Iterating the Cybernetic Loops in Anti-Phishing Behavior: A Theoretical Integration

Alaa Nehme
Iowa State University
anelhme@iastate.edu

Joey F. George
Iowa State University
jfgeorge@iastate.edu

Abstract

As phishing emails represent continuous attack vectors, users’ continuance in anti-phishing behavior is highly significant. This paper extends the previous literature on phishing and information security. We develop a hierarchical inter-looped cybernetic model that integrates protection motivation theory, expanded prominence interpretation theory, and risk analysis. Our model explores (1) continuous interdependence of avoidance and adoption cognitive systems in anti-phishing behavior, and (2) continuous interdependence of education, awareness and training. The conceptual foundation, derived propositions and contributions are discussed.

Keywords

Phishing avoidance, anti-phishing, cybernetics, email credibility, email deception, behavioral security

Introduction

Phishing emails pose immense threats to individuals and organizations. Ninety-one percent of cyberattacks start with a phishing email (PhishMe 2016), and organizations have lost billions of dollars due to phishing attacks (Jensen et al. 2017). Most recently, Business Email Compromise (BEC), a variant of spear-phishing, has led to a 1300% increase of financial losses in organizations (FBI 2017). BEC, also referred to as ‘CEO fraud’ and ‘whaling,’ typically manifests a fraudulent email attack, by which the deceiver (i.e. email scammer) impersonates a senior executive (e.g. a CEO or CFO) and requests specifically targeted organizational employees to make wire transfers or reply with personally identifiable information such as W-2 tax forms (Symantec 2017).

In response to phishing threats, organizations have adopted various technical solutions (e.g. DKIM and DMARC) to filter suspicious emails as spam or fraudulent (Derouet 2016). Yet, these solutions have not ultimately prevented phishing attacks, as phishers continuously improve their bypass techniques (Smadi et al. 2018). As such, the role of employees (i.e. phishing email receivers) in detecting phishing emails is highly significant. Additionally, this role is governed by security education, training and awareness (SETA) programs and behavioral controls in organizations (Landress et al. 2017; Winkfield et al. 2017).

Driven by this role, previous research has examined the factors that affect users’ success in detecting phishing (e.g. Wang et al. 2016, 2017; Wright and Marett 2010). These factors relate to emails’ structural properties and individual differences. Most of the IS research on phishing has employed cross-sectional research designs and neglected long-term continuance in anti-phishing behavior (Steinbart et al. 2016). However, IS continuance (i.e. the continued usage of systems) is a prominent factor in user behavior (Bhattacherjee 2001). Additionally, most studies related to phishing either examine the role of one element in SETA programs or examine SETA as a whole and do not account for the interrelatedness among its three elements (i.e. education, training and awareness). For instance, Jensen et al. (2017) study the effect of training, independent of awareness and education, on security behavior in the phishing context. Another attribute in the information security literature is the avoidance vs. adoption approaches (see Liang and Xue 2009). Avoidance manifests avoiding information threats, and adoption manifests adopting security measures. How are avoidance and adoption behaviors continuously interdependent? How are SETA...
elements inter-related in a continuous manner? These research questions guide this paper. As such, our objective is to develop a holistic framework that illustrates the continuous interdependence of (1) users' adoption and avoidance behaviors, (2) SETA elements and (3) 1 and 2. On that basis, we draw upon the cybernetics, information security, deception detection and risk analysis literatures.

This paper proceeds as follows. First, we introduce the conceptual foundation that informs our research questions. Next, we propose the Inter-Loop Anti-Phishing Model (ILAPM), an inter-looped cybernetic model that draws on Protection Motivation Theory (Rogers 1975), Expanded Prominence Interpretation Theory (George et al. 2016) and risk analysis. We derive several propositions from ILAPM and conclude with this paper’s contributions and suggestions for future research.

**Conceptual Foundation**

**Cybernetics**

Cybernetics (i.e. control theory) is a theoretical framework for understanding self-regulating systems in human behavior (Wiener 1948). It has been employed across different behavioral disciplines (e.g. mental health and organizational behavior). For instance, control theory has been used in psychology to examine emotion regulation (e.g. Etkin et al. 2015; Sheppes et al. 2015). Recently, the Information Systems Security literature has deployed the cybernetics framework to examine users' security behavior (e.g. Liang and Xue 2009; Steinbart et al. 2016). The main premises of control theory postulate that individuals regulate their behavior through a feedback loop (Carver and Scheier 1982). This loop may be negative or positive. A negative loop reduces a discrepancy between a present state and an objective one. On the contrary, a positive feedback loop increases the discrepancy. As depicted in Figure 1, a disturbance factor in the environment or a change in the reference value activates the system process. The input function inputs the perceived environment as a signal and sends it to the comparator. The latter compares the signal value to the reference value generated by the objective. In case of a negative feedback loop system: if the comparison results in a discrepancy between the two values, the output function outputs a behavior that reduces the discrepancy. In the case of a positive feedback loop, the output behavior increases the discrepancy. This behavior impacts the environment to reach the desired state.

![Figure 1. The Cybernetic Loop (Carver and Scheier 1982)](image)

![Figure 2. Protection Motivation Theory Overview (Rogers 1975)](image)

**Protection Motivation Theory and Security Behavior**

To examine security behavior, Information Systems Security (ISSec) studies have widely adopted fear appeal models from the public health literature (for a detailed review, see Boss et al. 2015). Protection motivation theory (PMT) has acted as a foundational constituent of such frameworks (Boss et al. 2015). PMT postulates that individuals engage in two cognitive processes, threat appraisal (TA) and coping appraisal (CA), and consequently adopt protective behavior (Figure 2; Rogers 1975). TA, a function of threat
susceptibility (i.e. the perceived degree of risk to threats) and threat severity (i.e. the perceived degree of threat harm), induces coping behaviors, through which individuals adopt problem-solving and danger-control measures to protect themselves against threats (Rogers 1975). CA, a function of self-efficacy (i.e. the perceived ability to take protective measures), response efficacy (i.e. their perceived effectiveness of protective measures) and response costs (i.e. the perceived costs associated with taking protective measures), engages individuals in appraising their ability to cope with threats.

In the information security context, studies have employed PMT to explain users' coping with threats in two distinct but complementary approaches: (1) to explain (or predict) users' adoption of security measures (e.g. Anderson and Agarwal 2010), and (2) to explain users' avoidance of information threats (see Liang and Xue 2009). To entrench the second approach and explain the difference between avoidance and adoption, Liang and Xue (2009) develop the technology threat avoidance theory (TTAT), which adjusts PMT to the avoidance approach. As such, the literature implies that security behavior is a function of both threat avoidance and protective (i.e. anti-threat) measures adoption.

**Expanded Prominence Interpretation Theory and Deception Detection**

The Expanded Prominence Interpretation Theory (EPIT, George et al. 2016), a synthesis of the Prominence Interpretation Theory (PIT, Fogg 2003)) and Interpersonal Detection Theory (IDT; Buller and Burgoon 1996)), provides a general theoretical framework of deception detection in computer-mediated communication (Figure 3). Deception refers to “a message knowingly transmitted by a sender to foster a false belief or conclusion by the receiver” (Buller and Burgoon 1996). EPIT’s primary tenets pose that the communication-medium between a sender and a receiver determines how the latter perceives the message's (or sender's) credibility. Further, this perceived credibility impacts the receiver’s success in detecting deception. In other words, high credibility deceptive messages may be more believed than low credibility authentic messages. Also, EPIT takes into account temporality, by which credibility assessments change over time. Perceived credibility is a function of an interaction between prominence (i.e. the likelihood that an element is noticed or perceived) and interpretation (i.e. a user’s judgment about the noticed element). In sum, the communication medium provides the “lens” through which users (i.e. receivers) determine an element’s prominence and interpretation, which impact the credibility assessment of messages and deception detection (George et al. 2016).

![Figure 3. Expanded Prominence Interpretation Theory (George et al. 2016)](image)

**Risk Analysis**

Multiple definitions of risk analysis exist in risk research. The Australian and New Zealand Standards define risk analysis as an application of “a systematic process for understanding the nature of the risk involved and determining its level, and risk evaluation consists of assessing the significance of this risk by comparing its level with some kind of standard or terms of reference” (Corvellec 2010). The US Department of Homeland Security estimates risk as a product of threat, vulnerability and consequence (Cox 2008). In information security, several risk analysis methods have been developed to handle information threats (e.g. Feng et al. 2014; Karabacak and Sogukpinar 2005).

**ILAPM and Propositions**

The proposed cybernetic model explores the continuous interdependence among avoidance and adoption behaviors, and security awareness, training and education. In alignment with Carver and Scheier (1982),
we develop Inter-Loop Anti-Phishing Model (ILAPM) as a hierarchically organized system, which has superordinate and subordinate goals. At each hierarchical level, the objective provides the comparator in the level below with the reference value (Carver and Scheier 1982). In our model, the first-level objective, safety, provides the comparator in the first-level system with the reference value. Similarly, the second- and third-level objectives provide the comparators in the second- and third-level systems with the relevant reference values. Additionally, we extend the concept of multi-level goals to disturbance factors. As such, we propose that each level passes a disturbance factor to the level above. This disturbance factor is what activates the control system at the next level. Our proposition aligns with the Perceptual Control Theory (PCT; Powers 1960) in the psychology discipline, which poses that perceptions are formed in a hierarchical level and that disturbances come from within the cognitive system.

ILAPM includes three inter-looped control systems. The first-level system (lowest level) maintains the protection state in the environment. It draws upon PMT, which is aligned with the expectancy value theory (Atkinson 1964) and as such subsumes the adoption approach in information security (Boss et al. 2015). The second- and third- (highest) level systems maintain resistance against phishing and risk-free environments. The two systems draw upon EPIT and risk analysis respectively. The latter relies on predictive models (Morgan 1993) and as such subsumes the avoidance approach. EPIT involves credibility assessment and deception detection. Credibility assessment involves factors such as involvement (i.e. motivation) and intentions (George et al. 2016). As such, credibility assessment may be viewed to align with the adoption approach. On the other hand, deception detection may be viewed to align with the avoidance approach. Therefore, we view EPIT to be incorporative of the adoption and avoidance approaches in information security. Additionally, we consider anti-phishing behavior to lie on a continuous spectrum of avoidance and adoption behaviors (Figure 4). PMT and risk analysis lie at the adoption and avoidance positions respectively. EPIT interlinks the two approaches as it embraces both. We call the three level systems (bottom-up) in ILAPM (Figure 5): the protection control system, the resistance control system, and the risk control system. The three-level systems are inter-looped, such as each level system inputs an objective reference value to the system level below and a disturbance factor to the system level above. ILAPM yields a set of axioms, upon which we derive multiple propositions illustrated by the variance model in Figure 6.

![Figure 4. Anti-Phishing Behavior Spectrum](image)

**Axiom** 1. The anti-phishing cognitive system has three layered-systems with three-level objectives and disturbance factors.

In an information security cybernetic loop, the process begins when malicious IT poses threats to the environment (Liang and Xue 2009). At ILAPM’s lowest level, phishing threats activate the first-level system (i.e. protection control system).

**Axiom 2.** Phishing email attacks emerge as a disturbance factor in the environment.

When threats arise in the environment, users engage in the threat appraisal process (Rogers 1975; Liang and Xue 2009). Hereby, users evaluate their susceptibility to phishing and its severity. Following, threat appraisal acts as an input function in the protection control system’s cybernetic loop and assigns a value (i.e. degree: high vs low) to the perceived phishing threat, an aggregate of susceptibility and severity.

**Axiom 3.** After the disturbance, users engage in the threat appraisal input function, which rates their perceived threat.

---

1 The use of the term *axiom* in this paper aligns with its usage in modern logic and thermodynamics. An axiom “is assumed to be true without proof for the sake of studying the consequences that follow from it” (Suh et al. 1978).
In the hierarchical system, the objective from the level, above the current, passes down the standard of comparison (Carver and Scheier 1982). In this instance, the current level is the protection control system, and the superordinate objective is deception detection success. As the central function of a feedback system is not creating behavior but maintaining the perception of a desired condition (Carver and Scheier 1982), the reference value holds the value of low threats. The objective of succeeding in deception detection inputs the value of low threats into the comparator of the protection control system. Following, the comparator checks if the perceived threat value, signaled from threat appraisal, is different than the reference value (i.e. if the assigned value to perceived threat is high).

Axiom 4. Deception detection, the third level objective, passes the threat reference value (value=low) to the comparator in the protection control system.

Axiom 5. After threat appraisal sends the threat perception value (i.e. degree) to the comparator, the latter compares the perceived threat to the threat reference value, which is set as low.

If the perceived threat is high (i.e. if there is a discrepancy), then the comparator triggers the output function. The output function outputs the necessary behavior to regulate the system. In this instance, the output function outputs coping-behavior. In PMT, performing protective behavior manifests individuals' coping behavior. It is a problem-solving-oriented conduct, by which individuals counter potential threats through the danger control process (Rippetoe and Rogers 1987). In the phishing context, coping behavior
includes adopting protective measures, such as an email authentication system or following rules to identify phishing emails.

Axiom 6. If there is a discrepancy between the perceived threat value and the reference value (i.e. if the perceived threat is high), users engage in coping behavior.

Axioms 4, 5 and 6 suggest that the intention (i.e. objective) of succeeding in deception detection indirectly affects users’ adoption of protective behavior against phishing through the comparator function. This function is conditional upon the perceived level (i.e. value) of phishing threat. If threats are perceived to be high, users engage in coping behavior; else, they don’t. Following, the level of perceived threat impacts the relationship between the purpose of detecting deception and adopting protective behavior.

Proposition 1. The intention to detect deception positively influences users’ coping behavior.

Proposition 2. Threat Appraisal strengthens the relationship between deception detection and coping behavior.

Axiom 7. Coping behavior reduces the discrepancy in the loop and fosters protection against phishing in the environment.

Proposition 3. Coping behavior improves the level of protection in the environment.

In ILAPM, each level induces a disturbance factor to the level above. At the first level, the protection control system produces a disturbance factor when threats are perceived as high (i.e. when there is a discrepancy in the system). Forewarning, which is analogous to high threats, induces resistance (Cameron et al. 2002; Wright et al. 2014). As such, perceived high threats disturb the level of resistance against phishing attacks in the environment.

Axiom 8. High threats disturb the environment in the resistance control system. This disruption activates the resistance control system.

The resistance control system deploys EPIT as its feedback loop mechanism. Its superordinate objective is phishing avoidance. Phishing avoidance passes the reference value of credibility to the comparator. As the goal is avoidance, the objective reference value of credibility is high. In alignment with TTAT (Liang and Xue 2009), the resistance control system has a positive feedback loop, since it deals with avoidance. After users appraise the threat of phishing attacks, they engage in credibility assessment, defined as users’ evaluation of emails’ believability. Credibility assessment acts as the input function that sends perceived credibility as a signal to the comparator. The system’s objective is to maintain the perception of a low credibility email when it is not credible. As such, the feedback loop iterates to increase the discrepancy between perceived credibility and the standard of comparison (i.e. reference value).

Axiom 9. Credibility assessment acts as an input function by which perceptions of the credibility of emails take place.


Axiom 10. Phishing avoidance, the second-level objective, feeds the comparator high credibility as the reference value.

Axiom 11. After credibility assessment sends perceived credibility to the comparator, the latter compares the perceived credibility value to the reference value, which is set at high.

To maintain the perceived credibility as low (i.e. increase the discrepancy in the system), the output function outputs deception detection behavior. In the phishing context, users’ deception detection refers to their success in identifying phishing emails. This behavior generates resistance, as an impact on the environment. In other words, successful deception detection generates resistance against phishing attacks.

Axiom 12. To maintain a discrepancy between the low perceived credibility value and the reference value (high), users engage in successfully detecting deception.
Iterating the Cybernetic Loops in Anti-Phishing Behavior

Figure 6. ILAPM Propositions

Axioms 10-12 suggest that the intention (i.e. objective) of avoiding phishing indirectly affects users' success in deception detection through the comparator function. This function is conditional upon the perceived level of credibility. If credibility is perceived to be low, users engage in deception detection behavior; else, they don't. Thus, the level of perceived credibility impacts the relationship between the purpose of avoiding phishing and detection deception in phishing emails.

Proposition 5. The intent to avoid phishing positively affects deception detection success.

Proposition 6. Credibility Assessment strengthens the relationship between phishing avoidance and deception detection.

Axiom 13. Successful deception detection fosters resistance against deception in the environment.

Proposition 7. Deception detection behavior improves the level of resistance against phishing in the environment.

Emails perceived to be of low credibility disrupt a risk-free (or a low risk environment) and activate the risk control system. As such, users engage in risk analysis to assess the risk associated with deceptive emails. The superordinate objective at this level, safety, generates the reference value and inputs it into the comparator. The reference value is set as low risk. The feedback loop in the risk control system is negative. Thus, when risks are perceived to be high, phishing avoidance, the output behavior, is generated to control the risk in the environment.

Axiom 14. Low credibility emails impose risk on the environment. Hereby, they disturb the environment.

Axiom 15. Users analyze the risk of deceptive emails in the environment.


Axiom 16. User safety, the first-level objective, inputs the reference value of risk to social engineering into the comparator.

Axiom 17. Users compare their perceived knowledge about the subject matter with the reference value, which is set as high.

Axiom 18. Users engage in phishing avoidance to reduce the discrepancy.

Axioms 16-18 suggest that the intention (i.e. objective) of being safe indirectly affects users' success in avoiding phishing through the comparator function. This function is conditional upon the perceived level of risk. If risk is perceived to be high, users engage in phishing avoidance behavior; else, they don't. Thus, the level of perceived risk impacts the relationship between the purpose of maintaining safety and avoiding phishing.
Proposition 9. Safety intentions positively affect phishing avoidance.

Proposition 10. Risk Analysis strengthens the relationship between safety and phishing avoidance.

Axiom 19. Phishing avoidance reduces the risk of being phished in the environment.

Proposition 11. Phishing avoidance behavior improves risk control against phishing in the environment.

Each of the three-level systems in ILAPM outputs a behavior that impacts the environment. As the three level systems, each with a higher-level objective, are interdependent, the state of the environment at each level impacts the state at the next level through the cybernetic loops.

Proposition 12. Risk control influences resistance against phishing in the environment.

Proposition 13. Resistance influences the level of protection against phishing in the environment.

ILAPM is a general theoretical framework that may act as a social cognition model (i.e. a theory of behavior) or a behavioral change theory. A social cognition model refers to a mental representation that elicits social behavior (Smith and Semin 2007), and a behavioral change theory refers to model that aims to elicit behavioral change in response to an intervention. Security education, training and awareness (SETA) programs have played an important role in motivating users to take security actions (Posey et al. 2015). Yet, an understanding of the mechanism of SETA’s impact on cognitive behavior is still lacking. SETA mainly incorporates a learning process. Learning “can be abstractly thought of as evolving in three distinct but interleaved levels, which together form a continuous rather than a discrete process” (Katsikas 2000). The bottom level contains awareness, and the middle and top levels contain training and education respectively (Katsikas 2000). Awareness activities manifest attracting individuals to a subject. Training activities require learners to be more active and is more formal than awareness. Training manifests producing the relevant and needed security skills. Education manifests creating expertise (Katsikas 2000). Each of the elements in phishing SETA programs (i.e. education, training and awareness) acts inter-dependently. Through ILAPM, we propose that each element activates a control system at each of the three levels. As awareness mainly draws the attention of individuals to a subject matter (Katsikas 2000), and threat appraisal is triggered by appeals (Rogers 1975), or urgent requests of attention, we pose that awareness induces threat appraisal and as such motivates users to take protective actions. Through the route of threat appraisal to credibility assessment to risk analysis, awareness also indirectly affects latter two.


As discussed earlier, credibility assessment requires prominence and interpretation. Both require a set of skills. As such, training activities positively impact assessing the credibility of emails and success in deception/phishing detection. Similar to awareness, training also has an indirect effect through the loops on threat appraisal and risk analysis.

Proposition 15. Training programs positively affect email credibility assessment.

Education fosters knowledge about social engineering and phishing techniques. As discussed earlier, risk analysis requires an in-depth understanding of threats, vulnerabilities and consequences. Consequences of cyberattacks may extend to environments outside organizations. Education positively impacts the risk analysis of phishing.

Proposition 16. Education programs positively affect phishing risk analysis.

Discussion and Conclusion

In this paper, we have developed ILAPM, a hierarchical cognitive cybernetic system that integrates three theoretical frameworks (i.e. protection motivation theory, expanded prominence eminence theory, and risk analysis) in the phishing context, and we have presented conceptual propositions. The model and propositions formulate the inter-related cognitive systems when users engage in anti-phishing behavior. It depicts the continuous interdependence among avoidance and adoption security behaviors. Additionally, the model serves as a behavioral change theoretical framework for parties that implement SETA programs.
Contributions and Future Research

The contributions of this paper are twofold. First, on a theoretical level, this paper develops a holistic interrelated system that explains anti-phishing behavior and more generally information security behavior. The holistic model includes three inter-related theories that are associated with phishing detection. To our knowledge, this is the first security behavior model that incorporates interdependent cognitive subsystems. This also is the first anti-phishing model that integrates a full deception detection theory. Integrating constructs from the deception detection literature in behavioral phishing security models furthers our understanding of why users avoid (or fall victim to) deceptive and malicious communications. Also, it is the first model that segregates awareness, training and education as three independent but interconnected elements. Additionally, ILAPM depicts the continuance mechanisms in anti-phishing behavior and SETA. Second, on a practical level, this paper sets a framework for organizations related to phishing attacks. We propose that organizations should (1) positively impact employees’ perceptions of phishing threats through awareness programs, (2) train employees to assess the credibility of emails, and (3) educate them about phishing, social engineering and the risks thereof. Organizations should execute the aforementioned activities in a continuous manner. ILAPM also suggests that organizations should promote for safety as a goal in their environments, as the latter is the source of reference values in ILAPM. Lastly, ILAPM may also inform policy makers and (non)governmental organizations that seek to raise citizens’ security awareness through awareness campaigns.

Further research may be conducted to test and expand the ILAPM. Researchers may empirically test the propositions, illustrated in Figure 6. Further research may also expand ILAPM through integrating models that examine dynamics between phishing attacks and organizations (i.e. defense mechanisms). Developing ILAPM is a first step towards research on hierarchical cybernetic models in information security. Further research on integrated cybernetics and the interdependence of avoidance behavior, adoption behavior and SETA elements in information security is needed.

REFERENCES


