ADVERT: An Adaptive and Data-Driven Attention Enhancement Mechanism for Phishing Prevention

Linan Huang, Shumeng Jia, Emily Balcetis, and Quanyan Zhu

arXiv:2106.06907v3 [cs.HC] 8 Jul 2022

***Abstract*—Attacks exploiting the *innate* and the *acquired* vulnerabilities of human users have posed severe threats to cybersecurity. This work proposes ADVERT, a *human-technical solution* that generates adaptive visual aids in real-time to prevent users from inadvertence and reduce their susceptibility to phishing attacks. Based on the eye-tracking data, we extract *visual states* and *attention states* as system-level sufficient statistics to characterize the user's visual behaviors and attention status. By adopting a data-driven approach and two learning feedback of different time scales, this work lays out a theoretical foundation to *analyze*, *evaluate*, and particularly *modify* humans' attention processes while they vet and recognize phishing emails. We corroborate the *effectiveness*, *efficiency*, and *robustness* of ADVERT through a case study based on the data set collected from human subject experiments conducted at New York University. The results show that the visual aids can statistically increase the attention level and improve the accuracy of phishing recognition from 74.6% to a minimum of 86%. The meta-adaptation can further improve the accuracy to 91.5% (resp. 93.7%) in less than 3 (resp. 50) tuning stages.**

***Index Terms*—Attention management, phishing mitigation, reinforcement learning, Bayesian optimization, eye tracking, human vulnerability, cybersecurity.**

I. Introduction

HUMAN is often considered the weakest link in cybersecurity. Adversaries can exploit human errors and vulnerabilities to launch deceptive attacks (e.g., social engineering and phishing) that lead to information leakages and data breaches. Moreover, these attacks often serve as the initial stages of sophisticated attacks (e.g., supply chain attacks and advanced persistent threats) that inflict tremendous damage on critical infrastructures. We classify human vulnerabilities into *innate vulnerabilities* (e.g., bounded attention and rationality) and *acquired vulnerabilities* (e.g., lack of security awareness and incentives). Previous works have mitigated the acquired vulnerabilities through security training [1], rule enforcement [2], and incentive designs [3], [4], but these methods are less than sufficient to deal with the innate ones, especially due to the unpredictability and heterogeneity of human behaviors. To this end, there is a need for *security-assistive technologies* to deter and adaptively correct the user misbehavior resulting from the innate vulnerabilities.

In this work, we focus on inattention, one type of innate human vulnerability, and use phishing email as a prototypical scenario to explore the users' visual behaviors when they determine whether a sequence of emails is secure or not. Based on the users' eye-tracking data and phishing recognition results, we develop ADVERT[1] to provide a human-centric data-driven attention enhancement mechanism for phishing prevention. In particular, ADVERT enables an adaptive visual-aid generation to guide and sustain the users' attention to the right content of an email and consequently makes users less likely to fall victim to phishing. The design of the ADVERT contains two feedback loops of attention enhancement and phishing prevention at short and long time scales, respectively, as shown in Fig. 1.

The bottom part of Fig. 1 in blue illustrates the design of adaptive visual aids (e.g., highlighting, warnings, and educational messages) to engage human users in email vetting. First, as a human user reads emails and judges whether they are phishing or legitimate, a covert eye-tracking system can record the user's eye-gaze locations and pupil sizes in real-time. Second, based on the eye-tracking data, we abstract the email's Areas of Interest (AoIs), e.g., title, hyperlinks, attachments, etc., and develop a Visual State (VS) transition model to characterize

Manuscript received 3 February 2022; revised 23 May 2022 and 24 June 2022; accepted 29 June 2022. This work was supported in part by the National Science Foundation (NSF) under Grant ECCS-1847056, Grant CNS-2027884, Grant CNS-1720230, and Grant BCS-2122060; in part by the Army Research Office (ARO) under Grant W911NF-19-1-0041; and in part by the DOE-NE under Grant 20-19829. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andrew Beng Jin Teoh. (Corresponding author: Linan Huang.)

L. Huang, S. Jia, and Q. Zhu are with the Department of Electrical and Computer Engineering, New York University, Brooklyn, NY, 11201, USA. E-mail:{lh2328, sj3233, qz494}@nyu.edu

Emily Balcetis is with the Department of Psychology, New York University, New York, NY, 10003, USA. E-mail:eb107@nyu.edu

Digital Object Identifier 10.1109/TIFS.2022.3189530

[1]ADVERT is an acronym for ADaptive Visual aids for Efficient Real-time security-assistive Technology.

Copyright © 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

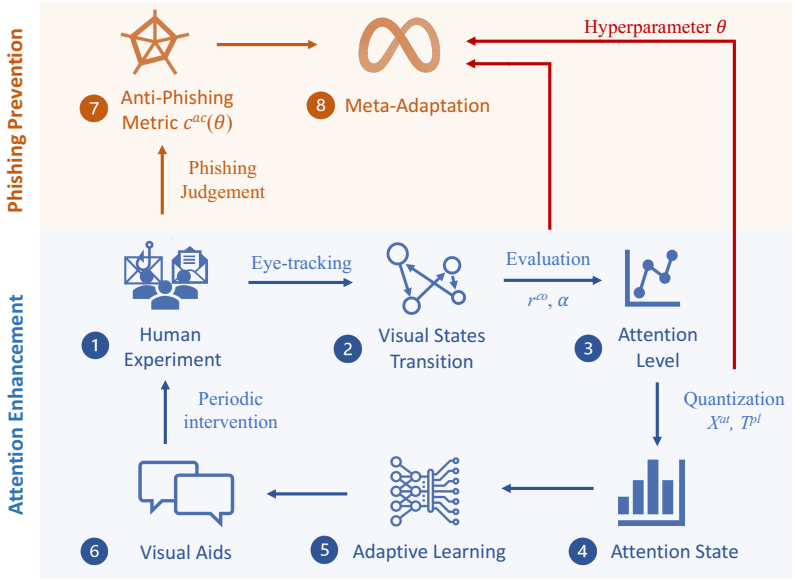


Fig. 1: The design diagram of ADVERT. The adaptive learning loops of the attention enhancement mechanism and the phishing prevention mechanism are highlighted using juxtaposed blue and orange backgrounds, respectively. Since a user needs to persistently pay attention to an email to make a phishing judgment, the meta-adaptation feedback in orange updates less frequently than the feedback of attention enhancement in blue.

the eye-gaze dynamics. Third, we develop system-level attention metrics to evaluate the user's attention level based on the VS transition trajectory. Then, we quantize the attention level to obtain the Attention State (AS) and develop adaptive learning algorithms to generate visual aids as feedback of the AS. The visual aids change the user's hidden cognitive states and lead to the set of eye-tracking data with different patterns of VS transition and AS, which then updates the design of visual aids and enhances attention iteratively.

The attention enhancement loop serves as a stepping-stone to achieving the ultimate goal of phishing prevention. The orange background in the top part of Fig. 1 illustrates how we tune the hyperparameters in the attention enhancement loop to safeguard users from phishing emails. First, we create a metric to evaluate the user's accuracy in phishing recognition under the current attention enhancement mechanism. Then, we iteratively revise the hyperparameters to achieve the highest accuracy. Since the accuracy evaluation depends on the implementation of the entire attention enhancement loop, the evaluation is costly and time-consuming. Thus, we leverage Bayesian Optimization (BO) to propose an efficient meta-level tuning algorithm that improves the accuracy.

The contributions of this work are threefold. First, we provide a holistic model of the human-in-the-loop system for email vetting and phishing recognition. By abstracting the complex human processes of sensing, thinking, and acting as a stochastic feedback control system of various parameters, we establish a system-level characterization of human attention and security judgment. Such characterization focuses on the interaction between the human and the technical systems, especially the inputs (e.g., visual aids) and the outputs (e.g., gaze locations, attention status, and security decisions) of the human system. Moreover, we propose new attention metrics to quantify the impact of hidden attention status on observable performance metrics, e.g., accuracy of recognizing phishing. These metrics enable a real-time modification of the human attention process through the adaptive visual-aid generation.

Second, we provide an adaptive technology called ADVERT to counteract inattention and improve the human recognition of phishing attacks. Two algorithms are developed to illustrate the design, where the *individual adaptation algorithm* improves the visual aid design for each individual user, and the *population adaptation algorithm* further learns the optimal visual aid for the user population. Since the data-driven approach achieves customized solutions in terms of the users and the content of the emails, ADVERT can be applied to various security threat scenarios caused by inattention. Since the feedback learning framework enables an adaptive and systematic design of the optimal visual aids, ADVERT can be applied with insufficient domain knowledge.

Finally, we corroborate the *effectiveness*, *efficiency*, and *robustness* of ADVERT through a case study based

on the data set collected from human subject experiments conducted at New York University [5]. The results show that the visual aids can sufficiently enhance the attention level and improve the accuracy of phishing recognition from 74.6% to a minimum of 86%. When we further tune the hyperparameters, we manage to improve the accuracy of phishing recognition from 86.8% to 93.7% in less than 50 tuning stages, while the largest accuracy improvement happens within 3 tuning stages. The results have also provided insights and guidance for the ADVERT design; e.g., the attention threshold for visual-aid selection (resp. the period length for visual-aid generation) has a small (resp. periodic) impact on phishing recognition.

A. Notations and Organization of the Paper

Throughout the paper, we use subscripts to index time and stages. Calligraphic letter $\mathscr{S}$ defines a set and $|\mathscr{S}|$ represents its cardinality. The indicator function $\mathbf{1}_{\{A\}}$ takes value 1 if condition A is true and value 0 if A is false. The rest of the paper is organized as follows. The related works are presented in Section II. We elaborate on the two feedback loops of Fig. 1 in Section III and IV, respectively. Section V presents a case study of ADVERT for email vetting and phishing recognition. Section VI discusses the limitations, and Section VII concludes the paper.

II. Related works

A. Phishing Attack Detection and Prevention

Phishing is the act of masquerading as a legitimate entity to serve malware or steal credentials. The authors in [6] have identified three human vulnerabilities that make humans the unwitting victims of phishing.

- Lack of knowledge for computer system security; e.g., www.ebay-members-security.com does not belong to www.ebay.com.
- Inadequacy to identify visual deception; e.g., the phishing email can contain an image of a legitimate hyperlink, but the image itself serves as a hyperlink to a malicious site. A human cannot identify the deception by merely looking at it.
- Lack of attention (e.g., careless users fail to notice the phishing indicators, including spelling errors and grammar mistakes) and *inattentional blindness* (e.g., users focusing on the main content fail to perceive unloaded logos in a phishing email [7]).

Many works have attempted to mitigate the above three human vulnerabilities to prevent phishing attacks. First, security education and anti-phishing training, e.g., role-playing phishing simulation games [8] and fake phishing attacks [9], have been used to compensate for the user's lack of security knowledge and increase users' security awareness. Second, detection techniques based on visual similarities [10] and machine learning [11] have been applied to help users identify visual deception. Modern web browsers and email clients also provide security indicators (e.g., the protocol used, the domain name, and the SSL/TLS certificate) to assist users in decision-making [12]. Third, passive warnings (i.e., do not block the content-area) and active warnings (i.e., prohibits the user from viewing the content-data) have been developed empirically to draw users' attention and prevent them from falling victim to phishing [11], [13]. Our work lays out a foundation to compensate for the third human vulnerability of inattention systematically and quantitatively.

B. Counterdeception Technologies

Adversarial cyber deception has been a long-standing problem. It is easy for an attacker to deceive yet much more difficult for regular users to identify the deception given the universal human vulnerabilities. Previous works have mainly focused on *human solutions* (e.g., security training [1]) or *technical solutions* (e.g., defensive deception technologies [14]–[16]), to deter, detect, and respond to deceptive attacks. This work focuses on designing a *human-technical solution* through eye-tracking data, visual aids, and learning techniques to counteract adversarial cyber deception.

Biosensors, including eye trackers and electroencephalogram (EEG) devices, provide a window into an analytical understanding of human perception and cognition to enhance security and privacy [17]. In particular, researches have investigated the users' gaze behaviors and attention when reading Uniform Resource Locators (URLs) [18], phishing webs [19], and phishing emails [5], [20], [21]. These works illustrate the users' visual processing of phishing contents [18]–[20], [22] and the effects of visual aids [21]. The authors in [19] further establish correlations between eye movements and phishing identification to estimate the likelihood that users may fall victim to phishing attacks. Compared to these works that *analyze* human perception, we use eye-tracking data to *design* visual aids and *modify* the human perception process for better security decisions. Moreover, we use biometric data at different granularities. Compared to previous works that exploit the *statistics* of the biometric data (e.g., the number of fixations and gaze duration distributions), we use the *dynamic transitions* of the eye-tracking data to extract attention metrics for corrective measures in *real-time*.

C. Human Vulnerability Quantification and Learning

Human plays significant roles in cybersecurity. It is challenging to model, quantify, and affect human behaviors and their mental processes such as reasoning, perception, and cognition. Therefore, various modeling and

learning approaches are developed to mitigate human vulnerabilities in cyberspace, as shown in the following two paragraphs, respectively.

The authors in [23], [24] use Signaling Detection Theory (SDT) to quantify phishing susceptibility and prioritize behavioral interventions for reducing phishing risk, respectively. Adopting SDT, they treat the phishing risk management as a *vigilance task*, where individuals monitor their environment to distinguish signals (i.e., phishing emails) from noises (i.e., legitimate emails). Their approaches investigate phishing on a detailed level based on varying factors, including task, individual, and environmental ones. We adopt a system-level characterization, where system-scientific tools such as feedback, Reinforcement Learning (RL), and BO are used to adapt to these varying factors.

Due to the modeling challenges and the unpredictability, RL [25] has been used to characterize and mitigate human vulnerabilities, including bounded rationality [26], prospect theory [27], incompliance [3], and bounded attention [28], [29]. Using RL to detect, evaluate, and compensate for risks induced by human vulnerabilities is still in its infancy, but it is a promising direction as RL provides a quantitative and adaptive solution.

III. Attention Enhancement Mechanism

As illustrated by Step 1 of Fig. 1, we consider a group of M human users who vet a list of N emails and classify them as phishing or legitimate. As a user $m \in \mathscr{M} := \{1,\cdots,M\}$ reads an email $n \in \mathscr{N} := \{1,\cdots,N\}$ on the screen for a duration of T_m^n, the eye-tracking device records the vertical and the horizontal coordinates of his eye gaze point in real-time. To compress the sensory outcomes and facilitate RL-driven attention enhancement solutions, we aggregate potential gaze locations (i.e., pixels on the screen) into a finite number of I non-overlapping Areas of Interest (AoIs) as shown in Fig. 2. We index each potential AoI by $i \in \mathscr{I} := \{1,2,...,I\}$.

Each email does not need to contain all the AoIs, and the AoI partition remains unknown to the users. Previous works [18]–[20] have identified the role of AoIs in helping human users recognize phishing, and different research goals can lead to different AoI partitions. For example, the main content AoI (i.e., area 5 in Fig. 2) can be divided into finer AoIs based on the phishing indicators such as misspellings, grammar mistakes, and threatening sentences. We refer to all other areas in the email (e.g., blank areas) as the *uninformative area*. When the user's eyes move off the screen during the email vetting process, no coordinates of the gaze location are available. We refer to these off-screen areas as the *distraction area*.

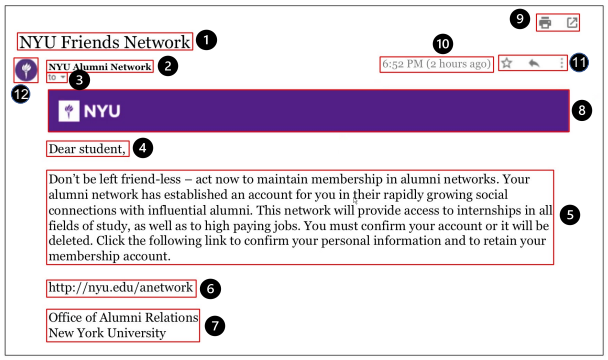


Fig. 2: A sample email with 12 AoIs. In sequence, they are the email's title, the sender's information, the receiver's information, the salutation, the main content, the URL, the sender's signature, the organization logo, the 'print' and 'share' buttons, the timestamp, the 'bookmark' and 'forward' buttons, and the sender's profile picture. The AoI partition in red boxes and their index numbers in black circles are invisible to users.

A. Visual State Transition Model

As illustrated by Step 2 in Fig. 1, we establish the following transition model based on the AoI to which the user's gaze location belongs at different times. We define $\mathscr{S} := \{s^i\}_{i\in\mathscr{I}} \cup \{s^{ua}, s^{da}\}$ as the set of $I+2$ *Visual States (VSs)*, where s^i represents the i-th AoI; s^{ua} represents the *uninformative area*; and s^{da} represents the *distraction area*. We provide an example transition map of these VSs in Fig. 3. The links represent the potential shifts of the

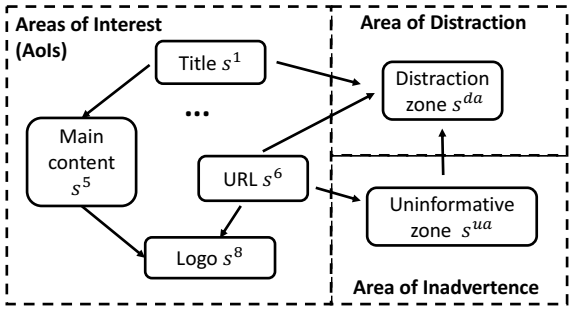


Fig. 3: Transitions among VSs in $\mathscr{S}$. The VS indices are consistent with the AoI indices in Fig. 2.

gaze locations during the email reading process; e.g., a user can shift his focus from the title to the main content or the distraction area. We omit most links for illustration purposes; e.g., it is also possible for a user to regain attention to the AoIs from distraction or inadvertence.

We denote $s_t \in \mathscr{S}$ as the VS of user $m \in \mathscr{M}$ vetting email $n \in \mathscr{N}$ at time $t \in [0, T_m^n]$. In this work, we do not distinguish among human users concerning their attention processes while they read different emails. Then,

each user's gaze path during the interval $[0, T_m^n]$ can be characterized as the same stochastic process $[s_t]_{t\in[0,T_m^n]}$. The stochastic transition of the VSs divides the entire time interval $[0, T_m^n]$ into different *transition stages*. We visualize an exemplary VS transition trajectory $[s_t]_{t\in[0,T_m^n]}$ in Fig. 4 under $I = 4$ AoIs and $T_m^n = 50$ seconds. As denoted by the colored squares, 40 VSs arrive in sequence, which results in 40 discrete transition stages.

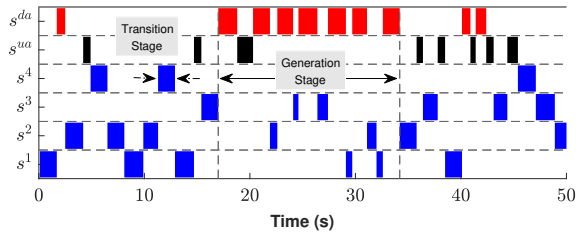


Fig. 4: An exemplary VS transition trajectory $[s_t]_{t\in[0,T_m^n]}$. The x-axis and the y-axis represent $T_m^n = 50$ seconds and $I + 2 = 6$ VSs, respectively. We denote VSs s^{da}, s^{ua}, and $\{s^i\}_{i\in\mathscr{I}}$ in red, black, and blue, respectively. Each generation stage can contain different numbers of transition stages.

B. Feedback Visual-Aid Design

Propel visual aids can help guide and sustain the users' attention. Previous works have proposed different classes of visual aids to enhance phishing recognition, including highlights of contents [21], [30], warnings of suspicious hyperlinks and attachments [13], [31], and anti-phishing educational messages [32]. These potential classes of visual aids construct the visual-aid library denoted as a finite set $\mathscr{A}$.

As illustrated by Step 6 in Fig. 1, different visual aids can affect the users' visual behaviors. The influence, however, can be beneficial (e.g., timely highlights prevent users from mind-wandering) or detrimental (e.g., extensive highlights make humans weary and less attentive to the AoIs). The effectiveness of visual aids for preventing phishing may not be straightforward, especially under different environmental (e.g., security indicator designs) and human factors (e.g., users' security knowledge and prior trust) [12]. In this paper, we focus on adapting visual aids to the human visual attention. We apply RL to learn the dynamic design of visual aids based on the real-time evaluation of the user's attention status detailed in Section III-C.

The sequence of adaptive visual aids is generated with a period of length T^{pl}, and we refer to the time interval between every two visual aids as the *generation stage* indexed by $k \in \mathscr{K}_m^n := \{1, 2, \cdots, K_m^n\}$, where K_m^n is the maximum generation stage during $[0, T_m^n]$; i.e., $K_m^n T^{pl} \leq T_m^n$ and $(K_m^n + 1)T^{pl} \geq T_m^n$. Then, we denote $a_k \in \mathscr{A}$ as the visual aid at the k-th generation stage. Fig. 4 illustrates how visual aids affect the transition of VSs in $K_m^n = 3$ generation stages divided by the two vertical dashed lines. During the second generation stage, an improper visual aid leads to more frequent transitions to the distraction area and also a longer sojourn time at the VS s^{da}. On the contrary, the proper visual aids during the first and the third generation stages engage the users and extend their attention spans, i.e., the amount of time spent on AoIs before a transition to s^{da} or s^{ua}.

C. Evaluation of Attention Status

From the VS transition trajectory (e.g., Fig. 4), we aim to construct the *Attention State (AS)* used as the feedback value for the adaptive visual-aid design. We define $\mathscr{X}$ as the set of all possible attention states. Previous works (e.g., [20], [22]) have defined attention metrics based on the AoIs, including the proportion of time spent on each AOI, gaze duration means, fixation count, and average duration. Compared to these *detailed-level* metrics extracted directly from raw eye-gaze data, we propose the following *system-level* metric of attention level based on the VS transition history as will be shown in Section III-C2. Such system-level metric serves as sufficient statistics to effectively characterize the attention status. Moreover, it preserves the users' privacy because the raw data of gaze locations can reveal sensitive information about their biometric identities, including gender, age, and ethnicity [33], [34].

To this end, we assign scores to each VS in Section III-C1 to evaluate the user's attention (e.g., gaze at AoIs) and inattention (e.g., gaze at uninformative and distraction areas). The scores can be determined manually based on the expert recommendation and empirical studies (e.g., [22]), or based on other biometric data (e.g., the pupil sizes in Fig. 8). Moreover, we can apply BO for further fine-tuning of these scores as shown in Section IV-B.

1) Concentration Scores and Decay Rates: Both the gaze location and the gaze duration matter in the identification of phishing attacks. For example, at the first glance, users cannot distinguish the spoofed email address 'paypa1@mail.paypaI.com' from the authentic one 'paypal@mail.paypal.com' while a guided close look reveals that the lower case letter 'l' is replaced by the number '1' and the capital letter 'I'. Therefore, we assign a *concentration score* $r^{co}(s) \in \mathbb{R}$ to characterize the sustained attention associated with VS $s \in \mathscr{S}$. Since the amount of information that a user can extract from a VS $s \in \mathscr{S}$ is limited, we use an exponential decay rate of $\alpha(s) \in \mathbb{R}^+$ to penalize the effect of concentration score as time elapses. Different VSs can have different concentration scores and decay rates. For example, the main

content AoI (i.e., area 5 in Fig. 2) usually contains more information than other AoIs, and an extended attention span extracts more information (e.g., the substitution of letter 'l' into 'I') to identify the phishing email. Thus, the main content AoI turns to have a high concentration score and a low decay rate, which is corroborated in Table I based on the data set collected from human experiments [5] as will be shown in Section V.

2) Cumulative Attention Level: We construct the metric for attention level illustrated by Step 3 in Fig. 1 as follows. Let $W_k \in \mathbb{Z}^+$ be the total number of transition stages contained in generation stage $k \in \mathscr{K}_m^n$. Then, we define $t_k^{w_k}, w_k \in \{1,2,\cdots,W_k\}$, as the duration of the w_k-th transition stage in the k-th generation stage. Take the gaze path in Fig. 4 as an example, the first generation stage contains $w_1 = 12$ transition stages and the first 7 transition stages last for a total of $\sum_{w_1=1}^{7} t_1^{w_1} = 10$ seconds. Based on the sets of scores associated with $s \in \mathscr{S}$, we compute the cumulative reward $u_k^{w_k}(s,t)$ at time t of the w_k-th transition stage in the k-th generation stage as $u_k^{w_k}(s,t) = \int_0^t r^{co}(s) e^{-\alpha(s)\tau} \cdot \mathbf{1}_{\{s=s^\tau\}} d\tau, 0 \le t \le t_k^{w_k}$. At generation stage k, we define $\bar{w}_k^t$ as the latest transition stage before time t, i.e., $\sum_{w_k=1}^{\bar{w}_k^t} t_k^{w_k} \le t$ and $\sum_{w_k=1}^{\bar{w}_k^t+1} t_k^{w_k} > t$. Then, we define the user's *Cumulative Attention Level (CAL)* $v_k(t)$ over time interval $[(k-1)T^{pl}, t]$ at generation stage $k \in \mathscr{K}_m^n$ as the following cumulative reward

$$v_k(t) := \sum_{s\in\mathscr{S}} \sum_{w_k=1}^{\bar{w}_k^t} u_k^{w_k}(s,t), 0 \le t \le T^{pl}, \tag{1}$$

We visualize the CAL of $K_m^n = 3$ generation stages in Fig. 5 based on the gaze path in Fig. 4.

Since $v_k(t)$ is bounded for all $k \in \mathscr{K}_m^n, t \in [0, T^{pl}]$, we can quantize it into X finite values to construct the set $\mathscr{X}$ of the attention states illustrated by Step 4 in Fig. 1. We represent the quantized value of $v_k(t) \in \mathbb{R}$ as $v_k^{qu}(t) \in \mathscr{X}$ for all $k \in \mathscr{K}_m^n, t \in [0, T^{pl}]$, and define the Average Attention Level (AAL) and Quantized Average Attention Level (QAAL) for each generation stage in Definition 1.

Definition 1. *Let $\bar{v}_k \in \mathbb{R}$ and $\bar{v}_k^{qu} \in \mathscr{X}$ denote the user's Average Attention Level (AAL) and Quantized Average Attention Level (QAAL) over generation stage $k \in \mathscr{K}_m^n$, respectively. They are measured by the improvement in CAL and the quantized value of the CAL improvement per unit time, i.e., $\bar{v}_k := v_k(T^{pl})/T^{pl}$ and $\bar{v}_k^{qu} := v_k^{qu}(T^{pl})/T^{pl}$, respectively.*

D. Q-Learning via Consolidated Data

In Section III-D, We elaborate on the adaptive learning block (i.e., Step 5 in Fig. 1). Since the inspection time of a user reading one email is not sufficiently long, we consolidate a group of email inspection data to learn the optimal visual-aid generation policy over a population.

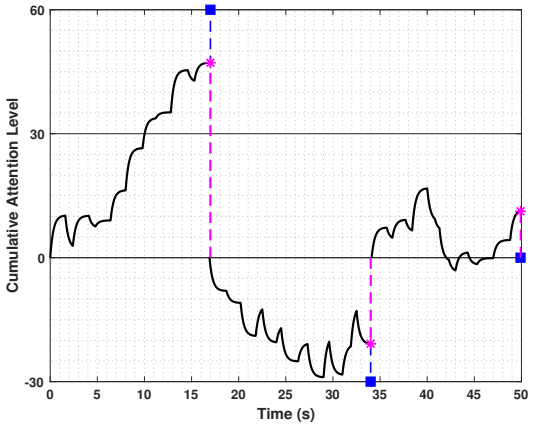


Fig. 5: The user's cumulative attention level $v_k(t-(k-1)T^{pl}), k \in \mathscr{K}_m^n, t \in [(k-1)T^{pl}, kT^{pl}]$, over $K_m^n = 3$ generation stages in $T_m^n = 50$ seconds. The horizontal lines quantize $v_k(t)$ into $X = 4$ values that form the finite set $\mathscr{X} = \{-30, 0, 30, 60\}$. The purple star and the blue square denote the values of $\bar{v}_k \cdot T^{pl}$ and $\bar{v}_k^{qu} \cdot T^{pl}$, respectively, at each generation stage $k \in \mathscr{K}_m^n$.

The QAAL $\bar{v}_k^{qu} \in \mathscr{X}$ represents the attention state at the generation stage $k \in \mathscr{K}_m^n$. Since the goal is to enhance the user's attention represented by the CAL, the reward function $R : \mathscr{X} \times \mathscr{A} \mapsto \mathbb{R}$ should be monotone concerning the value of $\bar{v}_k^{qu}$, e.g., $R(\bar{v}_k^{qu}, a_k) := \bar{v}_k^{qu}, \forall a_k \in \mathscr{A}$. In this work, we assume that each visual aid $a_k \in \mathscr{A}$ exerts the same statistical effect on the attention process regardless of different users and emails. Thus, we can consolidate the data set of $\bar{M} \in \{1,\cdots,M\}$ users and $\bar{N} \in \{1,\cdots,N\}$ emails[2] to learn the optimal visual-aid generation policy $\sigma \in \Sigma : \mathscr{X} \mapsto \mathscr{A}$ in a total of $\bar{K} := \sum_{m=1}^{\bar{M}} \sum_{n=1}^{\bar{N}} K_m^n$ stages. With a given discounted factor $\beta \in (0,1)$, the expected long-term objective can be represented as $\max_{\sigma\in\Sigma} \mathbb{E}[\sum_{k=1}^{\bar{K}} (\beta)^k \cdot R(\bar{v}_k^{qu}, \sigma(\bar{v}_k^{qu}))]$.

The Q-table $[Q_k(\bar{v}_k^{qu}, a_k)]_{\bar{v}_k^{qu}\in\mathscr{X}, a_k\in\mathscr{A}}$ represents the user's attention pattern at generation stage $k \in \bar{\mathscr{K}} := \{1,\cdots,\bar{K}\}$, i.e., the estimated payoff of applying visual aid $a_k \in \mathscr{A}$ when the attention state is $\bar{v}_k^{qu} \in \mathscr{X}$. Let the sequence of learning rate $\gamma_k(\bar{v}_k^{qu}, a_k)$ satisfy $\sum_{k=0}^{\infty} \gamma_k(\bar{v}_k^{qu}, a_k) = \infty$ and $\sum_{k=0}^{\infty} (\gamma_k(\bar{v}_k^{qu}, a_k))^2 < \infty$ for all $\bar{v}_k^{qu} \in \mathscr{X}, a_k \in \mathscr{A}$. Then, we can update the attention pattern at each generation stage $k \in \bar{\mathscr{K}}$ as follows, i.e.,

$$\begin{aligned} Q_{k+1}(\bar{v}_k^{qu}, \sigma_k(\bar{v}_k^{qu})) &= Q_k(\bar{v}_k^{qu}, \sigma_k(\bar{v}_k^{qu})) \\ &+ \gamma_k(\bar{v}_k^{qu}, \sigma_k(\bar{v}_k^{qu})) \cdot [R(\bar{v}_k^{qu}, \sigma_k(\bar{v}_k^{qu})) \\ &+ \beta \max_{a\in\mathscr{A}} Q_k(\bar{v}_{k+1}^{qu}, a) - Q_k(\bar{v}_k^{qu}, \sigma_k(\bar{v}_k^{qu}))], \end{aligned} \tag{2}$$

[2]When sufficiently large data sets are available, we can carefully choose these $\bar{M}$ users to share similar attributes (e.g., ages, sexes, races, etc.) and these $\bar{N}$ emails to belongs to the same categories (e.g., business or personal emails).

where the visual-aid generation policy $\sigma_k(\bar{v}_k^{qu})$ at generation stage $k \in \bar{\mathscr{K}}$ is an ε_k-greedy policy; i.e., with probability $\varepsilon_k \in [0,1]$, the visual aid a_k is selected randomly from $\mathscr{A}$ and with probability $1-\varepsilon_k$, the optimal visual aid $a_k^* \in \arg\max_{a\in\mathscr{A}} Q_k(\bar{v}_k^{qu}, a)$ is implemented. To obtain a convergent visual-aid generation policy, the value of ε_k gradually decreases from 1 to 0.

IV. Phishing Prevention Mechanism

The attention enhancement mechanism in Section III tracks the attention process in real-time to enable the adaptive visual-aid generation. By properly modifying the user's attention and engaging him in vetting emails, the attention enhancement mechanism serves as a stepping-stone to achieving the ultimate goal of phishing prevention. Empirical evidence and observations have shown that a high attention level, or mental arousal, does not necessarily yield good performance [35]. In the specific task of phishing recognition, recent works [36], [37] have also identified curvilinear relationships between phishing recognition accuracy and critical attentional factors, including a participant's cue utilization, cognitive reflection, and cognitive load. Thus, besides attention metrics, e.g., the AAL, we need to design anti-phishing metrics to measure the users' performance of phishing recognition as will be shown in Section IV-A.

In Section IV-B, we develop an efficient meta-level algorithm to tune the hyperparameters (e.g., the period length T^{pl} of the visual-aid generation, the number of attention states X, the attention scores $r^{co}(s), \alpha(s), \forall s \in \mathscr{S}$, etc.) in the attention enhancement mechanism. We denote these hyperparameters as one d-dimensional variable $\theta = [T^{pl}, X, [r^{co}(s)]_{s\in\mathscr{S}}, [\alpha(s)]_{s\in\mathscr{S}}] \in \mathbb{R}^d$, where $d = 2+2|\mathscr{S}|$. Let the i-th element θ^i be upper and lower bounded by $\bar{\theta}^i$ and $\underline{\theta}^i$, respectively. Thus, $\theta \in \Theta^d := \{[\theta^i]_{i\in\{1,\cdots,d\}} \in \mathbb{R}^d | \underline{\theta}^i \le \theta^i \le \bar{\theta}^i\}$.

A. Metrics for Phishing Recognition

As illustrated by Step 7 in Fig. 1, we provide a metric to evaluate the outcome of the users' phishing identification under a given hyperparameter $\theta \in \Theta^d$. After vetting email $n \in \{1,\cdots,\bar{N}\}$, the user $m \in \{1,\cdots,\bar{M}\}$ judges the email to be phishing or legitimate. The binary variable $z_m^n(\theta) \in \{z^{co}, z^{wr}\}$ represents whether the judgment is correct (denoted by z^{co}) or not (denoted by z^{wr}). We can reshape the two-dimension index (m,n) as a one-dimension index $\hat{n}$ and rewrite $z_m^n(\theta)$ as $z_{\hat{n}}(\theta)$. Once these users have judged in total of N^{bo} emails, we define the following metric $c^{ac} \in \mathscr{C} : \Theta^d \mapsto [0,1]$ to evaluate the *accuracy* of phishing recognition, i.e.,

$$c^{ac}(\theta) := \frac{1}{N^{bo}} \sum_{\hat{n}=1}^{N^{bo}} |\mathbf{1}_{\{z_{\hat{n}}(\theta)=z^{co}\}}|, \forall \theta \in \Theta^d. \tag{3}$$

The goal is to find the optimal hyperparameter $\theta^* \in \Theta^d$ to maximize the accuracy of phishing identification, i.e., $\theta^* \in \arg\max_{\theta\in\Theta^d} c^{ac}(\theta)$. However, we cannot know the value of $c^{ac}(\theta)$ for a $\theta \in \Theta^d$ a priori until we implement this hyperparameter θ in the attention enhancement mechanism. The implemented hyperparameter affects the adaptive visual-aid generation that changes the user's attention and the anti-phishing performance metric $c^{ac}(\theta)$. Since the experimental evaluation at a given $\theta \in \Theta^d$ is time-consuming, we present an algorithm in Section IV-B to determine how to choose and update the hyperparameter to maximize the detection accuracy.

B. Efficient Hyperparameter Tuning

We illustrate the meta-adaptation (i.e., Step 8 in Fig. 1) in Section IV-B. As illustrated in Fig. 6, we refer to the duration of every N^{bo} security decisions as a *tuning stage*. Consider a time and budget limit that restricts us to conduct L tuning stages in total. We denote θ_l as the hyperparameter at the l-th tuning stage where $l \in \mathscr{L} := \{1,2,\cdots,L\}$. Since each user's email inspection time is different, each tuning stage can contain different numbers of generation stages.

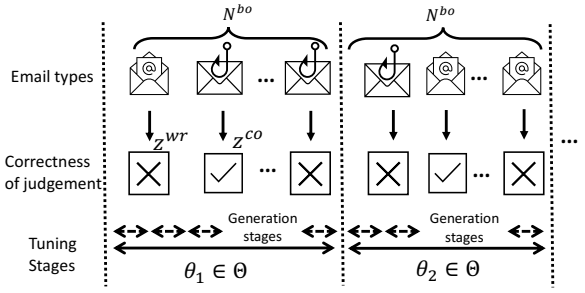


Fig. 6: Hyperparameter tuning based on the user's phishing recognition. Each tuning stage consists of N^{bo} emails and contains several generation stages.

To find the optimal hyperparameter $\theta^* \in \Theta^d$ within L tuning stages is challenging. The empirical methods (e.g., a naive grid search or a random search over $\Theta^d \subset \mathbb{R}^d$) become inefficient when $d > 1$. BO [38] provides a systematic way to update the hyperparameter and balance between exploration and exploitation. BO consists of a Bayesian statistical model of the objective function $c^{ac} \in \mathscr{C}$ and an acquisition function for deciding the hyperparameter to implement at the next tuning stage. The statistical model of $c^{ac} \in \mathscr{C}$ is a Gaussian process $\mathscr{N}(\mu^0, \Sigma^0)$ with a mean function $\mu^0(\theta) = \bar{\mu}^0$ and covariance function or kernel $\Sigma^0(\theta, \bar{\theta}) = \lambda^0 \cdot \exp(\sum_{i=1}^d \lambda^i (\theta^i - \bar{\theta}^i)^2)$ for all $\theta, \bar{\theta} \in \Theta^d$, where $\bar{\mu}^0$, λ^0 and $\lambda^i, i \in \{1,2,\cdots,d\}$, are parameters of the kernel. The kernel Σ^0 is required to be positive semi-definite and has the property that the points closer

in the input space are more strongly correlated. For any $l \in \mathscr{L}$, we define three shorthand notations $\mu^0(\theta_{1:l}) := [\mu^0(\theta_1), \cdots, \mu^0(\theta_l)]$, $c^{ac}(\theta_{1:l}) := [c^{ac}(\theta_1), \cdots, c^{ac}(\theta_l)]$, and

$$\Sigma^0(\theta_{1:l}, \theta_{1:l}) := \begin{bmatrix} \Sigma^0(\theta_1, \theta_1) & \cdots & \Sigma^0(\theta_1, \theta_l) \\ \vdots & \ddots & \vdots \\ \Sigma^0(\theta_l, \theta_1) & \cdots & \Sigma^0(\theta_l, \theta_l) \end{bmatrix}.$$

Then, the evaluation vector of $l \in \mathscr{L}$ elements is assumed to be multivariate Gaussian distributed, i.e., $c^{ac}(\theta_{1:l}) \sim \mathscr{N}(\mu^0(\theta_{1:l}), \Sigma^0(\theta_{1:l}, \theta_{1:l}))$. Conditioned on the values of $\theta_{1:l}$, we can infer the value of $c^{ac}(\theta)$ at any other $\theta \in \Theta \setminus \{\theta_{l'}\}_{l' \in \{1,\cdots,l\}}$ by Bayesian rule, i.e.,

$$c^{ac}(\theta) | c^{ac}(\theta_{1:l}) \sim \mathscr{N}(\mu^n(\theta), (\Sigma^n(\theta))^2), \tag{4}$$

where $\mu^n(\theta) = \Sigma^0(\theta, \theta_{1:l}) \cdot \Sigma^0(\theta_{1:l}, \theta_{1:l})^{-1} \cdot (c^{ac}(\theta_{1:l}) - \mu^0(\theta_{1:l})) + \mu^0(\theta)$ and $(\Sigma^n(\theta))^2 = \Sigma^0(\theta, \theta) - \Sigma^0(\theta, \theta_{1:l}) \cdot \Sigma^0(\theta, \theta_{1:l})^{-1} \cdot \Sigma^0(\theta_{1:l}, \theta)$.

We adopt *expected improvement* as the acquisition function. Define $c_l^* := \max_{l' \in \{1,\cdots,l\}} c^{ac}(\theta_{l'})$ as the optimal evaluation among the first l evaluations and a shorthand notation $(c^{ac}(\theta) - c_l^*)^+ := \max\{c^{ac}(\theta) - c_l^*, 0\}$. For any $l \in \mathscr{L}$, we define $\mathbb{E}_l[\cdot] := \mathbb{E}[\cdot | c^{ac}(\theta_{1:l})]$ as the expectation taken under the posterior distribution of $c^{ac}(\theta)$ conditioned on the values of l evaluations $c^{ac}(\theta_{1:l})$. Then, the expected improvement is $\mathrm{EI}_l(\theta) := \mathbb{E}_l[(c^{ac}(\theta) - c_l^*)^+]$. The hyperparameter at the next tuning stage is chosen to maximize the expected improvement at the current stage, i.e,

$$\theta_{l+1} \in \arg\max_{\theta \in \Theta^d} \mathrm{EI}_l(\theta). \tag{5}$$

The expected improvement can be evaluated in a closed form, and (5) can be computed inexpensively by gradient methods [38].

At the first $L^0 \in \{1, 2, \cdots, L\}$ tuning stages, we choose the hyperparameter $\theta_l, l \in \{1, 2, \cdots, L^0\}$, uniformly from Θ^d. We can use the evaluation results $c^{ac}(\theta_l), l \in \{1, 2, \cdots, L^0\}$, to determine the parameters $\bar{\mu}^0, \lambda^0$, and $\lambda^i, i \in \{1, 2, \cdots, d\}$, by Maximum Likelihood Estimation (MLE); i.e., we determine the values of these parameters so that they maximize the likelihood of observing the vector $[c^{ac}(\theta_{1:L^0})]$. For the remaining $L - L^0$ tuning stages, we choose $\theta_l, l \in \{L^0, L^0 + 1, \cdots, L\}$, in sequence as summarized in Algorithm 1.

Algorithm 1: Hyperparameter tuning via BO.

Implement the initial L^0 evaluations $c^{ac}(\theta_l), l \in \{1, 2, \cdots, L^0\}$;
Place a Gaussian process prior on $c^{ac} \in \mathscr{C}$, i.e., $c^{ac}(\theta_{1:L^0}) \sim \mathscr{N}(\mu^0(\theta_{1:L^0}), \Sigma^0(\theta_{1:L^0}, \theta_{1:L^0}))$;
for $l \leftarrow L^0$ **to** L **do**
  Obtain the posterior distribution of $c^{ac}(\theta)$ in (4) based on the existing l evaluations;
  Compute $\mathrm{EI}_l(\theta), \forall \theta \in \Theta^d$, based on the posterior distribution;
  Determine θ_{l+1} via (5);
  Implement θ_{l+1} at the next tuning stage $l+1$ to evaluate $c^{ac}(\theta_{l+1})$;
end
Return the maximized value of all observed samples, i.e., $\theta^* \in \arg\max_{\theta_l \in \{\theta_1, \cdots, \theta_L\}} c^{ac}(\theta_l)$;

V. Case Study

In this case study, we verify the effectiveness of ADVERT via a data set collected from human subject experiments conducted at New York University [5]. We elaborate on the experiment setup and the data processing procedure in Section V-A. Based on the features obtained from the data set, we generate synthetic data under adaptive visual aids to demonstrate the proposed attention enhancement mechanism and the phishing prevention mechanism in Section V-B and V-C, respectively.

A. Experiment Setting and Data Processing

The data set involves $M = 160$ undergraduate students ($n_{\text{White}} = 27$, $n_{\text{Black}} = 19$, $n_{\text{Asian}} = 64$, $n_{\text{Hispanic/Latinx}} = 17$, $n_{\text{other}} = 33$) who are asked to vet $N = 12$ different emails (e.g., the email of NYU friends network in Fig. 2) separately and then give a rating of how likely they would take actions solicited in the emails (e.g., maintain membership in Fig. 2). When presented to different participants, each email is described as either posing a cyber threat or risk-free legitimate opportunities to investigate how the above description affects the participants' phishing recognition.

While the participants vet the emails, the Tobii Pro T60XL eye-tracking monitor records their eye locations on a 1920×1200 resolution screen and the current pupil diameters of both eyes with a sampling rate of 60Hz. Fig. 7 illustrates the time-expanded eye-gaze trajectory of a participant vetting the sample email in Fig. 2. The z-coordinate of a 3D point (x, y, z) represents the time when the participant gazes at the pixel (x, y) in the email area. The participant's eye gaze locations move progressively from the points in warmer color to the ones in cooler color. Fig. 7 illustrates the zigzag pattern of the participant's eye-gaze trajectory; i.e., the participant reads emails from left to right and top to bottom. The participant starts with the title, spends the majority of time on the main content, and glances at other AoIs (e.g., the links and the signatures). There is also a small chance of revisiting the email content and looking outside the email area.

Fig. 8 illustrates the participant's pupil sizes of left and right eyes in red and blue, respectively, concerning the

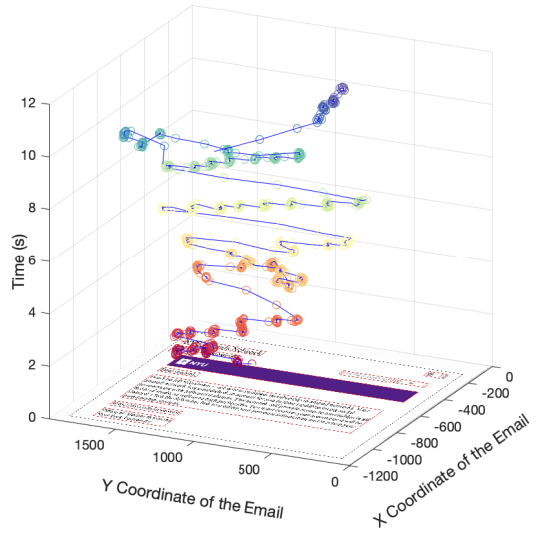


Fig. 7: A time-expanded plot of a typical eye-gaze trajectory with a sampling rate of 60 Hz. The x-y plane (in the unit of pixels) represents the email area. The z-axis represents the time (in the unit of seconds) of the participant's eye-gaze trajectory. The warmer color indicates a smaller value on the z-axis (i.e., an earlier gaze of the point).

same trial of the data set to generate Fig. 7. At different times, the average of the pupil diameters (resp. gaze locations) of the right and left eyes represent the pupil size (resp. gaze location). Following Section III-A, we obtain the 15 VSs illustrated by the grey squares in Fig. 8 based on the gaze locations of the email pixels in Fig. 7. Since the covert eye-tracking system does not require head-mounted equipment or chinrests, the tracking can occur without the participants' awareness. We refer the reader to the supplement materials of [5] for the survey data and the details of the experimental procedure[3].

1) Estimate Concentration Scores and Decay Rates based on Pupil Sizes: Empirical works in [39], [40] have demonstrated that pupils dilate as a consequence of attentional efforts. Building on the findings, we assume that the average pupil diameters of both eyes at time t of the generation stage $k \in \mathscr{K}_m^n$ is approximately proportional to the participant's attention level $\frac{dv_k}{dt}(t)$ at time t. We obtain the benchmark values of $r^{co}(s), \alpha(s), \forall s \in \mathscr{S}$, in Table I by minimizing the Mean Square Error (MSE) between the CAL in Section III-C and the cumulative pupil size through global optimization methods such as Simulated Annealing (SA) [41]. The results in Table I corroborate that the main content AoI $s^5 \in \mathscr{S}$ has the highest concentration score and the lowest decay rate.

[3]The processed data used in this manuscript, including the temporal transitions of AoIs and the pupil sizes, is available at https://osf.io/4y32d/. The raw eye-tracking data in the format of videos are available upon request.

AoIs	Meaning	$r^{co}(s^i)$	$\alpha(s^i)$
s^1	Title	9.48	2.17
s^2	Sender	3.55	4.04
s^3	Receiver	7.62	0.22
s^4	Salutation	13.76	0.57
s^5	Main Content	21.05	0.16
s^6	URL	7.84	10.90
s^7	Signature	6.47	5.46
s^8	Logo	6.44	5.16
s^9	Print& Share	4.86	13.91
s^{10}	Time	3.81	6.68
s^{11}	Bookmark& Forward	7.34	2.19
s^{12}	Profile	7.26	2.02
s^{13}	Attachment	4.74	3.46

TABLE I: The concentration score $r^{co}(s^i)$ and decay rate $\alpha(s^i)$ for $I = 13$ AoIs.

2) Synthetic VS Trajectory Generation under Visual Aids: In the case study, we consider $I = 13$ AoIs. The sample email in Fig. 2 illustrates the first 12 AoIs. The 13-th AoI is on the email attachment. Under visual aid $a \in \mathscr{A}$, we denote $P^{i,j}(a)$ as the probability of attention arriving at VS $s^j \in \mathscr{S}$ from VS $s^i \in \mathscr{S}$ and $\phi^i(a)$ as the average sojourn time at VS $s^i \in \mathscr{S}$. We specify the participants' VS transition trajectory $[s_t]_{t\in[0,T_m^n]}, \forall m \in \mathscr{M}, n \in \mathscr{N}$, as a semi-Markov transition process with probability transition matrix $P(a) := [P^{i,j}(a)]_{s^i,s^j\in\mathscr{S}}$ and exponential sojourn distribution of the scale parameter $\phi(a) := [\phi^i(a)]_{s^i\in\mathscr{S}}, \forall a \in \mathscr{A}$.

In particular, we consider a binary set of visual aid $\mathscr{A} = \{a^N, a^Y\}$, where a^N represents the benchmark case without visual aids and a^Y represents the visual aid of highlighting the entire email contents. Based on the VS transition trajectory from the data set, we obtain the probability transition matrix $P(a^N)$ and the sojourn distribution parameter $\phi(a^N)$ under the benchmark case a^N. The transition matrix $P(a^Y)$ and sojourn distribution $\phi(a^Y)$ under visual aid a^Y modify $P(a^N)$ and $\phi(a^N)$ based on the following observations. On the one hand, the visual aid a^Y decreases $P^{i,ua}(a^Y), P^{i,da}(a^Y), \forall s^i \in \mathscr{S}$; i.e., the participants will be guided by the visual aid to pay more frequent attention to the AoIs than the uninformative and distraction areas. On the other hand, the visual aid a^Y decreases $\phi^5(a^Y)$; i.e., the persistent highlighting makes participants weary and reduces their attention spans on the email's main content.

We illustrate $P(a^N)$ and $P(a^Y)$ using heat maps in Fig. 9a and Fig. 9b, respectively. In Fig. 10, we illustrate an exemplary transition trajectory of $I+2$ VSs under a^N and a^Y in blue and red, respectively. The trajectory corroborates that participants under visual aid a^Y incline to pay attention to AoIs yet have less sustained attention. Accurately quantifying the impact of the visual aid on the VS transition depends on many factors [42],

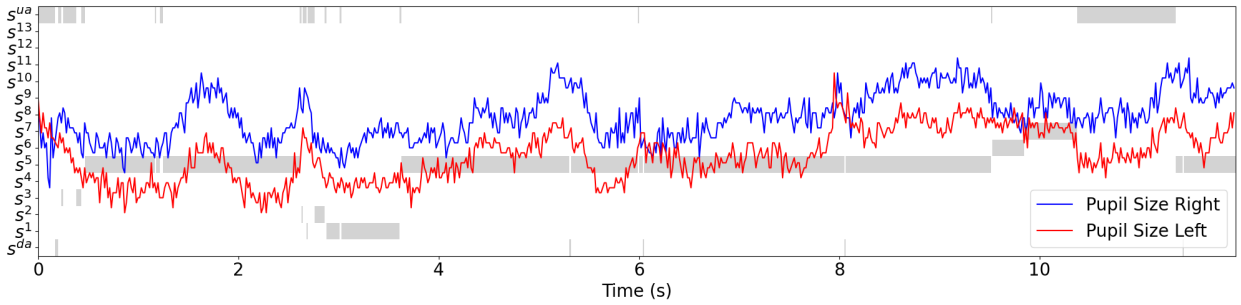
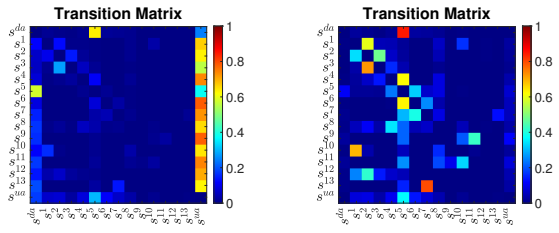


Fig. 8: Gaze locations and pupil sizes collected in the trial of the data set illustrated in Fig. 7. The grey squares illustrate the transition of 15 VSs. The red and blue lines represent the variations of the participant's left and right pupil sizes, respectively, as he reads the email. The x-axis represents the time (in the unit of seconds) during the email inspection.

including the graphic design, the human subject, and the cognitive task. In Section V-A2, we provide one potential estimation of the impact based on the human experiments to illustrate the implementation procedure and the effectiveness of the ADVERT framework.



(a) Under visual aid a^N. (b) Under visual aid a^Y.

Fig. 9: Heat maps of the transition matrices $P(a), a \in \mathscr{A}$. The row and the column represent the source and the destination of the $I+2$ VSs, respectively. Under a^Y, the participants tend to pay attention to AoIs rather than the uninformative and distraction areas.

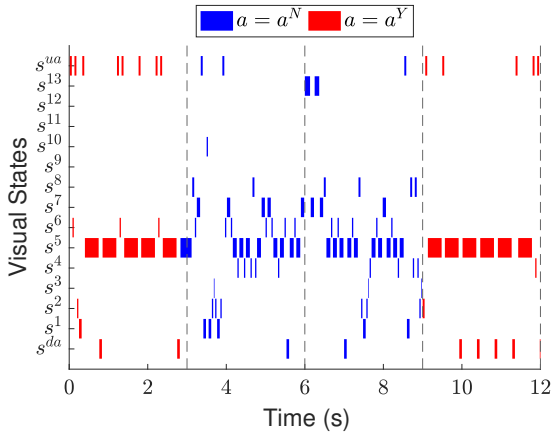


Fig. 10: The VS transition trajectory when the visual aids in four generation stages are a^Y, a^N, a^N, and a^Y, respectively. The inspection lasts for 12 seconds and the period length T^{pl} is 3 seconds.

B. Validation of Attention Enhancement Mechanism

Based on the benchmark attention score in Section V-A1, Fig. 11 illustrates the CAL of the VS transition trajectory shown in Fig. 10. We consider $X = 2$ attention states $\mathscr{X} = \{x^H, x^L\}$ with the *attentive state* x^H and the *inattentive state* x^L. Define $X^{at} \in \mathbb{R}$ as the *attention threshold*. If the AAL at generation stage $k \in \mathscr{K}_m^n$ is higher (resp. lower) than the attention threshold, i.e., $\bar{v}_k \geq X^{at}$ (resp. $\bar{v}_k \leq X^{at}$), then the attention state $x_k \in \mathscr{X}$ at generation k is the attentive state x^H (resp. inattentive state x^L). Fig. 12 further illustrates the impact of visual aids a^N and a^Y on the AAL in red and blue, respectively. The figure demonstrates that a^Y can increase the mean of AAL yet increase its variance.

In Algorithm 2, we present the Q-learning process for participant $m \in \mathscr{M}$ who reads email $n \in \mathscr{N}$ for T_m^n seconds. Define $\eta_k(x,a)$ as the total number of visits to attention state $x \in \mathscr{X}$ and visual aid $a \in \mathscr{A}$ up to generation stage k. Then, we choose the learning rate $\gamma_k(x_k, a_k) = \frac{\eta^0}{\eta_k(x,a)-1+\eta^0}$ for all $x_k \in \mathscr{X}, a_k \in \mathscr{A}$ to guarantee the asymptotic convergence, where $\eta^0 \in (0,\infty)$ is a constant parameter.

Based on the benchmark data set of $M = 160$ participants who inspect $N = 12$ emails in Section V-A, the inspection time $T_m^n, \forall m \in \mathbf{M}, n \in \mathscr{N}$, follows a *Burr distribution*; i.e., its cumulative distribution function is described by $F^{Burr}(t \mid \rho_1, \rho_2, \rho_3) = 1 - \frac{1}{\left(1+(t/\rho_1)^{\rho_2}\right)^{\rho_3}}$ with the scale parameter $\rho_1 = 11.7$, and the shape parameters $\rho_2 = 62.5, \rho_3 = 0.04$. The average inspection time of $M \times N$ samples is 18.7 seconds. During T_m^n seconds of the email vetting process, the eye-tracking device records the participant's gaze locations, which leads to the VS

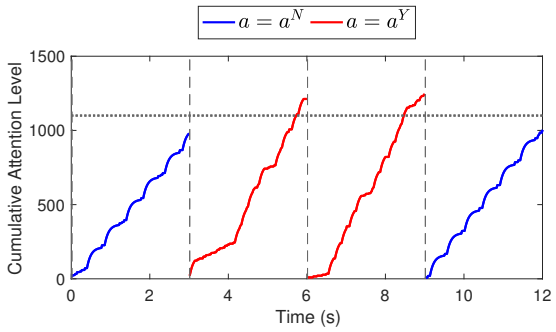


Fig. 11: The CAL of the VS transition trajectory shown in Fig. 10. The horizontal dotted line represents the attention threshold X^{at}. The visual aids in four generation stages are a^Y, a^N, a^N, and a^Y, respectively, and the resulting attention states are x^L, x^H, x^H, and x^L, respectively.

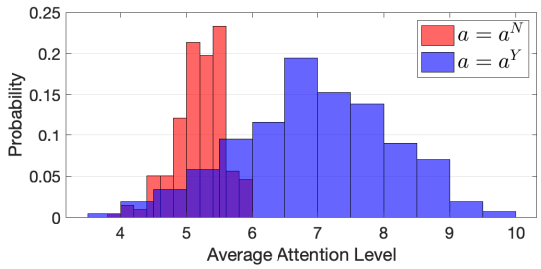


Fig. 12: The normalized histogram of average attention level under visual aids a^N and a^Y in red and blue, respectively.

transition trajectory. In Algorithm 2, we simulate the human email-reading process through the synthetic VS transition trajectory generated by the sufficient statistics $P(a_t)$ and $\phi(a_t)$. Every T^{pl} seconds, ADVERT updates the Q-matrix and the visual aid based on (2).

Following Section III-D, we develop Algorithm 3 to illustrate the entire attention enhancement loop that involves the consolidation of the data set from $\bar{M} \in \{1,\cdots,M\}$ participants and $\bar{N} \in \{1,\cdots,N\}$ emails. After the participant $m \in \{1,\cdots,\bar{M}\}$ finishes reading the email $n \in \{1,\cdots,\bar{N}\}$, Algorithm 2 returns the Q-matrix and the attention state at the final generation stage K_m^n. These results then serve as the inputs for the next email inspection until N^{bo} emails have been inspected.

Based on Algorithm 3, we plot the entire Q-learning updates with $N^{bo} = 100$ emails in Fig. 13 that contains a total of 609 generations stages. The learning results show that the visual aid a^Y outweighs a^N for both attention states and should be persistently applied under the current setting.

Algorithm 2: [Individual Adaptation] Optimal visual-aid learning and attention enhancement for participant $m \in \mathscr{M}$ vetting email $n \in \mathscr{N}$.

10 **Input:** Initial Q-matrix $[Q_0(x,a)]_{x\in\mathscr{X},a\in\mathscr{A}}$, initial attention state $x_0 \in \mathscr{X}$, the number of visits $\eta_k(x,a)$, and the hyperparameter $\theta = [X^{at}, T^{pl}]$;
11 **Initialize** time $t = 0$ and the inspection length T_m^n based on the Burr distribution F^{Burr};
12 **Set** the initial visual aid $a_0 \in \mathscr{A}$ based on the initial Q-matrix Q_0, the initial attention state x_0 and the ε_k-greedy policy in Section III-D;
13 **while** $t < T_m^n$ **do**
14 **Obtain** VS transition $s_t \in \mathscr{S}$ based on $P(a_t)$ and $\phi(a_t)$ (i.e., use synthetic visual data to achieve Step 2 in Fig. 1);
15 **Evaluate** the CAL $v_k(t)$ based on r^{co}, α as shown by Step 3 in Fig. 1;
16 **if** $t = kT^{pl}, k \in \mathbb{Z}^+$ **then**
17 **if** $\bar{v}_k \geq X^{at}$ *(shown by Step 4 in Fig. 1)* **then** attentive attention state $x_k = x^H$ **else** inattentive attention state $x_k = x^L$;
18 **Update** Q-matrix Q_k based on (2) as shown by Step 5 in Fig. 1;
19 **Implement** the visual aid $a_k \in \mathscr{A}$ based on the current Q-matrix Q_k and the ε_k-greedy policy (i.e., Step 6 in Fig. 1);
20 **if** $x_k = x, a_k = a$ **then** update the number of visits $\eta_{k+1}(x,a) \leftarrow \eta_k(x,a) + 1$;
21 **Output** the number of updates $K_m^n \leftarrow k$;
22 **end**
23 **end**
24 **Implement** the pre-trained neural network in Section V-C1 to estimate whether participant m has made the correct judgment concerning email n, i.e., $z_m^n(\theta) \in \{z^{co}, z^{wr}\}$ (i.e., use synthetic decision data to achieve Step 7 in Fig. 1);
25 **Return:** Q-matrix $[Q_{K_m^n}(x,a)]_{x\in\mathscr{X},a\in\mathscr{A}}$, final attention state $x_{K_m^n} \in \mathscr{X}$, number of visits $\eta_{K_m^n}(x,a)$, and $z_m^n(\theta)$;

C. Validation of Phishing Prevention Mechanism

After we obtain a participant's synthetic response (characterized by his VS transition trajectory) under the adaptive visual aids, we apply a pre-trained neural network to estimate whether the participant has made a correct judgment as shown in line 24 of Algorithm 2. In Section V-C1, we elaborate on the training process of the neural network based on the data set used in Section V-A. We apply BO in Algorithm 1 to evaluate the accuracy metric $c^{ac} \in \mathscr{C}$, as illustrated by Step 8 in Fig. 1. In Section V-C2, we show the results.

Algorithm 3: [**Population Adaptation**] Optimal visual-aid learning through a consolidated data set of $\bar{M} \in \{1,\cdots,M\}$ participants vetting $\bar{N} \in \{1,\cdots,N\}$ emails.

Input: Hyperparameter $\theta = [X^{at}, T^{pl}]$;
Initialize Q-matrix $[Q_0(x,a)]_{x\in\mathscr{X},a\in\mathscr{A}}$ as a zero matrix, $\eta_0(x,a) = 0, \forall x \in \mathscr{X}, a \in \mathscr{A}$, and initial attention state $x_0 \in \mathscr{X}$;
for *participant* $m \in \{1,\cdots,\bar{M}\}$ *vetting email* $n \in \{1,\cdots,\bar{N}\}$ **do**
  Implement Algorithm 2 with the inputs of $[Q_0(x,a)]_{x\in\mathscr{X},a\in\mathscr{A}}$, $x_0 \in \mathscr{X}$, and $\eta_0(x,a)$;
  Save the outputs of $[Q_{K_m^n}(x,a)]_{x\in\mathscr{X},a\in\mathscr{A}}$, $x_{K_m^n} \in \mathscr{X}$, $\eta_{K_m^n}(x,a)$, and $z_m^n(\theta)$;
  Cascade the outputs to the inputs of the next loop: $Q_0 \leftarrow Q_{K_m^n}$, $x_0 \leftarrow x_{K_m^n}$, and $\eta_0 \leftarrow \eta_{K_m^n}$;
end
Return: the accuracy metric $c^{ac}(\theta)$ based on (3);

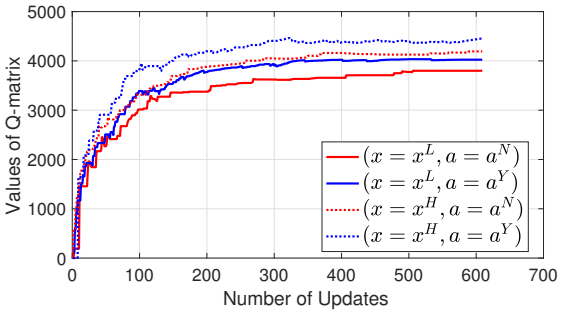


Fig. 13: The Q-learning updates under hyperparameters $X^{at} = 5.56$ and $T^{pl} = 3$ seconds. The red and blue lines represent the Q-matrix values under visual aids a^N and a^Y, respectively. The solid and dashed lines represent the Q-matrix values under attention states x^L and x^H, respectively.

1) Neural Network: In this case study, we regard the majority choice of the $M = 160$ participants as the email's true label. Without visual aids, these participants achieve an accuracy of 74.6% on average. Under the assumption that the hyperparameters affect the participants' phishing recognition only through their VS transitions, we construct a neural network with an LSTM layer, a dropout layer, and a fully-connected layer to establish the relationship from the sequence of VS transition trajectory $[s_t]_{t\in T_m^n}$ to the label of judgment correctness $z_m^n \in \{z^{co}, z^{wr}\}$. We split the entire trials of the eye-tracking data set into 1113 training data and 128 test data[4]. The trained neural network achieves a sensitivity of 0.89, a specificity of 0.21, an f1-score of 0.73, and an accuracy of 0.61.

2) Bayesian Optimization Results: As explained in Section IV, for each different application scenario, a meta optimization of the accuracy metric $c^{ac}(X^{at}, T^{pl})$ is required to find the optimal attention threshold X^{at} and the period length T^{pl} for visual-aid generation. To obtain the value of $c^{ac}(X^{at}, T^{pl})$ under different values of the hyperparameter $\theta = [X^{at}, T^{pl}]$, we need to implement the hyperparameter in Algorithm 3 and repeat for n^{rp} times to reduce the noise. Thus, the evaluation is costly, and BO in Algorithm 1 is a favorable method to achieve the meta optimization. We illustrate the BO for $L = 60$ tuning stages in Fig. 14. Each blue point represents the average value of $c^{ac}(X^{at}, T^{pl})$ over $n^{rp} = 20$ repeated samples under the hyperparameter $\theta = [X^{at}, T^{pl}]$. Based on the estimated Gaussian model in red, we observe that the attention threshold $X^{at} \in [1, 33]$ has a small impact on phishing recognition while the period length $T^{pl} \in [60, 600]$ has a periodic impact on phishing recognition. The optimal hyperparameters for phishing prevention are $X^{at,*} = 8.8347$ and $T^{pl,*} = 6.63$ seconds.

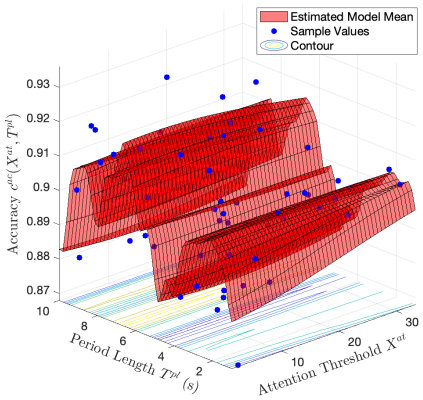


Fig. 14: The estimated Gaussian model of the objective function $c^{ac}(\theta)$ concerning the hyperparameter $\theta = [X^{at}, T^{pl}]$ in red with its contour on the bottom. The blue points represent the sample values of 60 tuning stages.

We illustrate the temporal procedure of BO for $L = 60$ tuning stages in Fig. 15. As we increase the number of tuning stages to obtain more samples, the maximized

[4]There are 1920 trials in total, and we carefully exclude the remaining 679 trials for two reasons. First, Tobii Pro T60XL records the participants' eye locations with a *validity level* ranging from 0 (high confidence) to 4 (eye not found). We exclude a trial if more than 70% of its vetting time has a validity value of 4. Second, we exclude trials of irresponsible participants who spend the majority (i.e., over 70%) of time in uninformative areas.

value of the accuracy metric $c^{ac} \in \mathscr{C}$ monotonously increases as shown in red. The blue line and its error bar represent the mean and variances of the sample values at each tuning stage, respectively. Throughout the $L = 60$ tuning stages, the variance remains small, which indicates that ADVERT is *robust* to the noise of human attention and decision processes.

Compared to the benchmark accuracy of 74.6% without visual aids, participants with visual aid achieve the accuracy of a minimum of 86% under all 60 tuning stages of different hyperparameters. The above accuracy improvement corroborates that the ADVERT's attention enhancement mechanism highlighted by the blue background in Fig. 1 effectively serves as a stepping stone to facilitate phishing recognition. The results shown in the blue line further corroborate the efficiency of the ADVERT's phishing prevention mechanism highlighted by the orange background in Fig. 1; i.e., in less than 50 tuning stages, we manage to improve the accuracy of phishing recognition from 86.8% to 93.7%. Besides, the largest accuracy improvement (from 88.7% to 91.4%) happens within the first 3 tuning stages. Thus, if we have to reduce the number of tuning stages due to budget limits, ADVERT can still achieve a sufficient improvement in the accuracy of recognizing phishing.

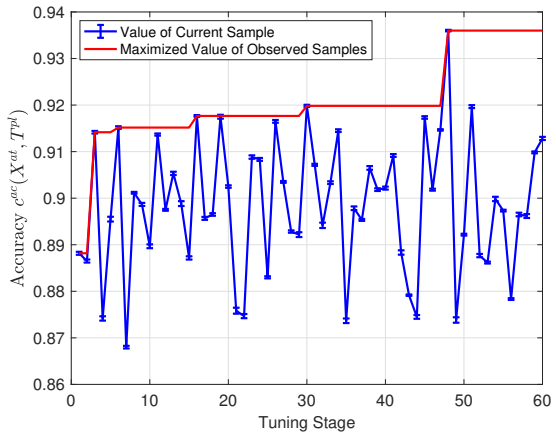


Fig. 15: Accuracy metric $c^{ac}(X^{at}, T^{pl})$ at $L = 60$ tuning stages. The blue line and its error bar represent the mean value of the samples and their variances, respectively. The red line represents the maximized value of the observed samples up to the current tuning stage.

VI. Limitations and Mitigation

The limitations of the data set and the data processing process are as follows. First, the demographic of the experimental subjects is limited to 160 undergraduate students. In the current work, we handle this issue by diversifying the participants (concerning their races, genders, and ages) and adopting the feedback loop of Bayesian optimization (that adapts to unconsidered user groups). To enable a more comprehensive study of the human behaviors that cover different user groups, we can recruit more diversified participants through crowdsourcing websites, including Amazon Mechanical Turk (MTurk). Second, the dataset contains 12 unique emails. They are certainly not meant to be comprehensive to cover all phishing scenarios. However, they are sufficient for this work, which focuses on the system-level control of human attention processes to improve the accuracy of phishing recognition. For each email, we conduct the vetting processes of $M = 160$ humans, which result in the distinct 1241 trials of eye-tracking trajectories. These eye-tracking trials are sufficient to reveal human attention patterns. Moreover, as a data-driven and system-level framework, ADVERT can adapt and generalize to unseen sets of emails. Third, we exclude approximately one-third of the eye-tracking data due to their low validity scores that arise from the limitation of the eye-tracking device and the imprudence of the participants, as stated in the footnote of Section V-C1. The reduced sample size may lead to overfitting issues. We can overcome it by improving the eye-tracking device, revising the experiment setting, and recruiting a sufficient number of participants.

VII. Conclusions and Future Work

As a prototypical *innate human vulnerability*, lack of attention is one of the main challenges to protecting users from phishing attacks. To address the challenge, we have developed a *human-technical solution* called ADVERT to guide the users' attention to the right contents of the email and consequently improve their accuracy of phishing recognition.

To enable a real-time evaluation of the user's visual behaviors, we have built AoIs from the entire email area and a transition model to compress the eye-tracking data into a representative VS transition trajectory. After assigning the concentration rewards and decay parameters to evaluate the user's CAL, we have defined *privacy-preserving* and *light-weight* metrics, i.e., AAL and QAAL, to represent the user's attention state at each time of visual-aid generation. These metrics enable us to apply model-free RL methods and generate the optimal visual aid for real-time attention enhancement. Using the above attention enhancement mechanism as a stepping-stone, we have designed an efficient algorithm to tune the hyperparameters related to the visual aid generation pattern and the attention evaluation parameters. The update of these hyperparameters at each tuning stage revises the visual aids, affects the users' attention, and consequently improves the accuracy of phishing recognition.

We have corroborated the effectiveness of ADVERT through a case study based on the data set collected from human subject experiments conducted at New York University. By abstracting the transition matrix and sojourn distribution from the data set as *sufficient statistics* of the stochastic VS transition, we have generated synthetic VS transition to simulate the participant's visual behaviors under visual aids. Meanwhile, we have trained a neural network to estimate the correctness of the participant's phishing recognition based on the VS transition trajectory. Finally, we have developed two algorithms to design visual aids that adapt to each individual and the population, respectively. For the attention enhancement mechanism, the results have shown that the visual aids can statistically increase the AAL and improve the accuracy of phishing recognition from 74.6% to a minimum of 86%. The meta-adaptation has been shown to be *effective* (e.g., improve the accuracy of phishing recognition from 86.8% to 93.7% in less than 50 tuning stages), *efficient* (e.g., the largest accuracy improvement happens within 3 tuning stages), and *robust* (e.g., the variances of $L = 60$ samples remain small). The results have also provided insights and guidance for the ADVERT design; e.g., the attention threshold (resp. the period length) has a small (resp. periodic) impact on phishing recognition.

The future work would focus on designing a more sophisticated visual support system that can determine when and how to generate visual aids in lieu of a periodic generation. We may also embed ADVERT into VR/AR technologies to mitigate human vulnerabilities under simulated deception scenarios, where the simulated environment can be easily repeated or changed. Finally, there would be an opportunity to incorporate factors such as pressure and incentives into the design by limiting the participant's vetting time and providing rewards for accurately identifying phishing, respectively.

ACKNOWLEDGMENT

The authors would like to thank Jennie W. Qu-Lee and Blair Cox for their help to export and interpret the eye-tracking data set housed on the experiment platform of the NYU Social Perception Action & Motivation (SPAM) laboratory. We thank Prof. Jonathan Bakdash and the other anonymous reviewer for the helpful comments on earlier drafts of the manuscript.

REFERENCES

[1] H. Aldawood and G. Skinner, "Educating and raising awareness on cyber security social engineering: A literature review," in *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pp. 62–68, IEEE, 2018.

[2] M. Alotaibi, S. Furnell, and N. Clarke, "Information security policies: A review of challenges and influencing factors," in *2016 11th International Conference for Internet Technology and Secured Transactions (ICITST)*, pp. 352–358, IEEE, 2016.

[3] L. Huang and Q. Zhu, "Duplicity games for deception design with an application to insider threat mitigation," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4843–4856, 2021.

[4] L. Huang and Q. Zhu, "Zetar: Modeling and computational design of strategic and adaptive compliance policies," *arXiv preprint arXiv:2204.02294*, 2022.

[5] E. B. Cox, Q. Zhu, and E. Balcetis, "Stuck on a phishing lure: differential use of base rates in self and social judgments of susceptibility to cyber risk," *Comprehensive Results in Social Psychology*, vol. 4, no. 1, pp. 25–52, 2020.

[6] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proc. of the SIGCHI conference on Human Factors in computing systems*, pp. 581–590, 2006.

[7] I. Baxter, "Fake login attack evades logo detection," 2020. https://ironscales.com/blog/fake-login-attack-evades-logo-detection.

[8] Z. A. Wen, Z. Lin, R. Chen, and E. Andersen, "What. hack: engaging anti-phishing training through a role-playing phishing simulation game," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[9] R. C. Dodge Jr, C. Carver, and A. J. Ferguson, "Phishing for user security awareness," *computers & security*, vol. 26, no. 1, pp. 73–80, 2007.

[10] A. K. Jain and B. B. Gupta, "Phishing detection: analysis of visual similarity based approaches," *Security and Communication Networks*, vol. 2017, 2017.

[11] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: a literature survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2091–2121, 2013.

[12] T. Kelley and B. I. Bertenthal, "Real-world decision making: Logging into secure vs. insecure websites," *Proceedings of the USEC*, vol. 16, no. 10.14722, 2016.

[13] S. Egelman, L. F. Cranor, and J. Hong, "You've been warned: an empirical study of the effectiveness of web browser phishing warnings," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1065–1074, 2008.

[14] E. Al-Shaer, J. Wei, W. Kevin, and C. Wang, "Autonomous cyber deception," *Springer*, 2019.

[15] L. Huang and Q. Zhu, "A dynamic games approach to proactive defense strategies against advanced persistent threats in cyber-physical systems," *Computers & Security*, vol. 89, p. 101660, 2020.

[16] J. Pawlick and Q. Zhu, *Game Theory for Cyber Deception: From Theory to Applications*. Springer Nature, 2021.

[17] C. Katsini, Y. Abdrabou, G. E. Raptis, M. Khamis, and F. Alt, "The role of eye gaze in security and privacy applications: survey and future hci research directions," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–21, 2020.

[18] N. Ramkumar, V. Kothari, C. Mills, R. Koppel, J. Blythe, S. Smith, and A. L. Kun, "Eyes on urls: Relating visual behavior to safety decisions," in *ACM Symposium on Eye Tracking Research and Applications*, pp. 1–10, 2020.

[19] D. Miyamoto, G. Blanc, and Y. Kadobayashi, "Eye can tell: On the correlation between eye movement and phishing identification," in *Int. Conf. on Neural Information Processing*, pp. 223–232, Springer, 2015.

[20] J. McAlaney and P. J. Hills, "Understanding phishing email processing and perceived trustworthiness through eye tracking," *Front. Psychol.*, vol. 11, p. 1756, 2020.

[21] A. Xiong, R. W. Proctor, W. Yang, and N. Li, "Is domain highlighting actually helpful in identifying phishing web pages?," *Hum. Factors*, vol. 59, no. 4, pp. 640–660, 2017.

[22] K. Pfeffel, P. Ulsamer, and N. Müller, "Where the user does look when reading phishing mails–an eye-tracking study," in *Int. Conf. on Human-Computer Interaction*, pp. 277–287, Springer, 2019.

[23] C. I. Canfield, B. Fischhoff, and A. Davis, "Quantifying phishing susceptibility for detection and behavior decisions," *Human factors*, vol. 58, no. 8, pp. 1158–1172, 2016.

[24] C. I. Canfield and B. Fischhoff, "Setting priorities in behavioral interventions: An application to reducing phishing risk," *Risk Analysis*, vol. 38, no. 4, pp. 826–838, 2018.

[25] Y. Huang, L. Huang, and Q. Zhu, "Reinforcement learning for feedback-enabled cyber resilience," *Annual Reviews in Control*, 2022.

[26] B. Shi, G. Liu, H. Qiu, Z. Wang, Y. Ren, and D. Chen, "Exploring voluntary vaccination with bounded rationality through reinforcement learning," *Physica A: Statistical Mechanics and its Applications*, vol. 515, pp. 171–182, 2019.

[27] A. Sanjab, W. Saad, and T. Başar, "A game of drones: Cyber-physical security of time-critical uav applications with cumulative prospect theory perceptions and valuations," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 6990–7006, 2020.

[28] L. Huang and Q. Zhu, "Combating informational denial-of-service (IDoS) attacks: Modeling and mitigation of attentional human vulnerability," in *International Conference on Decision and Game Theory for Security*, pp. 314–333, Springer, 2021.

[29] L. Huang and Q. Zhu, "Radams: Resilient and adaptive alert and attention management strategy against informational denial-of-service (IDoS) attacks," *arXiv preprint arXiv:2111.03463*, 2021.

[30] E. Lin, S. Greenberg, E. Trotter, D. Ma, and J. Aycock, "Does domain highlighting help people identify phishing sites?," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2075–2084, 2011.

[31] D. Akhawe and A. P. Felt, "Alice in warningland: A large-scale field study of browser security warning effectiveness," in *22nd USENIX Security Symposium (USENIX Security 13)*, pp. 257–272, 2013.

[32] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs, "Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 373–382, 2010.

[33] D. J. Liebling and S. Preibusch, "Privacy considerations for a pervasive eye tracking world," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 1169–1177, 2014.

[34] J. L. Kröger, O. H.-M. Lutz, and F. Müller, "What does your gaze reveal about you? on the privacy implications of eye tracking," in *IFIP International Summer School on Privacy and Identity Management*, pp. 226–241, Springer, 2019.

[35] M. I. Posner and O. S. Marin, *Attention and performance XI*. Routledge, 2016.

[36] G. Nasser, B. W. Morrison, P. Bayl-Smith, R. Taib, M. Gayed, and M. W. Wiggins, "The role of cue utilization and cognitive load in the recognition of phishing emails," *Frontiers in big Data*, p. 33, 2020.

[37] M. Ackerley, B. Morrison, K. Ingrey, M. Wiggins, P. Bayl-Smith, N. Morrison, *et al.*, "Errors, irregularities, and misdirection: Cue utilisation and cognitive reflection in the diagnosis of phishing emails," *Australas. J. Inf. Syst.*, vol. 26, 2022.

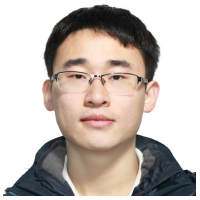
[38] P. I. Frazier, "Bayesian optimization," in *Recent Advances in Optimization and Modeling of Contemporary Problems*, pp. 255–278, INFORMS, 2018.

[39] M. P. Janisse, "Pupil size and affect: A critical review of the literature since 1960.," *Canadian Psychologist/Psychologie canadienne*, vol. 14, no. 4, p. 311, 1973.

[40] O. E. Kang, K. E. Huffer, and T. P. Wheatley, "Pupil dilation dynamics track attention to high-level information," *PloS one*, vol. 9, no. 8, p. e102463, 2014.

[41] P. J. Van Laarhoven and E. H. Aarts, "Simulated annealing," in *Simulated annealing: Theory and applications*, pp. 7–15, Springer, 1987.

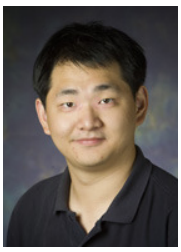
[42] C. D. Holland and O. V. Komogortsev, "Complex eye movement pattern biometrics: the effects of environment and stimulus," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pp. 2115–2126, 2013.



Linan Huang received the B.Eng. degree (Hons.) in Electrical Engineering from Beijing Institute of Technology, China, in 2016 and the Ph.D. degree in electrical engineering from New York University (NYU), Brooklyn, NY, USA, in 2022. His research interests include dynamic decision-making of the multi-agent system, mechanism design, artificial intelligence, security, and resilience for cyber-physical systems.

Shumeng Jia received the B.Eng. degree in rail traffic signaling and control from Bejing Jiaotong University, China, in 2020, and the M.S. degree in electrical engineering from New York University in 2022. She worked at the Laboratory for Agile and Resilient Complex Systems, Tandon School of Engineering, New York University, NY, USA, as a graduate assistant while working on her graduate degree. Her past research has focused on machine learning and epileptiform EEG data.

Emily Balcetis received a BA (honors) in Psychology and a BFA in Music Performance from the University of Nebraska at Kearney in 2001 and a PhD in Social Psychology from Cornell University in 2006. She is currently an Associate Professor of Psychology at New York University (NYU), and faculty affiliate of NYU's Institute for Human Development and Social Change. Her current research interests include motivation, decision-making, and visual experience.



Quanyan Zhu (SM'02-M'14) received B. Eng. in Honors Electrical Engineering from McGill University in 2006, M. A. Sc. from the University of Toronto in 2008, and Ph.D. from the University of Illinois at Urbana-Champaign (UIUC) in 2013. After stints at Princeton University, he is currently an associate professor at the Department of Electrical and Computer Engineering, New York University (NYU). He is an affiliated faculty member of the Center for Urban Science and Progress (CUSP) and Center for Cyber Security (CCS) at NYU. His current research interests include game theory, machine learning, cyber deception, and cyber-physical systems.