa, 3-6 January 2007, 70. https://doi.org/10.1109/hicss.2007.238

[47] Pinquart, M. and Sörensen, S. (2003) Associations of Stressors and Uplifts of Caregiving with Caregiver Burden and Depressive Mood: A Meta-Analysis. *The Journals of Gerontology*: *Series B*, **58**, P112-P128. https://doi.org/10.1093/geronb/58.2.p112

[48] Iloga, S. (2019) Apprentissage a l'aide des modèles de markov cachés et applications. Université de Maroua.

[49] Li, X., Lee, T.S. and Liu, Y. (2011) Hybrid Generative-Discriminative Classification Using Posterior Divergence. *CVPR* 2011, Colorado Springs, 20-25 June 2011, 2713-2720. https://doi.org/10.1109/CVPR.2011.5995584

[50] Sargant, W. (1957) Battle for the Mind. Prabhat Prakashan.

[51] Halton, C. (2019) Emotional Neutrality: What It Is, How It Works, Example. https://www.investopedia.com/terms/e/emotional-neutrality.asp

[52] Kantzavelou, I. and Katsikas, S. (2010) A Game-Based Intrusion Detection Mechanism to Confront Internal Attackers. *Computers & Security*, **29**, 859-874. https://doi.org/10.1016/j.cose.2010.06.002