Poster: Snooping Online Form Choice Privacy in Video Calls

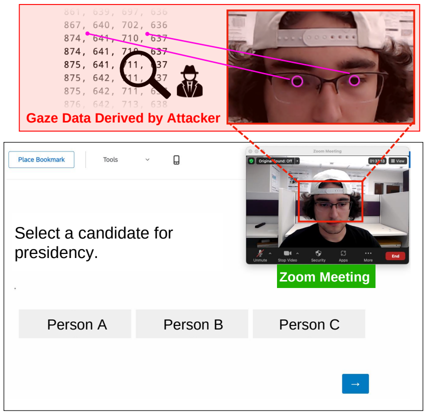
Steven Seiden*
Dept. of Computer Science
Louisiana State University
Baton Rouge, LA, USA
sseide3@lsu.edu

Long Huang
Dept. of Computer Science
Louisiana State University
Baton Rouge, LA, USA
lhuan45@lsu.edu

Chen Wang†
Dept. of Computer Science
Louisiana State University
Baton Rouge, LA, USA
chenwang1@lsu.edu

***Abstract*—In video calls, a user's eye gaze patterns can inadvertently disclose sensitive information, such as their online form choices and personal preferences. This is especially true when the on-screen content is predictable, such as in Zoom-proctored exams and online conference call polls. If users are expected to turn on their camera during a video call, such sensitive eye gaze information could be leaked from their own video feed. In this work, we propose a method to infer the user's choices made for online forms by analyzing the user's eye gaze captured in video calls. In particular, the pupil positions are derived from the webcam's video stream, after which unique eye motion features are derived to capture the user's eye behaviors of making different choices. The extracted features are then fed into an Support Vector Machine (SVM)-based algorithm to predict which choice has been made by the user. Preliminary results show our method infers online form choices achieving an 87% accuracy without requiring the training data of the target user.**

***Index Terms*—privacy, eye tracking, choice inferring**



Select a candidate for presidency.

Fig. 1. Revealing the user's online voting privacy using video calls.

I. Introduction

Since the global COVID-19 pandemic, the usage of online web conference applications has increased dramatically. One of the most popular web conference applications, Zoom, had a 383% increase in revenue and a 2900% increase in users [1]. Many workplaces and schools utilize applications such as Zoom, Microsoft Teams, Webex and Skype to host meetings or teach classes remotely. This introduces a new issue of a potential violation of privacy, especially when the user's webcam is being utilized. During online video calls, many users find themselves needing to multitask, putting them at risk of having their eye gaze pattern captured, revealing private information such as the choices made in an online form and personal preferences towards the content shown on the screen.

Traditional approaches to eye-tracking use a costly table-mounted (e.g., Tobii Pro Fusion) or a head-mounted (e.g., Tobii Pro Glasses 3) eye tracker to precisely estimate the eye gaze positions and directions. Web cameras have also been explored for implementing lower-cost eye tracking. Papoutsaki *et al.* used webcams to study user gazing behavior during web searches [2]. Khan *et al.* captures and analyzes students' attention during online classes by tracking the gaze points with a webcam [3]. In this work however, we study the potential to reveal a user's choices to online forms by analyzing the eye gaze behaviors captured by their webcam during a video call. The attacking scenario is illustrated in Figure 1.

To execute this attack we experimented with several eye-tracking algorithms. Initially, we attempted to utilize an open-source OpenCV algorithm [4] that that tracks a user's pupil position within a video feed. However, the results were too noisy to be used for our analysis. We next experimented with Brown HCI Research Group's WebGazer, which uses Google TensorFlow to map the features of a user's face to estimate the area of the screen that the user is gazing at. [5]. While highly accurate, WebGazer's application in a realistic attack scenario is limited due to the need for manual calibration by the user being tracked.

We are now utilizing Google's MediaPipe Iris [6] eye-tracking algorithm. This algorithm has the benefit of both a high tracking accuracy and not requiring calibration by the user. Nevertheless, webcam-based eye-tracking still has limited accuracy as it relies purely on computer vision analysis, introducing errors in the gaze positions reported. To mitigate these errors we utilize key points in the recorded data to normalize eye gaze positions based on the layout of the online form, after which unique eye motion features are derived and fed into an SVM-based algorithm to infer which choice has been made by the user.

*Steven Seiden is an undergraduate researcher at LSU.
†Chen Wang is the corresponding author.

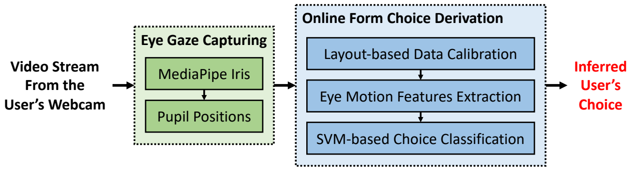

Video Stream
From the
User's Webcam
Eye Gaze Capturing
MediaPipe Iris
Pupil Positions
Online Form Choice Derivation
Layout-based Data Calibration
Eye Motion Features Extraction
SVM-based Choice Classification
Inferred
User's
Choice

Fig. 2. The architecture of our system.

II. System Design

The architecture of our system is shown in Figure 2. The video stream from the user's webcam is taken as the input. The *Eye Gaze Capturing* is performed first, using MediaPipe Iris to get pupil positions from the video feed. The resulting position sequence is processed by the *Online Form Choice Derivation*, where unique eye motion features are extracted and fed into an SVM-based algorithm to infer which choice has been made by the user. This situation assumes the host of a conference call has asked everyone in the meeting to fill out an online poll. Thus, the attacker is aware of the poll's format.

Eye Gaze Capture. We utilized MediaPipe Iris to extract the user's two pupils' positions from the video frames of the webcam, which results in a four-dimension (i.e., left pupil $\{L_x, L_y\}$ and right pupil $\{R_x, R_y\}$) pupil position sequence.

Online Form Choice Inference. The expected pupil position sequence is first determined based on the layout of the online form. For example, if the user is filling an online form with the question title located at the screen's top-left corner and the submit button located at the bottom-right corner (as illustrated in Figure 1), their eye gaze will likely start at the top-left corner and end at the bottom-right corner. For this case, the pupil position sequence will be normalized so that each dimension's start point is 0 and end point is 1. After this, unique eye motion features are extracted from the calibrated pupil position sequence to describe the user's behaviors when making different choices. In particular, we evaluate multiple features and select the moving variance, skewness, and position itself as the final features. The extracted eye motion features are then fed into an SVM-based algorithm to infer the online form choice.

III. Initial Findings

We recruited five participants to conduct experiments. IRB approval has been obtained. The participants were asked to sit in front of a computer screen displaying an online Qualtrics poll [7]. This poll included three answer choices, which we will call A, B, and C, to refer to the location of the three answer choices. The poll format can be seen in the *User's Screen* section of Figure 1. During the experiments, the user would make a total of sixty answer selections, with twenty selections for each of the three choices. As the users responded to the poll, their video feed was captured by a webcam positioned in front of their faces above the monitor. This live video feed was fed into MediaPipe Iris, allowing us to record where within the video feed the user's pupils are.

Online Form Choice Inferring Performance. We first evaluate the system performance when all five users are included in the training set. We use half of each user's data for training and the other half for testing. As shown in Figure 3, our system infers the users' choices with an overall accuracy of 89.3%. More specifically, the inferring accuracies for choice A, B, and C are 86%, 86%, and 96%, respectively. The results indicate our system is able to infer the user's online form choices with relatively high accuracy.

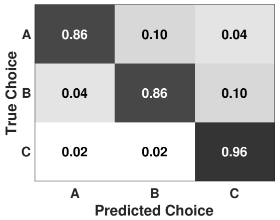


Fig. 3. Online form choice inferring performance.

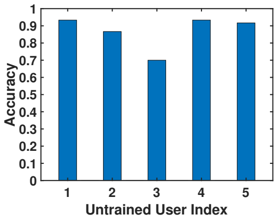


Fig. 4. Performance on untrained users.

Target Victim Not in Training Set. We consider a more practical scenario where the target user is not in the training set. To evaluate our system in this scenario, we iteratively selected one of the five users as the target user and trained the SVM model with the other four users' data and tested it with the target user's data. The results are presented in Figure 4. We can observe our system achieves accuracies of 93.3%, 86.7%, 70%, 93.3%, and 91.7% for each one of the five users selected as the target user. The results indicate our system generally works for inferring the choices made by untrained users.

IV. Conclusion

In this research we studied the danger of sharing your video feed in an online conference call, revealing possibility of inferring a user's answer choices to an online poll. We showed that if an attacker has access to the user's video feed, and the poll's interface, they can determine the user's responses to the poll. In the future we would like to study how a web camera's position affects the outcome of our method. While already minimized due to the normalization we apply, ways to combat issues could include building a 3D mesh of a user's face and tracking their pupils position relative to their head.

References

[1] M. Iqbal. (2023) Zoom revenue and usage statistics (2023). [Online]. Available: https://www.businessofapps.com/data/zoom-statistics/

[2] A. Papoutsaki, J. Laskey, and J. Huang, "Searchgazer: Webcam eye tracking for remote studies of web search," in *Proceedings of the 2017 conference on conference human information interaction and retrieval*, 2017, pp. 17–26.

[3] A. R. Khan, S. Khosravi, S. Hussain, R. Ghannam, A. Zoha, and M. A. Imran, "Execute: Exploring eye tracking to support e-learning," in *2022 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 2022, pp. 670–676.

[4] A. Lame, "Gaze tracking," https://github.com/antoinelame/GazeTracking, 2022.

[5] M. Sabra, A. Maiti, and M. Jadliwala, "Zoom on the keystrokes: exploiting video calls for keystroke inference attacks," *arXiv preprint arXiv:2010.12078*, 2020.

[6] Google, "Mediapipe iris," https://github.com/google/mediapipe/blob/master/docs/solutions/iris.md, 2023.

[7] Qualtrics. (2023) Survey template. [Online]. Available: https://www.qualtrics.com/marketplace/survey-template/