ISSN No:-2456-2165

Cybercriminal Behavior Model to Counteract Phishing in the Financial Ecosystem in Colombia

Ortiz Ruiz, E
Bogotá DC, Colombia

Abstract:- Nowadays, the anomalies found in the prevention of fraud are aimed at detection tools that exist in themarket of each one of the financial entities; many of them are focused onalgorithms predisposed for their use and implementation. However, some of them do not have an effective learning in the face of the needs of the phishing phenomenon as a crucial element in the face of the challenges of these sites in Colombia, this is a reason that does not allow their parameterization and and then prosecute criminal behaviour. The adaptive algorithm model allows detecting and investigating phishing from a tactical orientation that allows identifying these characteristic patterns.

Keywords: *Machine Learning, model, social engineering, artificial intelligence,cyber-research, OSINT.*

I. INTRODUCTION

Website spoofing has become one of thetechniques most used by cybercrime to be able to generate the massive and individual capture of information. One ofthe principal aspects that do not generate confidence on the entities, it isbased on the impersonation of the brandand the affectation of customer data when they interact with the banking ecosystem.

Several studies are guide on checking ofthe cost vs the impact caused by this kind of problems, however, many entities have opted for the implementation of detective tools against fraud as the possible solution to this problem. Considering the currentdynamics in the face of the increase in accesses generated to the client to digitize it, financial institutions have managed to guide the user against enrollment in the face of anti-fraud tools aimed at the detection and prevention of new techniques defined by cybercriminals. However, none of them facilitate the adaptation of a general model for the financial system, in relationto the identification of phishing patterns to counteract the main forms and actionscarried out by cybercriminals.

Some investigations in the United Kingdom (Cambridge University, 2012) have defined that the increase of the cost of cybercrime is due to the lack of tools and research elements necessary to counteract the types of the attacks; this is due to the shortage of effective tools for the development andidentification of possible new threats against current phenomena bycybercrime organizations.

- *Troublesome*

The booming trend of website spoofing is increasing, doubling the figure for the first four months of the year (APWG,2021), compared to 2020. The principal reason: the main modalities oriented in techniques like the BEC (Business Email Compromise) with losses exceeding $ 48,000 compared to Q3 of 2020.

The contact that cybercriminals have with users through social networks has also been the main axis in the construction of the vectors of phishing and spear phishing (targeted phishing).

	January	February	March
Number of unique phishing Web sites detected	245,771	158,898	207,208
Unique phishing email subjects	172,793	112,369	39,918
Number of brands targeted by phishing campaigns	430	407	465

Fig 1 Seguimiento orientado por APWG sobre lastécnicas usadas.

Figure 1 shows the main targeted elements (APWG, 2021) for the monthsof January, March and April, in relation to the increase detected: *Spoofed sites, Phishing related to emails and number of target brands of phishing campaigns.*

Most of these consequences are mixed and they are related to other techniquescoupled with web mail, therefore, cybercriminals use those attributes entered by each one of the financial entities at an individual level; however, the limitations of being able to counteract and mitigate fraud are limited to the use of certain analysis tools on the same individualized patterns.

An erroneous answer to the exercise of prosecuting fraud using external tools; Itconsists in that each one has a differentmodel focused on recognizing internal fraud and the different phishing techniques used for the internal dynamics of a Bank.

Circumstances vary when there are different characteristics that cannot be determined around the needs that a system must have that facilitates thecoordination and the orientation to reorient the typologies and schemes of phishing campaigns that are nowadays redirected to different users or clients.

Digital channels are the best scenarios for the cybercriminal to take advantage of the benefits executed by the users ofthe financial system; particularly in relation to the public access to banked websites.

Regarding the verification of these links, websites or spoofing mechanisms, they are reported by the victim to the financialinstitution and in other cases they are detected by the same entity with its digital surveillance tools.

Before writing this paper and being ableto focus results, it is necessary to intervene the types of supervised data that it requires against the patterns or anomalies observed by each of the financial entities; Also, in the same way,refocus the needs to be able to obtain the information that each one of these anti-fraud tools provides.

In this way, it is important to determine the typologies in which mitigation and containment measures are being carriedout in order to establish the orientationsderived from the model.

II. ANALYSIS FOR THE MEASUREMENT OF THECYBERCRIME

The cost of cybercrime has increased due to the ease that cybercriminals haveto be able to execute novel typologies and others already used in the same entities and ecosystems; based on this, it is important that the limitations continue to persist due to the lack of knowledge and the curve of complaints from clients who have captured financialdata from their financial institution.

A graph that exemplifies this dynamic has a behavior of interest that can be examined:

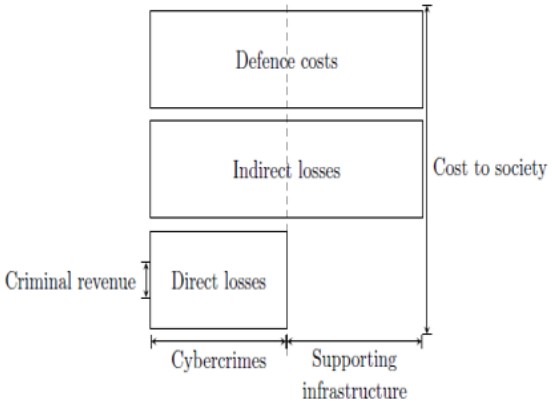

Fig 2: Tomada de (Universidad de Cambridge, 2012)pág. 3 relacionada con marco de trabajo del análisis delcosto del cibercrimen.

In 2012, the European Communications Commission subdivided the incrementalcost factors of cybercrime into four fundamental aspects:

- *The traditional form of cybercrimeor fraud committed by electronic or communication means.*
- *Posting illegal content and child sexual abuse.*
- *The exclusive crimes of electronic networks against information systems.*

There are two aspects to be highlightedin Figure 1: on the one hand, defense costs and indirect losses, both of which are related to the first quadrant in relation to the cost of society, which will have to bear these types of losses. On the other hand, there are the criminal revenues, a factor that involves direct losses, with the constant of the differentiated activities ofcyber crime.rente.

The reason for defining these three aspects and thus analyzing several aspects that involve much more specific characteristics is to elaborate joint actions in order to make the pursuit of cybercrime more efficient and to guide the gathering of valuable information to really contain the threat.

The criminal income as the generalized element that involves phishing supportsa very accurate theory before the currentdynamics of criminal organizations; oneof them based on the "trust" that the userhas at the time of performing routine activities before the bank. Factor that increases the exercise of being able to build an adaptable model to the fraud schemes that banks and financial institutions currently have. Os This is why the direct losses perceived by the Entity play a significant role in themoment when the client decides to withdraw from the service due to the"trust and credibility" indicator, which is why the theory of indirect losses from cybercrime gains strength.

This theory also has an impact on Gross Domestic Product (GDP), a factor that involves certain economic dynamics to reestablish confidence mechanisms in the financial ecosystem.staIn the same sense, the direct loss has to do with the damage suffered by the victim, which includes the loss of data inthe face of the established phenomenon, which impacts theindicator of loss of trust and credibility.

This same comparative mechanism of two indicators that seriously impact the financial ecosystem, translates into possible tools that facilitate adaptation to the different platforms that banks and financial institutions currently have; first starting with "brand or website impersonation", an aspect thatcontinues to be one of the headaches forsecurity and trust teams in traditionaland non-traditional banking users.

Within the model of the European cybercrime framework, there are other elements that are transversal to innovation and development of new elements that facilitate this exchange offactors in order to pursue and counteractfraud at the economic level. A studyconducted by (MIT, 2016) on how to renew the economy through disruptive ideas on innovation; involves several aspects taken into account in this article:

- The dramatic rise of technologies to mitigate or contain problems.
- The rise of emerging technologies to define essential market characteristics.
- Virtuality as an essential component indefining criteria for implementing newways to improve productivity.
- The combination of individual practices to promote a common good.
- The benefit of innovation in the face ofthe cost generated by external factors.

ISSN No:-2456-2165

These aspects imply several mechanisms to guide disruptive actions that involve the interaction of mechanisms that allow to effectively measure a problem that so far has not incurred any decrease in direct orindirect cost to the financial ecosystem.

III. ADAPTABILITY OF A MODEL IN THE CONTEXT OF CYBERSPACE

Many studies, link (Lukas Brenner, 2019), the importance of being able to determine learning models and understanding of the behaviors of the victims in the front of fraud, howeverthere is no tool that tangibilizes certain actions in favor of collaborative research to be able to determine the causes and origins.

In such a way that the technological developments defined by (Lukas Brenner, 2019), guarantee theunification and adaptability of a model tobe able to define the second stage of implementation to be able to determine the causes and origin of the main factorsthat originate fraud, in first measure are oriented to cyber fraud (Ortiz Ruiz, 2020), guidelines that suppose reliable inputs within a triangular analysis of behaviors, at a general level:

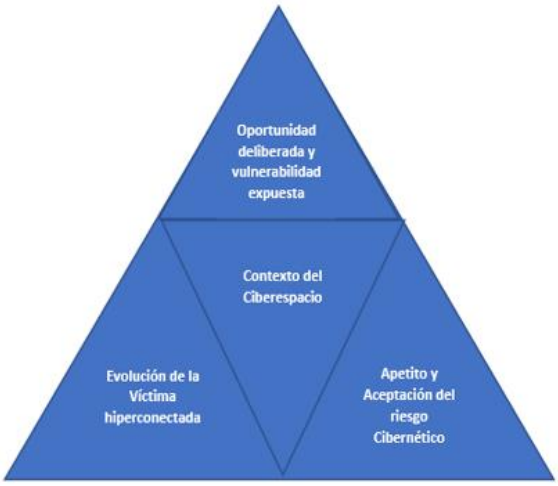

Fig 3:- Tomada de FRAUDE CIBERNÉTICO: METODOLOGÍA PARA SU IDENTIFICACIÓN Y RELACIÓN CON EL CIBERESPACIO, pág. 2

The oriented bases in terms of: the deliberate opportunity and vulnerability exposed, the evolution of the hyper- connected victim and the appetite and acceptance of cyber risk, possess certain facilitating characteristics of the possible elements that can distinguish the purpose of the model; which undoubtedly can favor the obtaining of the phishing schemes that are intendedto be evaluated.

Another task of the adaptability of the algorithm model to counter impersonation is based on the generation of individual and/or singular,or extended rules; which allows each of the information receivers to be adjusted to the parameterizable conditions of theinput method.

IV. CATEGORIZATION OFTECHNIQUES AS INFORMATION INPUTS

One of the main characteristics of the impersonation of a website or a domain is related to the categories that define some essential characteristics of the social engineering carried out by the cybercriminal, particularly in what deals with the construction or denomination inthe description (http or https), which have a certain classification:

- *Omission:*
examplecibernético.com

- *Adition*:
cyberneticexample.com

- *Substitution*:
exemplocibernético.com

- *Transposition:*
aplicaioclibernético.com

- *Separation*:
example-cibernético.com

- *Homoglyph:*
ExamplecIbernético.com

- *Homophone:*
examplecybernetico.com

- *Adition of keywords:*
ejemplocibernéticologin.com

Each of these features included in the model can focus on the structure of special characteristics for each of the exposed or impersonated phishing examples.

The techniques used are characterized by their different orientations, many of them focused on the novel construction by the criminal organizations, in terms ofthe variation of orientational patterns,which are limited only to their "base" characteristics, already mentioned.

Similarly, anti-phishing tools are oriented specifically towards the bank orfinancial institution, which can detect phishing with 100% accuracy, but there is no real comparison of the essential characteristic patterns.

Many of the campaigns contain similar vectors not only against the e-mail based structure but also with the use of other vectors and approaches.

Recently the campaign oriented on impersonation of known brand tools and platforms used by employees or potential victims, also relate to the function of each of each of the operatingsystems that determine the execution ofother types of cyber-attacks.

V. MODEL SUITABILITY ACCORDING TO THE TECHNIQUES USED IN MULTIPLE PLATFORMS

There are 7 types of targets for cybercriminals, and how to build the respective techniques and choice of components (APWG, 2021) that are being exploited:

- Financial Institutions with 24.9%.
- Social Media 23.6%
- Cryptocurrency 2.0%
- Logistics Chains and Purchasing5.8% ü eCommerce and Retail 5.8% ü Retail 7.8%
- eCommerce and Retail 7.6%.
- Payment 8.5% ü SAAS Web mail
- SAAS Web mail 19.6% ü Social Media
- Social Media 23.6

Each of these factors are involved in themodel, as well as the most usedtechniques for this type of objective. This allows the association of indicatorsrelated to construction characteristics,techniques used, and procedures executed.

This last aspect is linked at the momentof being able to identify the factors related to the matrix also stabilizes the main objective of this model, which dealswith the definition and stabilization ofsome viable algorithms to be able to determine the objectives on the characterized information mentioned in points 2 and 3.

In relation to this, there is a figure of timeexecuted by cybercrime that lies in the measurement of a sequence carried outby cyber attackers defined as follows:

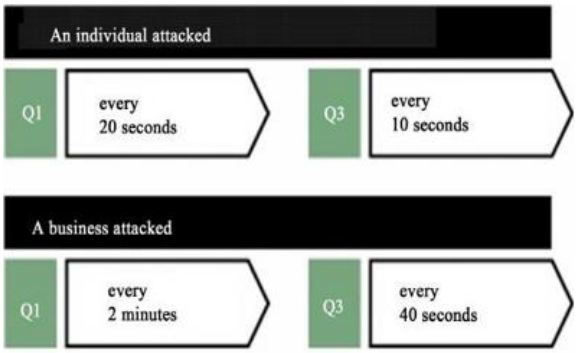

Fig 4: Tomada de (Amro, 2018) página 3

The measurement of the frequency of each of these attacks does not have a fixed characteristic related to the detection per tool, which implies, understanding the context of each of the factors delimited in the objectives of the campaigns and subsequently the factors that influence the execution of these.

An aspect that is adapted in the implementation of the model is to be able to determine quantitatively and qualitatively how the frequency of cyber-attacks mobilized in each of the campaigns impacts in order to determine the victimological expansion of phishing,once it is released to the different actors.

The adaptation to multiple platforms refers to the propagation through mobile components or cellular equipment usedto propagate malicious links or disseminated campaigns; a factor that increases the possibilities of identity theft and identity fraud.

Similarly some factors become very meaningful when establishing the propagation methods (Amro, 2018) that increase the spectrum of infection in front of mobile devices; and it is directlyrelated to the operating system that involves a certain number of increase infront of its analysis:

- Android: 37%
- iOS: 63%

Interestingly, this analysis can determine that the distribution of the techniques used by cybercriminals is oriented more to devices of this category; which implies that the creation of black or white lists is not exactly a mitigation tool that brings all the benefits for the pursued objective.

The difficulty that exists for the user or client to verify the URL within the validation context is one of the original components of the easy capture of information, which implies a high degreeof effectiveness in the use of thetechniques used by the attacker.

VI. MACHINE LEARNINGTECHNIQUES APPLICABLETO THE MODEL

Within the techniques used in this sameway, there are two lines to quoteaccording to their purpose:

- Unsupervised Learning: Based on this, initially no structured or classified data is expected, Example: Document Analysis.
- Supervised Learning: Informationadjusted to a respective label is expected, e.g. Email Spam Detection.
- Reinforcement Learning: Adjustable model of state and action learning.

These three elements can be used to determine actions to the proposed model; however, each of them can yield specific information for each of the datainteraction needs From this, each of the factors influencingin the data input resulting from each of the financial institutions and their tools (are understood as supervised information) however it is necessary to establish the necessary characteristic patterns to "match" with the patterns of identification, detection or characterization of phishing.

One of the techniques to be used in the model proposed in this paper is reinforcement learning, which will allow:

- feed decision making in the face of random and indiscriminate phishing campaigns.
- Establish tactical control elements for technical operators and pre- established rules.
- Establish guidelines or strategies tomitigate or repress the execution of phishing campaigns.

This model also relaxes the need toclassify entries related to the types of phishing schemes across the financial ecosystem, in addition to verifying through a possible identification of the content of each of them.

In this case, clustering or associationactivities may be needed to focus on possible events related to malspam campaigns or other events related tomass mailing campaigns.

VII. CONSOLIDATION OF THEDATA SET AND TRAINING SET

In other works, it is highlighted the composition of aspects of hybrid detection studies, certain proportion of algorithm mechanisms such as; the useof gradient (CF-g) [15], which is derivedfrom the characteristics of the data sets, in order to establish focused parameters in relation to predictive models. In the same way it is highlighted that there are other elements that relate each of these parameters or characteristics of phishing with an approximation [16] towards the elements taken into accountby ML tools, using decision tree, randomforest algorithm to be able to conclude that the introduced characteristics allow to approve or disapprove a legitimate ornon-legitimate site.

In the same way there are other sourcesof information such as [17] which deals with detection methods using ML techniques for the detection of phishing URLs. Models that allow detection from a preventive system by means of a blacklisting or whitelisting system of IPsor URLs, heuristics, ML content visualization and hybrid approaches

In the same vein, feature classification allows to determine qualitative indicators of phishing structures, as determined byartificial neural networks [18] EML (Extreme Machine Learning) and subsequently establish a classification relationship of fraudulent sites. Classification trees offer some key elements to be able to guide the classification of decision trees [19] and other algorithms that allow this learning to be continuous over time such as the validation of the number links containedin an illegitimate site.

Some features related to functionalalgorithms to detect phishing towards financial consumers are often associated with Internet components, one of them involves [20] Random Forest (RF) testing. As seen above it is only limited to observe equivalence, (positive or negative validation), using CANTINA [21] as a method of approaching heuristic detection of phishing websites.

Otherwise it is pertinent to relate that there are other types of algorithms that facilitate these three tasks, such as detection, prevention and correction; however, a calculation that guarantees how cybercriminals execute specific tasks employed in the different phishingcampaigns in Colombia is notparameterized.

VIII. CONLUSIONES

- For each of the aspects to betaken into account for the implementation of the model in itsfirst stage, it requires an essentialtest period to be able to test or test each of the algorithms with the related inputs.
- It is necessary that financial institutions interested in testingthe model in its second phase, can have a structured component at hand to show the different outputs of quantifiable products applying the model.
- In stage II of structuring the model, it is necessary to obtain the greatest amount of data that allows to consolidate the results focused on the prevention and reduction of fraud in the financial ecosystem.

FUTURE WORK

Carry out in conjunction with the Javeriana University the technical position for the materialization of the model applicable to the financial sector;focused on the guidance schemes that are required, in order to determine the second phase and implementation phase of the model presented.

At this time the preprint article isassociated with a DOJ document to guarantee and reserve copyright.

REFERENCES

[1]. Amani, A., Norah, A., Bashayr , A., & Dr.Aram , A. (s.f.). *Detecting Phishing WebsitesUsing Machine*.
[2]. Amro, B. (2018). *Phishing Techniques in Mobile Devices.* Journal of Computer and Communications. doi:https://doi.org/10.4236/jcc.2018.62003
[3]. APWG. (2021). *Phishing Activity Trends Reports.* Obtenido de https://apwg.org/trendsreports/
[4]. Kang , L., Choon , L., KokSheik, W., & Kelvin S.C., Y. (2019). *A new hybrid ensemble feature selection framework for machine learning-based phishing detection system*.
[5]. Lukas Brenner, T. M. (2019). Consumer fraud victimization and financial well-being.*Journal of Economic Psychology*, 35.
[6]. MIT. (2016). Wired for Innovation: How Information Technology Is Reshapingthe Economy. 153.
[7]. Ortiz Ruiz, E. E. (2020). *FRAUDE CIBERNÉTICO:METODOLOGÍA PARA SU IDENTIFICACIÓN Y RELACIÓN CON EL CIBERESPACIO.* Bogotá: ResearchGate. doi:10.13140/RG.2.2.15063.91047
[8]. Ozgur , K. S., Ebubekir , B., Onder , D., & Banu, D. (2018). *Machine learning based phishing detection from URLs.* Estambul: Elsevier.
[9]. Patil, V., Tushar Bhat, Pritesh Thakkar, & ChiragShah. (2018). *Detection and Preventionof Phishing Websites using Machine Learning Approach.*
[10]. Shyni, C. E., Anesh D Sundar, & G.S.Edwin Ebby. (2018). *Phishing Detection in Websites using Parse Tree validation.*
[11]. Sönmez, Y., Hüseyin Gökal, Türker Tuncer, & Engin Avci. (2018). *Phishing Web SitesFeatures Classification Based on Extreme Learning Machine*.

[12]. SuperFInanciera. (s.f.). *www.superfinanciera.gov.co*. Obtenido de https://www.superfinanciera.gov.co/jsp/16071
[13]. SuperSociedades. (2014). Guia para la adopción de un Sistema de Gestión del Riesgo de Lavado de Activos y de Financiación del Terrorismo. Obtenido de http://www.odc.gov.co/Portals/1/publicaciones/pdf/delitos-relacionados-drogas/CR1032017_guia_adopcion_sistema_gestion_riesgo_lavado_sector_transporte_terrestre_carga.pdf: http://www.odc.gov.co/Portals/1/publicaciones/pdf/delitos-relacionados drogas/CR1032017_guia_adopcion_sistema_gestion_riesgo_lavado_sector_transporte_terrestre_carga.pdf
[14]. Universidad de Cambridge. (2012). Measuring the Changing Cost of Cybercrime. 32.
[15]. Yue, Z., Jason, H., & Lorrie, C. (s.f.). CANTINA: A Content-Based Approach to Detecting Phishing Web Sites.