Proceedings from the Document Academy

Volume 8
Issue 2 *Proceedings from the 2021 Annual Meeting of the Document Academy*

Article 8

2021

Modeling Deception: A Case Study of Email Phishing

Abdullah Almoqbil
Imam Mohammad Ibn Saud Islamic University, abdullahalmoqbil@my.unt.edu

Brian C. O'Connor
Visual Thinking Laboratory, College of Information, University of North Texas, brian.oconnor@unt.edu

Richard Anderson
rich.anderson@untsystem.edu

Jibril Shittu
JibrilShittu@my.unt.edu

Patrick McLeod
patrick.mcleod@untsystem.edu

Recommended Citation

Almoqbil, Abdullah; O'Connor, Brian C.; Anderson, Richard; Shittu, Jibril; and McLeod, Patrick (2021) "Modeling Deception: A Case Study of Email Phishing," *Proceedings from the Document Academy*: Vol. 8 : Iss. 2 , Article 8.
DOI: https://doi.org/10.35492/docam/8/2/8
Available at: https://ideaexchange.uakron.edu/docam/vol8/iss2/8

I. Why Do Phishing Attempts Still Work?

Despite numerous efforts to raise awareness about phishing scam emails, the number of phishing attacks continues to grow significantly each year. According to the Federal Bureau of Investigation (FBI), between 2014 and 2018 phishing complaints increased more than 20%, with the financial losses increasing more than 100%, amounting to a total loss of $7.45 billion (FBI, 2018). The email phishing attacks of today are an evolution of similar techniques that can be traced back at least to the 19th century. Exploring the history of pre-internet swindling schemes helps draw a bigger picture of the current phishing and scamming methods. One of the most common 19th-century techniques was the "Spanish Prisoner Letter" where the scammer made a request to their target audience for token amounts of money to help the prisoner to retrieve their treasures. The Spanish Prisoner Letter scam continued to evolve with changing storylines and targets through the regular mail services until the postal inspectors noticed the flood of scam mail items and started warning people about it (Train, 1910).

We generated an idea for modeling key factors of digital phishing, such as prevalence, frequency, and effectiveness. However, finding a corpus of recent phishing emails proved to be almost impossible due to privacy and user information concerns. Companies are reluctant to provide researchers with phishing emails targeted at their server and users because they may hold sensitive information. After the Facebook–Cambridge Analytica scandal data came to light, companies and organizations became more stringent and wary about releasing data to researchers. We contacted several data centers and information researchers about sharing phishing emails for research purposes, but they all refused for the above-mentioned reasons. We turned to our home institution, the University of North Texas (UNT) email server.

II. Background for a Model

a. Distinguishing Deception from Lying

Lying is an act of delivering a false statement to a victim with the intention of making the target believe the statement. Deceiving someone requires intentionally causing the victim to believe a premeditated set of actions and speeches that will shape a false notion (Mahon, 2015). The two terms are often used interchangeably, though we rely here on a subtle difference – where "to lie" simply means to tell an "untruth," "to deceive" has its roots in "to ensnare, to trap." Deception is usually done with illusions and tricks of facts to make targets fall into the trap (O'Connor, Copeland & Kearns, 2003). According to one study, humans lie at least once a day (Feldman, Forrest & Happ, 2002).

b. Psychology of Deception

Phishing attacks typically rely on the psychology of deception. Deception in its various forms results from pervasive and adaptive phenomena. Dating back to the 1920s, entomologists (Wheeler, 1926) began studying and attempting to categorize levels of social behavior in animal species by observing parenting styles and mating rituals of different insects and animals. Socio-biologists (Wadsworth, Wilson & Barker, 1975) further expanded the species by looking at social behavior patterns of insects and animals into rudimentary forms of deceptive activity within and among different species. Knowledge of these rudimentary forms of deceptive activities led to a *deception hypothesis* in comparative psychology. Non-humans display different deception practices according to their surroundings, ranging from household pets seeking attention, to wild animals evading predators, and hierarchical groupings of apes and wolves. In humans a progression of these deceptive activities advances into a manipulative type of premeditated deception commonly seen in humans.

Premeditated deception is a behavior deliberately planned for a personal gain of advantage over another by hiding the truth or manipulating facts. In a study with fifty infants, Reddy (2008) found that infants as young as six months old pretend to cry just to attract the mother's attention (Reddy, 2008). Knowledge of deception dates, at least, back to the creation myths of the Abrahamic religions (Qur'an 2:35–36). Nowadays, technology has helped in creating new techniques and methods of deception that are enhanced by the expansion of human interaction. Engaging with targets is no longer limited to personal contact or slow and costly distant contact with a fairly small number of people; the digital environment essentially erases most boundaries and barriers.

c. Email Deception Theory

An electronic mail (email) vector is a specially crafted and distributed method of enticing targets to perform actions that will make their personal data available to attackers. The email vector is advantageous to the phishers because of the ease of distribution to a large quantity of recipients at very low cost. Furthermore, it conceals the geographical location of the sender. The FBI has modeled the typical steps in the phishing lifecycle (Figure 1). This lifecycle can vary in duration depending on the final goal or attack method of the phishing perpetrator. Email phishing is typically initiated with a target acquisition process which begins when an individual visits any online location which has less than optimum security features, as in step 1. Perhaps the most critical part of the lifecycle is step 2 (grooming), because the success of the entire phishing attack relies on successfully gaining the trust of the target. Step 3 is characterized by the exchange of

information between the target and the attacker. The attackers gather information relevant to the final goal by asking questions, seeking sympathy, implanting malware on the target's communications device, or directing the target to a specific online location that surreptitiously harvests information such as passwords and user information, or sometimes seeking nude photos for a future blackmail. In step 4, if the overall goal of the phishing attempt is solely monetary extortion as is the case of individual petty phishing scam, the attacker demands payment via wire transfer of funds to a location of the attackers choosing. If the phishing attack is only a portion of a larger scheme, further information or direction is given to the target with a laundry list of expectations to fulfill. Varying technical approaches are used at different stages of these phishing attacks.

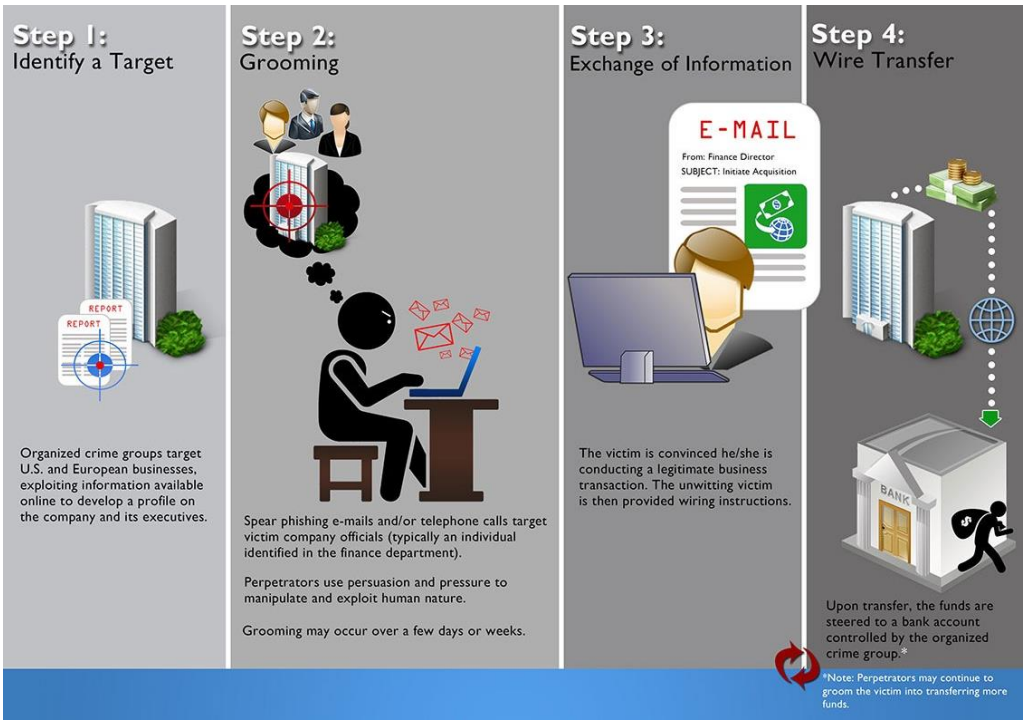

Figure 1: Email Phishing Steps (FBI, 2018)

III. Foundational Theory for Modeling Successful Phishing

We used the functional ontological construction (FOC) model, which was proposed by Anderson (2006) as a pragmatic approach to understanding the relationship between human behavior and the information environment. The model emerges from the application of behavior analytic theory to problems in information science. FOC is a multicomponent model which relies heavily on empirical history and behavioral analysis. The first component of the model is based on a binary model of documents inspired by Shannon and Weaver's information theory, Eco's theory of semiotics, Wittgenstein's notion of language games, Skinner's theory of verbal behavior, and Dawkins' theory of memes (Dawkins, 2014; Eco, 1986; Shannon & Weaver, 1948; Skinner, 1957; Wittgenstein, 1953). The second component of the model is based on the functional ontological space that engages both the user and document in a common ontological context. This ontological context covers behaviors such as information seeking by users, and the preference for documents that satisfy user needs. The third component of the model is based on the functional ontological implications of the model. User interaction with documents has a selective function on the user behavior, and in turn the user behavior has a selective function on the document. The model is sometimes referred to as ABC because it is formed around Antecedents, Behaviors, and Consequences.

IV. Framework and Methodology

Email communication offers large information carrying capacity and a nearly ubiquitous information sharing channel, especially in a higher education institution such as University of North Texas (UNT). UNT's email server, known as EagleConnect Email System, serves the official email communication needs of students, employees, retirees, and alumni; and as of March 10, 2021, the UNT email system has over 251,000 accounts and an average of close to 40,000 active monthly users. Also, UNT's staff and students send a monthly average of 10,000 emails and receive 110,000 emails, and the UNT's staff and students read close to 30% of the received emails. The email accounts are hosted on the Microsoft Office 365 platform, and therefore emails are automatically filtered for phishing attempts through the Microsoft Office 365 filtering system. Microsoft Office 365 has over 258 million active on-premises or cloud-based exchange users spanning across schools, healthcare, financial services, companies (70% of Fortune 500 companies), and governments. The phishing emails used in this study successfully bypassed the Microsoft office 365 phishing filtering systems to reach the email end user. The end users had reported the phishing emails to the network administrator who archived the email samples on a different email folder. The simple fact that

these phishing emails passed through a rigorous and well-regarded filtering system demonstrates that they were good at deception.

Over the period from October 17, 2018, to October 10, 2019, 432 phishing emails that passed the Microsoft Office 365 filtering system were reported and archived on the UNT server for a period of 12 months. The emails have different features, characters, length, context, and semantics; thus, the collected data is unstructured. While we had a significant sample of phishing emails, privacy policies presented the methodological challenge of not being able to reach out to the phishing email targets to gather detailed data from their end of the transactions.

V. Results and Analysis

Our explorations revealed that UNT staff and students are more heavily targeted with phishing emails in the Summer, and during the holiday season when students and staff are more likely to need extra money. Most successful deception emails were found to start by forming an engaging email subject line. We found that email scammers used the information theory gap as a bait to fool email recipients to open phishing emails. According to Ben-Haim (2006), information gap theory is "a non-probabilistic decision theory for prioritizing alternatives and making choices and decisions under deep uncertainty." For example, marketers often form a shadowy link title to draw viewers' attention to click on the link such as "how to become rich in few steps." In our corpus, scammers used a completely blank subject line in fifty emails (~15%) to deceive recipients into opening the phishing email. Ambiguity is a strong stimulus to humans, and it is being used by scammers to encourage human curiosity to open phishing emails (Livio, 2017).

In the quantitative analysis of the corpus, TF-IDF and LDA were useful in analyzing documents and raw text to provide insight into the topics of the document. The results from the LDA and TF-IDF analyses present similar outcomes despite the differences in statistical models. The results from both models show that the corpus has three primary topics: finance, jobs, and technology.

In the qualitative analysis of the corpus, we used the categories derived by Cofense, formerly PhishMe, a leading provider of human-driven phishing defense solutions and we manually categorized the emails into six categories: opportunity/reward, curiosity, urgency, fear, job, and social/entertainment. The graph shows the distributions for email per category (see Figure 2 below).

Also, we categorized the emails by reinforcer type – essentially promise of a reward or threat of a punishment. We found that most scammers used a positive reinforcer with 131 phishing emails in the corpus, a stimulus that promises positive feedback such as a job offering. Next, scammers used the penalty technique in 74 phishing emails where the target is being threatened with losing something; commonly the scammer asks recipients to update their login credentials, or they are

going to lose access to their email inbox. In 56 of the phishing emails there were general negative reinforcements where the type of the negative reinforcer is not specified. For example, a scammer might try to steal the identity of a person in authority and send an email to one of the employees with something like "Let me know when you are available.". Moreover, in 43 of the emails in the dataset scammers use the information gap theory to get a reply from targets; commonly, they send an email such as: "Are you available" to get a reply. Since it is not known what the scammer wants, these types of emails could fall into both positive or negative reinforcement category. The phishing emails that came under both categories have no indications or signs what the scammer is asking for unless the recipient of the phishing email replies to it and that is why it could be both positive reinforcer or negative reinforcer. In 22 of the emails the reinforcement was punishment negative, in which a scammer blackmails a target by threatening to expose them online unless there is payment with cryptocurrency.

In only 20 of the 432 phishing emails (~5%) involved any response exchange between email recipients and phishing attacker. The other 95% of phishing emails were reported to the information security team without communication with the attacker. Of those 20 responses, 14 were prompted by credit card phishing attackers posing as a superior requesting a subordinate employee to purchase redeemable gift cards for a company or personal event. The targets of the credit card attacks reported these phishing emails after at least one reply to the attacker, though the data show no email recipient actually made the requested purchase. The remaining six of the response-exchanged phishing emails were in three categories: malicious files, job offerings, and a request to pay a bill. Only two of the phishing email recipients from the twenty response-exchanged emails fell for the phishing attack. One fell victim to a false job offering and lost about $3,000 to the attacker, while the other employee fell for a malicious attack.

VI. Documenting Phishing Emails

We built an analysis tool by Elastic on the findings of this research for easier data collection in the future and to create automated reports about the phishing emails that the organization is targeted with. We used the programming language PHP to create an app to connect the IMAP server to Elasticsearch and parse the emails and attachments into meaningful fields in Elasticsearch. The analysis tool runs indexes to extracts the important fields from each email to Elasticsearch every night, as shown in Table 1. The collected emails in Elasticsearch can be easily extracted to CSV files or any other formats for future research and analysis.

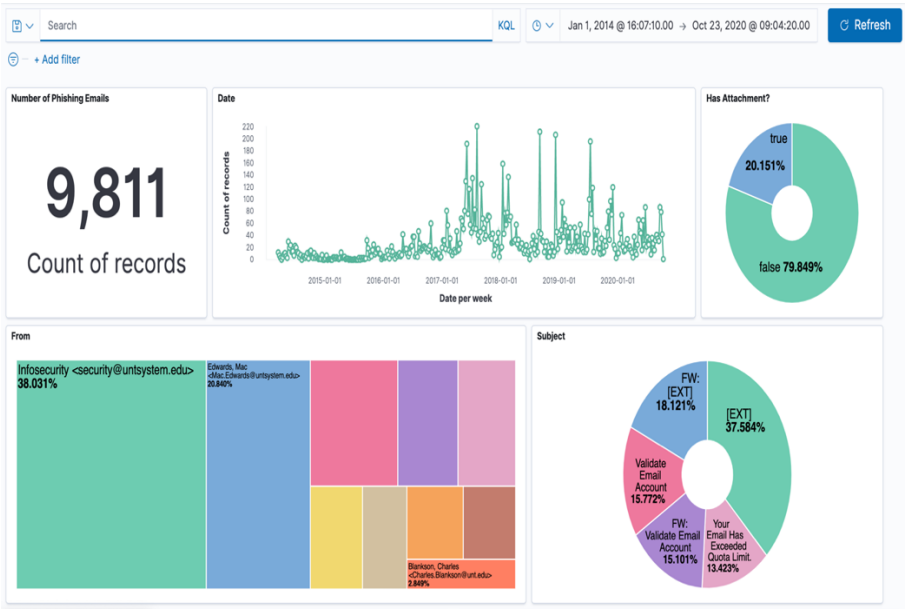

Figure 2: Visualization of the Indexed Phishing Emails

Table JSON

t _id	127858ea979788af9105578c59df7797e27916a0f339a7cf8c1a4dd4843a37b4
t _index	spam_emails
# _score	-
t _type	_doc
ⓘ attachments	⚠
t body	[redacted]
t cc	-
📅 date	Sep 16, 2020 @ 06:40:29.000
t from	"Wed, Sep 16, 2020 [redacted]
◕ hasAttachements	false
ⓘ hasExe	⚠ false
t subject	[EXT] Wednesday, September 16, 2020
t to	[redacted]

Table 1: Indexed Phishing Email

VII. Creating a Visualized Phishing Email Report with Kibana

Kibana is an add-on that can be installed on the top on Elasticsearch to add a user-friendly interface to Elasticsearch with no need of a knowledge of coding to use Elasticsearch. Kibana gives users the ability to customize slides with different metrics in with the option selecting time period. In a matter of seconds, the visualization metrics can be adjusted easily and can show phishing emails that dated back to March 2014 since the creation of the phishing emails inbox. Figure 2 is a screenshot of a visualization of the emails that the information security team received since 2014.

VIII. Deception in Phishing Emails Model

Figure 3 summarizes how scammers form their phishing emails. Firstly, if it is a spear phishing attack or a regular email phishing attack, scammers choose a topic depending on the pervious information they know about the target either as an individual or as an organization. Then, scammers pick from the five available categories to decide what type of reinforcement to use. Then they decide between encouraging the email recipients in a nice way or threatening them with a blackmail. Lastly, scammers form their phishing email with attractive wordings that fits their attack and words that bypass the respective filtering systems.

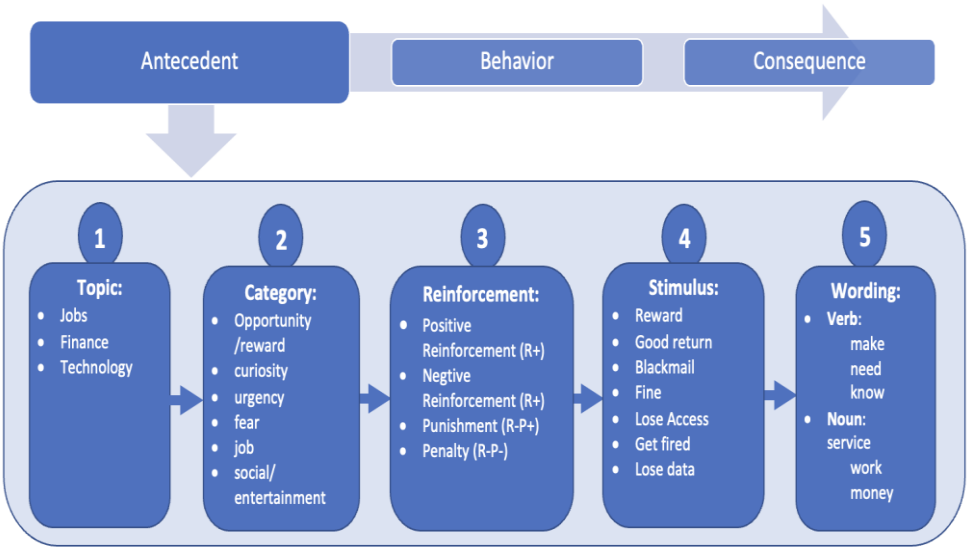

Figure 3: ABC model of Phishing Emails

IX. Conclusion

In this study, phishing emails that successfully passed through the Microsoft 365 email filtering system to the UNT staff and students were reported and archived in a phishing email inbox by UNT's security team as the primary dataset. Our dataset contained 432 phishing emails that were archived between October 2018 and October 2019. The phishing emails were studied from both technical and psychological angles to identify why an advanced filtering system such as Microsoft 365 email phishing defense system failed to prevent delivery of these phishing emails.

In this research, we used a mixed method to analyze the data. In the quantitative analysis, we used topic modeling and TF-IDF to get an overview of what are the topics of the studied email corpus. We found out that the emails could be categorized into three topics namely: jobs, finance, and technology. On the qualitative side of the analysis, we used Anderson's functional construction ontology theory to study the interaction between a human and a document, and we discovered that scammers use positive and negative reinforcement with different types of stimuli as a motivation to swindle email recipients. Furthermore, we categorized the emails into 6 different categories based on the reinforcements used: opportunity or reward, job, urgency, curiosity, fear, and social/entertainment.

Based on B. F. Skinner's reinforcement theory, we identified that all four different types of theoretical stimuli were used to defraud email recipients. The attackers used positive reinforcement to promise a prize or reward to the email recipient, and they used negative reinforcement to threaten the email recipients with unwanted potential consequences.

In addition, we observed that attackers used the information gap theory to scam email recipients by sending them either empty email subject or empty email body. This information gap technique accounted for 30% of the emails in our corpus having either an empty subject or body.

Moreover, a statistical visualization showed that the university staff received more phishing emails in the summer and winter holiday seasons. This seasonality was further explored to identify that some of the universities staff not paid during the summer months usually search for temporary jobs to fill the pay gap, and the attackers take advantage of this opportunity to offer fake job openings to their targets.

References

Anderson, R. L. (2006). *Functional ontology construction: A pragmatic approach to addressing problems concerning the individual and the informing environment* (unpublished Ph.D. thesis). University of North Texas.

Ben-Haim, Y. (2006). *Info-gap decision theory: Decisions under severe uncertainty*. Elsevier.

Dawkins, R. (2014, February 4). What's in a meme? *Richard Dawkins Foundation for Reason and Science.* https://richarddawkins.net/2014/02/whats-in-a-meme/

Eco, U. (1986). *Semiotics and the philosophy of language*. Indiana University Press.

FBI. (2018). Business email compromise contributes to large scale business losses nationwide. Internet Crime Complaint Center. https://www.ic3.gov/Media/Y2018/PSA180611

Feldman, R. S., Forrest, J. A., & Happ, B. R. (2002). Self-presentation and verbal deception: Do self-presenters lie more? *Basic and Applied Social Psychology*, *24*(2), 163–170.

Livio, M. (2017). *Why?: What makes us curious*. Simon and Schuster.

Mahon, J. E. (2015). The definition of lying and deception. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/win2015/entries/lying-definition/

O'Connor, B. C., Copeland, J. H., & Kearns, J. L. (2003). *Hunting and gathering on the information savanna: Conversations on modeling human search abilities*. Scarecrow Press.

Reddy, V. (2008). *How infants know minds*. Harvard University Press.

Shannon, C. E., & Weaver, W. (1999). *The mathematical theory of communication*. University of Illinois Press.

Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.

Train, A. (1910). The Spanish prisoner. *The Cosmopolitan Magazine*, 465–474.

Wadsworth Jr, A. P., Wilson, W., & Barker Jr, H. R. (1975). Reduction of state and trait anxiety by kind firmness attitude therapy. *Psychological Reports*, *37*(1), 23–29.

Wheeler, W. M. (1926). Emergent evolution and the social. *Science*, *64*(1662), 433–440.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe Trans.). Oxford University Press.