Empowering Security and Privacy-Preserving Interactions for Smart Device Users

by

Jingjie Li

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Electrical and Computer Engineering)

at the

UNIVERSITY OF WISCONSIN–MADISON

2023

Date of final oral examination: 06/14/2023

The dissertation is approved by the following members of the Final Oral Committee:
Younghyun Kim, Associate Professor, Electrical and Computer Engineering
Kassem Fawaz, Associate Professor, Electrical and Computer Engineering
Patrick McDaniel, Professor, Computer Sciences
Rahul Chatterjee, Assistant Professor, Computer Sciences

This dissertation is dedicated to my family.

ACKNOWLEDGMENTS

When putting down these words to mark the conclusion of this journey, my contemplation was interrupted by a mesmerizing afterglow filtering through the window. Such a moment seemed no more special. It resembled many other summer evenings I have spent in Madison. But I knew this moment was different, encapsulating numerous cherished memories that shaped me into who I am today. It is not a journey alone, throughout which I have been accompanied by so many people I am indebted to.

Foremost, I would like to express my greatest appreciation once more to my advisors, Prof. Younghyun Kim and Prof. Kassem Fawaz, for their invaluable mentorship. I call crossing paths with them a fortuitous turn in the course of my life. When I was in the last year of my undergraduate study, I had little idea of what I would expect in the forthcoming life chapter other than the passion to explore uncharted realms. Younghyun offered me an opportunity to embark on further studies in Wisconsin, a place ten thousand kilometers away. I am grateful that I learned from Younghyun by his own example of research professionalism in my early career. Beyond this, he nurtured an environment where we can truly enjoy our Ph.D. life. As a security and privacy researcher, I must also express my gratitude to Kassem for steering me into this vibrant community and helping me grow. The first thing I would like to make room for in my future office is a whiteboard – writing and drawing on it will always remind me of the days I spent with Kassem trying to exhaust our boundless creativity and ideas. I am privileged to learn the very best examples from both mentors across different aspects. While enumerating all their virtues would demand pages, I will dedicate my actions to echoing their mentoring. Moreover, their wholehearted support extended much beyond research, empowering me to overcome the obstacles through this six-year odyssey.

Next, I would like to extend my gratitude to the other members of

my Ph.D. defense committee and all the amazing faculty members who provided generous support for me: Prof. Kevin Butler, Dr. Christian Castro, Prof. Rahul Chatterjee, Prof. Earlence Fernandes, Prof. Patrick McDaniel, Prof. Joshua San Miguel, Prof. Florian Schaub, among others. I also want to give a big shout-out to my fellow collaborators, many of whom contributed to the work of this dissertation in particular: Sunpreet Singh Arora, Setareh Behroozi, Anna Bierley, Varun Chandrasekaran, Amrita Roy Chowdhury, Brittany Huff, Kyuin Lee, Mohsen Minaei, Kaiwen Sun, Di Wu, Yixin Zou, my labmates in the WISEST and WI-PI families, among others. It is a delightful journey to work with and make friends with them.

Then, I want to thank all my friends in Madison, particularly the companions of the "hotpot gang": Di Wu, Zongshen Wu, Bin Li, and Yue Gao, who sailed with me from the beginning of this journey till the end. I will always remember our Friday gatherings, where we cooked all the joy and bitterness of a week. Our friendship shall endure no matter where we are. I also want to express my sincerest gratitude to my friends all over the world, many of whom have accompanied me since my childhood. Thank you for being by my side despite the barriers of space and time.

Most profoundly, I reserve my utmost gratitude for my family, particularly my parents, for their selfless love. They have done so much to support me and my adventures while bearing numerous sleepless nights thinking about their child on the other side of the earth, whom they did not have the chance to meet face-to-face for years. Their love and encouragement have crafted a sanctuary where I find trust, solace, and love, even when we are physically apart. My achievements are theirs, though they are not expecting much more than their kid's happiness. Finally, I want to acknowledge myself for embarking on this adventure. It is a journey onward, where I sailed into the unknown through the winds and waves; it is a journey back to the origin, where I picked up the purest passion to embrace myself and this world.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Emerging smart devices, such as smart home and augmented/virtual reality systems, are reforming our living experience by automating our daily routines and interacting with us seamlessly. However, these smart devices also introduce new threat vectors to our security and privacy due to their advanced sensing and computational capabilities. Consequently, malicious actors can easily exploit these devices and users' private data, causing harm in various ways. Despite recent regulatory efforts aimed at granting users control over security and privacy, emerging smart device systems are often incompatible with proper security and privacy controls. Moreover, users still struggle to implement protective measures in real life due to a limited understanding of security and privacy information and the suboptimal cost-benefit tradeoffs of these measures. Thus, these challenges are calling for both technical and operational solutions to protect users' security and privacy in the "brave new world" of smart devices.

My dissertation solves the above challenges by empowering security and privacy-preserving interactions for smart device users. My research makes efforts from two directions to (1) enhance security and privacy controls by preserving utilities and interaction experiences and (2) understand how users perceive, assess, and react to security and privacy issues, which further informs solutions to help them protect security and privacy. To achieve this objective, my work leverages multiple methods of system design and user research. This dissertation presents the following contributions. Firstly, we introduce Kaleido, a usable privacy control that preserves the interactive utilities of eye-tracking systems in real-time. Secondly, we discuss Velody, a challenge-response biometric authentication that mitigates the security risk associated with biometric template breaches. Thirdly, we systematically present our findings in understanding how smart home users develop security and privacy considerations and

attitudes over time. Finally, we investigate how interaction experiences influence users' security perceptions of authentication in virtual reality. The dissertation concludes with my future research directions aimed at supporting users and society in adapting to forthcoming technologies with transparent, pervasive, and accessible security and privacy solutions.

1 INTRODUCTION

With emerging smart devices such as smart home, augmented and virtual reality (AR/VR) systems, the boundaries between the digital and physical worlds are rapidly becoming blurry. These devices automate our daily routines and create an immersive reality. The new experiences, however, accompany unprecedented security and privacy risks to consumers. Firstly, these advanced devices expose numerous attributes about users' physiological, psychological, and environmental traits that must be kept private. For example, users' health status can be revealed from eye movement data within seconds [130, 109]. Secondly, these interconnected devices often have access to users' critical digital and physical assets. Once such access is abused or compromised, users may suffer from physical, psychological, and financial harms [268, 12]. Regulations across the world, including EU's General Data Protection Regulation (GDPR) [290], California's California Consumer Privacy Act (CCPA) [41], and China's Personal Information Protection Law (PIPL) [181], have being taking effect to mitigate the security and privacy risks of using smart devices. One common theme shared in these regulations is that users should be in control of their personal information. However, we are still witnessing threats that happen in our real lives, for instance, massive smart cameras' video feeds being breached [99]. These concerns persist due to two main reasons. First, users' security and privacy needs often intertwine with the utilities and interaction experiences of smart devices. Second, users face real-world barriers, e.g., limited understanding of security and privacy information and technical incompetence, to protect their security and privacy when adopting and deploying smart devices. The above challenges often result in the misalignment between users' expectations for security and privacy and their inaction. Hence, these challenges motivate the overarching question to answer in this dissertation:

How can we improve system designs and practices for empowering users to preserve their security and privacy when interacting with and adopting smart devices?

Aiming to solve the question, the innovations this dissertation makes are in two folds. First, we design **solutions to enhancing security and privacy control** for different system stacks of smart devices. In particular, this dissertation addresses the security and privacy issues in using biosignals collected from smart device users. Nowadays, advanced smart devices collect massive amounts of biosignals that are security and privacy-sensitive, such as users' eye movements, to enable seamless interaction and various functionalities. However, it was previously an open challenge to control their security and privacy while preserving the utilities of these data types, because we have a limited understanding of their security and privacy semantics as well as growing requirements for better interaction experiences. To tackle this challenge, our research in this dissertation takes the following steps. We first advance the understanding of these data's security and privacy implications. Then, we develop solutions to preserve security and privacy that account for the functionality and utility requirements. These solutions are integrated into different system stacks, including the software processing layer, sensing frontend, and user interface. Furthermore, we engineer these software and hardware solutions and evaluate them with users in the loop.

Nevertheless, only providing the system infrastructures cannot guarantee that users will use them appropriately to protect their security and privacy in real life. Thus, the second aspect of this dissertation investigates **how users perceive, assess, and react to security and privacy issues** when they interact with and adopt smart devices. With such understanding, this dissertation further informs solutions and practices for multiple parties to improve users' security and privacy through different phases in their adoption journey, for example, when introducing new technologies to

them. The real-world context of technology adoption and interaction is complicated, which makes the above process not readily visible previously, especially for emerging applications of smart homes and VR. And many design spaces still remain unexplored to improve their security and privacy. My research leverages organic media in a systematic manner, including qualitative content analysis and technology probes, to analyze users' experiences and understand how they perceive, assess, and react to security and privacy issues.

The following chapters in this dissertation describe the following contributions around the above-mentioned two aspects. To enhance security and privacy control in smart devices, Chapter 2 first exemplifies how we **develop usable privacy control for real-time eye-tracking systems** that preserves interactive utilities. Next, Chapter 3 shows how to **secure user authentication using nonlinear vibration challenge-responses**, which improves security by minimizing the privacy exposure of biometric templates. On the other hand, Chapter 4 presents our systematic study to **understand how smart home users develop security and privacy considerations and attitudes** throughout their adoption journey based on their online discussion. Then, Chapter 5 demonstrates our work to **understand how interaction experiences influence security perceptions of VR authentication**, leveraging technology probes to provoke users' reactions during early adoption. Last, Chapter 6 concludes my dissertation by discussing the lessons learned from my research and future directions. We provide an overview of the four major contributions as follows.

Developing usable privacy control for real-time eye-tracking systems. Recent advances in sensing and computing technologies have led to the rise of eye-tracking platforms. Ranging from mobiles to high-end mixed reality headsets, a wide spectrum of interactive smart devices now employs real-time eye tracking, a prominent biosignal that unlocks hands-free

interaction. However, eye gaze data is a rich source of sensitive information that can reveal an individual's physiological and psychological traits, including users' implicit interests and health status. For example, my prior work demonstrates how users' personality traits can be revealed from eye-tracking data during multimedia experiences [34]. Yet, like many other emerging smart devices, current eye-tracking systems fail to offer practical privacy protections that accommodate users' utility requirements. Prior approaches to protecting eye-tracking data suffer from two major drawbacks: they are either incompatible with the current ecosystem for real-time eye tracking or provide no formal privacy guarantee [124, 244, 230, 175, 100, 261, 36]. This motivates the following research question:

> *How can we offer users a usable privacy control when preserving the real-time utilities of eye tracking?*

Proposed solution. In Chapter 2, we discuss our work Kalɛido that answers this question [164]. Kalɛido is among the first systems to provide users with a usable privacy control that preserves the interactive utility of eye tracking. Kalɛido bridges the formal privacy guarantees of local differential privacy to real-time eye tracking and seamlessly integrates with the existing ecosystems. It automates the configuration of privacy parameters for users to balance privacy and utility easily. By implementing Kalɛido in Unity AR/VR engine, we exemplify a method to quickly prototype privacy protection for smart devices. Kalɛido has been evaluated through comprehensive user studies and trace-based analysis, showing good usability and effectiveness against various attacks.

Securing user authentication using nonlinear vibration challenge-responses. Biosignals also make user authentication more convenient, benefiting usability by exploiting pervasive and collectible unique characteristics from physiological or behavioral traits of human, which are known

as biometrics. However, biometric templates are often exposed through insecure channels. Successful attacks on "static" biometrics, which reuse biometric templates such as fingerprints, have been reported. An adversary may acquire users' biometrics stealthily and compromise these non-resilient biometrics. This privacy exposure weakens the security of traditional biometric authentication: it is hard to know if the template is divulged, and users cannot revoke it when the system is compromised. Therefore, we attempt to answer the below question:

> *How can we make biometric authentication more resilient against the threats of template breach?*

Proposed solution. My work VELODY in Chapter 3 addresses this challenge by constructing unlinkable biometric challenge-responses to minimize privacy exposure in biometrics-based authentication [165]. VELODY distinguishes users based on their hand compositions, which result in differing echoes (responses) to an external vibration (challenge) from a commodity speaker. It makes biometric templates unlinkable by evoking nonlinear responses that are hard to predict and therefore preserves template privacy by renewing challenge-response pairs. Our user experiments further indicate VELODY's resiliency against impersonation, replay, and even sophisticated synthesis attacks. And VELODY retains good usability regarding its short authentication time and permanence over time.

Understanding how smart home users develop security and privacy considerations and attitudes. Smart home products, such as voice assistants, smart cameras, and robot vacuums, are a major category of smart devices that more and more users interact with on a daily basis. They offer many benefits to users, including automating their daily routines. Due to the growing market potential, smart home manufacturers are still eagerly releasing product updates with better sensing, networking, and analytic capabilities. Yet, they also carry complex security and privacy

implications that users often struggle to assess and account for during adoption and interaction. Investigating how users think of and react to the privacy issues of smart homes helps users overcome the barriers to incorporating security and privacy protection in real life. Consumers' experiences with a product are dynamic and reflective from pre-purchase to post-consumption [131]. However, prior work on users' security and privacy perceptions of smart homes often misses the real-life dynamics when adopting and interacting with them [83, 101, 320, 328]. Thus, we hypothesize that how smart home users think of (i.e., consideration) and react to (i.e., attitude) security and privacy is also a non-static process. This hypothesis drives our research to answer the following question:

> *How do smart home users develop security and privacy considerations and attitudes throughout their adoption journey?*

Proposed solution. Compared with prior studies that leverage surveys and interviews, we realize the opportunities in online discussion forums. Online discussion forums, where users exchange information, contain a rich source of such dynamics despite their complex nature that curbs prior research. Chapter 4 presents a systematic framework to reveal from online discussions how smart home users develop security and privacy attitudes in real life [166]. To cope with users' highly diverse security and privacy terminology, we design a natural language processing (NLP)-assisted pipeline in sampling 180 privacy or security-related discussion threads with 4,957 comments from a major smart home forum on Reddit [235]. Our content analysis of these discussions confirms our hypothesis and illustrates how users' attitudes, built on considerations of security and privacy concerns and protective strategies, evolve as they interact with products and exchange information in different phases. Our research reveals that users wrestle with the unavailability of security and privacy tools and resources, inaccessible security and privacy-related information, and social pressures. Based on our findings, we provide recommendations

to improve smart home designs, support users' attitude development, facilitate information exchange, and guide future research regarding smart home security and privacy.

Understanding how interaction experiences influence security perceptions of VR authentication. VR systems introduce novel and immersive interaction experiences to users, compared to other smart devices. Users embrace the rapid development of VR technology for everyday settings, such as gaming, social interactions, shopping, and commerce. VR systems access sensitive user data and assets to enable payment, which necessitates user authentication to secure such interaction. Despite further improving the security properties of authentication in VR [193, 150, 56, 330, 171], it is equally important to understand and improve users' security perception of VR authentication. As such, we can aid users in assessing the security risks better and taking protective actions. Recent research has shown that how users interact with the authentication systems and the context in which authentication is used affect users' perceived security [192]. Thus, we hypothesize that such interplay between users' interactions and their perceived security also exists for VR authentication. In this work, we attempt to answer this question:

> *How do smart home users' interaction experiences factor into their security perceptions of authentication in VR?*

Proposed solution. To explore the open design space in the early adoption of VR authentication, our work presented in Chapter 5 leverages "technology probes" to understand this question [163]. We design probes of authentication in VR to provoke participants' reactions from multiple angles. We embed these probes in the routine payment of a VR game, creating an organic study context. Our qualitative analysis reveals the interplay between participants' interaction experiences and security

perception. We show that our participants benefited from the intuitive virtualization of authentication and enjoyed the immersion in VR, despite encountering unique usability challenges in VR interaction. Then, we observe how these interaction experiences influence participants' transfer of prior authentication understanding into VR, leading to a discrepancy in perceived security. Further, we identify various tensions in participants' expectations, involving both their desire for an enjoyable VR experience and secure VR authentication. We propose recommendations to address these expectations and mitigate conflicts based on our observations.

In this dissertation, we demonstrate that designing security and privacy-preserving tools that account for smart device functionalities can enable better security and privacy controls for smart device users; additionally, studying how users interact with these devices at different stages of adoption informs design and practices to make their interactions more secure and private in real life.

2 DEVELOPING USABLE PRIVACY CONTROL FOR REAL-TIME EYE-TRACKING SYSTEMS

Recent advances in sensing and computing technologies have facilitated the rapid development of interaction modalities with smart devices. Among these modalities, eye tracking arises as a hands-free interface in augmented, virtual, and mixed reality settings. It offers users control over virtual components [279], events [168], and digital avatars [263], especially in settings where hand-based control is either impractical or infeasible [288]. Interactive systems are now capable of performing continuous eye tracking using off-the-shelf webcams [217], smartphones [203], tablets [112], desktops [204], wearable glasses [317], and mixed reality headsets such as the HTC VIVE and Microsoft HoloLens.

From a stream of eye gaze positions in a scene, eye-tracking applications precisely estimate what the user is viewing to trigger events, prefetch scenes, or perform actions in the virtual environment. One's eye gaze streams, however, are vulnerable to potential privacy threats. Previous research has demonstrated that psychological and physiological factors direct the formation of unique patterns in the user's eye gazes. For instance, researchers were able to infer insights about the user's behavioral traits [154, 249, 256], diagnose Alzheimer's disease and autism spectrum disorder [130, 109], understand the user's familiarity of a scene [260], infer mental status during social interaction [255], detect personality traits [34], and deliver personalized advertisements [86, 45, 314].

Third-party applications that use eye gaze streams can extract information beyond their intended core functionality, posing significant privacy threats to the users. For example, Figure 2.1(a) shows the heatmap of eye gazes on a web page from an individual user. While an application can help the user scroll up/down the web page, the aggregated eye gaze positions can reveal the user's interest. Unfortunately, the eye-tracking

No privacy (ε=∞) Low privacy (ε=3) High privacy (ε=0.5)

(a) Raw data (b) Noisy data from **Kalɛido**

Figure 2.1: Eye gaze heatmaps from an individual user with and without Kalɛido's noising effect on a web page.

platforms do not offer users the ability to control their privacy. They relay the raw eye gaze streams to the applications without much regard to the embedded sensitive information.

Researchers have developed privacy-preserving mechanisms for eye gaze streams [100, 37, 261, 175, 36] to alleviate these concerns. These mechanisms share a similar working principle: allowing access to only some high-level "features" of the eye gaze streams, possibly with some added noise, instead of the raw gaze streams. While some of them provide formal privacy guarantees [261, 175, 36], they are mostly impractical to deploy due to multiple limitations. First, they require modification of the eye-tracking application programming interfaces (APIs) since the applications expect to receive a sequence of raw eye gaze positions, not just features. Second, processing eye gaze streams to extract features does not happen in real-time, affecting the user experience. Third, they require the user to control a set of parameters that are hard to understand for most users. In short, the question of how to provide a backward-compatible, easy-to-use privacy-preserving system for real-time eye tracking is still an open one.

In this chapter, we design, implement, and evaluate Kalɛido, a usable

privacy control that preserves the real-time utilities of eye tracking, as an affirmative answer to the above question. Kalεido provides a formal privacy guarantee based on differential privacy (DP) [78], the de-facto standard for achieving data privacy. To the best of our knowledge, Kalεido is the first system to (1) provide a privacy guarantee on raw eye gazes, (2) seamlessly integrate with the existing eye-tracking ecosystem, and (3) operate in real-time. Kalεido offers the following advantages:

• **Formal privacy guarantee.** Kalεido uses a differentially private algorithm to release noisy eye gaze streams to the applications, which *protects the spatial distribution of a gaze trajectory that is formed within any window of a specific duration* (as determined by the users). Kalεido achieves this objective by bringing the privacy semantics from two distinct contexts, absolute location data and streaming event data, into the domain of eye gaze data (Section 2.2). Figure 2.1(b) shows Kalεido's privacy protection in action.

• **Seamless integration with the eye-tracking ecosystem.** As Kalεido operates on raw eye gaze streams, it fits within the existing ecosystem of eye-tracking applications. It is also platform- and application-agnostic; it operates on popular eye-tracking platforms and requires no modification of the applications, making it more practical to deploy.

• **Ease of use.** As the parameters of Kalεido's privacy guarantee are a function of the visual feed semantics, it reduces the burden of complex privacy configuration on the user.

We integrate Kalεido as a Unity [95] plugin; it acts as a protection layer between untrusted applications and trusted platforms. Unity is the mainstream engine for gaming and mixed reality applications; it supports various peripherals, such as eye-tracking sensors. Kalεido's architecture comprises four major components: (1) *context processing core*, which extracts scene semantics from keyframes of dynamic visual feed; (2) *configuration manager*, which automatically configures the parameters of the

DP guarantee based on scene semantics and user preferences; (3) *noisy gaze generator* which generates noisy gaze streams; and (4) *noisy gaze processor*, which performs local post-processing on the noisy gaze streams. The Kalɛido plugin leverages off-the-shelf APIs and computing blocks, providing backward compatibility across a broad spectrum of applications and platforms.

We conducted a user study and trace-based analysis to evaluate Kalɛido. To understand perceived utility, we investigated the user experience of a real-time eye-tracking game with Kalɛido. The quantitative and qualitative feedback indicates a minor impact on users' game performance and satisfaction. The users showed a high incentive to adopt Kalɛido and its control knob for eye-tracking privacy. Furthermore, we validate that Kalɛido can successfully thwart various adversarial analytics, aiming to identify unique traits from users' eye gazes. Even with modest privacy levels, Kalɛido can drive the attacker's accuracy close to random baselines.

2.1 Background on Eye Tracking

Properties of Eye Gaze

Eye gaze data, commonly represented as a stream of gaze positions projected onto a visual scene, reflects how people explore and process the visual content. Typically, eye gaze data is abstracted as a *scanpath*, which captures the characteristics of the user's visual attention [222]. A scanpath is a time sequence of *fixations* that are separated by *saccades* [30, 275]. Fixations represent clusters of gazes concentrated around specific regions in the scene (such as an object). Saccades denote gazes traveling rapidly from one fixation to another. A region in the scene space that attracts human attention [196] is referred to as a *region of interest* (ROI). Figure 2.2 illustrates fixations, saccades, and ROIs in a scene.

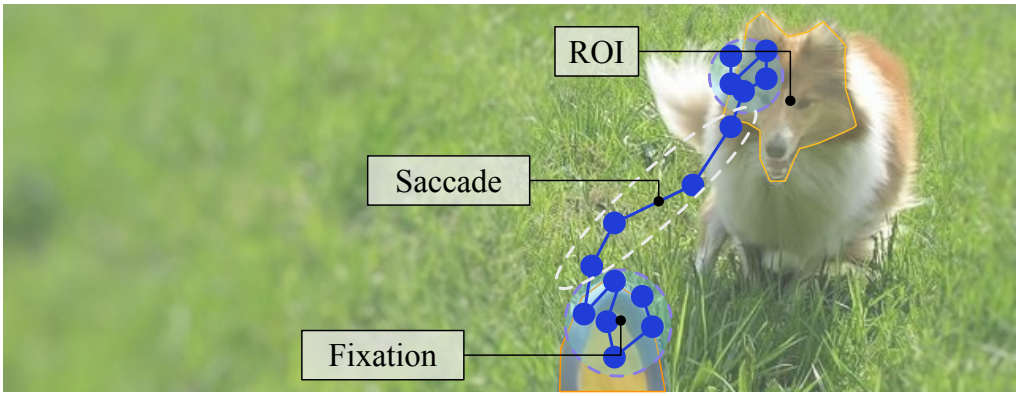


Figure 2.2: Example of fixations, saccades, and ROIs in a scene [174], where the blue dots represent individual gazes and purple (grey) dashed circles represent fixations (saccades).

Eye-Tracking Platform

Two of the most popular techniques for acquiring real-time eye gaze [187] are: *vision-based tracking* and *infrared pupil-corneal reflection tracking*. The former estimates gaze positions from the captured images of the eyes; the latter projects infrared light onto the eyes and estimates the point of gaze from the pupil and corneal reflections. The raw measurement data is represented as a stream of tuples $\langle x, y, t \rangle$, where x and y represent the 2D coordinates of its location on the visual scene (corresponding to a pixel of the image), and t is the associated timestamp [277, 283, 151].

Eye-tracking platforms [123, 142] incorporate eye-tracking with development engines, such as Unity. The platform exposes eye gaze streams to user applications through predefined APIs. An application *session* is the duration of user interaction with the platform to perform a task, such as playing a game or browsing a document. Each session is a series of *scenes* where the visual content remains relatively unchanged (e.g., part of the same panoramic view).

Each application defines its interaction semantics based on the eye gaze streams. Examples include eye gaze-based input and selection [279], active event triggering by eye gaze gestures [168], automatic scene switching during browsing [147], foveated rendering [220, 21], and virtual social interaction using digital avatars [263].

Privacy Threats

Eye gaze patterns inherently reflect human traits and carry sensitive information about the user. While the applications would primarily process eye gaze streams for user interaction purposes, accumulating the data over multiple sessions can result in privacy threats. Below, we discuss some examples of possible psychological and physiological inferences that can be drawn from eye gaze streams.

Absolute gaze distribution on a scene. The spatial distribution of absolute gaze positions on a scene can reveal insights about the individual's cognitive process of exploring specific visual content. Fixations and saccades within and between ROIs reflect how an individual's attention moves within a scene – revealing cues about one's interest. For example, gaze patterns on merchandise can enable precision marketing and personalized recommendations in consumer research [86, 45, 314]. Other researchers have attributed individuals' fixation patterns to their psychological state, such as lying about recognizing a face [260, 202]. Further, individuals with different physiological and cultural backgrounds demonstrate distinguishing characteristics depending on the ROI features such as color, texture, and semantics [6, 231].

Aggregate statistics on gaze distribution over time. The statistical characteristics or features of scanpaths computed over a period of time, such as fixation duration/rate and saccade speed/acceleration, can reveal sensitive information about an individual. For example, the length of saccades can help in categorizing fixations into different functional groups, including

"locating," "guiding," "directing," and "checking," which reveal one's behavioral traits while performing daily tasks, such as interpersonal communication [154, 249, 256]. Diseases such as autism spectrum disorder [109] and Alzheimer's [130] can also be diagnosed from fixation features. Additionally, fixation and saccade features can be utilized as biometrics for user identification and authentication [80, 113] because of their uniqueness to individuals. These features can also reveal information about a user's physiological conditions, such as vision correction conditions [206].

2.2 Privacy Model

As discussed in Section 2.1, we observe that the privacy threats to eye-tracking data arise either from the analysis of the absolute spatial distribution or the aggregate statistics of gaze positions over time. Thus, the spatial information of the gaze positions is the primary source of sensitive information. Hence, in Kalɛido, we choose to provide our formal guarantee (Definition 2.5) on the spatial information of the gaze positions. In what follows, we start with some background on differential privacy, followed by the privacy definition for Kalɛido and its implications.

Differential Privacy Preliminaries

For Kalɛido's formal privacy guarantee, we leverage two variants of differential privacy: geo-indistinguishability [17] and w-event differential privacy [132].

Geo-indistinguishability. Geo-indistinguishability is a specialization of differential privacy that provides privacy guarantees for geographical information in 2D space. It is formally defined as follows:

Definition 2.1 (**(ϵ, r)-geo-indistinguishability**). *A mechanism $\mathcal{M} : \mathcal{X} \mapsto \mathcal{Z}$ is defined to be (ϵ, r) - geo-indistinguishable iff for all pairs of inputs $(x, x') \in \mathcal{X} \times \mathcal{X}$ such that $d(x, x') \leqslant r$,*

$$\forall S \subset \mathcal{Z}, Pr[\mathcal{M}(x) \in S] \leqslant e^{\epsilon} Pr[\mathcal{M}(x') \in S] \tag{2.1}$$

where $d(\cdot, \cdot)$ denotes the Euclidean metric.

We refer to the pair (x, x') in the above definition as the r-Euclidean neighboring. Intuitively, the above definition protects all pairs of r-Euclidean neighbors[1].

***w*-event differential privacy.** As discussed above, eye gaze data in real-world interaction interfaces is obtained in the form of streaming data. Hence, we also use a variant of the *w-event differential privacy* guarantee [132], which is defined in the context streaming data. In this context, the user's behavior breaks into a set of "events," corresponding to data updates in the stream due to user actions. Intuitively, this privacy guarantee protects all event sequences of length w in a stream.

Let S be a stream of an infinite tuple $S = (D_1, D_2, \cdots)$ where every data point D_i at time stamp i is a database with d columns and arbitrary rows (each row corresponds to an unique user). Let S_t denote a stream prefix of S up till time stamp t, $S_t = (D_1, D_2, \cdots, D_t)$, and $S_t[i], i \in [t]$ denote the i-th element of S_t, D_i.

Definition 2.2 (w-Neighboring Stream Prefixes [132]). *Two stream prefixes S_t, S'_t are defined to be w-neighboring, if*

- *for each $S_t[i], S'_t[i]$ such that $i \in [t]$ and $D_i = S_t[i] \neq S'_t[i] = D'_i$ it holds that, D'_i can be obtained from D_i by adding or removing a single row, and*

[1] We introduce some notational change from the original work [17]. Our privacy parameter ϵ is equivalent to the term $\epsilon \cdot d(x, x')$ from the original definition (see Section 2.2 for details). We adopt this change to improve readability, which does not affect the semantics of the definition.

- *for each* $S_t[i_1], S_t[i_2], S'_t[i_1], S'_t[i_2]$ *with* $i_1 < i_2, S_t[i_1] \neq S'_t[i_1]$ *and* $S_t[i_2] \neq S'_t[i2]$, *it holds that* $i_2 - i_1 + 1 \leqslant w$.

Using the above definition, w-event differential privacy is defined formally as follows:

Definition 2.3 (w-Event Differential Privacy [132])**.** *A mechanism* $\mathcal{M} : \mathcal{S} \mapsto \mathcal{C}$, *where* $\mathcal{S}$ *is the domain of all stream prefixes, satisfies* w*-event differential privacy if for all pairs of* w*-neighboring stream prefixes* $\{S_t, S'_t\} \in \mathcal{S} \times \mathcal{S}$, *we have*

$$\forall O \subseteq \mathcal{C}, \forall t, Pr[\mathcal{M}(S_t) = O] \leqslant e^{\epsilon} Pr[\mathcal{M}(S'_t) = O] \tag{2.2}$$

Note that w refers to the count of distinct "events" in a stream in the above definition. In our definition, w refers to the duration of the event window (as in Definition 2.5).

Privacy Definitions in Kalϵido

We now discuss how the aforementioned privacy definitions are used for protecting eye gaze streams. We observe that in a 2D scene, the eye gaze data is analogous to geographical information as modeled in the geo-indistinguishability framework [17]. Specifically, we can use the Euclidean distance as a metric for gaze data points. Keeping this in mind, we model the eye gaze time series as a stream of an infinite tuple $S^g = (\langle g_1, t_1\rangle, \langle g_2, t_2\rangle, \cdots)$, where each data point $g_i = \langle x_i, y_i\rangle$ gives the corresponding 2D gaze position, and t_i is the associated timestamp. Let S^g_k denote a stream prefix of S^g of length k, i.e., $S^g_k = (\langle g_1, t_1\rangle, \langle g_2, t_2\rangle, \cdots, \langle g_k, t_k\rangle)$. Using this model of eye gaze positions, we present our notion of (w, r)-neighboring for gaze stream prefixes.

Definition 2.4 (**(w, r)-neighboring gaze stream prefixes**)**.** *Two gaze stream prefixes* $S^g_k = (\langle g_1, t_1\rangle, \cdots, \langle g_k, t_k\rangle), S^g_{k'} = (\langle g'_1, t'_1\rangle, \cdots, \langle g'_k, t'_k\rangle)$ *are defined to be* (w, r)*-neighboring, if*

- *the timestamps of their elements are pairwise identical: for* $i \in [k]$*, we have* $t_i = t'_i$*;*
- *the gaze positions of their elements are* r*-Euclidean neighboring: for each* g_i, g'_i *such that* $i \in [k]$*, it holds that* $d(g_i, g'_i) \leqslant r$*; and*
- *all of the neighboring gaze points can fit in a window of time duration at most* w*: for each* $g_{i_1}, g_{i_2}, g'_{i_1}, g'_{i_2}$*, with* $i_1 < i_2, g_{i_1} \neq g'_{i_1}$ *and* $g_{i_2} \neq g'_{i_2}$*, it holds that* $t_{i_2} - t_{i_1} \leqslant w$*.*

Leveraging the notion of neighboring gaze stream prefixes, we present our formal privacy definition as follows. This definition is a variant of the *w-event differential privacy* guarantee [132].

Definition 2.5 (**(ϵ, w, r)-differential privacy for gaze stream prefixes**)**.** *A mechanism* $\mathcal{M} : \mathcal{S}^g \mapsto \mathcal{C}^g$*, where* $\mathcal{S}^g$ *is the domain of all stream prefixes, satisfies* (ϵ, w, r)*-differential privacy if for all pairs of* (w, r)*-neighboring gaze stream prefixes* $\{S_k^g, S_k^{g'}\} \in \mathcal{S}^g \times \mathcal{S}^g$*, we have*

$$\forall O \in \mathcal{C}^g, \forall k, \Pr[\mathcal{M}(S_k^g) = O] \leqslant e^{\epsilon} \cdot \Pr[\mathcal{M}(S_k^{g'}) = O] \tag{2.3}$$

Based on this definition, we present a result that enables a (ϵ, w, r)-differentially private mechanism to allocate a privacy budget of ϵ for any sliding window of duration w in a given stream prefix.

Theorem 1. *Let* $\mathcal{M} : \mathcal{S}^g \mapsto \mathcal{C}^g$ *be a mechanism that takes as input a gaze stream prefix* $S_k^g = (\langle g_1, t_1 \rangle, \cdots, \langle g_k, t_k \rangle)$ *and outputs a transcript* $O = (o_1, \cdots, o_k) \in \mathcal{C}$*. Additionally, let* $\mathcal{M}$ *be decomposed into* k *mechanisms* $\mathcal{M}_1, \cdots, \mathcal{M}_k$ *such that* $\mathcal{M}_i(g_i) = o_i$*, and each* $\mathcal{M}_i$ *generates independent randomness while achieving* (ϵ_i, r)*-geo-indistinguishability. Let* $l \in [1, i-1]$ *represent an index such that*

$(t_i - t_l) = w$. *Then,* $\mathcal{M}$ *satisfies* (ϵ, w, r)*-differential privacy if*

$$\forall i \in [k], \sum_{j=l}^{i} \epsilon_j \leqslant \epsilon \tag{2.4}$$

The proof of Theorem 1 follows directly from the proof of Theorem 3 in Kellaris et al. [132].

Discussion of privacy semantics. The idea behind (ϵ', r)-geo-indistinguishability (Definition 2.1), in the context of eye-tracking data, is that given a gaze position g, all points within a circle of radius r centered at g (i.e., all r-neighbors of g) would be "indistinguishable" to an adversary who has access to the corresponding "noisy" location. Thus, this privacy guarantee provides a cloaking region of radius r around g. (ϵ, w, r)-differential privacy (Definition 2.5) extends this guarantee to gaze stream prefixes. Specifically, an adversary cannot distinguish[2] between any two gaze stream prefixes, which (1) differ in gaze positions that are within a distance of r from each other, and (2) all such differing pairs occur within a window of duration w.

Additionally, from Theorem 1, we observe that a (ϵ, w, r)-differentially private mechanism can achieve two goals: for every subsequence of duration w in the gaze stream S^g_k, it (1) allocates up to ϵ privacy budget, and (2) takes budget allocation decisions considering the *entirety* of the subsequence. Thus, this privacy definition *protects the spatial distribution of any gaze trajectory that is formed over any window of a duration* w.

Further, we define and prove another result, which shows that the privacy guarantee degrades gracefully if the r-Euclidean neighbors in both stream prefixes are separated by more than w duration.

[2] with probability higher than what is allowed by the privacy parameter ϵ

Theorem 2 (Composition over multiple windows theorem). *Let* $\mathcal{M} : \mathcal{S}^g \mapsto \mathcal{C}^g$ *be a mechanism that takes as input a gaze stream prefix* $S_k^g = (\langle g_1, t_1\rangle, \cdots, \langle g_k, t_k\rangle)$*, and outputs a transcript* $O = (o_1, \cdots, o_k) \in \mathcal{C}$*. Additionally, let* $\mathcal{M}$ *be decomposed into* k *mechanisms* $\mathcal{M}_1, \cdots, \mathcal{M}_k$ *such that* $\mathcal{M}_i(g_i) = o_i$*, and each* $\mathcal{M}_i$ *generates independent randomness while achieving* (ϵ_i, r)*-geo-indistinguishability. Then for two stream prefixes* S_k^g *and* $S_k^{g'}$*, such that:*

- *for all* $i \in [k]$, $t_i = t_i'$;
- *for each* g_i, g_i' *such that* $i \in [k]$ *and* $g_i \neq g_i'$ *it holds that* $d(g_i, g_i') \leqslant r$*, i.e.,* $(g_i, g_{i'})$ *are* r*-Euclidean neighboring; and*
- *for each* $g_{i_1}, g_{i_2}, g_{i_1}', g_{i_2}'$*, with* $i_1 < i_2, g_{i_1} \neq g_{i_1}'$ *and* $g_{i_2} \neq g_{i_2}'$*, it holds that* $t_{i_2} - t_{i_1} \leqslant m \cdot w, m \in \mathbb{N}$;

we have

$$\forall O \in \mathcal{C}^g, \forall k, Pr[\mathcal{M}(S_k^g) = O] \leqslant e^{m\cdot\epsilon} \cdot Pr[\mathcal{M}(S_k^{g'}) = O]. \tag{2.5}$$

Proof. Let $m = 2$ and i_1 be the least index such that $g_{i_1} \neq g_{i_1}'$ and i_2 be the highest index such that $g_{i_2} \neq g_{i_2}'$. Additionally, let $i* \in [i_1, i_2]$ such that $time(i*) - time(i_1) = w$. Let $S_{i*}^g = (\langle g_1, t_1\rangle \cdots \langle g_{i*}, t_{i*}\rangle)$,$S_{k*}^g = (\langle g_{i*+1}, t_{i*+1}\rangle \cdots \langle g_k, t_k\rangle)$ and $O = O_1||O_2, |O_1| = |S_{i*}^g|, |O_2| = |S_{k*}^g|, O \in \mathcal{C}^g$. Now using the independence of noise generation for each gaze position,

$$\begin{aligned} Pr[\mathcal{M}(S_k^g) = O] &= Pr[\mathcal{M}(S_{i*}^g) = O_1] \cdot Pr[\mathcal{M}(S_{k*}^g) = O_2] \\ &\leqslant e^{\epsilon} \cdot Pr[\mathcal{M}(S_{i*}^{g'}) = O_1] \cdot e^{\epsilon} \cdot Pr[\mathcal{M}(S_{k*}^{g'}) = O_2] \\ &= e^{2\epsilon} \cdot Pr[\mathcal{M}(S_k^{g'}) = O] \end{aligned}$$

The rest of the proof follows trivially using induction using the above case as the base. □

Another important result for differential privacy is that any post-processing computation performed on the noisy output does not cause any privacy

loss. Thus, once Kalєido releases the noisy gaze streams, all subsequent analyses by the adversary enjoy the same privacy guarantee.

Theorem 3 (**Post-processing**). *Let the randomized mechanism* $\mathcal{M} : \mathcal{S}^g \mapsto \mathcal{C}^g$ *satisfy* (ϵ, w, r)*-differentially privacy. Let* $f : \mathcal{C}^g \mapsto \mathcal{R}$ *be an arbitrary randomized mapping. Then* $f \circ \mathcal{M} : \mathcal{S}^g \mapsto \mathcal{R}$ *is* (ϵ, w, r)*- differential private.*

Privacy Implications of Kalєido

In the following, we discuss the implications of the formal privacy guarantee of Kalєido (Definition 2.5).

Choice of Parameters

The aforementioned privacy guarantee involves three parameters – the privacy budget, the window length, and the radius of location indistinguishability:

Privacy budget ϵ. ϵ captures the privacy requirements of the user which can be set at the user's discretion [116, 160, 4].

Window length w. As explained above, the proposed privacy definition protects the spatial distribution of a gaze trajectory that is formed within any window of duration w. In a typical eye-tracking setting, gaze trajectories are formed over individual visual scenes. Thus, a good choice for w could be average scene lengths in a visual feed. Over the whole session, which spans multiple windows, the resulting privacy guarantee degrades gracefully (by Theorem 2).

Radius of location indistinguishability r. Recall that eye gaze streams be abstracted to a series of *fixations* and *saccades* within and between ROIs. Hence, we propose the following two choices for the value of parameter r:

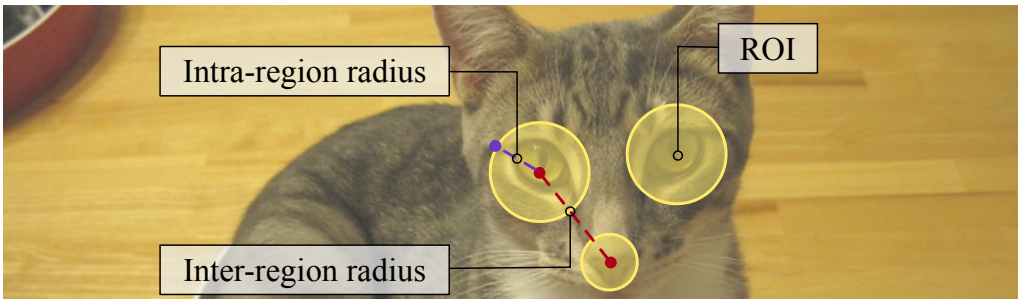


Figure 2.3: Illustration of the two choices for the radius of location indistinguishability parameter [174].

- **Intra-region radius r_{intra}.** This measure captures the radius of a single ROI (approximated by a circular area) and is catered to protect gaze data positions corresponding to fixations.
- **Inter-region radius r_{inter}.** This measures the distance between a pair of ROIs (approximated by circular areas) and protects gaze positions corresponding to inter-ROI saccades.

The two radii are illustrated in Figure 2.3. As a general rule, the larger the value of r, the greater the privacy is enjoyed (at the cost of lower utility). Note that we assume that the visual feeds are publicly available (see Section 2.3).

Thus, in a nutshell, Kaleido's privacy guarantee ensures that an adversary cannot learn about the distinguishing features of a user's spatial distribution. Specifically, if r is chosen as r_{intra}, then an adversary cannot distinguish[3] between two users gazing at the same ROI, within any window of length w. Similarly, if r is chosen as r_{inter}, then the adversary cannot distinguish two users such that (1) user 1's gaze moves from ROI_1 to ROI_2, and (2) user 2's gaze moves from ROI_1 to ROI_3, within any window of length w.

[3]with probability higher than what is allowed by privacy parameter ϵ

Discussion on Temporal Information of Eye Gaze

Kalεido's formal privacy guarantee focuses solely on the location information of eye gaze streams. However, as discussed in Section 2.1, some privacy attacks utilize both location and temporal information (aggregate statistics) of gaze streams. In these cases, the location information contained in the aggregate statistics constructed over noisy gaze positions (Definition 2.5) will also be noisy (Theorem 3) – thereby reducing the efficacy of the attacks. Our evaluation results in Section 2.5 provide *empirical evidence* for the above: Kalεido is able to protect against analyses that exploit such spatio-temporal statistics. Additionally, a formal guarantee on the temporal information would require interfering with the timeliness of the release of gaze data points (noisy or otherwise), which might adversely affect the utility [97]. Nevertheless, Section 2.6 discusses a possible extension of Kalεido for providing a formal guarantee on the temporal information of eye gaze streams.

Contributions of Kalεido's Privacy Definition

Here, we discuss the contributions of Kalεido's formal privacy definition (Definition 2.5).

First, this definition combines the privacy semantics from two distinct contexts: absolute location data and the streaming of event data. Specifically, Definition 2.5 provides (ϵ, r)-geo-indistinguishability guarantee for every gaze position within a window of duration w in a gaze stream.

Second, there are certain semantical differences in the use of location perturbation techniques (such as (ϵ, r)-geo-indistinguishability guarantee) in the contexts of geographical information and eye gaze data. Typically, ROIs (also known as points of interest) for geographical information include physical units such as restaurants, shopping malls, or schools. On the other hand, ROIs in the eye-tracking context are characterized by visual stimuli such as the scene's color and texture. Consider a case where only

a single ROI is located within a circle of radius r centered at the true user location (or eye gaze position). In the case of geographical information, the adversary can conclude that the user is visiting the particular ROI. Thus, this completely violates the user's location privacy. However, the above-described scenario corresponds to a fixation event (r_{intra}) in the context of eye-tracking, and eye movements, even within a single ROI are a rich source of sensitive information [231] (as discussed in Section 2.1). Thus, even if the adversary learns the ROI's identity, the perturbation still provides meaningful privacy protection.

Additionally, for the standard geo-indistinguishability guarantee [17], the privacy guarantee enjoyed is parameterized by the multiplicative term $\epsilon \cdot d(x, x')$, i.e., the privacy guarantee degrades with the distance between the pair of points $\{x, x'\}$. This makes the task of choosing the value of ϵ tricky for geographical data [215]. The reason behind this is that, for any given value of ϵ, if the distance $d(x, x')$ becomes too large, then the subsequent privacy guarantee provided ceases to be semantically useful. Hence, deciding on the size of the cloaking region ($d(x, x')$), such that any two points within the region are sufficiently protected, is difficult for geographical data in practice. However, in the context of eye gaze data, sensitive information is captured in the form of fixations and saccades. Thus here, we are primarily concerned about protecting pairs of gaze positions that are bounded by a specific distance (r_{intra} and r_{inter} as discussed in Section 2.2). Hence, our formulation (Definition 2.1) explicitly parameterizes the size of the cloaking region, r, and its privacy parameter, ϵ, is equivalent to the term $\epsilon \cdot d(x, x')$ (equivalently, $\epsilon \cdot r$ where $d(x, x') \leqslant r$) from the original definition. This ensures that all pairs of gaze positions within a distance of r from each other enjoy a privacy guarantee of at least ϵ, thereby mitigating the aforementioned problem.

2.3 Kalϵido System Design

We introduce the system design of Kalϵido, starting with the threat model followed by design goals. Next, we present the architectural overview followed by detailed descriptions.

Threat Model

The software stack of real-time eye tracking comprises two major parties: the *eye-tracking platform* and the *third-party application* (Section 2.1). In our threat model, we assume the eye-tracking platform to be trusted (a common assumption in prior works [124, 244]) and consider the untrusted third-party application to be the adversary. The application can perform analysis on the gaze streams to learn sensitive information about the user (as described in Section 2.1). Additionally, we assume that the visual feeds (image or video scenes users look at) are publicly available. This assumption holds in most practical eye-tracking applications such as movies and VR games. Thus, attackers (untrusted third-party applications) can access visual feeds and noisy gazes (output of Kalϵido), but not raw gazes.

Kalϵido Design Principles

Kalϵido relies on the following three design principles.

- **Seamless integration with existing eye-tracking interfaces.** Kalϵido seamlessly integrates with the current eye-tracking ecosystem. Specifically, it interacts with the different components of the eye-tracking framework using their existing interfaces.
- **Real-time system.** Kalϵido is capable of generating noisy gaze streams (satisfying Definition 2.5) in real-time that is suitable for interactive eye-tracking interfaces.

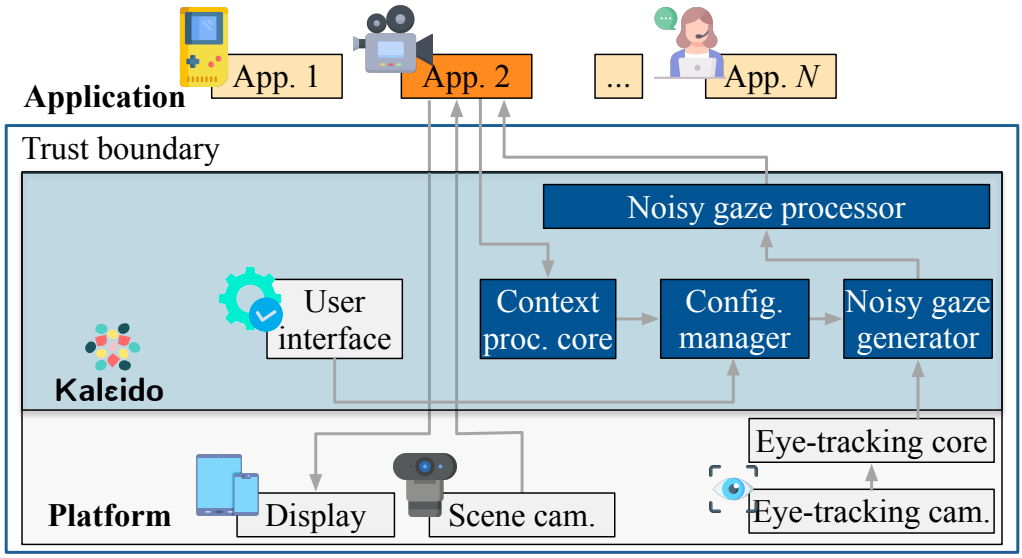


Figure 2.4: Architectural overview of Kaleido.

- **Automatic privacy parameter configuration.** Kaleido automatically configures the privacy parameters, namely w and r, based on the properties of the visual feed.

Architectural Overview

Figure 2.4 depicts the high-level architecture of the eye-tracking framework with Kaleido. It comprises three layers: the eye-tracking platform, Kaleido, and the applications. Kaleido is an intermediary layer in this stack that defines the trust boundary.

Eye-tracking platform. The eye-tracking platform includes a display, the eye-tracking camera, the eye-tracking core, and potentially a scene camera. Users consume the visual feed via the platform-specific display, generated either entirely digitally (VR platforms) or from the scene camera (augmented reality platforms). The eye-tracking camera captures eye image

frames, from which the eye-tracking core generates raw gaze streams.
Kalεido**.** Kalεido processes the raw gaze stream obtained from the eye-tracking platform in a privacy-preserving manner. Based on the information from the visual feed and user-specified guidelines, it automatically configures the parameters required for the privacy guarantee of Definition 2.5. It then perturbs the raw gaze stream, sanitizes it, and feeds it to the applications. Section 2.3 elaborates the design of Kalεido.
Applications. The applications use eye gaze streams for their functionalities. They receive gaze streams (albeit noisy) from Kalεido using the original APIs. Therefore, they need not be modified in any way to be compatible with Kalεido.

Kalεido System Modules

Kalεido views user interaction with the eye-tracking platform as a set of sessions with dynamic scenes. We elaborate on Kalεido's modules and how it achieves its privacy guarantee.

Context Processing Core

The context processing module extracts the size and locations of the ROIs from individual frames (still images of a scene) of the visual feed. Kalεido adopts off-the-shelf region and object detectors [178, 298] for ROI extraction. However, these detectors are computationally heavy, and continuously running them results in a high computational overhead that might hinder real-time operation. Kalεido solves this challenge by incorporating a threshold-based keyframe detector. As frames remain relatively consistent over short periods, Kalεido invokes the object detector only in the instances of a scene change.

Configuration Manager

The configuration manager module automatically configures the privacy parameters to satisfy the privacy guarantee of Definition 2.5. It accepts as inputs the processed scene information from the context processing core and the user's privacy preferences, and it configures the parameters as follows:

Privacy budget ϵ. For setting the value of ϵ, Kalεido provides the users with a privacy scale ranging from no privacy (releases raw gaze streams) to high privacy (releases noisy gaze streams). Users can adjust this knob during an active session through the configuration manager's UI, and Kalεido interpolates the corresponding value of ϵ in the background.

Window length w. As discussed in Section 2.2, w is set according to scene lengths. Each scene corresponds to a period during which the visual content, e.g., a video, remains relatively static as defined in Section 2.1. The configuration manager can compute this value either on the fly from the context processing core's scene detectors or offline profiling and video metadata. Small values of w (of the order of a few seconds) usually work well as most real-world interactive scenes are rapidly changing and spatially heterogeneous.

Radius of location indistinguishability r. The configuration manager module sets the value of r based on either r_{intra} or r_{inter} according to the user's preference. It uses the set of detected ROIs for each scene to compute r as follows. Let $\{ROI_i\}, i \in [N]$, denote the set of ROIs for a given scene where N is the total number of ROIs. Let a tuple $\langle x_i, y_i, d_i^w, d_i^h \rangle$ represent the output of the object (or region) detector, where (x_i, y_i) is the position of a reference point (for example, the centroid) of the bounding box of ROI_i, and (d_i^w, d_i^h) is its width and height, respectively. Thus, ROI_i can be approximated by a circular area centered at (x_i, y_i) and its radius

that is computed from the diagonal of the bounding box:

$$r^i_{intra} = 0.5 \times \sqrt{d_i^{w^2} + d_i^{h^2}} \tag{2.6}$$

For any pair of regions of interest (approximated by circular areas) ROI_i and ROI_j $i, j \in [N], i \neq j$, we have

$$r^{i,j}_{inter} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{2.7}$$

After computing the radii of all ROIs, the configuration manager has two default modes for r: r_{small}, which is the median of $\{r^i_{intra}\}$, and r_{large}, which is the median of $\{r^{i,j}_{inter}\}$.

Noisy Gaze Generator

The noisy gaze generator module perturbs the raw gaze streams generated by the eye-tracking core. This perturbation entails allocating a privacy budget for each gaze position and then generating its corresponding noisy position in a (ϵ, w, r)-differential private manner (Definition 2.5).

The raw measurement frequency is very high (~ 120 Hz), especially for interactive settings. Even for low values of w, the number of individual gaze positions could be relatively high. Therefore, naive budget allocation strategies such as uniform allocation or fixed-rate sampling are likely to provide poor utility [132]. To this end, we use an adaptive budget allocation strategy that considers the dynamics of the human eye gaze. We observe that the *human gaze is relatively localized during fixations*. Based on this observation, we identify two optimizations for the budget allocation strategy. Let g' denote the last published noisy gaze position.

- Gaze data points generated in quick succession of g' can be skipped over.
- The last released g' can be used as a proxy for data points that lie in its

spatial proximity.

These optimizations are akin to (1) performing a simple fixation detection (in a privacy-preserving manner) based on the spatio-temporal gaze data points, and (2) publishing a noisy gaze position only when a new fixation is detected. This requires the privacy budget to be distributed between two tasks: testing the proximity of the gaze positions and the publication of noisy gaze positions. The temporal check (for skipping data points) consumes no privacy budget since our formal guarantee (Definition 2.5) applies to spatial information only.

Kalεido uses an *adaptive budget allocation* strategy that (1) starts with a total privacy budget ϵ for every window of duration w, (2) allocates no budget for the gaze data points to be skipped over, (3) allocates a fixed budget for testing all other data points, (4) distributes publication budget in an exponentially decreasing manner to the data points which have been decided to publish, and (5) recycles the budget spent in timestamps falling outside the active window. Algorithm 1, based on the BD algorithm [132], outlines the above method; similar ideas have also been presented in the context of location sequences [51].

Algorithm 1 Adaptive Budget Allocation

Parameters: w - Time duration of a single window in seconds (s),
ϵ - Total privacy budget per window of size w
p_{raw} - Rate of raw gaze data generation in $samples/s$
l_{thresh} - Threshold for distance
t_{skip} - Time duration for skipping after every gaze data point testing
r - Radius of indistinguishability
h - Ratio of privacy budget used for testing

Initialization:
$n_{raw} = w \cdot p_{raw}$ ▷ Number of raw gaze data points generated in a single window
$n_{test} = \lceil w/t_{skip} \rceil$ ▷ Number of raw gaze data points tested in a single window
$\epsilon_{test} = \epsilon/(h \cdot n_{test})$ ▷ Privacy budget allocated for every test in a single window
$i_{test} = \varnothing$ ▷ Timestamp of the last tested gaze position
$i_{pub} = \varnothing$ ▷ Timestamp of the last published noisy gaze position

Input: g_i - True gaze position for timestamp i
$g'_{i_{pub}}$ - Output for the last timestamp, initialized to $\varnothing$ when $i_{pub} = \varnothing$
$\{\epsilon^{pub}_{i-n_{raw}+1}, \cdots, \epsilon^{pub}_{i-1}\}$ - Privacy budget consumed for publication in last n_{raw} timestamps, initialized to 0 if $i < n_{raw}$

Output: g'_i - Noisy gaze position released for timestamp i
ϵ^{pub}_i - Privacy budget consumed in publications

Stage I: Check whether to skip or test the gaze data point ▷ Fixation detection based on timestamp of data

if $(i_{test} \neq \varnothing$ and $time(i) - time(i_{test}) < t_{skip})$ **then**
  $g'_i = g'_{i_{pub}}$ ▷ Reuse last published gaze position
  $\epsilon^{pub}_i = 0$
  Return $\{g'_i, \epsilon_i\}$

Stage II: Test whether current gaze data point should be published ▷ Fixation detection based on location of data

$i_{test} = i$
$l_{dis} = d(g_i, g'_{i_{pub}})$ ▷ Euclidean distance between last published gaze position and current gaze position with $d(\cdot, \varnothing) = \varnothing$
$\eta \sim \mathsf{Lap}(1/\epsilon_{test})$ ▷ $\mathsf{Lap}(\cdot)$ denotes the Laplace distribution
if $(l_{dis} \neq \varnothing$ and $l_{dis} \leqslant l_{thresh} + \eta)$ **then** ▷ Test whether current gaze position is in the proximity of the last published gaze position
  $g'_i = g'_{i_{pub}}$
  $\epsilon^{pub}_i = 0$
  Return $\{g'_i, \epsilon^{pub}_i\}$

Stage III: Publish noisy gaze point

$i_{pub} = i$
$\epsilon_{rem} = \epsilon - \epsilon/h - \sum_{k=i-n_{raw}+1}^{i-1} \epsilon^{pub}_k$ ▷ Remaining privacy budget for the active window
$\epsilon^{pub}_i = \epsilon_{rem}/2$
$g'_i = \mathsf{PlanarLap}(g_i, \epsilon_i/r)$ ▷ $\mathsf{PlanarLap}(\cdot)$ is a geo-indistinguishable mechanism from [17]
Return $\{g'_i, \epsilon^{pub}_i\}$

Adaptive budget allocation. The algorithm proceeds in three stages. In the first stage (Steps 1–4), every gaze position that is generated up to duration t_{skip} after i_{test} is skipped, where i_{test} denotes the timestamp of the last tested gaze position. A good choice for t_{skip} can be the minimum duration of fixations ≈ 50 ms [153]. Thus, this stage reuses the last published noisy gaze ($g'_{i_{pub}}$) and consumes no privacy budget (Step 3).

The second stage (Steps 5–11) is the testing phase, where all the "not-skipped" gaze positions are tested for their proximity to $g'_{i_{pub}}$. Specifically, it checks whether the current gaze position g_i (not-skipped) is within a certain noisy threshold $(l_{thresh} + \eta)$[4] from $g'_{i_{pub}}$ (Steps 6–8). In case this is satisfied, the algorithm again reuses $g'_{i_{pub}}$. The total privacy budget allocated for testing for any window duration of w is ϵ/h. Each individual test consumes a budget $\epsilon_{test} = \epsilon/(h \cdot n_{test})$, where n_{test} is the number of gaze positions to be tested per window, and h is a parameter with a value greater than 2. The first two stages of the algorithm can be interpreted as a simple $(\epsilon/h, w, r)$-differentially private fixation detection scheme.

Finally, in the third stage (Steps 12–16), the algorithm publishes a noisy gaze position corresponding to g_i only if it is sufficiently distant from $g'_{i_{pub}}$. For this, it computes the remaining budget for the active window (Step 13) as follows

$$\epsilon_{rem} = \underbrace{\epsilon}_{\substack{\text{Total privacy}\\ \text{budget for}\\ \text{each window}}} - \underbrace{\epsilon/h}_{\substack{\text{Budget consumed}\\ \text{for testing in}\\ \text{the active window}}} - \underbrace{\sum_{k=i-n_{raw}+1}^{i-1} \epsilon_k^{pub}}_{\substack{\text{Budget consumed}\\ \text{for noisy publication in}\\ \text{the active window}}}$$

Next, the algorithm assigns half of it ($\epsilon_{rem}/2$) for the noisy publication (Step 14). Thus, the publication budget is allocated in an exponentially decreasing manner. The rationale behind this is that investing a high budget (i.e., injecting low noise) in the current measurement g'_i would result

[4]The value of l_{thresh} impacts utility and is chosen empirically depending on r.

in better approximation (test and reuse) for future ones. Additionally, note that ϵ_{rem} considers the budget consumed only in the active window $[i - n_{raw} + 1, i]$. Thus, the publication budget of older timestamps (preceding the active window) is recycled for future usage. The generation of the noisy gaze position is done via the PlanarLap() mechanism (Step 15), which satisfies geo-indistinguishability [17] (with the notational difference of using ϵ_i^{pub}/r as the privacy budget).

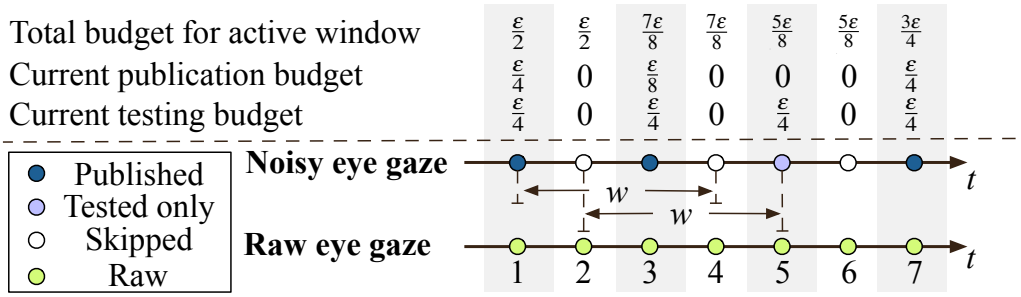


Figure 2.5: Illustrative example of Kalεido's budget allocation ($n_{raw} = 4$, $n_{test} = 2$, $h = 2$).

Illustrative example. Figure 2.5 presents an illustrative example of Algorithm 1. Here we consider $n_{raw} = 4$, $n_{test} = 2$ and $h = 2$. Hence, the budget for testing per gaze position is $\epsilon/4$. For the first window (timestamps 1-4), the algorithm publishes at timestamps 1 and 3 and skips at timestamps 2 and 4. Hence, timestamps 1 and 3 consume budget $\epsilon/4$ each for testing. Additionally, the publication budgets are $\epsilon_1 = (\epsilon/2 - 0)/2 = \epsilon/4$, $\epsilon_3 = (\epsilon/2 - \epsilon/4)/2 = \epsilon/8$ and $\epsilon_2 = \epsilon_4 = 0$. Thus, the total privacy budget consumed in this window is $\epsilon/2$ (budget for testing) $+ \epsilon/4 + \epsilon/8 = 7\epsilon/8 \leqslant \epsilon$. For the second window (timestamps 2-5), the algorithm reuses g_3' at timestamp 5. Hence, its total privacy budget is $\epsilon/2 + \epsilon/8 = 5\epsilon/8 \leqslant \epsilon$. For the third window (timestamps 3-6), the algorithm skips the gaze position at timestamp 6 and the total privacy budget is $\epsilon/2 + \epsilon/8 = 5\epsilon/8 \leqslant \epsilon$. A noisy gaze position is published at timestamp 7 in the fourth window

(timestamp 4-7) with $\epsilon_7 = (\epsilon/2 - 0)/2 = \epsilon/4$. Thus, the total privacy budget for this window is $\epsilon/2 + \epsilon/4 = 3\epsilon/4 \leqslant \epsilon$.

Theorem 2.6. *Algorithm 1 satisfies* (ϵ, r, w)*-differential privacy.*

Proof. First, note that Stage I (Steps 1–4, Algorithm 1) does not consume any privacy budget. Next, from Fact I in [51], Stage II consumes privacy budget ϵ_{test} for every test. Specifically, the output of the test mechanism (Step 8) is a binary decision and hence, its sensitivity is 1. Finally, Stage III consumes budget $\epsilon_i^{pub} = 1/2(\epsilon - \epsilon/h - \sum_{k=i-n_{raw}+1}^{i-1} \epsilon_k^{pub})$ if it publishes, and 0 otherwise. Next, we prove that the total budget consumed in every window is at most ϵ. For this, note that the total budget consumed for testing is ϵ/h. Hence, it suffices to show that $0 \leqslant \sum_{k=i-n_{raw}+1}^{i-1} \epsilon_k^{pub} \leqslant \epsilon - \epsilon/h$ which follows directly from the proof of Theorem 4 in Kellaris et al. [132]. □

Noisy Gaze Processor

The noisy gaze processor takes as input the noisy gaze streams generated in real-time and performs post-processing operations on it before releasing it to the applications. This module is identical to any local post-processing unit existing in current eye-tracking systems, except for noisy inputs. Examples of such post-processing include data sanitization, such as bounding of off-screen points and data smoothing. Moreover, Kalɛido's noisy gaze processor can support local feature extraction similar to that in the "recognizer" framework [124] (Section 2.7). Kalɛido is thus compatible with applications with APIs expecting specific features as input, such as fixation/saccade statistics. By Theorem 3, this step does not impact the privacy guarantee of Kalɛido.

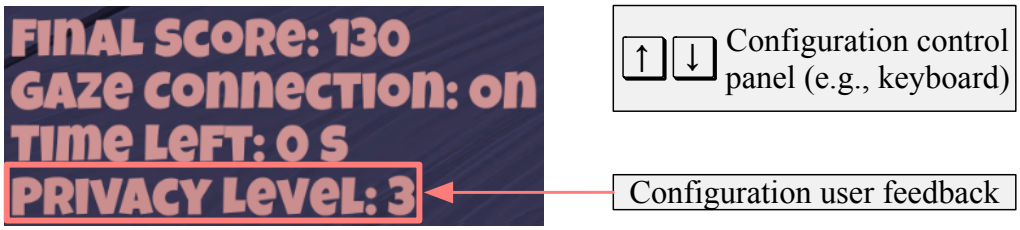


Figure 2.6: Basic template of Kaleido's user interface.

2.4 Implementation

We implemented Kaleido as a C# plugin in Unity [95], a cross-platform engine for developing interactive applications, such as games and mixed reality content. Unity allows developers to integrate plugins that generate visual content and communicate with peripherals, including eye trackers. In our implementation, Kaleido acts as an intermediate protection layer between applications and the platform.

Stream acquisition. Kaleido acquires real-time eye gaze streams from the eye-tracking core and forwards them to the noisy gaze generator. To synchronize these gaze streams, we implement the eye gaze receiver using the TCP/IP protocol, which is the most common communication channel for off-the-shelf eye-tracking cores, such as Tobii [277], GazePointer [92], and PupilLab [151].

ROI extraction. Kaleido identifies the instances of scene change and extracts the ROIs from each scene. For deterministic visual content (such as movies), Kaleido acquires the timing of keyframes (instances of scene changes) from either the video decoding process or keyframe properties obtained from Unity's Animation feature or content providers [285]. As for online content, Kaleido identifies the keyframes using an on-the-fly scene change detector [318]. In particular, we implement a threshold-based real-time keyframe detector using the mean absolute frame difference method. First, Kaleido fetches the current frame from Unity's rendering

process. Next, it takes the pixel-wise difference between the current frame and the last keyframe. Kalɛido detects a new keyframe by comparing the pixel values of the binarized difference matrix against a pre-calibrated threshold. We set the default update interval of keyframe detection to 500 ms, which is the typical response latency of human attention to visual stimuli [40].

Kalɛido identifies the spatial information of ROIs for digitally rendered frames using Unity's GameObject API. For all other types of frames, Kalɛido uses YOLOv3-tiny [238], a light-weight neural network. To study the impact of YOLO on real-time performance, we make an exception and use it for digitally rendered frames as well in our user study.

User Interface. Kalɛido offers the users an interface to adjust their privacy-utility trade-off. Users can control the privacy budget ϵ on-the-fly through pre-defined triggers, such as keypress, as illustrated in Figure 2.6. We chose a basic interface for our prototype implementation since UI design is not the focus of this work.

2.5 Evaluation

Table 2.1: Properties of eye gaze traces, with a video dataset highlighted.

Dataset	Num. of stimuli	Num. of users	Sampling rate (Hz)	Avg. duration (s)
Natural [305]	10	19	100	6.0
Web page [305]	10	22	100	16.8
Human [126]	10	60	100	3.7
VR video [14]	12	13	120	64.9

We evaluate three aspects of Kalɛido: (1) user-perceived utility, (2) real-time performance, and (3) effectiveness against spatio-temporal attacks. We performed a trace-based evaluation to measure the effectiveness of Kalɛido against attackers using four popular eye-tracking datasets. These

datasets, described in Table 2.1, include the scenarios of natural environment, web pages, human, and virtual reality (VR) videos. In particular, our evaluation answers these questions:

Q1: *How do users perceive the utility of real-time interactions with* Kaleido*?*
We conducted a remote user study with 11 participants to assess the user-perceived utility while playing a real-time PC game with Kaleido.

Q2: *How much latency overhead does* Kaleido *incur?*
We measured the latency overhead of the main modules of Kaleido to assess its real-time performance.

Q3: *Can* Kaleido *thwart attacks that rely on spatio-temporal analysis of eye gaze streams?*
We performed a trace-based evaluation of Kaleido on popular eye-tracking datasets. We investigate the effectiveness of Kaleido's formal privacy guarantee against real-world adversarial analytics.

User Perception in Real-Time Interaction

We conducted a user study to evaluate Kaleido's impact on utility, as perceived by the users, while playing a real-time PC game. Our objective is to understand the impact of Kaleido on user experience at different settings of privacy. To this end, we adapted the game "Survival Shooter" [284] from Unity to be eye-tracking compatible. Participants shot targets (Zombie Bunnies) by gazing at the target position on a computer screen, as shown in Figure 2.7. They used the keyboard to move their digital avatar in the game. We used this PC game because of the requirement to perform the study remotely at the users' places. An in-person lab session with state-of-art eye-tracking or virtual/augmented reality was not possible during the study[5].

[5]We conducted this study during the state of Wisconsin's Safer at Home order due to the COVID-19 pandemic.

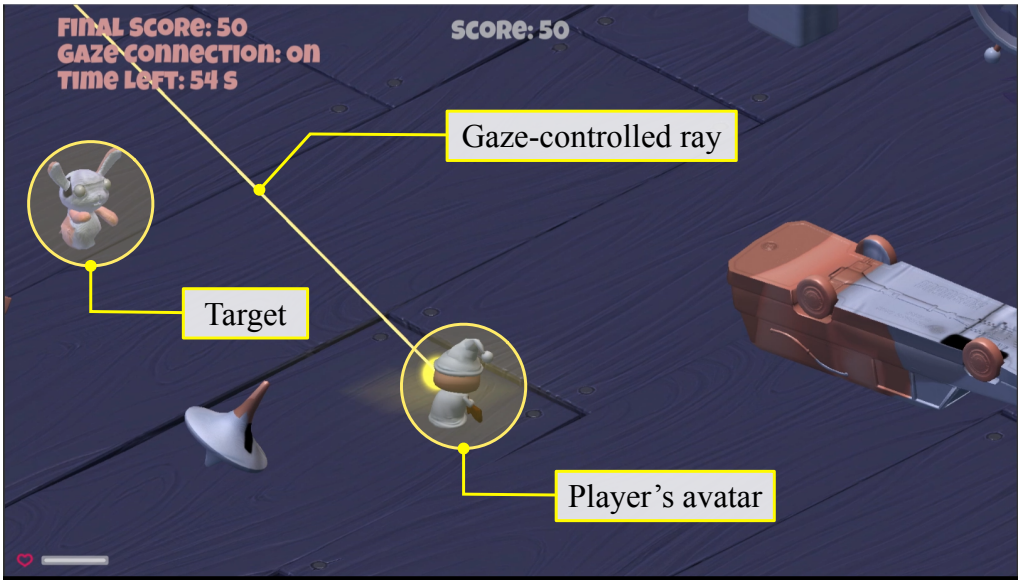


Figure 2.7: A scene of the "Survival Shooter" game with the player's avatar, target, and gaze-controlled ray annotated.

Setup. To accommodate a commodity PC setup, we utilize the webcam-based eye-tracking core, GazePointer [92], for detecting the participant's gaze on the screen. The remote user study design was approved by the Institutional Review Board (IRB) of our institution. We recruited 11 individuals from the mailing list of our department. The recruitment email provided no details about the study's privacy objectives and mentioned only user experience with eye-tracking games. Each remote session took 35 minutes on average, and we provided each participant with $15 worth of supplies as a token of appreciation for participating.

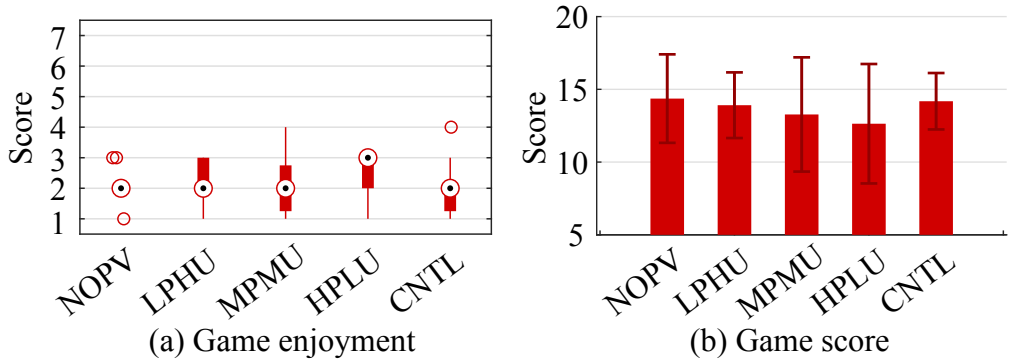
Limitations. We acknowledge the following limitations in our study setup resulting from the imposed lockdown. First, the demographic diversity of the participants, as well as the number of participants, might be limited. Hence, one caveat is that the confidence interval of the quantitative analysis is relatively large. Thus, we treat our presented results as a prelim-

inary study. Second, an in-person study using state-of-the-art eye-tracking devices was not possible, which hindered our ability to study diverse scenarios, such as foveated rendering in VR and video watching. We carefully designed our study protocol to reduce the impact of the low accuracy of the webcam-based eye-tracking core; its accuracy is sensitive to posture and lighting conditions. Before starting every new session, the participants were instructed to calibrate the eye tracking using GazePointer's panel. Finally, the constraints of a remote user study also hindered us from conducting a qualitative study via in-person interviews and behavioral observation. An additional caveat is that we did not perform coded analysis for the qualitative study of user responses (via techniques such as open or axial coding [265]) of the free text.

Design. Each study session consisted of five tasks (conducted over a video call using a separate device). The first is a *pre-study survey* to collect the participant's demographic information using a Qualtrics survey. The second is the *calibration* of the webcam-based eye-tracker to map the eye gazes to the computer screen using GazePointer's calibration interface. The participants were asked to familiarize themselves with the game by practicing eye gaze-based shooting until they felt confident. The third covers the *within-subject evaluation* sessions. The fourth task tests the *privacy control knob*. The last task is the *post-session survey*.

To reduce individual differences in gaming behavior and perception, we conducted the within-subject study [48] to test four game settings: (1) *No privacy* (*NOPV*) — Kalɛido layer disabled; (2) *Low privacy-high utility* (*LPHU*) — $\epsilon = 3$, $w = 0.5$ s, r_{small}; (3) *Medium privacy-medium utility* (*MPMU*) — $\epsilon = 1.5$, $w = 1.5$ s, r_{small}; and (4) *High privacy-low utility* (*HPLU*) — $\epsilon = 0.5$, $w = 2$ s, r_{large}[6]. Each setting lasted for

[6]These values were chosen based on a parameter sweep to represent different points along the privacy-utility spectrum (Figure 2.14). In the trace-based analysis of offline datasets, the root mean square error (RMSE) serves as a proxy for measuring application-specific utility loss.



(a) Game enjoyment (b) Game score

Figure 2.8: Scores obtained in different conditions.

90 s[7], and we randomized their order for every participant. Additionally, the participants had no knowledge about the setting to which they were exposed. After the completion of each setting, we recorded: the subjective game enjoyment [191] as a 7-item Likert scale, the game score, and the qualitative feedback.

After the four randomized settings, the objective of Kalɛido was revealed, and the participants were offered an adjustable knob to control the tradeoff between privacy and utility. We asked each participant to interact with the control knob; we observed how frequently they adjusted the knob and solicited qualitative feedback about their experience. This part of the study follows a *technology-probe*-based approach [119]. Our objective is to probe the participants to elicit their opinions about the missing design elements that need to be introduced.

Results. We asked the participants to report their subjective experience to evaluate the validity of our game's adaptation. To this end, we asked each participant to report their level of agreement (or disagreement) with this statement: "*You enjoyed the game in this session.*" on a 7-item Likert scale with 1 being "*Strongly Agree*" and 7 being "*Strongly Disagree*". Figure 2.8(a)

[7]The interval value was chosen during calibration to balance the validity of the session and user fatigue.

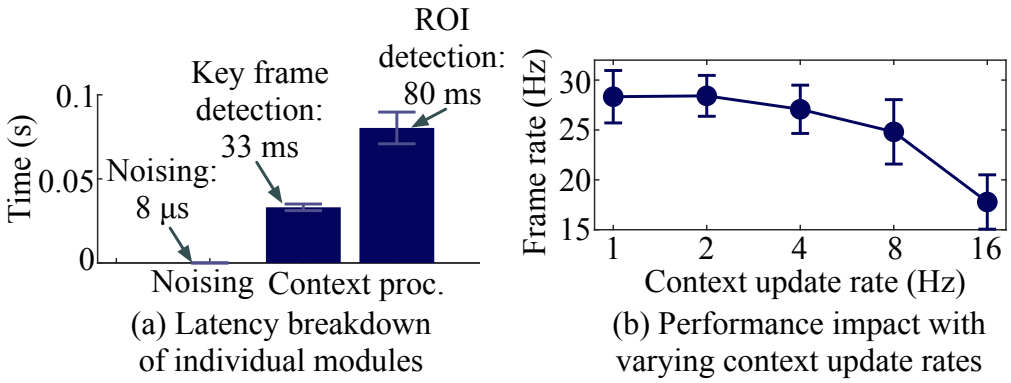
shows that for all of the game settings, the participants enjoyed their experience – at least 82% of them reported a score of 3 or lower.

Next, we study the effect of the privacy level on the participants' game scores. Figure 2.8(b) shows these scores for the different settings. We observe that the game scores decrease with a stronger privacy guarantee. However, the decrease in the score is not significant from the no privacy (NOPV) setting to the low privacy (LPHU) setting (only 3.2%). Even the decrease from the NOPV setting to the high privacy (HPLU) setting is modest (12.0%). These results show that Kalϵido's noise does not adversely affect users' utility in this scenario.

The qualitative feedback that we obtained from the users aligned with our quantitative observations. Some participants were unable to distinguish between the LPHU and NOPV settings – (**P8**: "The second (NOPV) and third (LPHU) configurations are almost the same for me.") The majority of the participants found the highest privacy (HPLU) setting to be the hardest to control. Some participants had a surprisingly different view. For example, **P7** enjoyed the conditions with higher noise because it was more challenging to play.

Finally, we performed a preliminary analysis of the privacy control knob (setting: CNTL). In the last task of the study, we introduced the control knob to the participants and asked them to control the privacy level as per their desired level of utility. Figure 2.8(b) shows that the adjustment of the control knob does not affect the game scores. However, we find a large variation in the frequency of knob adjustment and the privacy level (ϵ) across the participants.

The qualitative feedback also indicated that while such a knob might be useful, they had some suggestions for improvement. For example, **P8** and **P11** proposed adding flexibility for an offline calibration of the privacy level for each application. Other participants commented that frequently adjusting the knob during intense gameplay is suboptimal.



(a) Latency breakdown of individual modules

(b) Performance impact with varying context update rates

Figure 2.9: Performance breakdown and trend. ROI detection is the most expensive operation. The frame rate remains relatively steady even for a high context update rate of 8 Hz.

System Performance

We evaluated Kalεido's real-time performance and measured its processing delay on a commodity PC with an Intel i7-7700 CPU and Nvidia GTX 1080 GPU. Figure 2.9 shows the latency overheads incurred by the three main operations of Kalεido: noisy gaze generation (noising), keyframe detection, and ROI detection. We ran 100 trials for each of the operations and reported the average running time. The latency of the noising operation is only 8 µs, and thus, has no discernible impact on the user's real-time experience.

ROI detection takes 80 ms on average, but it only runs when a new keyframe is detected. Based on our offline game calibration, a new keyframe is detected only every 2.3 s (similar to the timing from the VR videos dataset). Thus, the overall impact of ROI detection in Kalεido is not significant.

Keyframe detection takes 33 ms on average. The frequency of keyframe detection (context update rate) is comparatively higher (2 Hz in our implementation). Figure 2.9(b) shows its performance impact on effective frame rates of the game used in the study. We observe that, even with a

high context update rate of 8 Hz, the frame rate degrades only slightly to 25 Hz.

In this chapter, we evaluate a research prototype of Kalɛido, which shows its real-world potential. Nevertheless, to deploy in scale, Kalɛido can leverage various performance optimizations, such as GPU offloading, model compression, and resource sharing. These optimizations would enable fast context processing even on resource-constrained platforms.

Effectiveness Against Attacks

Recall that post-processing operations on the outputs of a DP algorithm do not result in additional privacy loss (Theorem 3). Thus, Kalɛido's formal DP guarantee for the spatial information of gaze streams holds for every attacker (even for one with full knowledge of Kalɛido's protocols). However, Kalɛido does not provide a formal guarantee on the temporal information of gaze streams (Section 2.2). Hence, we performed a trace-based evaluation to study the effectiveness of Kalɛido against spatio-temporal attacks using the datasets in Table 2.1. These attacks exploit the spatio-temporal features of gaze streams, such as fixation durations and saccade velocity [110, 248]. We select two representative analyses of gaze streams: (1) similarity and outlier analysis of a scanpath for an individual, and (2) biometric inferences. We use (1) MultiMatch [71] for computing the scanpath similarity scores, and (2) F1 score, which considers both precision and recall, to measure attackers' classification accuracy.

Note that the attackers considered in this section are knowledgeable; they have complete knowledge of the target visual scenes and Kalɛido's noise generation protocols. Further, they use a noise-robust fixation detection [110]. Additionally, all the classifiers used in this section are trained on noisy gaze streams from Kalɛido (for the same privacy configurations).

Similarity and Outlier Analysis of Scanpath

Given a dataset of gaze streams for single scenes, this attack constructs a feature vector of the scanpath for each individual in the dataset. Since the visual stimulus is the same, the hypothesis is that the differences in the scanpath features arise from distinguishing psychophysiological traits. Thus, this type of analysis aims at distinguishing individuals based on their scanpath features [31].

Setup. We use the image datasets (the first three rows of Table 2.1: natural, web page, and human) to evaluate the distinguishability of the scanpath features on static image frames. This evaluation assesses the accuracy of the analysis of raw and noisy gaze streams. For each stream, we extract the scanpaths using an offline algorithm [110]. Next, we perform similarity analysis and outlier identification as follows.

Similarity analysis. The adversary here has a priori knowledge of a user's scanpath on a certain image. It attempts to re-identify the user by measuring the similarity between this scanpath and a newly observed one formed on the same image. For each dataset, we compute the similarity between the scanpaths of the same user, before and after adding noise. We use the standardized similarity metric, MultiMatch[71], which ranges from 0 to 1. This score measures scanpath similarity by considering features about the shape (the length, shape, and direction of saccade vectors) and the spatial distribution (position and duration of aligned fixations) of gaze data.

Outlier identification. In this attack, the adversary tries to identify the outlier users whose scanpath features are significantly different from that of the rest. This attack utilizes a density-based clustering model DBSCAN [98], where inter-scanpath distances are computed via dynamic time warping (DTW) over the scanpaths on a single image. We use the F1 score to report the attacker's success in identifying the outlier users from the dataset containing noisy gaze streams. We show the F1 scores of out-

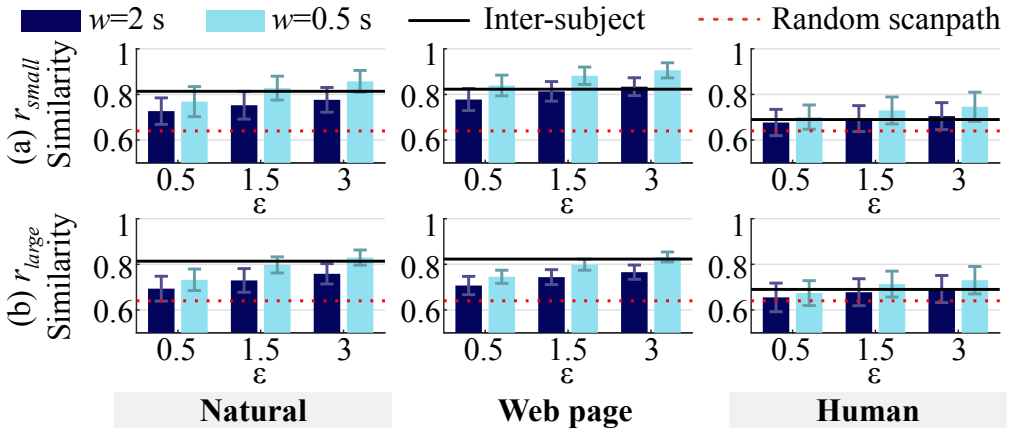


Figure 2.10: Similarity scores between noisy and raw scanpaths. Kaleido reduces the similarity scores to be close to the inter-subject threshold (black lines) even at low privacy configurations (r_{small}). The scores are reduced further to be close to the random scanpath baseline (red dash lines) at high privacy configurations ($\epsilon = 0.5$, r_{large}, and $w = 2$ s).

lier identification compared to random guessing as a baseline ("Random guess").

Results. *Similarity analysis.* In Figure 2.10, we compare the measured similarity with two thresholds: (1) mean inter-subject similarity score ("Inter-subject") in each dataset, and (2) the similarity of two randomly synthesized scanpaths presented in [71] ("Random scanpath"). Figure 2.10 shows a consistent trend in all three image datasets: the scanpath similarity decreases with higher privacy level (i.e., smaller ϵ, larger w, and larger r). Kaleido degrades the similarity score below the inter-subject threshold, even though it perturbs the spatial data only; at $\epsilon = 0.5$, Kaleido brings the similarity score close to the random scanpath baseline.

Outlier identification. As observed from Figure 2.11, Kaleido degrades the effectiveness of outlier identification for all of the privacy settings. For the natural and human image datasets, Kaleido reduces the attacker's F1 scores to the random guess using r_{large} with ϵ as high as 3. Although the

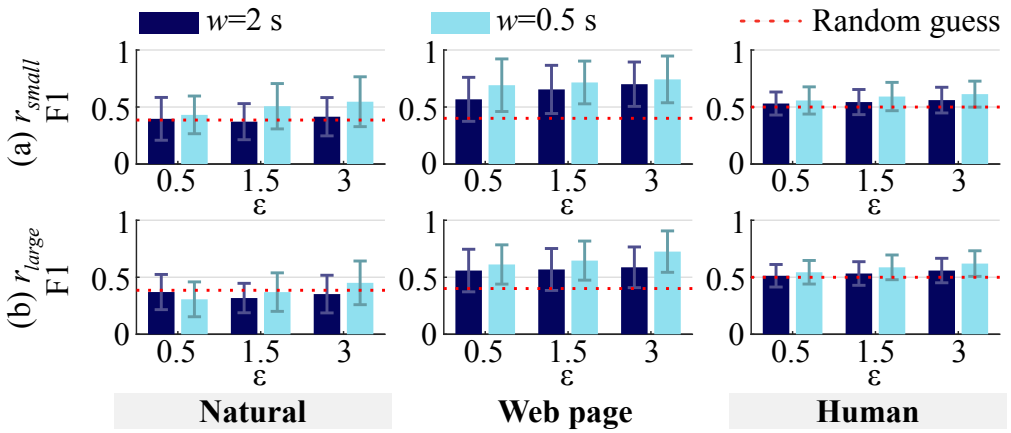


Figure 2.11: F1 scores of outlier identification among scanpaths. At high privacy configurations (low values of ϵ, r_{large}, and $w = 2$ s), Kaleido thwarts outlier identification attacks in all three datasets by reducing F1 scores to be close to the random guess baseline (red dash lines).

attacker's F1 score remains relatively high in the web page dataset, it is reduced significantly for $\epsilon = 0.5$.

Biometric Inferences

Setup. We construct attacks that attempt to predict (1) users' identities and (2) whether the users wore contact lenses for vision correction (use of contact lenses leads to distinguishing eye gaze patterns [206]).

For this experiment, we use the VR video dataset (last row in Table 2.1). The associated classification labels are provided in the dataset. This attack uses aggregate statistics of fixation/saccade features over several VR video sessions as training data and predicts users' identities and vision conditions for an unseen session. Specifically, each video session uses a different VR context for the same user. Hence, the evaluation of biometric inferences here assesses Kaleido's effectiveness against linkability attacks across different contexts (this has been exploited in prior work [79]). We

adopt the features suggested by the Cluster Fix toolbox [140], which are then used to train a discriminant analysis classifier [64]. This evaluation includes 11 users from the VR video dataset who comfortably completed all 12 video sessions. Additionally, the training and test sets correspond to the same privacy configuration, i.e., either raw gaze streams or noisy gaze streams. We report the F1 scores for leave-one-out cross-validation.

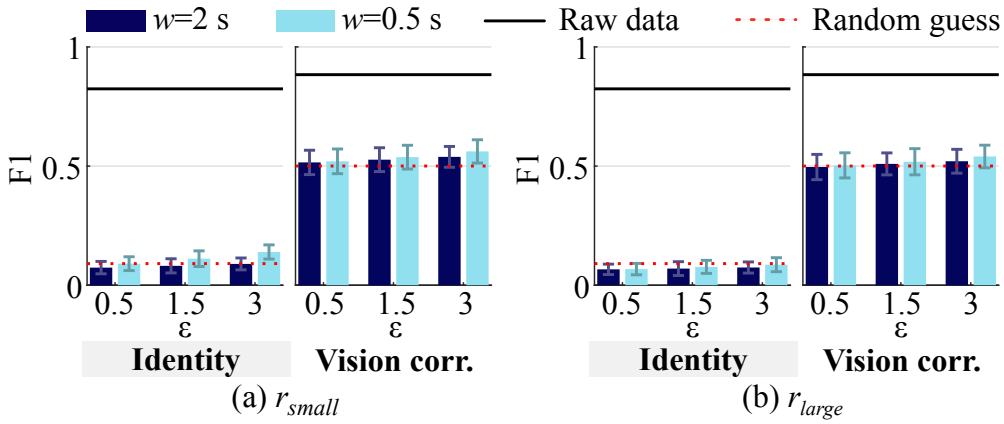


Figure 2.12: F1 scores of predicting user identity and vision correction. Kaleido reduces the F1 scores of biometric inferences to be close to random guess baselines (red dash lines) even for low privacy configurations (high values of ϵ or r_{small}).

Results. Figure 2.12 shows the F1 scores obtained from both the raw and noisy gaze streams. For both classifiers (identity and vision correction), the raw gaze streams enable accurate classification – the F1 score is close to 1 ("Raw data" in Figure 2.12) and is much higher than that of random guess. This indicates that the attacker can successfully predict users' identities and vision correction conditions, even across different contexts. On the other hand, we observe that Kaleido significantly degrades the attacker's classification accuracy to be close to the random baseline even for low privacy configurations (high values of ϵ or r_{small}).

2.6 Discussion

Kalεido is a first step toward designing real-time eye-tracking systems that provide a formal privacy guarantee. Here, we discuss several possible avenues for future research:

Support for more data formats and types. An eye-tracking platform may offer eye-tracking data in various formats such as 2D gaze positions and 3D gaze positions. Currently, Kalεido is designed for 2D gaze streams and supports head-and-eye gaze streams as well.

Extension to 3D gaze streams is possible and would involve extending the PlanarLap mechanism (Algorithm 1 to 3D positions. Additionally, some eye-tracking cores collect data including blink timing and pupil dilation. Kalεido's scope of privacy can be further broadened to address these data types.

Privacy guarantee for temporal information. Kalεido can be extended to protect the temporal information of eye gaze streams by interfering with the timeliness of gaze releases. For example, for fixation duration (a popular aggregate statistic), Kalεido can decide on a predefined threshold T based on standard human gaze fixations [115]. Next, stage I and II from Algorithm 1 can be replaced by a sophisticated fixation detection approach such as online differentially private clustering [185, 136], which (1) releases a single noisy position in the first T duration of a fixation and (2) stops any further data release for the given fixation. This ensures that the duration for all fixation events in the noisy gaze stream is fixed to T.

Optimization for long scenes. Although visual content in an eye-tracking application is typically dynamic, it might remain relatively static for long periods in some cases. Such long scenes that span multiple windows may lead to a large privacy budget consumption. Techniques including noisy data caching can be used to help address this issue. Specifically, Kalεido can check online if the current ROI has been visited previously, and it can

reuse the corresponding noisy gazes from recent history. Additionally, for applications where interactions are sporadic, Kalɛido can skip releasing new gazes for scenes when the user is inactive to save the privacy budget.

Optimizations for context processing. One interesting future direction can be optimizing Kalɛido's context processing core. The overhead of Kalɛido's context processing can be reduced by sharing the detection module with other applications. Kalɛido can leverage other models for ROI detection, including Selective Search [282] and Faster R-CNN [241], which may be implemented by the platform already. For instance, eye-tracking platforms, such as Hololens [201], provide certain context information that Kalɛido can use directly for performance optimization. Additionally, smart calibration of the frequency of key frame detection can also reduce the overhead of context processing.

Optimizations for privacy budget allocation. In this chapter, the presented composition theorem (Theorem 2) is based on the simple k-fold composition of the DP guarantee [78]. However, a tighter analysis might be possible via advanced composition [78] and moment-based accounting [1].

Evaluation of other utility metrics. In this chapter, we primarily focus on qualitatively evaluating Kalɛido's utility for the use case of a real-time game (as demonstrated in Section 2.5). However, as mentioned in Section 2.1, eye-tracking data is used for diverse purposes. Hence, an important future direction is to investigate user perception for other online applications and quantitatively evaluate Kalɛido's utility for offline gaze data analysis. For example, in some cases, the application utility might require extracting the saliency maps [127, 25] from users' fixations. Figure 2.13 shows Kalɛido's impact on the saliency maps. We compute the correlation coefficient, a standard metric for saliency map similarity [44], between each user's clean and noisy maps. For all the datasets, Kalɛido's accuracy (higher correlation coefficient) [35] increases with increase in the privacy

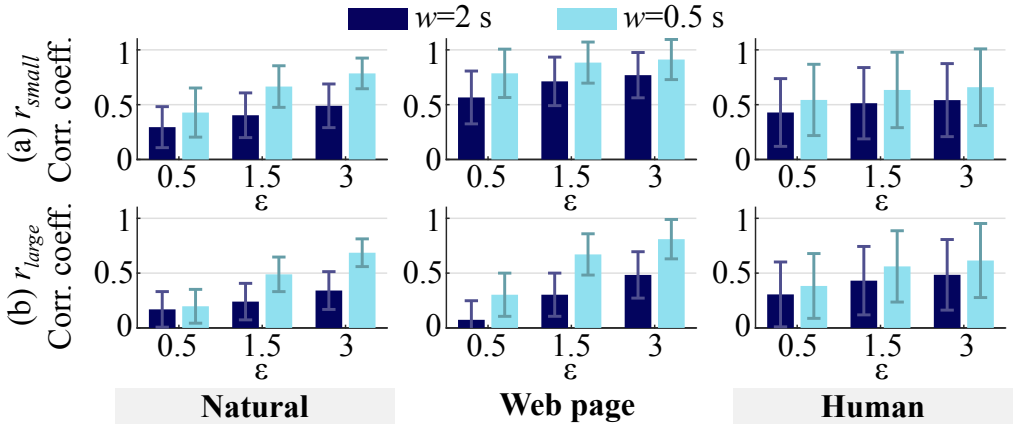


Figure 2.13: Kalɛido's impact on saliency map at varying privacy configurations.

budget ϵ or decrease in window duration w. At the same value of ϵ and w, using r_{large} gives lower accuracy than r_{small}. These results are consistent was Kalɛido's premise: it attempts to hide the spatial patterns of the user's fixations. A lower value of ϵ would result in less accurate extraction of the saliency maps. Another direction could be exploring application-specific utility optimizations. For instance, data-smoothing techniques can be used to improve the accuracy of the noisy gaze streams.

In addition, we study the privacy-accuracy trade-off for varying configurations of Kalɛido in Figure 2.14. These results serve as a proxy to understand Kalɛido's utility impacts across applications. The utility is measured by the root mean square error (RMSE) in pixel. We vary the parameters as follows: $\epsilon \in \{0.5, 1, 1.5, 2, 2.5, 3\}$, $w \in \{0.5, 1, 1.5, 2\}$ and $r \in \{r_{small}, r_{large}\}$. We generate 100 random trials for each combination and report the mean observation. In all the datasets, we observe a clear trend of accuracy improvement (lower RMSE) with increasing privacy budget ϵ or decreasing window duration w. At the same value of ϵ and w, using r_{large} gives lower accuracy than r_{small}. We also show the privacy-accuracy trade-off for Kalɛido for head-and-eye gaze data for the VR video

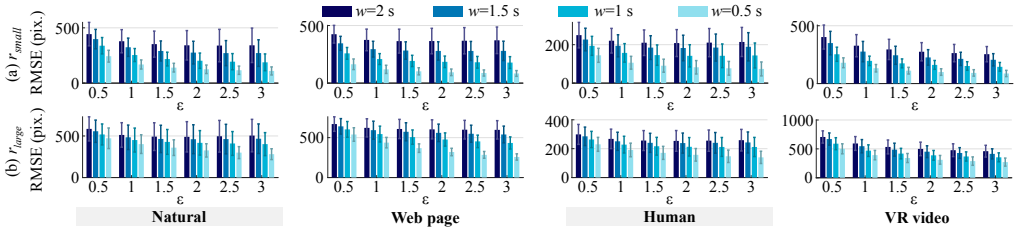


Figure 2.14: Privacy-accuracy trade-off of Kalεido on eye gaze data.

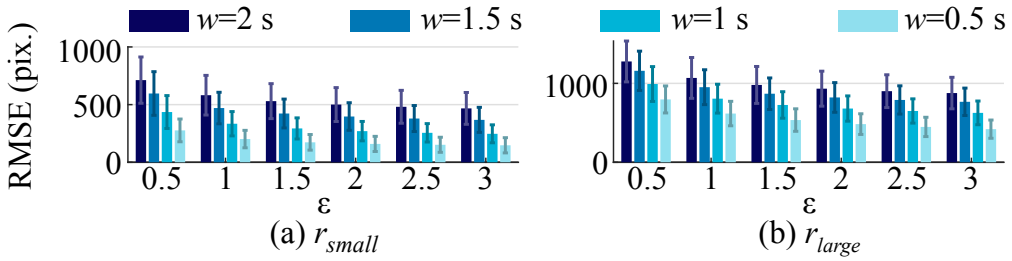


Figure 2.15: Privacy-accuracy trade-off of Kalεido on head-and-eye gaze data.

dataset in Figure 2.15.

2.7 Related Work

In this section, we provide a summary of the related work. One line of work proposed "recognizer" systems that process a sensor stream, such as a video, to "recognize" predefined objects or features [124, 244, 230]. The principle underlying these systems is to send only abstract features from the data stream (possibly after obfuscation) to the untrusted applications in place of the raw stream. However, this approach suffers from a set of shortcomings when applied in the context of real-time eye tracking. First, APIs of current user applications expect, as inputs, raw eye gaze streams directly or basic gaze events such as fixations. Second, this approach does not provide a formal privacy guarantee and cannot defend against attacks

that consume only coarse-grained measurements (that can be computed from the features) [175]. Last, such systems introduce complications for permission control for both users and application developers.

Another line of work used adversarial machine learning-based approaches to protect the raw eye gaze data [100]. However, such techniques operate on predetermined data streams and require training. Hence, these solutions are not practically feasible for real-time interactions. Additionally, they did not offer any formal privacy guarantee. In another work, Bozkir et al. [37] used randomized encoding to privately train an SVR model for gaze estimation. However, this method would require significant changes, such as communication with a third-party server, to existing eye-tracking ecosystems.

Differential privacy has been proposed in the context of eye tracking [261, 175, 36]. However, the major problem with the existing works is that they release noisy high-level features, such as heatmap [175] and ratio of saccades [261, 36]. Moreover, their workflow involves collecting the dataset of eye gaze streams from a group of users and then performing noisy feature extraction from it – the data release cannot be performed in real-time. Also, the computation of the sensitivity [78] of the features in two of the works [261, 36] is dependent on the dataset, leading to additional privacy leakage [210]. Further, Bozkir et al. [36] adopted the central differential privacy setting that requires the presence of a trusted data aggregator, an infeasible proposition for most eye-tracking applications.

Thus, the solutions above are not directly comparable to Kaleido, aiming to provide a formal privacy guarantee for raw gaze streams in real-time interactions. Instead, we have designed a scheme based on a variant of local differential privacy known as geo-indistinguishability [17]. This variant is suitable for real-time privacy protection and was originally intended for preserving geo-location privacy, such as GPS data. Subsequent research on geo-indistinguishability has extended its application to protecting mobil-

ity trajectories and incorporating different geographical semantics within their privacy settings [51, 117, 50, 327], such as points of interest. We will later discuss how we adapt this framework to accommodate the unique semantics of eye tracking.

2.8 Conclusion

We have designed and implemented Kalεido, an eye gaze processing system that (1) provides a formal privacy guarantee on the spatial distribution of raw gaze positions, (2) seamlessly integrates with existing eye-tracking ecosystems, and (3) is capable of operating in real-time. Kalεido acts as an intermediary protection layer between the eye-tracking platform and the applications. Our evaluation results showed that users enjoy a satisfactory level of utility while deploying Kalεido for an interactive eye-tracking game. Additionally, it is successful in thwarting real-world spatio-temporal attacks on gaze streams.

3 SECURING USER AUTHENTICATION USING NONLINEAR VIBRATION CHALLENGE-RESPONSES

In Chapter 2, we have discussed how to better control the privacy of biosignals when preserving interactive utilities. This chapter will discuss the use of biosignals as biometrics in user authentication, a utility for security purposes, and how we can make biometric user authentication more secure. In addition to privacy issues, the mass proliferation of "smart" devices has created unprecedented security concerns to their users. One of the significant security concerns comes from unauthorized entities accessing and controlling user devices. Stronger access control goes a long way towards alleviating security and privacy threats to users and their devices. User authentication, where a user has to prove their identity to a system, is one core mechanism to achieve adequate access control.

Biometric user authentication, which relies on the unique physiological or behavioral traits of the user to verify their identity, has been touted as the solution that meets both security and usability goals. Thanks to its low cognitive burden, it is more attractive to the users who wish to authenticate themselves to their devices without having to memorize a password or use an additional security device.

Several commercial and research solutions have been proposed or deployed to achieve biometric authentication. These solutions range from the traditional approaches such as fingerprints [135] and iris scan [129] to the more advanced modalities such as human touching [55, 258, 254], human speech [324, 323, 242], eye movement patterns [138, 133, 257], electrophysiological measurements [23, 325, 289], and vibration responses [177, 180, 54]. Of these modalities, vibration response has emerged as an attractive method due to its compatibility with commodity devices. Consumer devices, such as smartphones and watches, are commonly equipped with vibration motor, microphone, accelerometer, and gyroscope which can

generate and measure the vibrations off the human body.

Typical biometric approaches rely on what we refer to as "static" biometrics. An initial training phase collects physiological or behavioral information from the user, such as a gesture, fingerprint, or voice print. At the authentication phase, the user proves their identity by reusing the same template every time. The problem lies in that **human biometrics are non-resilient** [216, 184, 312, 207]: once the biometric template has been breached and compromised, the user cannot recover. Some biometric methods such as gesture-based vibration [177] can scale to multiple traits corresponding to a specific gesture. Their usability, however, will degrade significantly as a cost due to increased training effort and mental burden.

In this work, we attempt to answer this question: *How can we make biometric authentication more resilient against the threats of template breach?* We answer this question in the affirmative and argue that the key to answering this question is to consider a dynamic view of human biometrics. The human body is a complex and dynamic system that reacts differently to different physical stimuli. If through some training phase, an authenticating service knows the responses to a large set of stimuli, then it can play a new and disposable stimulus at each session. It collects the response and attempts to match it to the previously recorded response. Instead of reusing the same biometric template to authenticate the user, an authenticating service can use a new biometric, which is unlikable to the used ones, for each authentication session and never use it again. We refer to this model as **challenge-response biometric authentication**. This model is akin to physically unclonable functions (PUFs) that are popular in hardware security [266].

In this chapter, we present VELODY, a system that adopts a challenge-response protocol for biometric authentication. It leverages the nonlinear and complex nature of hand-surface vibration. Figure 3.1 illustrates the use case of VELODY. It has access to a pool of pre-collected challenge-response

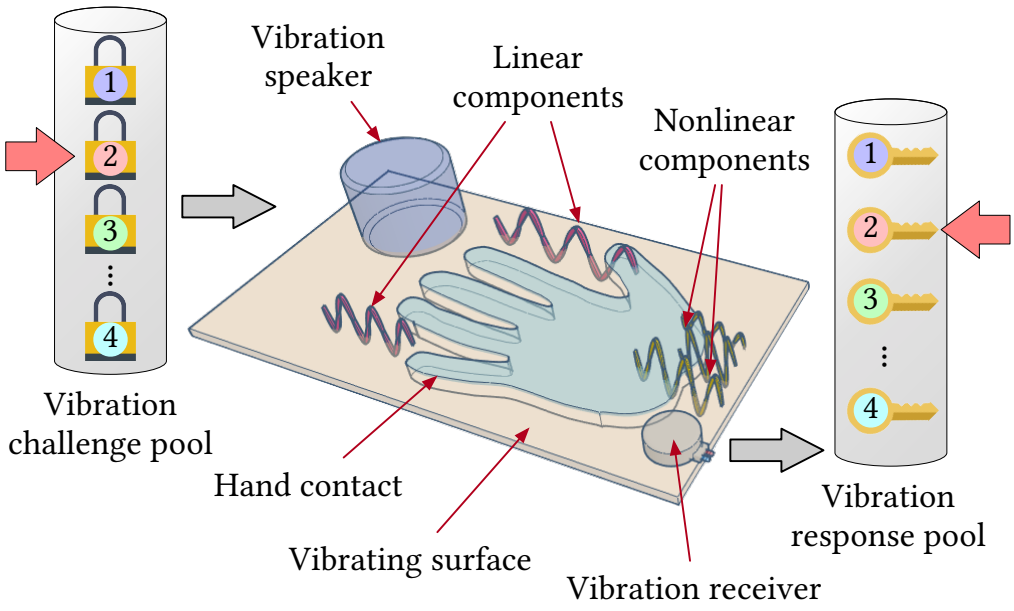


Figure 3.1: Illustration of VELODY.

pairs from a user. The challenge refers to a vibration stimulus to the user's hand through a surface, and the response is the collected vibration. Due to the properties of the user's hand contact, each response is unique per-challenge and per-user. At each authentication session, VELODY plays a disposable challenge and uses a classifier to decide whether the measured response matches the pre-collected one. By design, VELODY is resilient to an attacker replaying previously used biometric information.

To realize VELODY, we have to design two core components: (1) the challenges to play and (2) the classifiers to compare the collected and pre-collected responses.

Challenge design: A challenge is a vibration stimulus that comprises different spectral components. First, to maximize the user-distinguishability as a biometric, a frequency sweep is used to capture the frequency selec-

tivity contributed by the physiological traits in human hands. Second, combinations of sinusoidal waves with random frequencies act as stimuli along with the frequency sweep in disposable challenges to elicit the user-distinct and varying degrees of complicated nonlinearity in vibration responses, including harmonics and intermodulation, which are hard to model and predict for unseen responses.

Response classification: VELODY is a per-user system; a VELODY user does not have access to other users' response data for privacy and security considerations. This requirement constrains VELODY's classifier design as it cannot obtain negative samples from other users. To address this issue, we utilize the one-class k-nearest neighbor (OC-kNN) classifier, which relies on the similarity between inference-time observations and training instances. VELODY trains one classifier for each challenge. We devised a novel mechanism to set the matching threshold of the classifier per-user as to reduce the misclassification rate.

We implement VELODY using off-the-shelf speaker and accelerometer. Our evaluation via 15 individuals shows the following:

- VELODY exhibits a favorable performance in terms of security and usability with an EER at 5.8% evaluated using long-term authentication session against impersonation attacks.
- VELODY can reject 97.3% impersonation samples and 100% replay and synthesis attacks with reasonable effort in passive enrollment and an extremely short 200-ms vibration challenge in one authentication session.
- VELODY's challenge-response design is resilient to variations in the challenge design. Using shorter challenges with fewer spectral stimuli still maintains a satisfactory EER.

3.1 Background on Hand-Surface Vibration Response

In this section, we introduce two properties of hand-surface vibrations that enable the operation of VELODY: user distinguishability and nonlinearity.

User-Distinct Vibration Response

A human hand exhibits unique physiological features such as geometry, bone shape, bone-muscle ratio, bone density, which have been utilized as a static biometric for a while [26]. These features lead to the human-distinguishable characteristics of acoustic dispersion, absorption, and reflection when a person places his/her hand on a vibration surface. Specifically, the contact area between a hand and the vibration surface affects the reflection and absorption of the surface vibration. Differences in the contact area (due to different hand geometry of different users) contribute to different vibration propagation paths and varying constructive or destructive interferences at different frequencies – leading to frequency-selective vibration responses. Moreover, the differences in hand's damping and acoustic absorption relate to composition, the force and distribution of contact between the hand and surface, contributing to vibration responses that are user-distinct, too [73].

One can naively model the vibration response of a hand using a spring-mass-damper system. Such a model, however, ignores several practical issues, including the multipath-induced frequency selectivity dependent on the hand-surface contact and the nonlinear spectral interactions. As a result, an accurate user-specific model for hand contact interaction is extremely hard to build even by state-of-the-art 3D finite-element (FE) modeling techniques [272, 74].

Nonlinear Effects in Vibration Response

The second property that VELODY utilizes is the nonlinearity in the vibration responses of the hand-surface system, which is difficult to model and predict [179, 291, 319]. Previous studies have demonstrated that a hand itself, due to its geometry and composition, is a nonlinear medium for acoustic propagation [139, 74].

Here, we show a model of nonlinear acoustics to explain the complexity of vibration responses of the hand-surface system. For a linear system, the output signal S_{out} is a linear combination of the input signals S_{in} , which can be represented as:

$$S_{out} = A \cdot S_{in}. \tag{3.1}$$

The complex gain only affects the phase and amplitude of the inputs, and no new frequency component appears in the response of the linear system. In a nonlinear system, however, like the hand-surface system, the response will contain new frequency components. For simplicity, we model the nonlinear response as a power-series of inputs with different gains at each term:

$$S_{out} = \sum_{n=1}^{\infty} A_n \cdot (S_{in})^n. \tag{3.2}$$

For example, if the input is a single sinusoidal wave at a frequency f_1, different orders of *harmonics* $(n \cdot f_1)$ will appear in the response. For an input composed of two signals, the output of this nonlinear system exhibits *intermodulation*:

$$S_{out} = A_1 \cdot (S_{in,1} + S_{in,2}) + A_2 \cdot (S_{in,1} + S_{in,2})^2 ... \tag{3.3}$$

For example, the second order term in Eq. 3.3 has a product of signals resulting in new frequency components at $f_1 - f_2$ and $f_1 + f_2$. We can rewrite

the second-order term of the output in the equation above as follows.

$$\begin{aligned} S_{out,2} =& a_1^h \sin(2\pi \cdot 2f_1) + a_2^h \sin(2\pi \cdot 2f_2) \\ &+ a_1^m \sin(2\pi(f_1 + f_2)) + a_2^m \sin(2\pi(f_1 - f_2)), \end{aligned} \tag{3.4}$$

where a_i^h are the gains for harmonics and a_i^m are those for the intermodulation.

The harmonic gains depend on the medium properties and the frequency, while the intermodulation gains depend on several factors including the material coefficients between f_1 and f_2, the amplitudes of both f_1 and f_2, which are sensitive to the structure of vibration medium [179] – the hand-surface system in our case. The system creates more complicated intermodulation interactions for higher order terms which are hard to predict.

Note that this simplified model does not convey the dynamics and component interactions of a nonlinear system as the nonlinear responses are highly input-dependent within the same nonlinear system. The model fails to describe the non-analytic responses like complicated energy exchange between different frequencies as well as temporal dependencies of nonlinear coefficients [291, 319]. Other nonlinear effects include nonlinear attenuation rates at different frequencies depending on the input excitation level [13]. Due to this complex and nonlinear nature of vibration responses in a hand-surface system, precise modeling or prediction of arbitrary responses preserving individual traits is highly implausible. It is very hard to predict the hand response for a previously unobserved input signal, to the best of our knowledge.

Motivational Example of Hand-Surface Vibration

We take an exemplification approach to motivate the distinct and nonlinear hand-surface vibration. We record the vibration responses of a

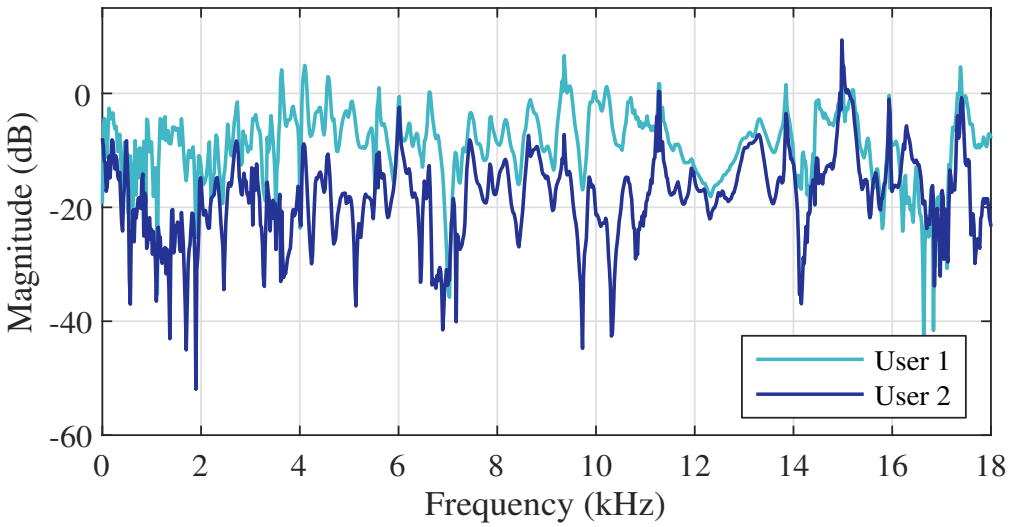
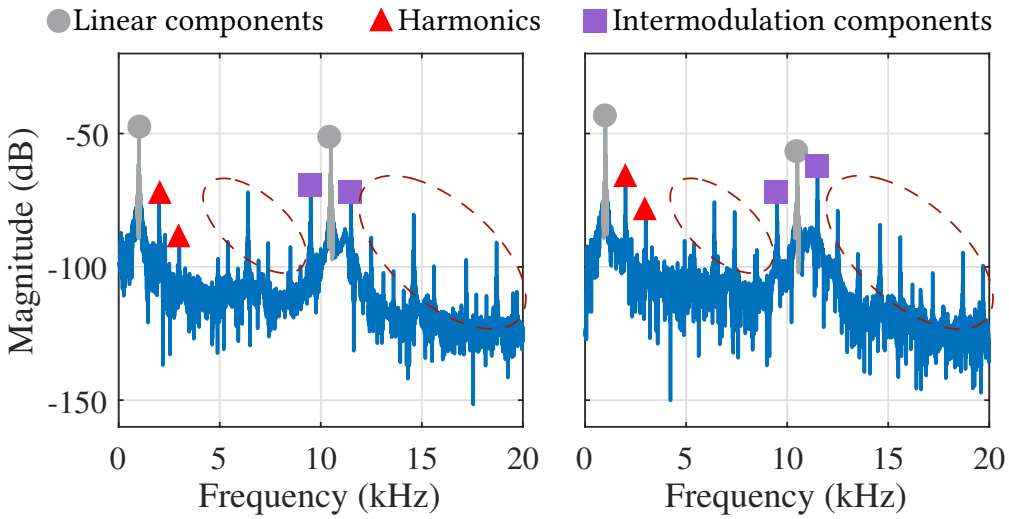


Figure 3.2: Vibration responses of two difference users.

hand-surface system to provide an intuition about our model. We use a portable vibration speaker (Vib-Tribe Troll Plus) to generate an input vibration and we collect the responses using a contact microphone (BU-27135 accelerometer) from a vibrating copper surface (setup similar to Figure 3.1).

User distinguishability: We first examine the user distinguishability of frequency responses. Two users place their hands on the vibration surface with the same gesture (relaxed with spreading fingers). Meanwhile, the vibration speaker plays a sweeping sinusoidal vibration from 0.2 to 18 kHz for a duration of 200 ms. Figure 3.2 shows the frequency response of the transfer function of each user, illustrating the attenuation at different frequencies. It is evident from the figure that the responses of the two users are easily distinguishable. The transfer function does not capture all sources of nonlinearity like harmonics and intermodulation which result



(a) With a hand on surface (b) Without a hand on surface

Figure 3.3: Nonlinearity in hand-surface measurement.

in more distinguishability.

Nonlinearity: To visualize the nonlinearity in hand-surface system, we play two sinusoidal waves at 1 kHz and 10.5 kHz simultaneously. We show the frequency response of the raw recorded signals (not the transfer functions as before) with and without a hand placed on the vibration surface in Figures 3.3(a) and 3.3(b), respectively. We mark the major frequencies in grey dots, some representative harmonics in red triangles, and intermodulation components as purple squares. The spectral locations of the newly-generated frequencies match the anticipated harmonics and intermodulation results in both scenarios. The intermodulation components are significant in both cases and even comparable with the major frequencies. Also, the hand exhibits distinguishable modification of nonlinear components as evident from components marked and circled in

Figures 3.3(a) and (b).

The findings above show an intuition that the vibration responses of hand-surface system are distinct between users (Figure 3.2), and the nonlinear effects are significant (Figures 3.3), too. Both observations are critical to the design of VELODY.

3.2 System and Threat Models

In this section, we describe the system and threat models for VELODY.

System Model

Figure 3.4 shows an overview of the system model, including the involved parties. We assume a general scenario where VELODY is employed to authenticate a user (U) to use smart devices (D). The authenticator service (S) grants permission for the user (U) to use smart devices (D) and access to authorized contents. The user (U) requests authentication and permission through the VELODY terminal (V), which is associated with an interface consisting a surface, a vibration speaker, and contact microphones. For example, V can simply refer to laptop or a smartphone paired with a smartwatch that has a high bandwidth accelerometer [155]. V generates a vibration signal according to a challenge assigned by S, collects the response, and sends it to S. We assume a secure training phase during which S collects all vibration challenge-response pairs securely for future verification.

For each authentication request, S randomly selects one disposable vibration challenge and sends to V, which collects the hand-surface response. The response is sent back to S to verify the claimed identity U. Note that V may not only verify the identity solely relying on vibration challenge-responses but also on other factors like password in a multi-factor authentication scenario. Once U is verified and authenticated, the

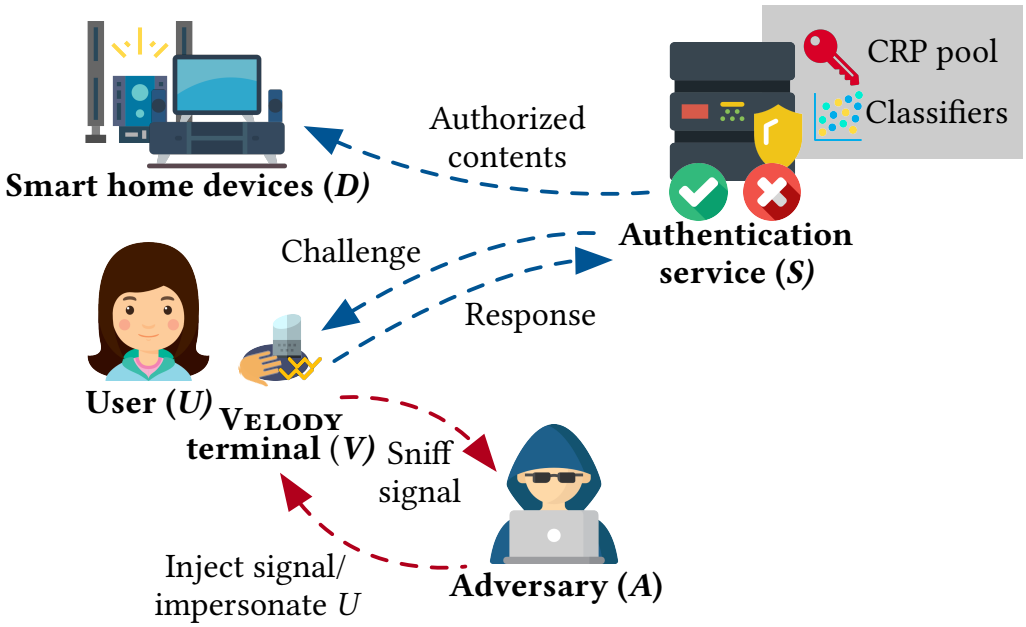


Figure 3.4: System and threat model.

requested D will be activated, and the authorized contents, such as a video stream, will be distributed.

Figure 3.4 depicts the involved parties in our system model as separate entities, just for visualization. There is nothing preventing V, D and S to be part of the same device, such as a laptop, desktop, or even a smartphone.

Threat Model

The goal of the adversary (A) is to deceive S to grant the access to the victim, U. In addition to the attacker capabilities that have been typically assumed in previous work, such as physical access to the authentication devices, we take one step further and assume that the active attacker is able to observe previously used responses and replay raw or synthesized response corresponding to an unknown challenge through a side

channel. This side channel could refer to (1) a compromised networking interface between V and S or to (2) the attacker collecting responses through a placed/compromised device in the same environment. In this chapter, we assume a strong adversary that is capable of recording the exact challenge-response pairs. By considering a strong adversary model capable of recording and replaying biometric information, we avoid the pitfalls of previous defense approaches. Under this scenario, we consider the following attack scenarios.

- **Zero-effort attack.** In this scenario, A only bypasses the password and tries to authenticate opportunistically by vibrating an empty surface without hand contact using the authentication-time challenge assigned by VELODY.
- **Impersonation.** In this scenario, A has access to V, bypasses other authentication factors like password, claims the identity of U, and places his hand on the vibration surface to impersonate legitimate U using the same gesture.
- **Raw signal replay attack.** In this scenario, A acquires previously-used vibration challenge-responses from U and replays an arbitrary raw response to S during an authentication session through a compromised wireless channel.
- **Synthesis attack.** More advanced than simply replaying raw signal, A attempts to predict the response of a specific challenge by modeling from previously observed responses and inject the synthesized signal in real time. We consider the implementation of multiple synthesis methods in our evaluation.

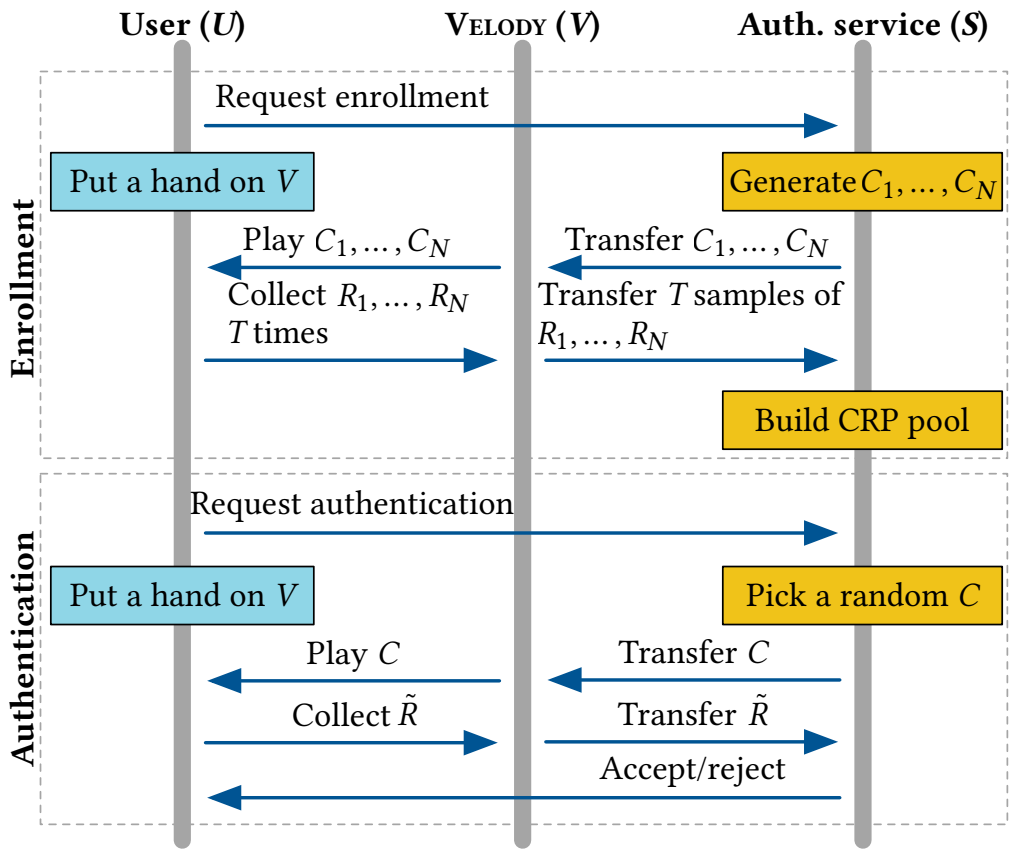


Figure 3.5: Authentication protocol of VELODY.

3.3 VELODY Protocol and Framework

In this section, we present the design details of VELODY.

Authentication Protocol

Challenges and responses: VELODY employs a challenge-response protocol as illustrated in Figure 3.5. At each authentication session, S sends the user a challenge and receives a response. Only after matching the

measured response to the previously recorded response is the user authenticated. Each challenge-response pair (CRP) is disposable; a challenge will not be reused in other authentication sessions.

A vibration challenge (C) is a specially designed acoustic signal played by V. S collects a challenge-specific and user-distinct response for *verifying the user identity*. The n_{th} challenge $C_n = (f_{crp}, f_n^1 ... f_n^M)$ can be characterized by M randomly selected distinct spectral stimuli (sinusoidal waves) appearing at different slots within the entire time period of the challenge, and f_{crp} is the time-varying frequency of a chirp signal. For each challenge C_n, the response is measured T times. R_n, the response to challenge C_n has T elements: R_n^i $(i = 1, 2, \ldots, T)$. As explained earlier, each response is a function of the challenge as well as the nonlinearities associated with playing the challenge to the user's hand. The nonlinear dynamics are challenge-dependent and user-specific; each challenge produces a unique response for each user.

Enrollment: The enrollment phase of VELODY is initiated when requested by U, or CRPs are depleted. S generates N new random challenges C_1 to C_N that are not previously used for authentication. V plays each C_n with the user's hand placed on the panel and records the corresponding response R_n. This procedure is repeated T times to generate a robust training set. After receiving the responses R_1 to R_N, S trains the classifiers for the new CRPs; VELODY trains one classifier for each CRP. We employ one-class k-nearest neighbors (OC-kNN) classifier for verifying the response corresponding to a challenge. During training, a threshold Th_n is computed for each classifier corresponding to each challenge. We assume that the enrollment phase takes place in a secure setting (attacker cannot record/alter the recorded responses).

Authentication: After the enrollment is completed, U can request authentication to S. Upon receiving an authentication request, S randomly

chooses a challenge C_n from unused challenge pool, which is sent to V. While U places their hand on the vibration surface, V plays the challenge C_n collects a response $\tilde{R_n}$, which is sent to S. S performs the feature extraction and decision making.

The authentication decision D on $\tilde{R_n}$ corresponding to C_n is described as follows:

$$D = F_n(\tilde{R_n}, R_n, Th_n), \tag{3.5}$$

where F_n represents the process starting at feature extraction and ending at the OC-kNN-based classification. R_n represents T training responses collected during enrollment; and Th_n is a challenge- and user-specific threshold. The challenge used in the current session, C_n, is disposed to ensure security against replay attack. The detailed decision process is discussed in Section 3.3.

Framework Overview

The processing framework of VELODY is illustrated in Figure 3.6 including is major stages. The collected responses during the enrollment session (i.e., $R_n^1, R_n^2, \ldots, R_n^T$) and authentication session (i.e., $\tilde{R}$) is synchronized and segmented first; then, filters and normalization are applied on the raw response segments. VELODY extracts effective spectrotemporal features from the raw time-domain response. For each CRP on normalized feature vectors, an OC-kNN classifier is built. An authentication decision is made based on the comparison of the CRP-specific threshold and the OC-kNN distance between observed features of response $\tilde{R}$ and the templates. The advantage of using OC-kNN as a classifier is that training can be conducted per-person, without the need to collecting data from multiple people.

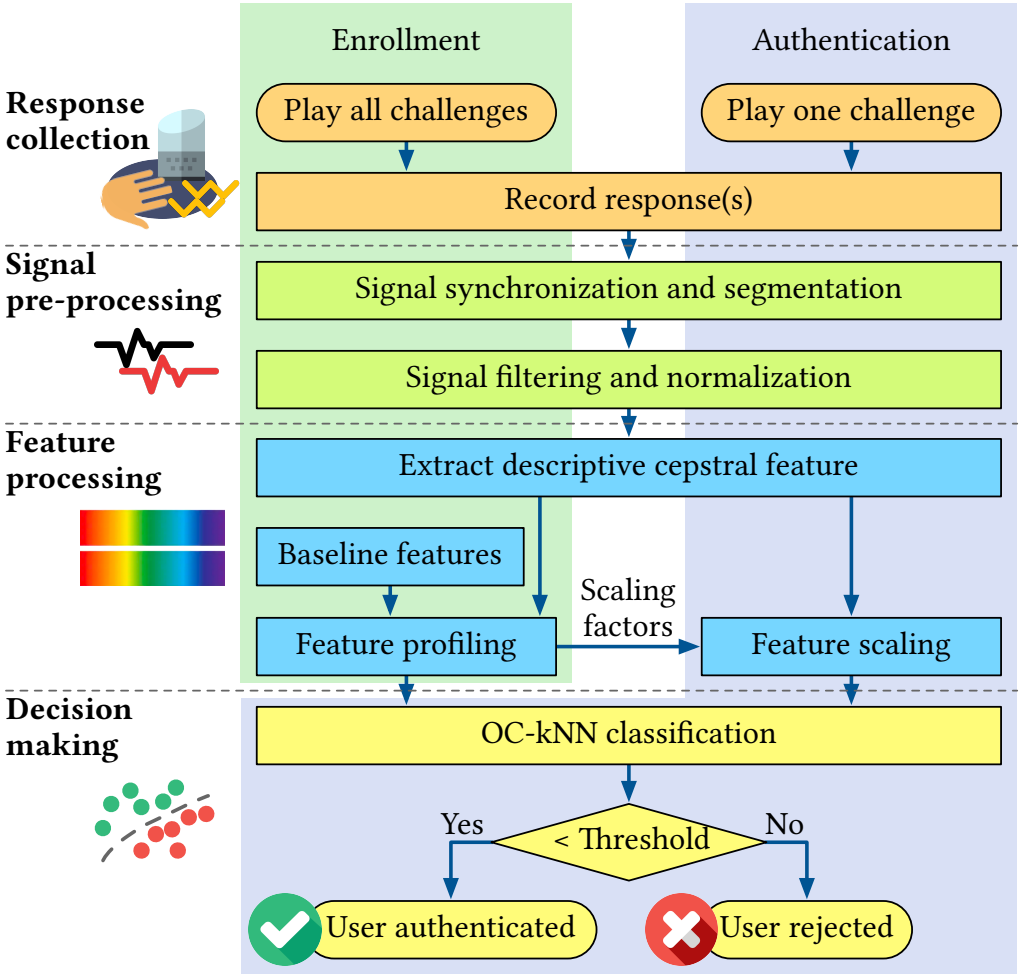


Figure 3.6: Processing framework of VELODY.

Vibration Challenge Design

We have two requirements from Velody's vibration CRPs: (1) distinguishability between the users of the system and (2) distinguishability as well as unpredictability from previously observed CRPs. These requirements necessitate the careful design of the challenges.

To meet the first requirement, we adopt a chirp vibration signal (frequency sweep) to capture the frequency selectivity contributed by the physiological characteristics of human hand in a short time. We meet the second requirement by designing each challenge to evoke a unique vibration response each time. The period of entire challenge is divided into several time slots, and in each slot, VELODY superimposes a sinusoidal wave at a random frequency onto the chirp instance to make the response unpredictable. The superimposition of the chirp signal with a sinusoidal wave generates complex harmonics and intermodulation interactions of different orders simultaneously, which is practically unpredictable from previously observed CRPs.

The vibration challenge signal $C_n(t)$ as a function of time t is expressed as:

$$C_n(t) = S_{crp}(t) + \sum_{i=1}^{I} S_{sin,i}(t). \tag{3.6}$$

The linear chirp signal $S_{crp}(t)$ is constructed by:

$$S_{crp}(t) = A_{crp} \sin\left(2\pi f_{crp}(t)t + \phi_{crp}\right), \tag{3.7}$$

where A_{crp} and ϕ_{crp} denote the amplitude and phase of the chirp signal, respectively; and $f_{crp}(t)$ is the frequency of the chirp, which linear changes from f_b to f_e over time:

$$f_{crp}(t) = \frac{f_e - f_b}{T_{crp}}t + f_b. \tag{3.8}$$

The random component $S_{sin,i}(t)$ in (3.6) is defined as:

$$S_{sin,i} = \begin{cases} A_i \sin(2\pi f_i t + \phi_i) & \text{if } (i-1)\frac{T_{crp}}{I} \leqslant t < i\frac{T_{crp}}{I} \\ 0 & \text{otherwise,} \end{cases} \tag{3.9}$$

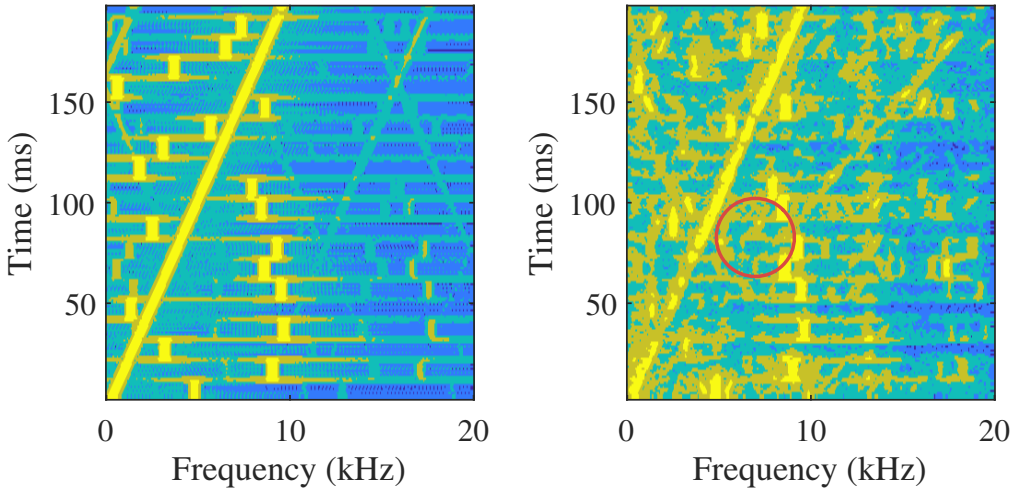
where A_i is the amplitude of the sinusoidal wave in the i-th time slot, $(i-1)\frac{T_{crp}}{I} \leqslant t < i\frac{T_{crp}}{I}$; and f_i is the random frequency.

In our prototype, the chirp S_{crp} changes from f_b =0.5 kHz to f_e =10 kHz, in which the vibration speaker generates stable vibration and hand-surface responses preserve useful information for distinguishing different users. The duration T_{crp} is set to 200 ms, short enough to avoid annoying the user during enrollment and authentication. The changeable stiumli of each challenge consist of 20 different sinusoidal waves of random frequencies (i.e., $I = 20$), uniformly distributed over in a range between 0.5 kHz to 10 kHz to ensure diversity of both linear and nonlinear components. The amplitudes of sinusoidal stimuli, A_i, is also randomly determined for challenge diversity.

In Figure 3.7, we show two spectrograms: one from a challenge and one from its corresponding response. From Figure 3.7(b), we can clearly observe some nonlinear components, such as the highlighted ones, including harmonics and intermodulation, which are widely spread over a wide frequency range.

Feature Processing

Signal pre-processing: First, we perform signal alignment and segmentation to minimize bias for feature extraction, resulting from imperfect hardware synchronization. We align the measured response with the challenge by finding the time lag that maximizes the cross-correlation between them. Second, we apply a bandpass filter between 0.3 kHz and 20 kHz to remove external vibration induced by motion. Also, we apply



(a) Challenge (b) Response

Figure 3.7: Comparison of challenge and response spectrograms. The challenge contains the chirp as well as superimposed sinusoidal signals at different frequencies. Some non-linear components are highlighted in the response.

multi-band spectral subtraction to clean the in-band noise due to measurement. Finally, we apply Z-score normalization on each response signal to reduce the variability from gesture inconsistency.

Cepstral feature extraction: The cepstral features are widely adopted for acoustic modeling of music, human speech, and structural damage, etc. which are of complex or nonlinear nature. Intuitively, cepstral coefficients describe the dynamics among the different frequency bands of a signal, including the contribution of linear and nonlinear spectral components. Cepstral coefficients are calculated by applying discrete cosine transform (DCT) on the complex logarithm of the Fourier transform of a time-domain signal. A sliding window is used to extract cepstral coefficients over the duration of a signal to model its temporal dynamics.

The Mel-frequency cepstral coefficient (MFCC) is the most frequently

used cepstral feature for human speech modeling and recognition since the Mel-scale filter banks are optimized for human speech and perception frequency. Instead of using the Mel-band, VELODY applies linearly allocated filter banks before calculating the coefficients. We argue that unlike human speech where high-frequency components contribute less to human perception, the nonlinear vibration responses of VELODY are spread more widely across the spectrum. Specifically, the band edges of overlapped filter banks are separated by 0.25 kHz, and we take 40-th order cepstral coefficients at each time window of 10 ms, with a window overlap of 8 ms to capture fine-grained dynamics. Moreover, the delta and delta-delta of the cepstral coefficients are also computed to capture more fine-grained spectral dynamics within a short time frame. A cepstral feature map combines all the cepstral coefficients with the log energy and first/second order delta energies per window.

Statistical feature extraction: Raw cepstral features exhibit inconsistencies brought by several factors such as circuitry randomness, gesture variation, and imperfect signal segmentation. To overcome this issue, we extract statistical features for each coefficient channel. Each coefficient channel is defined as the sequence of the values of cepstral coefficients over signal duration.

Besides mean, variance, entropy, and power, which are standard metrics in characterizing a random variable or its distribution, we adopt other metrics to assess the distribution of cepstral coefficients over the signal period. Skewness measures the degree of symmemtry of left and right parts of a distribution; kurtosis estimates the 'tailedness' of one distribution compared to normal distribution; and crest factor examines the significance of the extreme peak in the distribution [169]. The final feature vector comprises statistical features describing the cepstral, delta-cepstral, and delta-delta-cepstral coefficients as well as log frame energies. This

results in a feature vector with 1722 elements per response in this work.

Classification

VELODY is a per-user system; a VELODY user does not have access to other users' CRPs for privacy and security considerations. This requirement constrains VELODY's classifier design as it cannot obtain negative samples from other users. OC-kNN is an instance-based classifier that relies on the similarity between inference-time observations and training instances. VELODY trains one OC-kNN classifier for each CRP; the underlying assumption is that the response to a challenge for a user is different from those to other challenges. It is also different from responses to the same challenge from other users. The authenticator service passes the features from the response to the CRP's OC-kNN that decides whether the response is valid or not for the played challenge. The two major steps in OC-kNN decision making are distance calculation and threshold comparison.

Distance calculation: Recall that during enrollment, VELODY plays each challenge T times, so that it collects T copies of the response. Each response is associated with a feature vector. For the rest of this discussion, $x^i_{n,j}$ refers to the i_{th} feature of a training response (R^j_n) to the challenge C_n. To keep the notation simple, we use x^i_n instead of $x^i_{n,j}$, except for Eq. 3.12.

We first normalize each feature to the same scale by min-max normalization for the fairness of the distance-based OC-kNN:

$$\hat{x^i_n} = \frac{x^i_n - min(x^i_n)}{max(x^i_n) - min(x^i_n)}, \tag{3.10}$$

where the min and max are taken for a feature value over the T responses.

Given an unseen feature vector z_n at the authentication phase, VELODY scales it using the min and max factors computed during training: $\hat{z^i_n} = \frac{z^i_n - min(x^i_n)}{max(x^i_n) - min(x^i_n)}$. We observe that different features have varying sensi-

tivity to system or gesture randomness. We introduce a weight for each feature so that the more consistent features have higher weights [54]:

$$w_n^i = \frac{max(E(\hat{x_n^i})) - E(\hat{x_n^i})}{\sum_{i=1}^{1722}(max(E(\hat{x_n^i})) - E(\hat{x_n^i}))}. \tag{3.11}$$

The expectation is taken over the T training samples (responses to a single challenge during enrollment). The min and max are taken over the 1722 features.

The weights are applied to both the training and test instances. The ℓ_1 distance is calculated between the weighted test instance $\hat{z_n}$ and all T training instances $\hat{x}_{n,1...T}$ as:

$$d_j = d(\hat{z_n}, \hat{x}_{n,j}) = \sum_{i=1}^{1722} \left| (\hat{z_n^i} - \hat{x_{n,j}^i}) \cdot w_n^i \right|. \tag{3.12}$$

The final distance of the test instance to the challenge is calculated by averaging the k smallest d_j values. Comparing the final distance to a threshold Th yields the final classification result.

Threshold estimation: The major obstacle in VELODY's classification is determining a proper Th for each user and each CRP. An ideal Th accepts legitimate samples while rejecting all illegitimate samples. The ℓ_1 distances from the classifier show great diversity among users and CRPs, hence, a fixed threshold for every user and CRP is not ideal. Nevertheless, we notice that distances between training instances and baseline responses collected from vibrating surface without hand contact correlate with those of illegitimate distances ($\rho > 0.5, p = 0.000$) for each user. VELODY utilizes these baseline samples available to every user during enrollment to estimate Th_n corresponding to the n_{th} challenge of one user. VELODY calibrates Th by leave-one-out cross validation based on training and baseline samples. More specifically, one training instance is held out at a fold, and

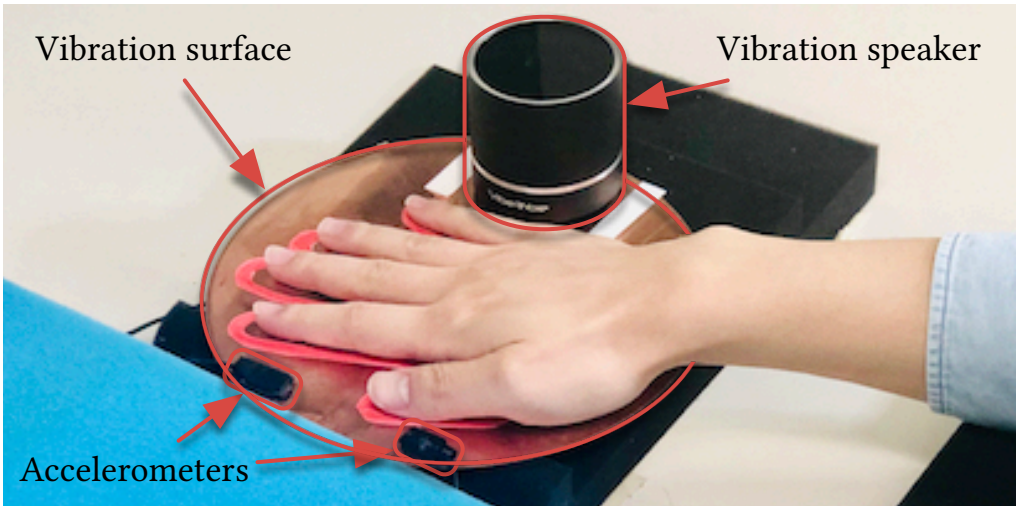


Figure 3.8: VELODY prototype setup.

its kNN distance $d_{n,pos}$ as well as distance of baseline samples $d_{n,bl}$ are computed using the rest training instances. Then, the threshold Th_n is determined by

$$Th_n = E(d_{n,pos}) + \alpha \times (E(d_{n,bl}) - E(d_{n,pos})) \tag{3.13}$$

where the expectation is taken through all folds and α is a global tuning factor, the usability of which on all CRPs will be evaluated instead of determining thresholds by exhaustive search.

3.4 Prototype and Data Collection

Hardware Prototype

A prototype of VELODY was built, as shown in Figure 3.8. A commercial off-the-shelf vibration speaker Vib-Tribe Troll Plus is used to play challenges. It is attached to a vibration surface, which is an 8-inch copper plate laying

on a polymer foam pad. The speaker has an effective frequency range between 80 Hz to 18 kHz and a signal to noise ratio (SNR) of 80 dB. Two contact microphones (BU27135 accelerometer) are attached on two different locations of the vibration surface to measure vibration responses. The BU27135 is an analog accelerometer with a wide effective spectrum and a high sensitivity. Since VELODY relies on the physiological properties of human hand instead of behavioral traits, we fix the gesture for all the users: all users are asked to put the right hand with fingers spread on the vibration surface, where we draw a hand shape for consistent alignment in evaluating impersonation attack. As a proof-of-concept, a PC is used to output all challenges through a built-in sound card and collect responses through a dual-channel USB sound card, sampling at 48 kHz. MATLAB's data acquisition (DAQ) toolbox is used.

We argue transferring challenges and collecting responses can be done remotely via wireless protocol, such as WiFi and Bluetooth. in a real-world use case. The duration of each challenge is set to 200 ms. We generate 100 challenges, and these challenges are kept unchanged for all users for establishing impersonation attacks.

Data Collection

We recruited 15 subjects with body mass index (BMI) ranging from 17.5 to 29.6 with a median of 22.2. The entire course of data collection took place over one and a half months, during which each participant was involved in three data collection sessions. The first two sessions were performed within one day with a time gap of at least 30 minutes. This was to verify intra-day (short-term) consistency and to establish baselines of consistency. The third session was arranged at least five days after the first two sessions to collect data for verifying inter-day (long-term) consistency.

Each session took about 20 to 30 minutes, including introduction, orientation, surveying, and data collection. After explaining the consent form,

having user's agreement and signature and collecting basic information about the user, each participant was demonstrated with how to interact with VELODY interface and take a good gesture. For each challenge, responses were measured for 15 trials. In between two consecutive trials, the user was asked to remove the hand from the plate and relax to ensure the diversity of the data set. Each trial took 30 seconds, including short intervals of 100 ms between two consecutive challenges. No complicated task or gesture for enrollment or authentication was needed. In a real use case, each authentication session will take only 200 ms, which is short enough to ease user's burden. The user study is approved by the Institutional Review Board (IRB) of our institution.

The total number of collected responses is 67,500 (3 sessions $\times$ 15 users $\times$ 100 challenges $\times$ 15 trails). Additional 15 responses were collected from empty vibration surface for threshold estimation and attack evaluation.

As for impersonation attack, for each user, we consider all other 14 users as active impersonators. Therefore, we use $3\times15\times100\times14 = 63{,}000$ samples for impersonation attack against each user. As for replay attack of raw signal attack, we use responses collected for challenges other than the legit one. For each participant, the number of raw signal replay samples is $99\times100 = 9900$. For each user, we also conduct benchmarking sessions for evaluating the attack using modeling and synthesis.

3.5 Evaluation

In this section, we evaluate the VELODY framework focusing on answering two questions about its usability and security aspects.

Q1: How well does VELODY authenticate legitimate users? The major factor impacting the usability of biometric authentication is its success rate of verifying true users (true positive), which is typically compared against

the possibility that an illegitimate user is falsely accepted (false positive), where we adopt responses from other users performing the same gesture while being stimulated by the same challenges as impersonation samples.

More specifically, four detailed usability aspects need to be analyzed to answer *Q1* comprehensively, for which we vary VELODY's configuration like threshold, training set size, and CRP complexity, and interpret results of FNR, FPR, and EER.

- How sensitive is VELODY to system parameters such as k in OC-kNNs and threshold factor α?
- How consistent is VELODY's accuracy in the long term?
- How much training data do we actually need?
- How scalable are the CRPs of VELODY?

Q2: How robust is VELODY against various attacks? The security evaluation focuses on examining and comparing the attack success rate of zero-effort attack, impersonation attack, raw signal replay attack, and synthesis attacks. The following question will be answered in this regard.

- What is the most effective attack modality, and why?

Evaluation metrics: The major metrics used for quantitatively analyzing the system's usability and security are as follows:

- False negative rate (FNR): The rate of mistakenly rejecting legitimate users, as a function of classification threshold. It is a usability metric.
- False positive rate (FPR): The ratio of how many illegitimate samples are accepted, as a function of classification threshold. It is a security metric.

- Equal error rate (EER): The rate when FPR equals to FNR for a certain classification threshold. It is a widely adopted metric to assess the overall accuracy and how well usability and security are balanced in an authentcaion system.

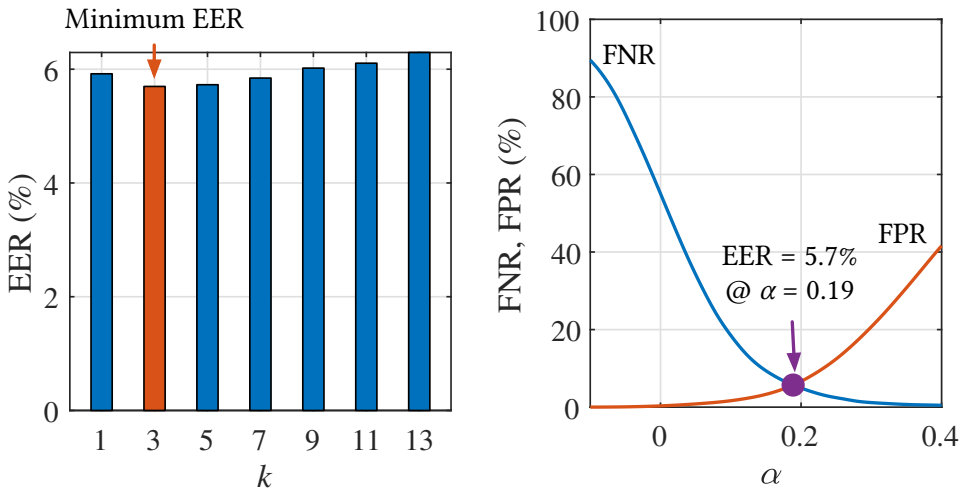
Accuracy of Authenticating Legitimate Users

System parameter baselining using intra-day sessions

One of the challenges in implementing VELODY's classification scheme is tuning the large number of OC-kNN classifiers corresponding to many CRPs with a minimal effort since it is not practical to exhaustively search the optimal configuration for each classifier of every CRP. To this end, we evaluate whether two major parameters, OC-kNN component k and global threshold tuning factor α, are sufficient to achieve a good overall authentication accuracy.

For each user, two separate sessions are used for evaluating system performance. Though physiological characteristics of human hand are relatively consistent, we argue that multiple factors, such as gesture, posture, and contact force, which may not be well controlled by users without concentration across different sessions, may influence the authentication success rate. The system configuration of VELODY should be robust against these variations.

Setup: We use two sessions within one day (intra-day) but 30 minutes apart for all 15 users and 100 CRPs to establish a baseline for authentication accuracy. One session is used as a training set, and another acts as a test set. Each session includes 15 trials for every CRP. For each user, 30 trials of both two sessions from all 14 other impersonators are used as illegitimate samples for the classifier of each CRP. We evaluate $k = 1, 3, 5, \ldots, 13$, which



(a) EER vs. k (b) FNR and impersonation FPR vs. α (k = 3)

Figure 3.9: Authentication performance of intra-day sessions.

are fixed for both threshold estimation and OC-kNN testing. Tuning factor α is varied from -0.1 to 0.4 with a step of 0.02.

Results: The impact of k in OC-kNNs is evaluated first. A very small k may lead to noisy classification results and unstable performance; on the other hand, if k is too large, it will cause under-fitting and the decision boundary will be overly smoothed. Figure 3.9(a) shows the average EERs of all the users and classifiers with various k values, which are calculated by finding the crossover of interpolated FNR and FPR data points at varying discrete α. We can see that VELODY is able to achieve a satisfactory EER below 6.3% within a wide range of k from 1 to 13. The minimum EER of 5.7% is attained when $k = 3$.

The trend of FNR and FPR with varying threshold factor α from -0.1 to 0.4 is shown in Figure 3.9(b) at an optimal $k = 3$, which is fixed for following experiments. Both FNR and FPR change smoothly and monotonically with α as a larger α accepts more legitimate samples while misclassify-

ing more impersonation samples as well, which is intuitive regarding the distance-based OC-kNN classification. FNR and FPR intersect at $\alpha = 0.19$ when EER is 5.7% (marked with a purple dot in Figure 3.9(b)). VELODY performs satisfactorily within a broader range of α. For example, if $\alpha = 0.14$ is chosen, Velody can reject over 97.1% of attacks while maintaining a FNR at 10.7%.

Hence, we verify that VELODY's classification can achieve a good overall authentication accuracy with a large pool of CRPs without tuning parameters in a brute-force manner, and it is capable of handling inter-session variation of intra-day tests.

Long-term consistency evaluation on inter-day sessions

To verify long-term consistency and strengthen our usability argument, we collected the third session, following the same experimental procedure, but five days later than the first two sessions for each user. In daily usage, larger variation in vibration responses may occur due to behavioral changes by different cognitive and physical statuses, which may not be well considered by intra-day experiments.

Setup: We fix k to 3 and use the first two sessions, including $T = 30$ trials as the training set to authenticate the third session, which capture more variation of users due to inter-session behavioral inconsistency, as we observe that using training data collected in a single session for authenticating inter-day trials may not cover this variation perfectly, resulting a higher average EER of 7.9% by training on two individual sessions respectively.

Results: We show the varying FNR and impersonation FPR evaluated on inter-day sessions in Figure 3.10. We observe similar trend of FNR and FPR compared to intra-day verification results. A low EER of 5.8% can be achieved at $\alpha = 0.23$ (marked with a purple dot in Figure 3.10),

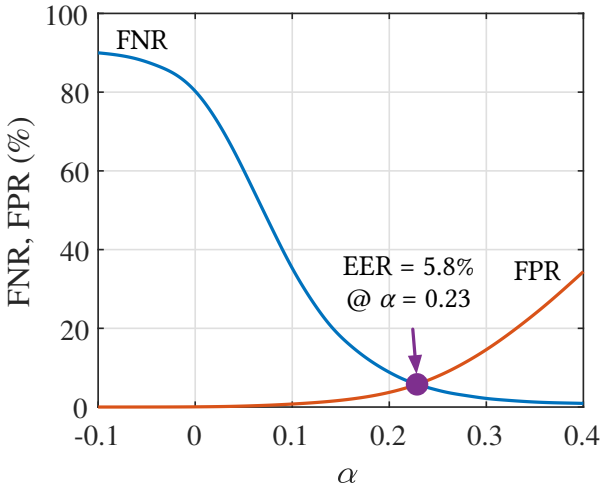


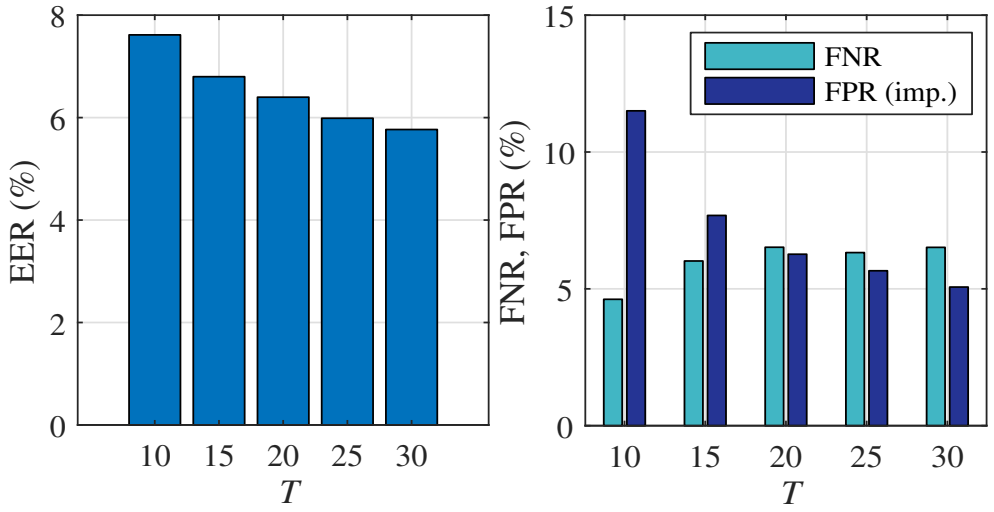
Figure 3.10: Authentication performance of inter-day sessions.

which indicates negligible difference compared to 5.7% from intra-day evaluation. Though the optimal α varies slightly, Velody still achieves low FNR and FPR of 11.8% and 2.7% respectively using an $\alpha = 0.18$, close to the interpolated EER point at $\alpha = 0.19$ of intra-day verification, indicating good consistency.

Therefore, we verify that Velody is robust to system and behavioral variation and attains good long-term consistency with reasonable training effort. We argue that physiological properties of human hand are relatively stable regarding to time despite that physical development process of children or aging may affect the properties [218], which can be addressed by updating the CRP pool.

Impact of training set size

Though Velody employs very short CRPs of 200 ms and almost passive enrollment/authentication sessions without performing complicated tasks, the size of training set influences usability in multiple angles such as duration of enrollment, the computation time for kNN at authentication phase, as well as data storage. To investigate the sensitivity of authentication



(a) EER vs. T (b) FNR, impersonation FPR vs. T (a = 0.22)

Figure 3.11: Authentication performance with different training set sizes.

performance to the number of instances used in training each classifier, we vary the number of training instances and examine accuracy of VELODY for each case.

Setup: We prune the training set from T $= 30$ instances of two intra-day sessions to 10 with a step of 5 by trimming those have larger average pairwise ℓ_1 distances to other training instances in the validation phase then test using 15 inter-day trials.

Results: In Figure 3.11, we show the variation of authentication performance ((a): EER, (b): FNR/FPR) with training set sizes. From the EER plot, we conclude that the performance is generally stable against different T, however, the more legitimate templates we have, the better VELODY's overall performance is, as the EER decreases from 7.6% to 5.8% by varying the number of instances from 10 to 30. Also, from Figure 3.11(b) we see both FPR and FNR do not vary substantially from 15 to 30 at a fixed α

of 0.22, meanwhile a smaller size T benefits consistency while sacrificing security slightly.

These findings indicate more training instances do improve system robustness, nevertheless, using fewer training instances around 15 is feasible to achieve comparable authentication performance while saving enrollment time if users keep good consistency, as well as data storage and computation at authentication time.

Scalability of VELODY CRP

The CRP pool of VELODY can be scaled by changing challenges in different domains like sinusoidal frequencies or complexity in terms of challenge duration and bandwidth of signal. We still anticipate that VELODY maintains its performance when a larger CRP pool is deployed for realistic usages with daily authentication activity, which is evaluated here.

Setup: First, for validating the variation in authentication success rate regarding different combinations of sinusoidal stimuli, we use the inter-session results and demonstrate the individual accuracies of all 100 200 ms-challenge.

Also, based on the same dataset we have, we can emulate the scenario when the challenge complexity is varied by truncating each 200-ms CRP in time domain to 100 ms and 50 ms respectively, starting from $t = 0$ which ensures that responses are not impacted by previous signal. Each truncated challenge-response has a narrower effective chirp bandwidth and fewer sinusoidal stimuli.

Results: The accuracy statistics of different vibration challenges are shown in Figure 3.12. The performances of vibration challenges of varying combinations of stimuli are quite consistent, and 99% of them have an average FPR lower than 10%. The average FNR per CRP is stable across var-

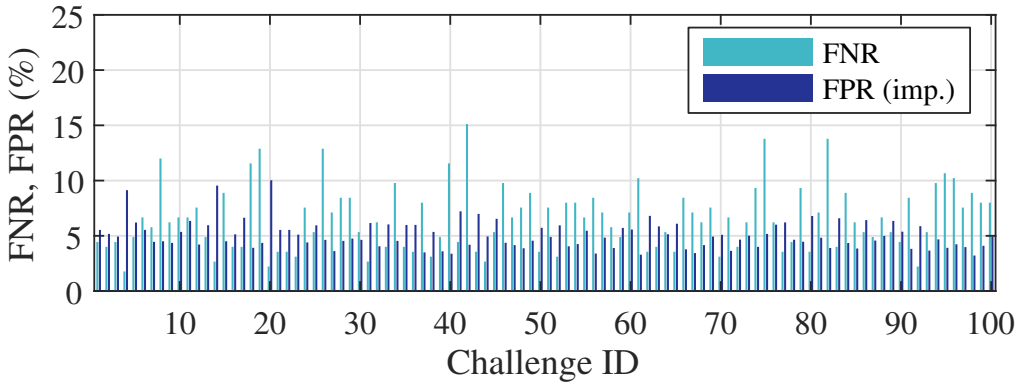


Figure 3.12: Authentication performance of different CRPs.

ious challenges, though more variant than FPR, and only a few challenges' (11%) FNRs are higher than 10%.

For verifying the efficacy of CRPs with reduced complexity, the threshold tuning factor α is varied for each case, and we also evaluate the FNR with FPR from impersonation attack, whose results are shown in Figure 3.13. From the plot, we observe that EER only degrades slightly from 5.8% to 9.1% and 10.4% when 200-ms, 100-ms, and 50-ms CRPs are used, respectively. Despite the observation that CRPs with reduced complexity lead to higher FNR while contributing to lower FPR with an α ranging from 0.15 to 0.4, and the thresholds to achieve equal error drift from that using 200-ms CRPs.

Revisiting the findings, we conclude that the design of VELODY vibration challenge is scalable and flexible. A user can enlarge the CRP pool by different approaches like updating the spectral stimuli, changing chirp bandwidth, and varying signal duration. Also, the enrollment and authentication time will be saved proportionally using decreased challenge duration with an insignificant penalty in system accuracy. However, VELODY also leaves the opportunities for improving the accuracy of different CRP designs by reconfiguring framework parameters such as the

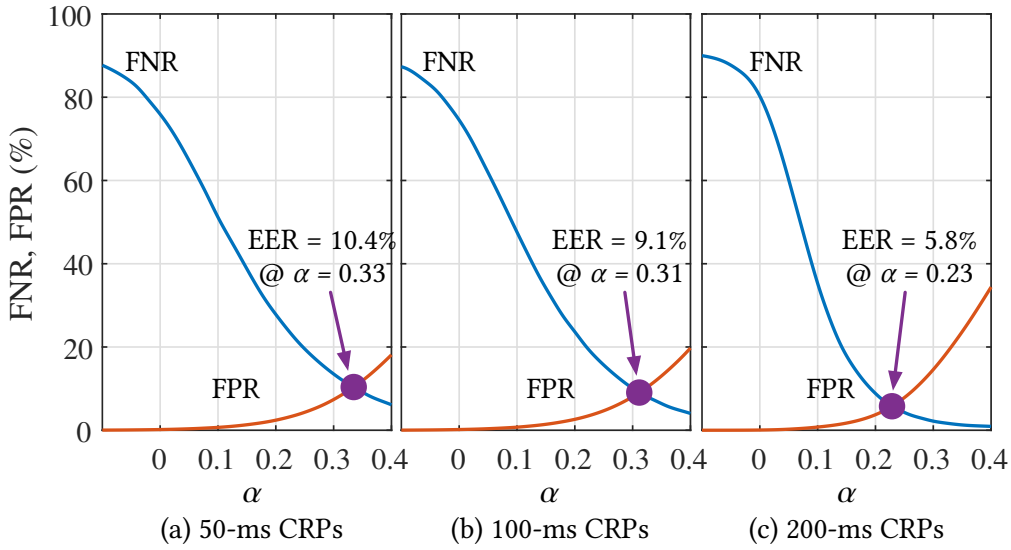


Figure 3.13: Authentication performance using CRPs with various complexities.

duration of sliding window, cepstral filter banks, etc., in feature extraction.

Robustness against Various Attacks

To answer Q2, we set up multiple attack scenarios with varying attacker capabilities and compare the results in respect to usability represented by FNR, whose results are shown in Figure 3.14.

Setup: The configuration detail of all evaluated attacks is explained below. Note that all classifiers are trained using 30 trials and k is set to 3. (i) *Zero-effort attack:* For evaluating this attack, we collected 15 responses from the vibrating Velody surface without hand contact to attack all 100 classifiers and all 15 users. (ii) *Impersonation attack:* This attack is evaluated with system consistency in previous section. Every classifier is attacked by responses of other 14 impersonators in all 3 sessions. (iii) *Raw signal replay attack:* We consider the worst case scenario that all

previous CRPs are overheard For each classifier, one raw response from every other challenge is replayed, resulting $99 \times 100 = 9900$ replay attacks per user. (iv) *Synthesis attack:* Based on resources assumed in raw signal replay attack, the adversary is capable of predicting users' responses in real time using different modeling methods.

Three modeling methods used in synthesis attack are as follows. (a) *Transfer function-based synthesis*: The adversary approximately models the nonlinear vibration system using transfer function. Chirp signal is frequently used for identifying vibration system [223]. The attacker cacu-lates the transfer function from the response of a linear frequency sweep between 0.2 kHz to 18 kHz with a duration of 200 ms, same as a legal challenge. The transfer function is computed by averaging 10 estimates. Two inputs are considered: raw/original challenge templates (TF-O in Figure 3.14) and responses acquired from the empty vibration surface (TF-E). Using the second input, the attacker focuses on modeling the effect contributed by contact of the user's hand. (b) *Nonlinear system identification-based synthesis*: The attacker adopts cascaded Hammerstein model, which is a well established method to identify nonlineartiy in vibration system [232]. In this method, nonlinear system is modeled as multiple branches of nonlinear static polynomial elements followed by a linear impulse response, which is computed by measurement from an optimized exponential frequency sweep. Similar to transfer function-based synthesis, we compute the Hammerstein model for each user by exciting the hand-surface system with a 0.2 kHz to 18 kHz optimized sweep of 200 ms, and attack 100 times for each user, considering two input sources same as (a) (NI-O, NI-E respectively). (c) *Feature-level synthesis*: Features of an unknown response is predicted by estimating a feature-level mapping between challenge and responses modelled by the least square solution x in $Ax = B$ where A is the feature vector extracted from responses of empty surface and B is that obtained from the corresponding hand-surface

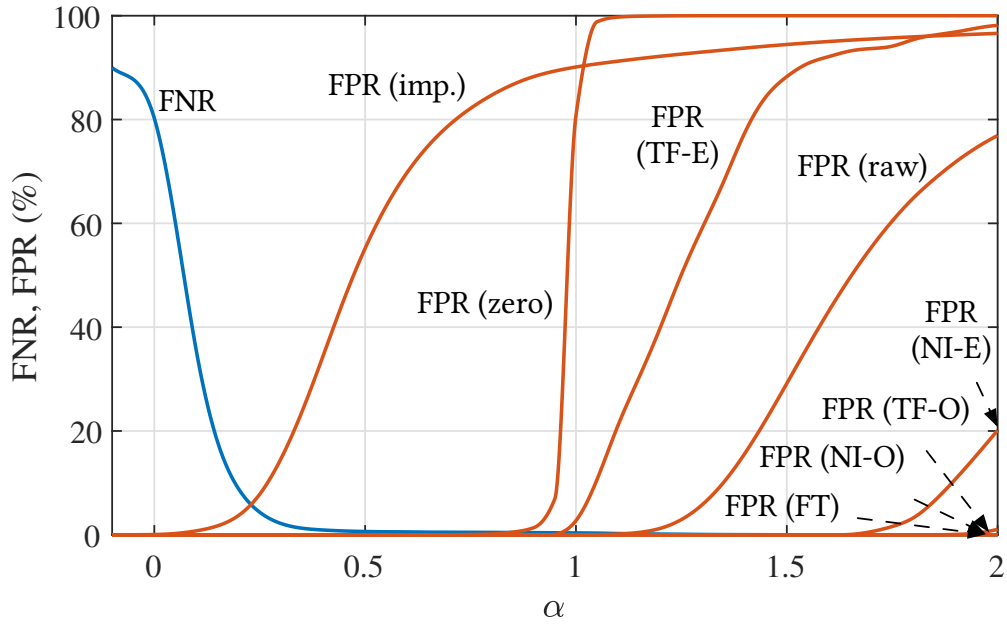


Figure 3.14: Robustness of VELODY against various attacks.

vibration response signal. The attack success rate is represented by FPR (FT) in Figure 3.14.

Results: Comparing various attack success rates in Figure 3.14, we conclude that impersonation attack is the strongest one. More specifically, when α below 0.8, none of the other modalities succeeds in attacking VELODY (0% FPR). We interpret the finding as follows. The failure of zero-effort attack is due to largely different force distributions and linear/nonlinear responses on the surface compared to impersonation. Replaying raw responses is not a feasible attack due to unique spectrotemporal characteristics of randomized stimuli in each challenge. The failure of synthesis methods attributes to the heavy nonlinearity in the vibration response introduced by either circuitry, vibration surface, or hand contact. Also, a 'corrupted' measurement consisting of complicated nonlinear responses

of different orders' harmonics and intermodulation even biases estimation by transfer function or Hammerstein model. These findings confirm that precise modeling and prediction in VELODY's scenario is very difficult because of multiple factors including non-analytic nonlinearity in real-world measurement. Hence, impersonation is the strongest attack in this case because of similar physical properties between hands and surface contact condition among multiple users.

So far, we have answered all questions post before. To summarize, VELODY authenticates legitimate users consistently across time with minimal effort in fine-tuning for many CRPs, minimal authentication effort, and reasonable training effort. VELODY's disposable CRPs are scalable for long-term usage. Security under various attacks is also guaranteed as VELODY achieves a low EER at 5.8% impersonation attack and stopping 100% of other attacks including replay and synthesis, benefiting from the unique spectrotemporal characteristics of nonlinear vibration responses.

3.6 Discussion

We have successfully demonstrated VELODY's usability and security against various attacks through extensive experiments and analysis. To further improve VELODY's practicality and security, the following issues are to be considered.

Deployment in various settings: In this work, we used a vibration speaker as a vibration source and a copper plate as a vibration media. We envision that VELODY can be deployed in a variety of settings with a different vibration source and vibration media as long as nonlinearity exists in vibration propagation. It could even be embedded in smart devices, such as laptops and smartwatches. To achieve this vision, a platform-specific challenge generation scheme and evaluation would be required.

Enriching VELODY's CRP pool: The most important security attribute of VELODY is its non-static and disposable biometric features. Other than the dimensions we discussed in the analysis, such as duration, random frequencies, and bandwidth, more aspects can be tuned to increase the the number of CRPs and improve distinguishability. Examples include the number and composition of spectral stimuli at each window and different gestures made by the user during enrollment and authentication, etc.

VELODY's training protocol balances between the effort in generating non-static biometrics and the size of the CRP pool to cover the user's authentication needs. According to a recent user study about daily authentication behavior [190], the average biometric authentication frequency is about 20 times per week for each user. VELODY can enroll 100 CRPs, each lasting for 200 ms, within 30 minutes. These CRPs can cover the user's authentication needs for 5 weeks.

Emerging attacks: Although we assumed an attacker with strong capabilities, except obtaining a precisely replicated physical model of the victim's hand, we cannot completely eliminate the possibility of more sophisticated attacks in the future. Existing methods of nonlinear system modeling like Hammerstein model, mostly work only in a constrained and controlled scenario. These methods rely on sufficient measurement, specially designed excitation, etc., for limited objectives, such as assessing the total harmonics distortion, instead of covering all nonlinear dynamics like non-analytic intermodulation. We can also consider neural network-based modeling methods, such as voice or music synthesis. However, they typically require a mature auditory model or sufficient training [85, 287], which require the adversary much more effort and stronger capabilities. We believe these attacks are applicable to VELODY's scenario.

Table 3.1: Comparison among biometric systems. (∗: zero-effort impersonator; †: reduced replay quality; ‡: static user features.)

Work	Protocol	Modality	FNR	FPR		
				Impersonation	Replay	Synthesis
Cardiac Scan [173]	Physiological	Radar-measured heart motion	4.42%	4.42%*	N.A.	N.A.
Wang et al. [297]	Physiological	Heartbeat-induced vibration	2.48%	2.48%*	N.A.	N.A.
BiLock [334]	Behavioral	tooth click sound	<5%	<1.5%	5.6%†	N.A.
BreathPrint [52]	Behavioral	Breathing gesture-induced sound	6%	2%	2%†	N.A.
Taprint [54]	Behavioral	Tapping-induced vibration	1.74%	1.74%	N.A.	N.A.
VibWrite [177]	Behavioral	Vibration response of dynamic gestures	10%	2%	N.A.	N.A.
Sluganovic et al. [257]	Challenge-response‡	Reflective eye movement	6.3%	6.3%	0.06%	N.A.
Brain Password [172]	Challenge-response	Electroencephalogram	2.503%	2.503%	0.789%	N.A.
VELODY (this work)	Challenge-response	Vibration response	5.8%	5.8%	0%	0%

3.7 Related Work

In this section, we revisit previous effort on biometric authetnication, where we both qualitatively and quantitatively compare VELODY with the state-of-art to show VELODY's contribution.

Traditional biometrics can be categorized into physiological biometrics and behavioral biometrics. Physiological characteristics like fingerprint, hand geometry, iris structure, or physiological signals like electroencephalogram (EEG), electrocardiogram (ECG), and electromyogram (EMG), have been used as biometrics [23, 325, 289]. Behavioral biometric refers to unique characteristics preserved in human dynamics such as gesture dynamics, speech, or gait [258, 254, 70, 122], which are easy to acquire.

In Table 3.1, we compare several state-of-art biometric authentication systems with VELODY. The works are divided by protocols, namely physiological, behavioral, and challenge-response. Note that the biometric-based challenge-response protocol here also relies on physiological properties of users but leveraging unique, passive, and varying responses to different stimuli. Following attributes are listed together: modality, FNR, FPR by impersonation, FPR by replay and synthesis. If the EER between falsely rejecting user samples and accepting impersonator is available, it is reported as FNR and FPR (impersonation) separately.

In Cardiac Scan [173], authors exploited the sensing capability of a DC-coupled continuous wave radar to sense unique motion patterns of users' hearts and achieve an EER as low as 4.42%. Note that the FPR (impersonation) reported here originates from zero-effort impersonators since it is not possible to mimic one's heartbeat. Similar characteristics of heartbeat are utilized in [297] with heartbeat-induced vibration captured by smartphones. The EER is as low as 2.48% against zero-effort impersonator. However, this protocol may not be applicable to defend against replay and synthesis attacks in VELODY's threat model where static biometic features may be leaked through a compromised channel.

The authors of [334] harvested unique sound from a tooth click recorded by commodity devices and achieved good consistency and security through a comprehensive user study and evaluation. With an increasing replay distance, the FPR of replay attack decreases to 5.6%. The authors of BreathPrint [52] utilized distinction in users' breath, and three types of breathing gestures–sniff, normal, and deep breathing are evaluated, whose FNR and FPR are 6% and 2% respectively. Chen et al. [54] designed a system named Taprint that uses vibration induced by finger tapping measured for user authentication, whose EER is as low as 1.74%. Liu et al. [177] leveraged the facts that varying user gestures will change the frequency response measured from a vibrating surface and designed a generalizable platform called VibWrite for authenticating users by password, lock pattern, and gesture input. We report the FNR and FPR by using password input. Even under imitation attack when the password is leaked, the FPR is as low as 2%. Though in [52, 334], the authors acknowledged and evaluated the security against replay attack of recorded noisy biometric samples, they are not applicable to VELODY's threat model where clean raw responses can be injected directly, since they discovered that the efficacy of replay attacks on these biometrics is highly dependent on the quality of replaying.

In terms of protocol, our work is most similar to [257, 172], where the

replay attack of raw responses can be stopped by adopting a challenge-response protocol with changing visual stimuli that elicits unique passive reflective eye movement for each challenge and each user. The system achieves an EER of 6.3% against impersonation, and it rejects almost all replay samples. Note that though a challenge-response protocol is used, the security against synthesis attack is guaranteed by the high complexity of synthesizing eye movement because features used to verify user identity from different responses are still static. Hence, this modality may not be suitable for VELODY's use case as well. Also, a similar protocol is implemented by using the event-associated electroencephalogram to generate vision-related challenge-response pairs, achieving good accuracy. The cognitive factors involved in the user enrollment, however, restricted the number of responses gathered within a satisfactory time [172]. VELODY takes advantage of the challenge-response protocol and the modality of hand-surface vibration response to achieve robust authentication, where the physiological characteristics of hand and the nonlinearity in hand-surface vibration responses are utilized to generate numerous disposable CRPs for defending against various attacks including raw signal replay and even strong synthesis attacks. VELODY attains low error rates while succeeding in rejecting all synthesized samples, too.

3.8 Conclusion

This chapter verified the feasibility of using the nonlinear response from hand-surface system for user authentication, relying on the unique physiological characteristics of human hand with a challenge-response protocol. By building the prototype of VELODY and conducting extensive user experiments, we validated several properties of VELODY regarding usability and security. First, VELODY is able to achieve an EER against impersonation as low as 5.8% in long term, showing a negligible loss with 5.7% using

short-term test trials, indicating good temporal permanence. Moreover, this result can be attained with reasonable training effort and negligible authentication time of a 200-ms challenge. Furthermore, we verified the scalability of VELODY's disposable CRPs by examining the FNR and FPR of individual challenges and challenges of different complexities. More importantly, VELODY succeeds in defending against all replay and synthesis attacks, benefiting from distinct features in each nonlinear response to a unique challenge.

Our findings suggested that VELODY's non-static biometrics are robust even when strong attackers are present. Nevertheless, to further improve the scalability, more effort should be put into investigating its performance in ubiquitous settings and the design space of CRPs.

4 UNDERSTANDING HOW SMART HOME USERS DEVELOP SECURITY AND PRIVACY CONSIDERATIONS AND ATTITUDES

Smart home technologies integrate a wide range of smart devices to support users' daily routines. With the wide adoption of smart home products such as smart speakers, thermostats, and door locks, users enjoy the conveniences of automated daily experiences and the reduction of repetitive menial tasks [332]. However, as these technologies impact users' lives in various aspects, they also present unprecedented security and privacy (S&P) threats to users and their environments [320, 101, 328].

In Chapter 2 and Chapter 3, we have presented how to make smart device systems more secure and private. Nevertheless, this may not guarantee that users can protect their security and privacy properly in real life during the interaction with and adoption of the products. In this chapter, we study how users think of and react to S&P issues of smart home products, which further informs solutions and practices for multiple stakeholders to support users in protecting security and privacy.

Existing work has looked into the role of S&P in users' adoption of smart home technology, especially during the acquisition and use stages [84, 83, 81, 328, 152]. Users factor S&P qualities of smart home products into their purchases, despite the observation that they may not be fully aware of S&P risks [152, 83]. Users may also come to realize S&P issues and implement reactive mitigation strategies during actual usage [84, 328].

Throughout the adoption journey, users' experiences with a product represent a reflective process from pre-purchase to post-consumption [131]. Considering S&P as a critical part of the user experience [320, 321], users exhibit varying S&P attitudes and concerns [149, 77]. While existing studies on users' S&P perceptions of smart home have primarily focused on singular timepoints in the adoption journey and are often conducted in controlled contexts using methods such as interviews and surveys

[83, 101, 320, 328]; these studies may miss the rich dynamics when users develop their S&P considerations and attitudes over time. Meanwhile, little research has investigated and holistically understood how users organically develop varying S&P considerations and attitudes throughout their adoption journey.

Recently, researchers have started leveraging online communities to study users' attitudes, including those on S&P-related topics, in vivo [167, 270, 271]. Online communities provide venues for many smart home users to seek product information and exchange S&P ideas. Members of such online communities collectively drive the topics and discussions based on their interests. As such, we choose a smart home-related online discussion forum to investigate *how smart home users develop S&P considerations, which shape their S&P attitudes during the adoption of smart home products.* We investigate our main research objective through three research questions:

- **RQ1:** [Consideration] What are users' S&P considerations in the adoption of smart home technologies?
- **RQ2:** [Attitude] What are users' attitudes toward S&P, and how do users' S&P considerations shape them?
- **RQ3:** [Discourse Influence] How does online discourse influence users' S&P considerations and attitudes?

We utilize Reddit (`www.reddit.com`), a major online platform of interest-based communities, as our research site. In particular, we analyze users' discussions in `/r/homeautomation`, [1] one of the largest forums for smart home users. This forum covers a broad range of specific and in-depth S&P topics, making it a suitable medium to study how smart home users develop S&P considerations and attitudes. We conducted a qualitative content analysis of 180 discussion threads, including 4,957 comments.[2]

[1] https://www.reddit.com/r/homeautomation/

[2] Reddit uses a tree-like structure called "comment thread" for online discussion. One user starts a thread with an initial post (root comment), and other users may leave comments under the post or other comments [59].

Our analysis contributes rich insights into users' dynamic considerations and attitudes. First, users develop two types of evolving and multi-dimensional S&P considerations: (1) S&P concerns regarding smart home technologies and (2) protective strategies during adoption. We observe that a set of interplaying contextual factors shape these considerations, including adoption phases and product factors (**RQ1**). Second, users' S&P considerations map to five categories of attitudes, namely dismissiveness, exploration, resignation, positive pragmatism, and devotion; each attitude combines the user's degree of S&P concern and level of incorporating protective strategies. We show that users' S&P attitudes are context-dependent and evolve according to the progression of considerations as they seek and gain information. However, their preconceptions may override a more objective assessment (**RQ2**). Third, while users exchange opinions and resolve ambiguity to develop S&P considerations and attitudes, they also wrestle with occasional social pressures and inaccessibility of accurate information during online discussion (**RQ3**).

Based on our findings, we provide recommendations to better support smart home users' evolving and dynamic S&P considerations and attitudes with improved designs and practices, S&P nudging, and information exchange. Finally, we inform future research to study smart home users' longitudinal attitudes from multiple angles, the geopolitical and cultural influences, and the impact of information access on smart home users' attitudes.

4.1 Method

We analyzed online discussions on Reddit to investigate how users' S&P considerations are discussed during their adoption of smart home products. This approach has the advantage of observing people's actual behaviors and information-seeking processes. A large body of research has

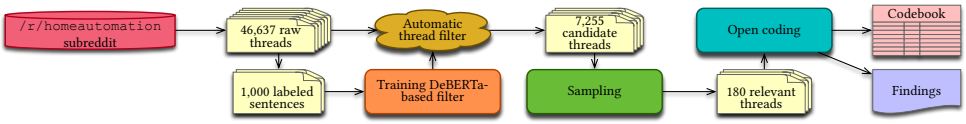


Figure 4.1: Our data analysis pipeline.

studied the discussion on Reddit platform, with increasing interest by S&P researchers [167, 33, 280, 295]. Reddit uses a threaded structure in discussion – each initial post (root comment) is followed by a series of comments over time. This feature provides an opportunity to observe discourse and interaction dynamics among discussants [167].

Reddit consists of subreddits, which are forums for specific topics. We looked for smart home-related subreddits that cover diverse products and integration levels. We decided to focus on `/r/homeautomation`, compared to other smart home-relevant subreddits suggested by Reddit's engine [234], for multiple reasons. First, it includes broad topics, diverse brands, products, and smart home ecosystems to support users' varying needs in different adoption and use phases, from seeking purchase advice to recommending customized automation. Existing work has leveraged the same subreddit to study smart home users but did not focus on the S&P aspects [141, 91, 16]. Second, `/r/homeautomation` has a large user base. Established in 2010, `/r/homeautomation` has about 1.6M members as of March 2022, much larger than other relevant subreddits, e.g., the second largest `r/homenetworking` (223K members) and the third largest subreddits `r/smarthome` (133K members) suggested by Reddit. We show a comparison of subreddits in Table 4.1. The comparison includes the top 10 subreddits ranked by their subscribers as of March 2022. `/r/homeautomation` is the most appropriate, because it has the largest user base and covers diverse products and integration levels. Note that while smart home S&P may also be discussed on other subreddits, we were interested in how S&P topics emerge organically from smart home-focused discussions. And we decided to focus on subreddits that

center around smart home topics (e.g., product information and adoption considerations). Third, `/r/homeautomation` has over 20 new threads per day, providing rich data to study.

Table 4.1: Comparison of smart home-related subreddits suggested by Reddit's engine.

Subreddit	Subscribers	Appropriate?	Explanation
r/homeautomation	**1,622,028**	Yes	
r/googlehome	587,788	No	Brand-specific
r/hue	227,287	No	Brand-specific
r/homenetworking	223,605	Yes	
r/homeassistant	166,711	No	Brand-specific
r/iota	146,161	No	Brand-specific
r/smarthome	133,135	Yes	
r/ubiquiti	124,902	No	Brand-specific
r/amazonecho	124,089	No	Brand-specific
r/homekit	119,568	No	Brand-specific

Dataset

Identifying discussion threads relevant to S&P from `/r/homeautomation` is non-trivial because of the massive amount of data and high diversity in the S&P terminology being used. Neither a manual nor a keyword-based approach is desirable due to the huge effort and lack of inclusiveness. Instead, we used a semi-automatic approach with a customized machine learning filter to increase the inclusiveness of data while easing human involvement. Our data collection and study received an 'exempt' determination from our Institutional Review Board. We downloaded data from `pushshift.io`, which maintains an up-to-date public archive for Reddit and complies with Reddit's terms of service in data collection and maintenance [225, 29]. `pushshift.io` ingests data through Reddit's official application programming interface (API) and handles removal requests,

although some removal requests may not be timely [224, 28]. Following prior work that used `pushshift.io` [167, 247], we neither de-anonymized users nor included sensitive data.

Initial corpus collection & cleaning

Leveraging `pushshift.io`, we downloaded all available comments on the `/r/homeautomation` subreddit between December 2010 and June 2021. By excluding threads and comments deleted or removed by the administrator or the user, the resulting corpus contains 46,637 threads with an average number of 12.72 comments per thread ($std = 19.98$).

Automated selection of candidate threads

To identify threads with S&P topics, we leveraged machine learning to process natural language text through fine-tuning a binary classification model to report a sentence's relevance to S&P on a pre-trained DeBERTa model [106]. We adapted it for our corpus and task since there is no perfectly sufficient off-the-shelf trained model.

Annotating the training data. Two authors, both experts in information and computer security, independently labeled 1,000 samples by their relevance to S&P-related topics. In annotation, we considered criteria adopted in multiple prior studies on S&P, such as the CMU taxonomy of Internet of Things S&P [84, 81, 83]. These aspects included data privacy, platform security, vulnerability, etc.

Evaluating sentence classification. We achieved a Cohen's Kappa κ of 0.92 from our annotation, showing high inter-coder agreement. We obtained 115 positive samples among the 1,000 coded samples. Then we randomly sampled 800 sentences for fine-tuning our machine learning filter and 200 samples for sentence testing. The model we trained attained a satisfactory F1 (micro-averaged) score of 0.965 on sentence classification.

Generating thread candidates. We considered a thread a candidate if the classifier labeled at least one sentence positive (i.e., related to security or privacy) within the thread. This strategy optimized for reducing the false-negative rate by capturing relevant discussions as much as possible. The same strategy will result in a higher false-positive rate as sentences might be taken out of context. However, we can still accommodate a non-low false positive rate as manual coding rules out false-positive threads. For validation, we sampled 50 threads from a pool, mixing an even number of positive and negative threads reported by our filter. The same annotators labeled these 50 threads ($\kappa = 0.82$) without prior knowledge about the prediction. As a result, we found 13 misclassifications among the 50 threads, with only one of them corresponding to a false negative, indicating that the classifier is unlikely to miss relevant threads. We opted to deal with false positives later on in the qualitative coding stage. In summary, the filter identified 7,255 candidate threads from the entire corpus.

Sampling

Following the guidelines in prior research, we randomly sampled and coded threads, from the period between 2010 and 2021, until we reached data saturation [250]. Threads not relating to smart home S&P were filtered out in the process. In total, we coded and reached saturation with 180 relevant threads (4,957 comments) among 303 random candidates sampled from all 7,255 threads.

Data Analysis

Our analysis considered all comments over time in each thread. First, two authors went through 28 threads to create the analytical memos that revealed initial codes. The research team discussed these codes, clarified definitions, resolved disagreements, and established an initial codebook.

Then, these two authors independently coded a subset of 15–20 threads randomly sampled from the dataset each time, while comparing codes and revising the codebook iteratively until high inter-rater reliability at the comment level was reached ($\kappa = 0.74$) at the 82nd thread. Using the revised codebook, the two authors then split the samples and coded independently until hitting saturation at the 180th thread. Then, we revisited all threads multiple times and conducted thematic analysis. We make our codebook available in Appendix A.1. From the 180 threads, we observed 2,181 users. Noticeably, 477 of them actively participated in S&P-related discussions.

Limitations

Our analysis has several limitations. First, there is sample selection bias as we collected data from `/r/homeautomation`. Presumably, users on this forum are more passionate and knowledgeable about smart homes than the general population. This is reflected in our observation, where many users demonstrated extensive knowledge of device functions and the associated S&P issues. Second, we did not have access to our sample's demographic data such as age, gender identity, education level, or occupation. So it is difficult to ascertain whether the demographic distribution of the sample is reflective of the general population. Future work may want to study how smart home S&P is discussed on other forums with different focuses or within other populations. Third, in our findings, we discuss various actions such as abandoning product ownership. As is common with self-reported behavior, users' discussions may not necessarily correlate with their actual actions. Fourth, our focus in this work is not the temporal relation between different threads, which potentially captures more dynamics of how users develop considerations and attitudes in a longer period of time. Finally, our methodology to detect S&P-related discussion using text classification is generalizable across different domains, e.g., Twitter, and future work may

leverage our content analysis and codebook. However, the quantitative results we report are less likely to generalize due to the Reddit population that is presumably more tech-savvy. Keeping these limitations in mind, our research still revealed significant trends in S&P discussions in a previously unstudied population, and it fills a gap in the literature about users' S&P considerations and attitudes.

Roadmap

In the following sections, we present the findings in correspondence with our research questions. Figure 4.2 shows our analysis framework with an example. We first reveal users' S&P considerations in Section 4.2 – how they assess their S&P concerns and incorporate S&P protective strategies – given the contextual factors, such as adoption phases and product factors. Then, in Section 4.3 we map the considerations to five major categories of S&P attitudes during smart home product adoption, namely dismissiveness, exploration, resignation, positive pragmatism, and devotion. Lastly, we discuss how online discourse influences users' S&P considerations and attitudes in Section 4.4. In addition, we show the prevalence and co-occurrence of themes and subthemes regarding the three research questions from Figure 4.3 to 4.8 to support our qualitative findings.

In the example of Figure 4.2, Alex started a thread to seek advice on buying a robot vacuum with potential privacy concerns due to its foreign manufacturer. Bob declared that they were aware but not concerned about privacy of the vacuum personally. Charlie offered an alternative to acquiring the local version that does not connect to the foreign server. Following Charlie, Alex confirmed their purchase of the local version and further sought advice to balance its privacy tradeoff with functionality. During both pre- and post-purchase, Alex showed considerations of privacy concerns and protective strategies, demonstrating an exploration attitude; Bob dismissed their concern.

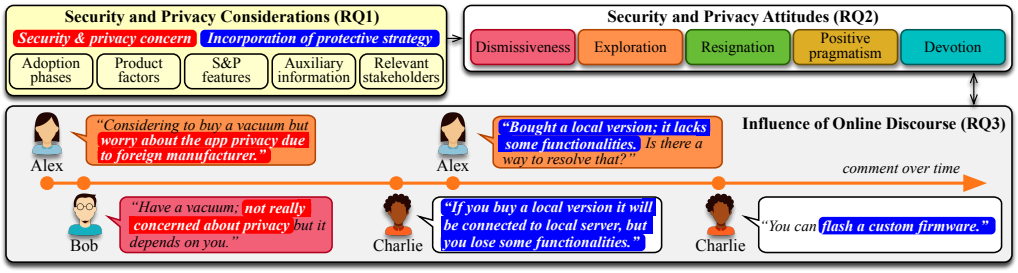


Figure 4.2: Our analytical framework with a paraphrased Reddit thread as an example. The texts that show either considerations of concern and protective strategies are color-coded with red and purple. The text box's color aligns with the user's S&P attitude in discourse, if exists.

4.2 RQ1: Security and Privacy Considerations

Our analysis of `/r/homeautomation` reveals two types of user considerations: (1) their S&P concerns regarding smart home technologies and (2) how they incorporate protective strategies during adoption. We observe that a set of interplaying contextual factors shape these considerations, including adoption phases and product factors. Next, we describe the contextual factors, explain how users consider S&P issues, and show how users incorporate S&P-protective strategies in smart home adoption. We note, however, that users do not necessarily fully develop their considerations during discussion or may only exhibit a subset of them in their comments.

Contextual Factors

During the discussion, users reference a range of factors affecting their considerations. Our thematic analysis reveals five themes of contextual factors: adoption phases, product factors, S&P features, auxiliary information, and relevant stakeholders.

Adoption phases. Users' S&P considerations evolve during the discussion

according to their adoption phase for a product. Consistent with prior work [84, 62, 199], we observe four major phases: *consideration of product acquisition*, e.g., purchase, inheriting, and sharing; *acquisition of the product but not in use*; *active use*, during which users may personalize the product; and *abandonment or transfer of product ownership*.

Product factors. Product-related factors during the discussion lead to users' awareness of S&P issues. We observe that users reference two kinds of product factors: *quality requirements* and *technology features*. Product quality requirements include compatibility, reliability, price, customization, and core functionality. Examples of the technology features of products include (open-source) software, cloud dependency, connectivity, user control, and data storage.

Security & privacy features. Users also reference specific S&P features of the product. These features fall into five categories: *account access* (e.g., resource authorization), *safety measures* (e.g., data backup), *system integration* (e.g., exposure to network), *privacy options* (e.g., rights to review, edit, and delete their data), and *security features* (e.g., encryption).

Auxiliary information. Users leverage auxiliary information about S&P aspects throughout their adoption journey. We identified two types of information sources: *public information channels* and *evidence from real-life interaction*. The former covers news and reports, social media, customer reviews, or privacy policies. The latter includes experiences of suspicious activities or communications with customer support.

Relevant stakeholders. In addition to external attackers, users recognize different stakeholders in the smart home ecosystem. These stakeholders include *companies* (manufacturers, vendors, and service providers), *governments*, *users*, and other *third parties*. Users associate stakeholders with different roles. For example, the government can serve as a regulator or a possible adversary. Similarly, third parties can provide compliance oversight or impose threats. Lastly, users discuss sharing devices in multi-user

smart home scenarios, with attention to special populations, e.g., children and the elderly.

Developing Security and Privacy Concerns

The first component of S&P considerations is developing S&P concerns. As users develop S&P concerns, they perform threat modeling that consists of three themes as shown in Figure 4.3: *security and privacy awareness* – recognition of potential concern, *threat identification* – mapping of potential to actual threats based on smart home products and associated stakeholders, and *risk assessment* – assessment of the likelihood and severity of threat influences. Next, we elaborate on the three themes and the subsequent subthemes; Figure 4.3 depicts the frequency of each subtheme in the coded threads.

Security and privacy awareness

The first theme describes how users' needs, the adoption phases, and information sources drive their S&P awareness. It includes two subthemes: how contextual factors contribute to awareness and how awareness evolves according to contextual factors. Figure 4.3 shows a noticeable contribution of contextual factors to S&P awareness, compared to other subthemes.

Contextual factors contribute to awareness. Users' S&P awareness arises from the smart home device features they deploy to address their needs. Users associate specific concerns (eavesdropping, spying, safety hazard, etc.) with distinctive product modalities, such as audio recording by voice assistants or room scanning by a robot vacuum. For example, one user expressed concern about their smart lock being tampered via *"the digital part than the actual deadbolt"* (U6-T33). Many concerns center around devices' dependency on the Internet or cloud to function, e.g., trigger-action services through MQTT, which may possibly *"expose something on your*

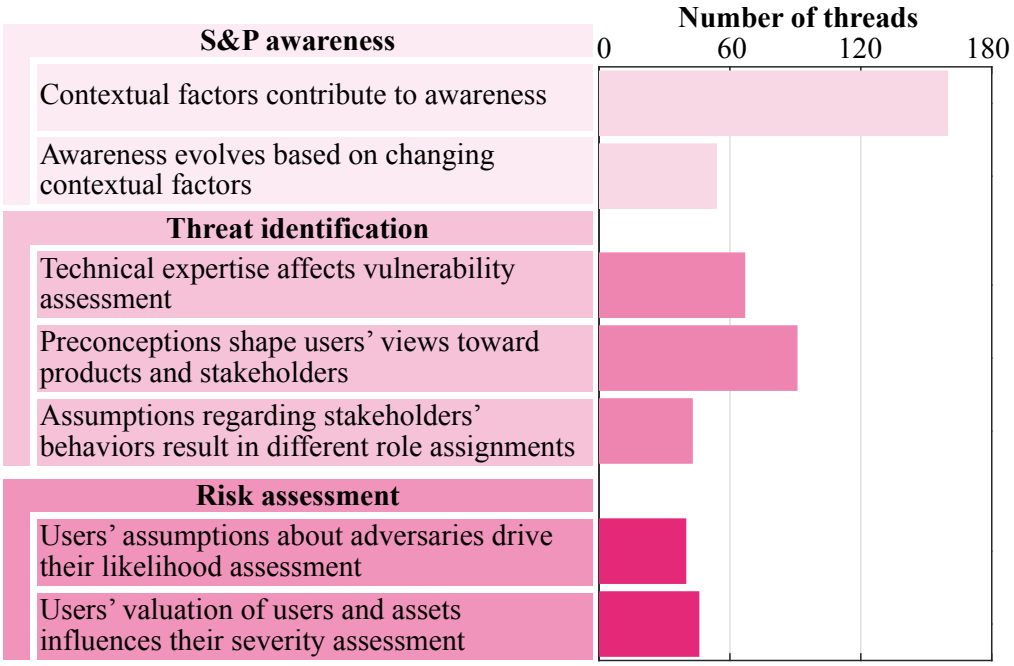


Figure 4.3: Three themes and the frequencies of seven subthemes of users' considerations in developing security and privacy concerns. Note that when we refer to "contextual factors" in a subtheme, it indicates a list of items, as we explain in Section 4.2, such as S&P features and auxiliary information.

home network to the internet" (U2-T9). Also, S&P information contributes to users' awareness. For example, one user was concerned about the *"(lack of) privacy policy"* prior to purchasing a product (U40-T29). Additionally, awareness arises from users' specific use cases, such as remotely allowing tenants to enter a rental home via smart entrance by giving them *"a one use access code to dial on pad"* (U1-T155).

Awareness evolves based on changing contextual factors. Users interact with stakeholders, products, and information sources across adoption phases, prompting an evolution of S&P awareness. We find that before acquiring a new product, users have abstract S&P awareness; they are

more concerned about product features. The awareness becomes more concrete after the purchase, when users describe specific concerns about the product. In one case, even before receiving the purchased device, a user developed a concern about a phone call they received about the product due to *"the information [shipping location, credit card, etc.]"* the caller asked for (U1-T8).

After acquiring the product, user awareness becomes more specific to their experience with the device or stakeholders. The example below shows that the user returned a thermostat after they had noticed its lack of privacy options:

> *"Yeah i had gotten a discounted nest thermostat from my power company but after seeing their reluctance to let me access *my own data* i returned it."* (U5-T69)

During adoption, auxiliary S&P information, such as media reports, further contributes to a user's awareness, as evident in one user's complaint about a company's response to a threat:

> *"8 months ago Ring's VP said he'd come back here to tell us when their firmware stopped sending data to China. He hasn't commented since."* (U1-T139)

Awareness about S&P issues persists post abandonment. For example, when transferring their device to others, one user recommended to *"wipe everything from my [their] accounts, setup new accounts...and hand it over"* (U4-T23).

Threat identification

The second theme of developing S&P concerns covers users identifying specific threats in their smart home, such as data misuse, unwanted data collection, and surveillance. This theme consists of three subthemes. First,

users' technical expertise affects their vulnerability assessment. Second, their preconceptions and assumptions – preconceptions are subjective while assumptions are more situational – of stakeholders shape how they define the adversary. Third, users assess the vulnerability in smart homes and place stakeholders as adversaries, victims, or good Samaritans.

Technical expertise affects vulnerability assessment. Although some users name specific attacks, several exhibit uncertainty regarding how particular products or technologies are vulnerable. For example, when discussing buffer overflow attacks on device firmware, one commentator confused the device firmware with the wireless protocol it employs:

> *"I assume Z-Wave doesn't suffer from this problem due to the certification process? Or are there attack vectors that could be leveraged against that particular tech?"* (U4-T133)

Further, insufficient technical understanding manifests in over- or under-estimation of the threat. One user described any device requiring an Internet connection as insecure when comparing products that rely on WiFi connectivity versus those on Zigbee and Z-Wave. In the same thread, another user clarified the threat model, comparing WiFi versus Zigbee and Z-Wave devices, that the latter are *"not IP routable so they are much more difficult to use as attack vectors"* (U19-T79).

Preconceptions shape users' views toward products and stakeholders. Users carry preconceptions, e.g., perceived trust, reputation, or reliability, which they *"hearken from the bad old days"* and reflect in their views toward products or stakeholders (U3-T51). These preconceptions can arise from previous experience with a stakeholder. For example, when considering a replacement purchase, one user warned others *"DO NOT buy a Skybell* [*smart doorbell they want to replace*]*"* by referencing their concern about the security and reliability of a smart doorbell from their prior use and issues *"that are well documented across Reddit"* (U1-T11).

Further, users exhibit differing trust and reputation preconceptions for the same stakeholder; for example, one thought *"Apple's reputation for privacy far outweighs Amazon's"*, while another was more positive about Amazon because *"historically speaking, Amazon is much more protective of user data* [*compared to Nest*]*"* (U2-T81, U30-T29).

However, we observe a predominant distrust toward Chinese smart home companies, not just particular brands, on privacy and quality. For instance, one user argued:

> *"Not specifically for Xiaoyi, but it is quite common for low-priced Chinese brands to have embedded backdoors or privacy-invading snooping by the company."* (U15-T58)

This distrust possibly arises from the perceived tie between Chinese manufacturers and authorities, sometimes described as the *"Chinese state overlords"* (U17-T119).

Assumptions regarding stakeholders' behaviors result in different role assignments. Users have different assumptions about the behaviors of stakeholders or products (such as security practices). These assumptions result in different assignments of adversarial roles to stakeholders. Users' assumptions also do not necessarily align with their views toward the stakeholder, as shown in the three distinct outcomes below.

First, users associate the for-profit business model with more vulnerable products because their manufacturers *"have zero incentives to patch holes in their older products"* (U6-T3). Second, users consider these companies as an adversary because of the *"blatant disregard for the moral and ethical responsibility that companies should have when they have access to such sensitive data"* (U16-T58). Third, the same for-profit business model, interestingly, leads to users assuming companies to be good Samaritans because enforcing data privacy improves their competitiveness:

> *"Well Google records all your data, but they are highly incentivized to keep it safe and not sell it, because having exclusive access to it is their core business."* (U97-T115)

Similarly, for authorities, some users note their positive role in regulating companies. Below we show an example:

> *"There are laws that vary depending on region, such as Europe's GDPR privacy laws, hence why Facebook is acting like they would pull out of Europe."* (U15-T151)

Others are, however, skeptical about governments' role in regulating smart home companies. One user thought that, because governments are not doing *"what's right about trade," "it's up to the consumer to be smart."* Another user sarcastically responded: *"and... we* [*users*]*'re doomed"* (U7-T147, U5-T147).

Risk assessment

After identifying S&P threats, the third theme involves evaluating the associated risks, i.e., how these threats might affect users' health, finances, and physical and digital assets. This assessment consists of two subthemes: the evaluation of the likelihood and severity of a threat.

Users' assumptions about adversaries drive their likelihood assessment. Users associate the likelihood of attacks with the required technical sophistication. For example, one user assumed that attackers can use a cheap radio to jam a wireless security system that does not employ frequency hopping:

> *"I can disable the system with a walkie talkie after using an SDR* [*software defined radio*] *to find the exact frequency the system is on and just blast it the entire time I'm in your house ...$20 baofeng* [*radio*] *will kill simplisafe since it uses 433*[*MHz*]*."* (U4-T45)

On the other hand, users doubt the likelihood of attacks that appear resource-consuming. For example, an attacker is unlikely *"sitting outside my [their] house for the next month trying to guess"* an 8-digit smart lock code from fingerprint dusting (U3-T100). Some users think certain attacks are unlikely since they bring low benefit to the attackers, e.g., attacking a smart vacuum to reveal *"how dirty your carpet is"* (U3-T125).

Users' valuation of users and assets influences their severity assessment. Even with an adequate threat model, users have different valuations of the associated risks. For instance, while all recognized the threat from a compromised voice assistant, some were concerned about sensitive conversations being recorded by the device when working from home:

> *"I need to deal with sensitive HR [Human Resources] issues from home, all of which should never be recorded without consent of a third party."* (U8-T68)

Similarly, other users elevated perceived risks when special populations interact with devices, due to severe consequences that *"[the elderly] can't escape 99 degree or higher heat"* (U1-T3).

However, others devalued the severity of recorded voice, e.g., chatting with family members, as they thought *"nothing I say in my home is important that I worry someone heard"* (U1-T68).

In contrast, users may assign possible attacks with lower severity based on the countermeasures in place. For example, one user felt their home would be safe as their security system *"is basically tamper proof"* (U46-T30). In that case, the device employed power backup and intrusion alarm against malicious power shutoff and wireless jamming.

Incorporating Protective Strategies

The second component of S&P considerations is users examining whether and how to incorporate S&P protective strategies into their smart home

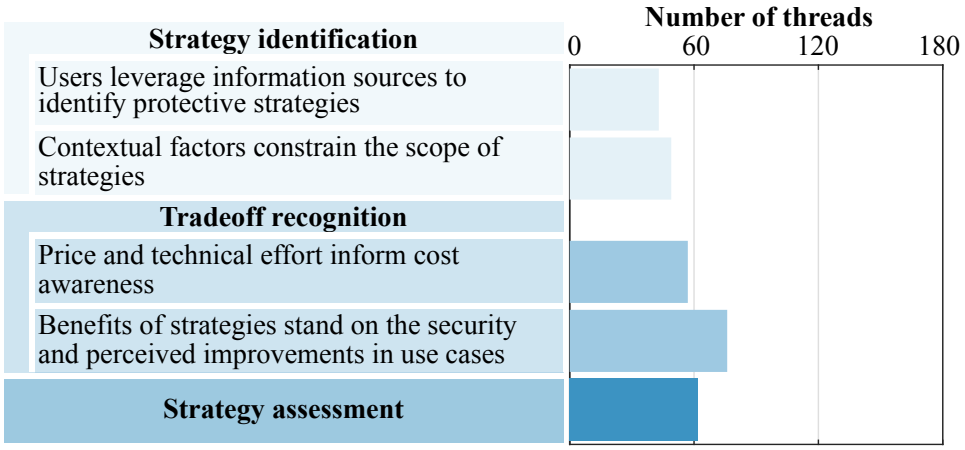


Figure 4.4: Three themes and the frequencies of five subthemes of users' considerations in examining protective strategies.

deployment. We observe two types of protective strategies. The first represents adoption decisions such as the purchase or abandonment of a specific product. The second includes product setup, customization, or configuration such as changing passwords, disabling the Internet, and DIY solutions. We identify three themes of how users arrive at protective strategies: *strategy identification*, *tradeoff recognition*, and *strategy assessment*. First, users identify potential strategies to alleviate S&P concerns (if any), given the adoption context. Second, users recognize the tradeoffs associated with the identified strategies. Last, users assess whether to incorporate the S&P protective strategies after assessing the identified tradeoffs. Figure 4.4 shows the five subthemes comprising the three themes; the distribution of the subthemes is relatively uniform in the coded threads.

Strategy identification

We identify two subthemes for strategy identification. First, users leverage available information sources to explore possible protective strategies. Second, facing the constraints from contextual factors, users narrow the scope of their strategies.

Users leverage information sources to identify protective strategies. Users leverage their knowledge and understanding to explore protective strategies. They build their knowledge or understanding from access to information sources, e.g., online discussions. For example, one user opened a poll for *"best practice advice"* as they sought to secure their smart home against external threats from the Internet (U1-T139). Additionally, users learn about possible strategies from auxiliary information. For instance, a user referenced an online open-source project about rooting a robot vacuum's firmware to alleviate a privacy concern (U3-T73). Stakeholders, including third-party organizations, represent another source of information about possible strategies. In one case, a user praised *"Consumer Reports"* for their *"good write up"* on alternatives for secure smart locks (U3-T100).

Contextual factors constrain the scope of strategies. First, stakeholders, such as companies or governments, might restrict the scope of possible strategies for marketing or legal reasons. For example, returning a product because of a security issue might be infeasible due to a restrictive return policy. In one case, a user asked if they should return a smart lock, within the return period, in response to a warning that indicates it has *"weaker security"* (U1-T122). Additionally, some products lack S&P configurations, forcing users to consider "all-or-nothing" strategies. In other cases, users were unaware of such configurations because they were inaccessible. One comment referred to an interface, which disables cloud access of a smart bulb, as *"*[*buried*] *in the app"* (U4-T72).

Second, being a secondary user limits available protective strategies, e.g., one user expressed a loss of control of using a voice assistant with *"3rd party always-on microphones"* in a hotel (U4-T20). Third, inadequate understanding or limited information contributes to constraining the scope of protective strategies. For example, one user mistakenly referenced the U.S. Federal Communications Commission (FCC)'s terms to question the

legitimacy of deploying WiFi access control by deauthentication flood in enterprise settings (U8-T128).

Tradeoff recognition

We observe two subthemes in how users recognize tradeoffs associated with the identified strategies. First, the price and technical effort of incorporating a protective strategy inform their cost awareness. Second, users identify the benefits of strategies based on perceived improvements, e.g., enhanced security.

Price and technical effort inform cost awareness. First, users reference the monetary value of a strategy as part of its cost. For example, when a user was wary about buying a security camera from a foreign manufacturer, a user responded:

> *"What's your budget? If you want the best Axis cameras are the most reliable cameras..."* (U12-T75)

Second, users evaluate how the technical effort associated with the customization strategies contributes to the cost. One Reddit user mentioned the technical knowledge and effort needed for a DIY alternative to commercial voice assistants:

> *"Actually if you know python writing a voice assistant capable of controlling your house and doing other basic functions is a matter of a couple days."* (U21-T29)

Benefits of strategies stand on the security and perceived improvements in use cases. The benefits of incorporating a strategy depend on users' use cases, e.g., their smart home setup or feature requirements. Users prefer compatibility between S&P improvements, brought by the strategy, and other desired qualities in their use cases. For example, a potential buyer

wanted a *"*very* secure lock"* that is compatible with other functions of HomeKit, Apple's smart home app (U9-T5).

However, users might not be fully aware of the S&P benefits of a strategy, e.g., due to its deployment. For example, one user sought confirmation about whether their exposed information can be limited to only *"details of the separate network, address, and room layout"* if they associate their smart vacuum with a dedicated phone on a separate network (U1-T90).

Strategy assessment

Finally, users assess the tradeoffs between cost and benefits to determine which strategies are viable. Users incorporate strategies when they think the benefits outweigh the cost, e.g., one thought a product that was a *"little pricey"* but *"worth it"* (U4-T18). The user's tradeoff assessment tilted toward the strategy when the use case is important to them. For example, one user preferred to purchase a more secure smart thermostat when the primary user is an older adult: *"[the elderly]'re unable to set their own thermostat, and their caretaker isn't close by to address the situation"* (U1-T3). Some users value the power of customization; even when they perceive a cost from the needed effort to incorporate, the added controls outweigh these costs:

> *"I agree that Homeseer's interface is not the prettiest one around, but so far it does far and away more, is stable, and exposes the "nerd knobs" needed to do darn near anything including secure z-wave."* (U9-T1)

Benefit valuation depends on the user's assumptions and understanding of the threat. Several users perceived an added security benefit when they cover their internal IP in a posted photo as they thought there was *"no reason to give our potentially vulnerable information,"* and *"if they compromise*

a router, internal IPs could be useful" (U1-T120, U8-T120). But others considered covering *"non internet routable"* IPs as *"diminishing returns"* since if the router is compromised, these IPs can be easily revealed by port scanning (U7-T120, U10-T120).

When they perceive the cost as outweighing the benefits, users are less motivated to incorporate the strategy. Cost manifesting in a high technical effort can dissuade them from a protective strategy, e.g., when they consider themselves *"not* [*technically*] *qualified"* (U3-T73). An extreme case is fatalism [313], where every protective strategy is perceived as ineffective since *"everything can be broken"* (U10-T100).

Takeaway-RQ1. We observe that smart home users' S&P considerations consist of S&P concerns and protective strategies. First, users develop S&P concerns through in-depth and tech-involved threat modeling: becoming aware of the S&P issues, identifying and assessing the actual S&P risks and threats. Second, users identify and assess S&P protective strategies while recognizing cost-benefit tradeoffs. As such, these considerations are multi-dimensional and depend on an interplay of contextual factors in adoption, as shown in Figures 4.3 and 4.4. We also find that users' considerations progress according to changing contextual factors, e.g., adoption phases and access to information. These takeaways contrast with prior findings, where users' assessment of threats and protective strategies can be single-dimensional and static [101, 320, 328].

4.3 RQ2: Security and Privacy Attitudes

We now discuss users' S&P attitudes toward the adoption of smart home products. These attitudes are the result of users' (1) S&P concerns and (2) incorporation of protective strategies. As shown in Figure 4.5, we identify five categories of attitudes: *dismissiveness* of S&P concerns; *exploration* of possible concerns and protective strategies; *resignation* to incorporating

S&P protective strategies; *positive pragmatism* in terms of incorporating protective strategies that balance cost and benefit tradeoffs; and *devotion* to incorporating protective strategies. We observe that one user may exhibit different S&P attitudes as the context varies; users' attitudes may evolve over time during exploration as they advance their considerations with the progressing context or better knowledge of it. As such, we do not segment users based on attitudes. Also, some boundaries between attitudes are blurry (e.g., some pragmatic traits shared in resignation), and users may not fully express their attitudes in online discussion. We complement qualitative insights with frequencies of attitudes to characterize their prevalence. The frequencies in Figure 4.5 characterize users on this subreddit: many are technically informed and devoted to incorporating protective strategies. Figure 4.6 shows the co-occurrence of users' attitudes and their considerations.

Dismissiveness as a contextual or general attitude. The dismissive attitude refers to users who exhibit low S&P concerns; this attitude leads to users' reluctance to further incorporate protective measures. First, some users exhibit low S&P concerns in general, sometimes regarded as *"wilfull* [*willful*] *ignorance"* by others (U12-T20). For example, some users were not concerned about privacy in the smart home because they felt *"privacy honestly isn't of utmost importance to all"* or they *"got nothing to hide"* (U18-T26, U28-T115). Interestingly, these users may rationalize their attitudes by making an analogy with physical privacy:

> *"Personally I'd go with something like, having a private conversation in public and notice the person a table over eavesdropping but don't really care."* (U9-T68)

Second, while some users have general S&P concerns, they might be less concerned about specific use cases. For example, one rationally dismissed their concern *"through some troubleshooting"* after they had determined that

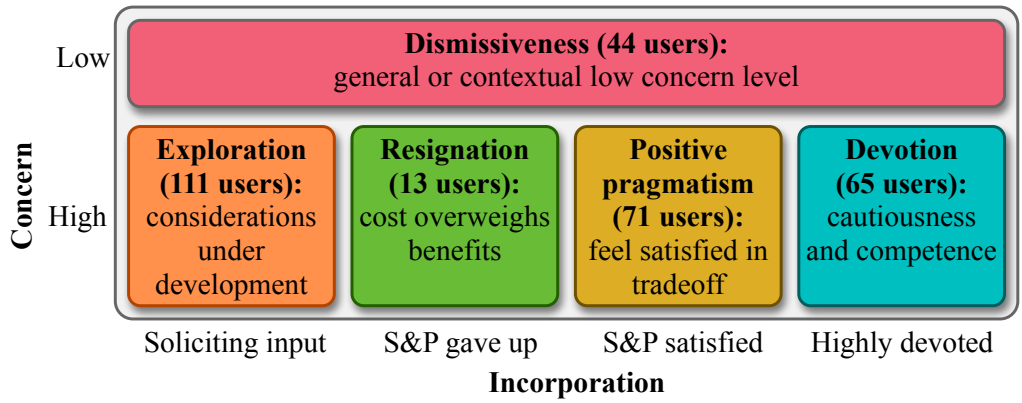


Figure 4.5: Users' S&P attitudes in adoption aligned with considerations of concerns and incorporating protective strategies. Representative traits are summarized with each category, and we report the frequency of each attitude among the 255 out of 477 users who revealed their attitudes in S&P-related discussion. Note that one user may hold more than one attitude. We observe more users carrying exploration attitudes than others. Consistent with prior findings [149, 77], we see many pragmatism users. The users devoted to incorporating protective strategies are noticeable, too.

the suspicious device traffic was due to repeated attempts of a firmware update (U5-T110). When dismissing a concern, users may appear to be inconsistent, e.g., by saying *"privacy is a big concern, and so is use case,"* when they were not *"personally worried about the privacy implications"* of smart speakers (U4-T180).

Exploration as an attitude while developing considerations. The exploration attitude features users' proactive needs to develop their understanding of S&P threats and protective strategies. Users may exhibit the exploration attitude at different adoption phases, depending on their evolving awareness. Users solicit input from others to educate themselves about possible S&P threats. For instance, one user was open to learning more about integration vulnerability and threat:

"Network security definitely isn't my strong suit, although it is a

> *priority of mine ... I'd definitely like to hear what other people have to say."* (U2-T34)

Others explore additional protective strategies; one user asked for help on Reddit when they noticed a mismatch in the password length limits between their WiFi and a smart AC control:

> *"Have you just shorten your WiFi password, returned the devices and/or find any other solution? What password length do you find both secure enough and compatible with almost any connected device?"* (U1-T41)

After exploration, a user's attitude may change, such as dismissing S&P concerns after troubleshooting (U5-T110).

Resignation as a result of cost outweighing security or privacy benefits. The resignation attitude highlights users who worry about S&P issues but tend to give up on incorporating protective strategies. This resignation results from their perception that the cost of protective strategies outweighs any potential S&P benefits. For instance, one user regarded S&P practices as *"too much of a hassle,"* when they had to include network separation that impacts connectivity (U7-T35). Similarly, another thought price and device availability contributed to their resignation as *"there's not a better option* [*of a smart door bell*]*"* compared to other brands (U41-T17).

Another motivation behind resignation is users' fatalism, previously mentioned in Section 4.2. They consider S&P threats inevitable due to the loss of control [65]. Such beliefs appear to be a result of users' continuous exposure to tracking, e.g., due to people *"walking around with a cellphone 24/7"* (U6-T38).

Positive pragmatism results from satisfactory tradeoffs. Similar to resignation, this attitude features the tradeoffs users make when considering protective strategies. The difference is that users value S&P more than the other factors. They feel satisfied with protective strategies that strike a

compromise between cost and benefits. For example, one voice assistant user *"values convenience over complete privacy"*; they thought opting out of data sharing suffices to protect privacy (U6-T68).

While pragmatic users can be technically competent, they may seek protective strategies with lower costs. One user, who *"program[s] by day,"* felt comfortable with setting up a smart home with network separation rather than setting up a secure smart home by DIY, as *"the last thing I [the user] want to do on the weekend is fiddle with Raspberry Pi"* (U4-T29).

Devotion as a result of caution and competence. Users with high S&P concerns tend to be devoted to S&P-protective strategies. While some users' devotion stems from their overreaction to an incomplete threat model, others are technically competent and enthusiastic about sharing their knowledge.

Devoted users thoroughly form their threat model, even by decompiling an app to examine its encryption standard:

> *"So far I've decompiled the app and found that it uses Ayla Networks IoT platform. Oh and the Cipher Suite for added security."* (U11-T121)

Others form their threat models from their prior experience; one user referenced their job when expressing security concerns about unpatched hubs:

> *"I once worked at a company with 8M customers. My mantra regarding problems was '1 in a million happens every 3 hours' "* (U1-T133)

Some devoted users are capable of incorporating sophisticated mitigation strategies, e.g., multi-layer security:

> *"I still have fallback if somethings fail so i'm never without some degree of protection. Personally I would and have layered the devices*

in 3 layers, A 'real' security panel for the vital parts where a burglar must pass..." (U3-T45)

However, devotion does not mean blindly adopting a more technically involved strategy. For example, one user concerned about unpatched hubs showed fatigue in reacting to vulnerabilities, e.g., by patching. They would rather invest in other brands if they were *"going to guess at the next target"* (U1-T133).

Takeaway-RQ2. We map the S&P considerations to five categories of attitudes; each attitude combines the user's degree of S&P concern and level of incorporating protective strategies (Figures 4.5 and 4.6). We observe that users' S&P attitudes are context-dependent and evolve according to the progression of considerations. Users do not hold a fixed S&P attitude; their attitudes change depending on the context (e.g., product factors and adoption phase), as many of them proactively seek and gain more information. Also, prior experiences or preconceptions about a product might shape a user's attitude, overriding a more objective assessment. These findings enrich literature where the focus has been traditionally on users' static S&P attitudes [149, 77].

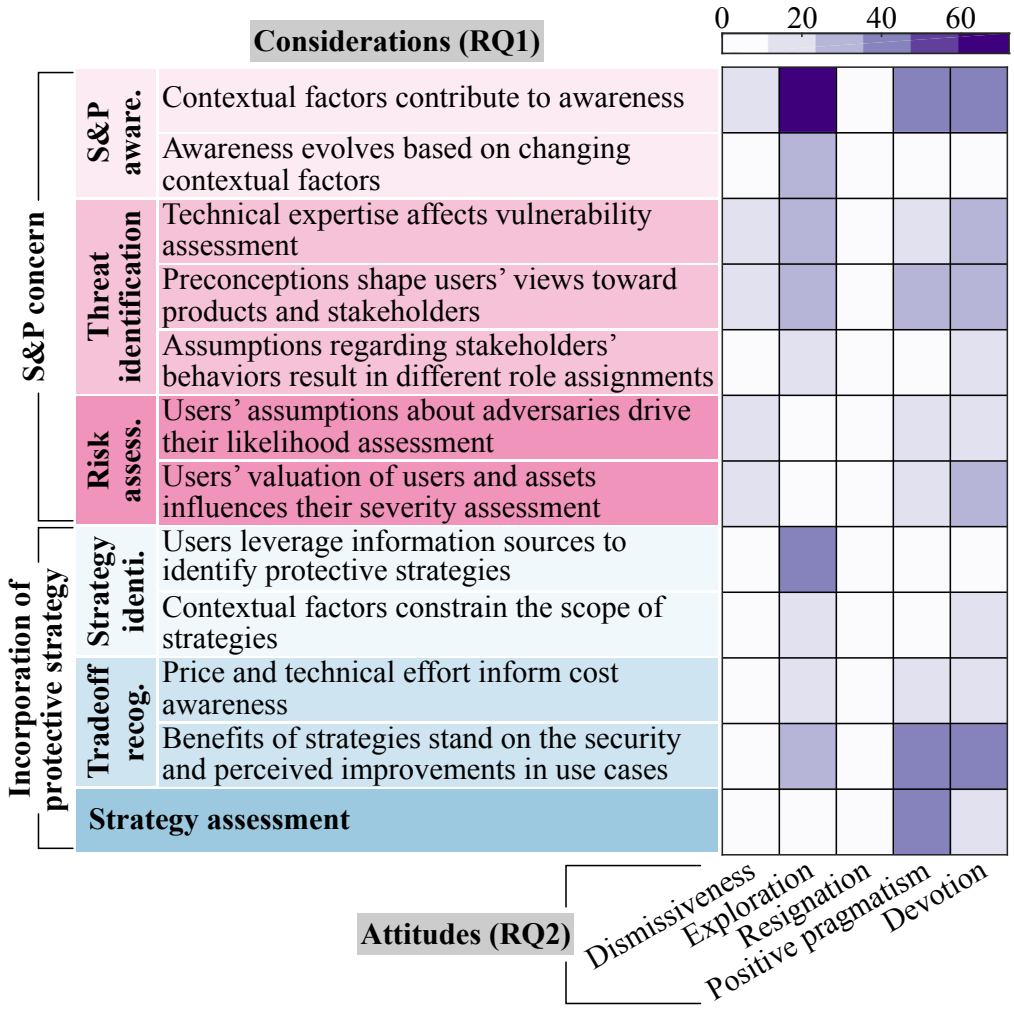


Figure 4.6: Co-occurrence of users' S&P attitudes with subthemes in considerations, which supports our attitude mapping. For example, users with dismissiveness attitudes rarely consider protective strategies; users demonstrate their devotion and pragmatism by the final assessment of concerns and protective strategies; users seldom reach the final assessment during exploration.

4.4 RQ3: Influence of Online Discourse

Discourse in an online forum fulfills users' information and social needs despite their varying attitudes. It also fosters users' development of considerations and attitudes. We identify three themes of active interactions in discussing smart home S&P-related topics. These themes are: users' *strategies to resolve ambiguity* in S&P-related discourse; *contributions of the discourse to users' attitude development*; and *the influence of users' varying attitudes on the discourse environment*. We identify seven subthemes, e.g., collaborative exploration through exploration. We show each subtheme's frequency and the co-occurrence of users' attitudes and these subthemes in Figures 4.7 and 4.8. These figures support our qualitative findings, e.g., the high occurrence of users with devotion attitudes informing others of alternative strategies, which contradicts prior findings that S&P fundamentalists are reluctant to help [77].

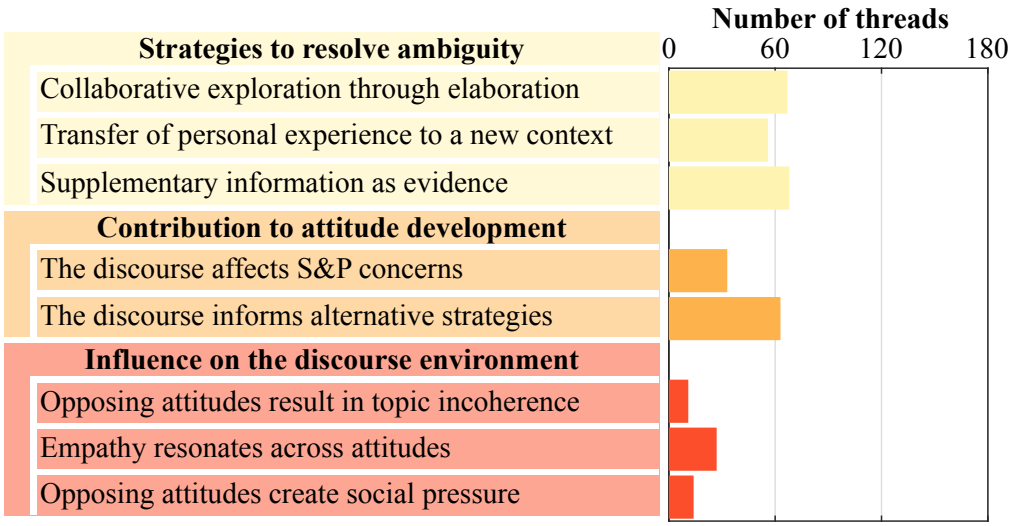


Figure 4.7: Three themes and the frequencies of eight subthemes of the online discourse's influences.

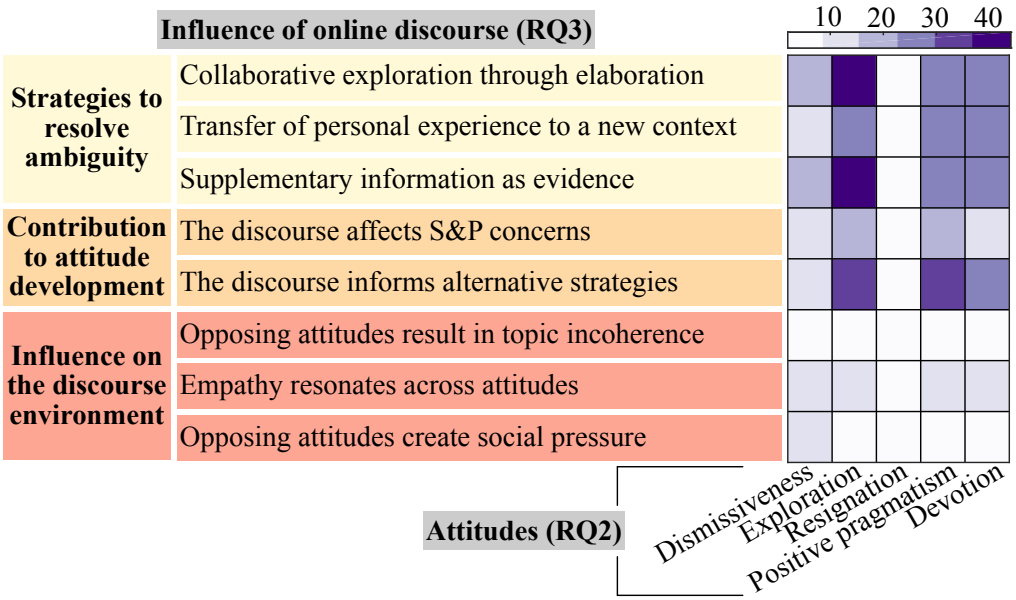


Figure 4.8: Co-occurrence of users' S&P attitudes with the subthemes in discourse influences per thread. The statistics highlight the participation of users carrying exploration, positive pragmatism, and devotion in resolving ambiguities for S&P-related discussion and their active contribution to attitude development, e.g., informing alternative strategies. In contrast to prior findings that S&P fundamentalists are reluctant to help [77], we observe users of high technical competence (devotion) proactively support others.

Strategies to Resolve Ambiguity

In this theme, we describe how users resolve problems and confusion collectively.

Collaborative exploration through elaboration. We find that users seek input to complete their understanding of S&P threats and better evaluate protective strategies. They try to resolve ambiguities through iterative elaboration together, which helps them better understand the sources of

others' confusion. For example, a user was skeptical of others' concerns about a smart remote and asked for elaboration: *"What are they recording me with? It's an IR blaster"* (U42-T170). After observing concerns about *"someone could tunnel in and eavesdrop on cameras,"* U42-T170 elaborated that this issue may not represent a threat as the device uses a local API.

Transfer of personal experience to a new context. Users become aware of others' backgrounds and contexts when exploring a problem together. So, they transfer personal experiences and understandings to support other people in new smart home contexts. For example, one user suggested blocking internal IPs based on their experience with a router. However, they were aware that this suggestion might not apply to owners of a robot vacuum, as it can hinder the vacuum's functionality.

> *"In my TPLink router, I have an option to block internal IP addresses from accessing the internet. Would it still work at all if it had NO connection?"* (U4-T125)

Supplementary information as evidence. During collaborative exploration, users often use supplementary information to strengthen their arguments. In a thread in which multiple users discussed their concerns about a smart lock's vulnerability, one user cited news about a product update that had potentially patched the vulnerability: *"I believe this was resolved in 2016 with a new version of smart key.* [*url of the news*]*"* (U6-T10).

We see other patterns to supplement information, e.g., forum moderators pinning a *"good discussion"* (U4-T139) about S&P for more visibility. This effort, however, seems ad-hoc, and users still have difficulty navigating S&P information on Reddit. For example, one user reused advice across similar threads about networking protocols and security since it is *"a recurring topic and reddit churns so much that reuse makes sense"* (U3-T51).

Contribution to Attitude Development

Above, we discussed users' strategies to collectively resolve ambiguity in the discourse related to S&P. Here, we present how users' attitudes change as a result.

The discourse affects S&P concerns. Users' discourse with others makes them more aware of S&P risks (Section 4.2), leading them to revisit their attitudes. For example, one user changed their attitude, from exploration to dismissiveness, about a presumed "credit card scam":

> *"Edit: You're right, I wasn't scammed, but if it is a scam, someone else could easily fall for this."* (U1-T8)

However, discourses do not always change opposing attitudes. In a previous example (Section 4.2), the intensive debate about the legitimacy of performing a deauthentication flood for access control did not change either of the parties' attitudes. One user described the others as *"knowing participant in a criminal conspiracy"* and rejected their interpretation of evidence from multiple news or legal documents (U5-T128).

The discourse informs alternative strategies. As discussed in Section 4.2, users inform others of S&P protective strategies. Rather than replicating the advice, users take inspiration from the collective wisdom and develop new strategies. For example, being inspired by others' advice on how to transfer smart device ownership securely, a user took a hybrid approach, moving from the exploration to the pragmatist attitude:

> *"UPDATE: I took a hybrid approach. I setup a gmail account for the house and moved all accounts to it. I removed one of my Wink Relays..."* (U1-T23)

A user, who had almost *"abandoned all hope"* in flashing custom vacuum firmware for privacy, benefited from discourse and posted their updates

for *"those who might search the forum"* in need (U1-T89). This example shows a user changing their attitude from the resigned to the pragmatist category.

Influence on the Discourse Environment

Facing the complex topics of smart home S&P in conjunction with other technical and personal issues, users hold various considerations and attitudes in the discourse. We observe that the sentiment created by users' consensus and disagreement in the discourse influences the discourse environment.

Opposing attitudes result in topic incoherence. When S&P discussions intertwine with other topics, users tend to go off-topic, especially when they have opposing attitudes. In one thread, two users discussed how security systems help thwart the adversary. One user questioned the security improvements of smart cameras compared to an alarm system because it is not temper proof, as *"simply disconnecting your cable will render your comms* [*communications*] *useless"* (U4-T16). The discourse then drifted to presumption about personal stances:

> *"We will just have to agree to disagree. I understand your profession as a security system employee is threatened by home automation, and rightfully so."* (U2-T16)

Empathy resonates across attitudes. Some users show empathy when bridging the gaps between different attitudes. Users who share similar considerations are more likely to create empathy and resonance. For example, many users shared similar complaints about voice assistants being activated by kids:

> *"... our son managed to get the Google Home Mini to recognize his 'HEY GOOGLE!!' this morning. People with small children and voice recognition.... Help?"* (U1-T78)

Resonance also spans different attitudes, such as sharing a negative view of companies. For example, one user, who fell into the devotion category, did not *"trust Google, Apple, or the NSA with my* [*their*] *naked pictures."* Similarly, another user, who exhibited resignation, thought *"everything is controlled by the megacorps of modern society"* (U11-T137, U3-T137).

Opposing attitudes create social pressure. There is a sense of social pressure created when users of different attitudes interact. Users with dismissive attitudes may view those who worry about WiFi thermostats as being *"paranoid"* (U15-T136). In another case, two users reinforced their stance on surveillance issues when discussing smart speaker deals.

> *"right?! anyone skeptical of government surveillance can GTFO my life!!"* (U8-T55)

On the opposite, when defending their view about spying activities from smart speakers, a user was passive-aggressive toward others who dismissed the concern by saying *"fill your house with internet-connected microphones"* (U12-T20).

Meanwhile, we observe self-censorship when some users stated their attitudes:

> *"Am I weird for thinking it's weird that someone would potentially have access to control many things in my home through that echo?"* (U1-T24)

Takeaway-RQ3. Compared to prior work that focused mainly on the topics and users' intent in online S&P discussions [167, 271], we show how attitudes and discussion patterns influence each other, supported by quantitative statistics in Figures 4.7 and 4.8. Facing complex smart

home S&P topics, users with different attitudes spontaneously resolve the ambiguity of information, contributing to attitude development. However, this process remains challenging for users due to complex topics and social pressures from opposing attitudes in other circumstances. We also identify an information gap in online discussion. Users rarely refer to reputable and understandable information sources about S&P properties of smart home products during discussions. Also, while community moderators highlight informative comments about S&P, this effort seems ad-hoc, and they are unlikely to correct misinformation at scale.

4.5 Discussion

Based on our findings, we provide three sets of recommendations in addition to future work. First, smart home companies should consider transparency, flexibility, and accessibility in product designs and practices to account for users' multi-dimensional S&P considerations (Takeaway-RQ1). Second, stakeholders (companies, governments, and third parties) should provide S&P nudges to help users develop S&P attitudes by improving users' awareness of S&P risks and appropriate protective strategies (Takeaway-RQ2). Third, online communities should facilitate the access and exchange of smart home S&P-related information, possibly with automated information retrieval and moderation (Takeaway-RQ3).

Incorporating users' multi-dimensional S&P considerations into smart home designs

Our findings reveal that users face challenges, such as limited information and product support, when deciding on S&P protective strategies; the associated tradeoffs force users to pick either a less secure or a less usable deployment. To mediate the conflicts between users' S&P considerations and other functional needs, it is important for smart home designs to

incorporate users' multi-dimensional considerations to help them assess S&P risks and the appropriate protective strategies. However, it can be challenging to incorporate users' multi-dimensional considerations due to the interplay of many contextual factors, users' lack of comprehensive understanding of S&P risks, and the diversity of use cases. Therefore, we recommend three product design and practice guidelines for smart home companies.

Companies should ***inform users about smart home operations and practices in a transparent and understandable manner***. Our findings indicate that users' incomplete views of stakeholders and products could result in bias and conflicts when they assess S&P concerns. When users are not sure *"how reputable* [*the companies*] *are,"* their privacy concerns may even conflict with a product's security settings; for example, one user hesitated to provide their WiFi password to a smart bulb app (U1-T74). Thus, we suggest that companies should first provide understandable information about their compliance with S&P standards and regulations. Then, companies should communicate with users transparently about S&P threats, e.g., how the company *"responds to zero days"* as well as how they enforce S&P protection institutionally (U3-T121) [101]. In addition, explaining the operation of smart home designs, e.g., how the device certification guarantees reliability or interoperability, could help users assess associated S&P concerns when developing tech-involved threat models.

Smart home companies should ***make S&P protective strategies available and flexible for products***. Our findings reveal that users may not adopt certain smart home products when S&P protective strategies are unavailable or inflexible to mitigate users' concerns. For example, the availability and flexibility could be achieved by making features that cause users' S&P concerns (e.g., becoming a part of a mesh network with neighbors) as opt-in rather than a default, as well as by allowing users to delete

data after device abandonment (U7-T87). For more tech-savvy users, companies could provide customizable designs that include the exposure of *"nerd knobs"* or allow DIY and open-sourced software to support their more sophisticated needs to personalize protective strategies (U9-T1). Companies may also offer controls in smart homes that allow users to balance between their S&P needs and utility requirements [229, 164]. In addition, companies may reduce the perceived cost of protective strategies by improving usability, e.g., via an *"easy to use GUI"* (U5-T101).

Smart home designs should ***accommodate considerations of different users, e.g., tech-savvy vs. novice, for specific use cases***. Users' use cases vary, especially in device-sharing contexts, which may lead to different assessments of S&P risks and protective strategies even for the same smart home product. For instance, users demanded accessible S&P controls to *"only activate* [*smart speaker*] *for approved people"* or prevent sensitive conversations from being recorded when working from home (U6-T78, U1-T68). Furthermore, though many users in our dataset are tech-savvy, we recognize their awareness of the technical cost and desire to support less tech-savvy users. Thus, smart home designs could offer a collaborative framework to support the S&P need of less tech-savvy users. In response, companies could establish "tech caregiving," which provides a software interface for technically informed users to help others such as the elderly [143, 326].

Supporting smart home users' attitude development with S&P nudges

We find that users' S&P attitudes, manifested in the level of S&P concerns and incorporation of protective strategies, change depending on the context. Users demonstrate varying assessments of S&P risks, with prior experiences and preconceptions often overriding objective assessments. As such, they develop attitudes that do not always result in secure and privacy-preserving behaviors. Companies, governments, and third parties

could provide S&P nudges – gentle interventions that direct users toward safer practices [8] – to help users' attitudes evolve toward preserving S&P.

Smart home products could ***nudge users' S&P attitudes with physical metaphors***. We observe that users leverage physical S&P metaphors, e.g., comparing an always-listening voice assistant to a person eavesdropping over a table, to rationalize their smart home S&P attitudes. This observation is logical given that users exhibit more developed S&P attitudes in the physical world [114, 240, 188]. As such, a physical metaphor can help inform users about S&P risks and better utilize protective strategies. For instance, the "privacy nutrition labels" take inspiration from food nutrition labels to provide understandable S&P information [84, 81, 83]. Another example by Teyssier et al. explored anthropomorphic smart home product designs that prompt privacy awareness through mimicking bystanders via a human-eye-liked camera [276]. Protective strategies that draw analogies from the physical world might also be more intuitive to users, such as physical webcam covers or voice assistant jammers [47].

Stakeholders, such as regulators and non-profits (e.g., Consumer Reports and Mozilla), could ***help deploy S&P nudges at scale through automated assessment of smart home products***. Deploying nudges at scale solely by companies could be challenging as we observe that users are disappointed by their lack of responsibility. Other entities that are potentially motivated to deploy respective nudges for users include the workplaces, as the prevalence of working-from-home may motivate them to inform their employees about smart home S&P concerns. Furthermore, third-party non-profits may leverage automated S&P assessment via natural language processing to audit smart home products based on their privacy policies and apps [104]. S&P nudges could include the assessment results and inform users about the S&P properties of diverse products, e.g., compliance with regulations.

Supporting smart home users in accessing and exchanging online S&P information

While users actively seek information and voluntarily provide advice in the online discussion, we find that they face several challenges in accessing and exchanging online information about smart home S&P. The various information types, e.g., news, reviews, and anecdotes, are complex in nature. Reputable and understandable information sources are not easily accessible to online users. Moreover, we identify that users' opposing attitudes add pressure and may amplify the difficulties in knowledge exchange. To address these challenges, we suggest that online communities improve information access potentially with the help of automation.

Discussion forums could ***highlight the access to credible S&P information and sources***. We find that forum moderators' efforts in highlighting S&P information seem ad-hoc, as some users had to repeat the same advice to others. There is also little S&P information at the time being on the `/r/homeautomation` wiki page [236], where smart home resources are more organized. Thus, we suggest moderators maintain an up-to-date section about S&P on the wiki page with other volunteers. Moreover, we observe only occasional references to credible third parties that provide accessible S&P assessments, such as Consumer Reports or Mozilla. As such, we suggest online forums, smart home companies, and credible third parties collaboratively maintain communication channels to bridge S&P information and users online. For example, companies may leverage these channels to share S&P information in time, such as patching notices. Companies and third parties may also leverage automated agents to share S&P information via online communities' APIs [183, 237].

Online communities and credible third parties could ***help mediate S&P discussion by detecting misinformation and moderation***. Our findings show that users sometimes perceive others' views of S&P information as conspiratorial. Online communities may moderate discussions to facilitate

more peaceful conversations. Moderation may ease tension resulting from opposing attitudes and detect inappropriate content, e.g., hate speech. Further, automated agents that process natural language may help mediate S&P discussion and ease the burden on moderators and third parties [189].

Directions for future work

Our findings, along with our methodology and limitations, e.g., the demographic bias on Reddit, motivate our suggestions for future research on smart home users' dynamic attitudes and considerations. Our method to detect S&P-relevant discussions and codebook may contribute to future research.

We suggest that research should ***consider users' dynamic and context-dependent S&P attitudes from multiple domains***. Future studies should consider a richer representation of S&P attitudes beyond associating individuals with a static S&P attitude. Such studies may capture users' adoption journeys and contexts from multiple domains, such as other social media platforms (Twitter, etc.) [182, 205], customer reviews [331], and even non-S&P related comments. Moreover, researchers could look into the alignment of users' attitudes between smart homes with other digital and physical S&P domains.

Researchers should ***study users' attitudes longitudinally at a community level***. Our findings on users' evolving attitudes motivate future longitudinal research to investigate attitude development between discussion threads over time in online communities. From `/r/homeautomation`, we observe such evidence of attitude shifts in the S&P discussions over the past decade. For example, in a 7-year-old thread, one comment suspected the smart home market might not take off due to interoperability problems, and there were few S&P concerns *"if we can't even build the system"* (U3-T136). However, we now observe discussions of S&P concerns about specific brands and products. Longitudinal research could track users'

attitudes at the community level, including the community's responses to certain major S&P "events." For example, the widespread publicity of the *"Mirai"* attack affected user attitudes toward smart cameras (U1-T104) [20].

Future work may ***investigate the underlying geopolitical and cultural influences on smart home users' S&P attitudes***. We notice other factors influencing users' S&P considerations and attitudes. For example, the common distrust in Chinese products possibly arises from national security and political perspectives. Prior research showed the influence of geopolitical and cultural factors on the adoption of digital products [66, 75, 152], e.g., Chinese consumers' transition to using mobile payment instead of physical "Red Packets" for transferring ceremonial money [253]. However, these influences are not fully revealed by users on `/r/homeautomation`, since some content, including racism, violates Reddit's content policy [233]. Therefore, researchers can potentially study these influences in conjunction with other platforms, e.g., social media in different countries. Furthermore, researchers could cross-compare different countries' attitudes toward each other, e.g., how Chinese users consider U.S. products and vice versa.

Another venue is to ***study the impact of different designs of online S&P discussion platforms or reviews on users' information access***. As users value information sources when considering smart home S&P, the question of how the sources' information presentation, interaction structures, and credibility may influence users' S&P considerations and discussions remains open. In addition, future work may couple smart home users' demographics and roles to online S&P information, including children and victims of intimate partner violence who have different intentions to access such information [280].

4.6 Related Work

Security and privacy concerns and smart home adoption. Prior research highlighted users' unique S&P concerns toward smart home products. From the security perspective, users worry about vulnerabilities and threats in smart home products and networking, such as malicious devices, adversarial control, and cloud insecurity [320]. They also fear security compromises that lead to physical safety hazards [102]. Concerns of smart home privacy issues arise when users' private activities and information such as conversation and precise location data are collected [57, 58, 328]. While some users are aware of respective risks, others lack a full understanding of certain sensitive practices and risks, including how data is exploited for analytics [3, 268]. Often, users' limited technical knowledge results in bias and lack of S&P concerns [320, 328].

Users are also concerned about the stakeholders involved in the smart home ecosystem, including users with different roles, companies, and government entities. For instance, in multi-user smart homes, different users' varying S&P concerns can remain unresolved due to current role-based access control approaches [321, 63, 93, 107]. In particular, less tech-savvy users, such as children, are treated as passive smart home users who encounter privacy and safety issues [267]. Users further think companies and governments should be responsible for addressing smart home privacy concerns [101].

Users do not develop all S&P concerns at once. While some studies found that users lack S&P awareness or concerns before purchase, others identified users' realization of S&P issues through use [84, 82, 267]. Researchers have quantified the effect of S&P attributes on purchase willingness in relation to risk perception and other concerns such as usability [83, 84, 82]. In hypothetical scenarios where users are prompted with threats, users show higher demand for S&P protective strategies [268].

Whereas in actual use, users would repurpose a product to mitigate S&P risks [46] or ultimately abandon a certain smart home feature or S&P control [118].

Unlike many studies that have investigated users' S&P concerns at a single point in time or in hypothetical scenarios, our analysis of users' S&P discussion on `/r/homeautomation` enables us to study how S&P considerations progress during adoption over time without researchers' intervention.

Security and privacy attitudes. Researchers studied users' S&P attitudes and the associated behaviors. Westin segmented users' privacy attitudes into three groups, corresponding to high, medium, and low levels of concerns [149]. However, Watson et al. found that users' S&P attitudes tend to be more complex [300]. Dupree et al. clustered users' attitudes according to how they are motivated to protect their privacy and their knowledge about privacy [77]. Users also present paradoxical privacy choices as their self-reported privacy attitudes and concerns are inconsistent with their actual behaviors [27].

This disconnect can be attributed to the complex context of S&P, which is often missing in attitude predictors [307]. While many users are "very concerned" about privacy, a myriad of factors impacts users' privacy behaviors [9, 10], e.g., the reward in trading off privacy for convenience [7, 18, 306], the trust of entities that request information [128], self-efficacy [159, 211, 303, 101], and social influence [11, 88]. Moreover, triggers such as social influences, external events, and active priming can change users' attitudes and behaviors [194, 299, 167, 67, 219, 227]. Compared to prior work that categorizes users according to their static S&P attitudes [149, 77], our study focuses on how users' attitudes evolve through the interplay of considerations about S&P concerns and protective strategies when users interact with each other in an online social discussion setting.

Online discussion of security and privacy. Social interaction influences

users' S&P preferences [82] and behaviors [67]. Online discussion offers people a platform to exchange opinions, learn from each other, and provide support. Users consider S&P in this collaborative environment [299, 269], but online discussions about S&P topics only recently started gaining attention [167, 302, 280, 271, 33]. This is possibly because users tend to focus more on functional requirements.

Despite the challenges in locating relevant discussions about S&P, researchers uncover insights of S&P from online discussion [167, 237, 295, 270]. Analysis of online discussions concerning intimate partner violence showed that such an approach is useful for studying issues of their safety and security [302, 280, 33]. Meanwhile, in the privacy realm, prior research using discussion forums has investigated software developers' questions about privacy [271], advice for privacy [270], and in-depth discussion about data practices [167]. However, how smart home users discuss S&P online remains elusive. To the best of our knowledge, our work is the first to leverage online discussion data (Reddit) to understand smart home users' S&P considerations and attitudes. Moreover, we study the interaction dynamics created by multiple users to show how attitudes and discussion patterns influence each other, other than the topics and intent of individual users' commenting [167, 271].

4.7 Conclusion

We analyze smart home users' S&P considerations and attitudes from a major Reddit forum, `/r/homeautomation`. Through our analysis of 180 threads, we discover that smart home users develop multi-dimensional considerations regarding the interplay of contextual factors. Users' S&P attitudes are shaped and further evolve with these considerations. We also study the influence of online discourse–users exchange knowledge and develop attitudes collectively. Accordingly, we propose recommendations

to support users' S&P considerations, attitude development, information exchange, and future research to study users' S&P attitudes from multiple angles.

5 UNDERSTANDING HOW INTERACTION EXPERIENCES INFLUENCE SECURITY PERCEPTIONS OF VR AUTHENTICATION

Virtual reality (VR) systems immerse individuals in a digital world, one that simulates real-world interactions with objects and characters [42]. In addition to specialized use cases (e.g., military training and healthcare [243]), VR technology is seeing widespread adoption in everyday settings, such as gaming, social interactions, shopping, and commerce [90, 259, 111]. Payment features in VR empowers these activities and contributes to the growth of VR economics [157].

To enable payments, VR systems access sensitive user data and assets, raising the need to authenticate users. VR service providers deploy user authentication methods borrowed from traditional platforms to verify users' identities, such as using passwords and personal identification numbers (PINs). However, VR presents a unique context where users interact with digital objects to perform routine activities, such as payment, which were once limited to conventional platforms. Recent research has shown that the context in which authentication is used (e.g., where and for what purpose) affects how users perceive the security of authentication. For example, users feel insecure when using an ATM in a crowded space [192]. There is a critical need to understand *how smart home users' interaction experiences factor into their security perception of payment authentication in VR.* With such understanding, we can guide the future design of authentication methods that are both secure and usable in the growing VR commerce ecosystem. Our work provides this understanding by investigating these related research questions.

- **RQ1 – Interaction Experiences:** What are the factors in users' interaction experiences that contribute to their security perception of

authentication in VR?

- **RQ2 – Influences on Security Perception:** How do users' interaction experiences influence their security perceptions of authentication in VR?
- **RQ3 – Understanding User Expectations:** What are the tensions in users' expectations for VR authentication in relation to **RQ1** and **RQ2**?

Compared to many smart home technologies discussed in Chapter 4, VR is still in its early adoption phase. As such, to answer these questions and explore the open design space, we leverage *technology probes* [120], where proof-of-concept interfaces uncover hidden phenomena in user interaction, to study user authentication in VR. We designed four probes pertaining to authentication interactions for payment authentication VR – two variants of entering a PIN, tapping a virtual card, and signing a signature – to evaluate the user interaction experience and perceived security. We embed these probes in a routine payment interaction for users when they play a VR archery game, which is an organic study context. These probes follow three interaction paradigms of authentication using something you know (e.g., PIN), something you have (e.g., token), and something you are (e.g., biometrics) [245]. The probes and the payment context of our VR game allow us to draw valuable insights into user perception of authentication in VR.

We conducted a user study with 24 participants and evaluated their experiences in the VR game with the probes. We qualitatively analyzed open-ended responses from the participants, which is also supported by quantitative ratings, e.g., the overall usability of our designs. Our analysis reveals these findings in response to the three research questions:

- **RQ1:** The interaction experiences of participants were associated with the perceived usability of authentication and their experience

in the gamified context of VR payment. Participants benefited from intuitive virtualization and seamless interactions in authentication. However, they faced unique challenges, such as motion control. Our participants exhibited feelings of high presence and engagement in the VR game and payment. At the same time, they were sensitive to the interruptions caused by payment authentication.

- **RQ2:** Participants found value in realistic interactions in VR authentication as they could transfer their real-world understanding to the virtual environment. However, usability challenges and limited knowledge about VR authentication jointly hindered this translation, such as losing the sense of ownership of their signature in VR. Additionally, the immersive nature of VR heightened participants' uncertainty about threats in both the virtual and physical realms. Moreover, the gamified VR context may decrease participants' sensitivity to security risks associated with payment authentication.

- **RQ3:** While participants prioritized usability, secure authentication remained a crucial consideration. However, they expressed contradictory expectations from a VR authentication method. These contradictions stemmed from a mismatch between perceived and actual security, the inability to detect threats in both VR and the physical worlds, and the difficulties in accurately assessing risks in playful VR experiences.

Based on our findings, we propose recommendations to meet participants' expectations for VR authentication. These include (1) exploring virtualized interaction metaphors to bridge the gap between perceived and actual security, (2) implementing system support to detect threats in both VR and the real world, and (3) enhancing security risk communication through multiplexed feedback channels. We also highlight the open challenges and research opportunities associated with our findings and

study setup, such as the optimal approach to aid security decision-making for different user groups.

5.1 Method

We use a "technology probe" approach to understanding how interaction experiences affect participants' perception of authentication in VR and extract design guidelines [120]. The idea of the "technology probe" approach entails using a set of proof-of-concept interfaces. As these interfaces package basic interactions, researchers can reveal phenomena otherwise hidden from user interactions [47] This approach is commonly used to evaluate emerging technologies, including VR and user authentication [192, 273]. Our probes implement the core interaction patterns of user authentication to elicit user responses in regard to their interaction experience and perception of security. We deploy our probes for participants to make routine payments in a VR game, which provides a context that mimics VR payments in real-world scenarios.

Authentication Probes

Here, we describe the process that led to the design of four authentication probes in VR.

Design Process

Our study design primarily aims to create a realistic payment authentication scenario to capture authentic reactions from the participants. Toward that end, we iteratively developed our design by considering the usage context and probes together. Our research team conducted regular meetings and tests to discuss, test, and refine our concepts and implementations.

To explore the probe design space, we drew insights from research literature and industry products, many of which are still in the early stage of commercialization. We identified three design dimensions of VR authentication: providing something you know, showing something you have, and proving who you are [245]. These dimensions cover unique interaction patterns and inherent security properties. By including these variations, our probes helped us maximally capture different user interactions and identify common themes. We then narrowed the design scope down to authentication based on PIN, token, and behavior (gesture).

Rather than replicating authentication interfaces in other digital contexts such as mobile apps, we designed the probes to establish a mapping of real-world experiences into VR. Such mapping has already become a popular design choice in other applications of VR for authentication [22]. To provoke participants, we opted not to optimize the usability of probes for everyone in our implementation. We, however, made our design *flexible* by allowing varying gestures to interact with the interface and modification of inputs.

Our context for authentication initially involved payment at a vending machine that dispenses virtual objects. However, we found, in pilot studies, that this task was not ideal for engaging people in VR. As a result, we shifted our focus to integrating payment authentication within VR games, a more popular scenario. Building on this context, we further refined our probe design of payment authentication to align with the virtualization of this context.

The Four Probes

We designed four probes: (1) floating PIN pad (`PIN-F`), (2) on-kiosk PIN pad (`PIN-K`), (3) tap-to-pay (`TAP`), and (4) signature (`SIGN`). We now describe these four probes, which are shown in Figure 5.1.

• **Floating PIN pad (`PIN-F`).** `PIN-F` resembles the default virtual input interface of many VR platforms, where participants interact with a floating PIN pad. `PIN-F` also conceptualizes the idea of giving participants a personal and isolated virtual experience in the authentication. Participants enter PINs by pointing to a floating PIN pad that follows the participants' viewport. `PIN-F` randomizes the PIN layout, which is a common security mechanism in digital PIN pads against observers [192]. Note that `PIN-F` serves more like a baseline authentication probe.

• **On-kiosk PIN pad (`PIN-K`).** In contrast to `PIN-F`, `PIN-K` presents a better mapping of physical-world experiences by rendering its PIN pad on the kiosk where participants initiate and confirm payment. As such, participants do not experience a gap in the transition between authentication and other payment tasks, e.g., selecting the items.

• **Tap-to-pay (`TAP`).** `TAP` represents how participants could own and use a personal virtual object as a unique token to authenticate. To pay, participants take out a virtual credit card from an inventory on their virtual body and tap the card on the kiosk display. The kiosk checks whether the card is in proximity and being held by the user.

• **Signature (`SIGN`).** `SIGN` represents how participants draw a unique signature to give consent and prove their identity. It virtualizes the real-world signing processes. Participants grab a virtual pen from the kiosk and sign in the designated area using the hand-held controller; Only after participants sign, they can proceed to check out.

Our design achieves the goals of the technology probe method to collect in-context information about the usage, test out the technology, and inspire future design by upholding the principles as proposed by Hutchinson et al. [120]. As the functionality of technology probes should be simple, our

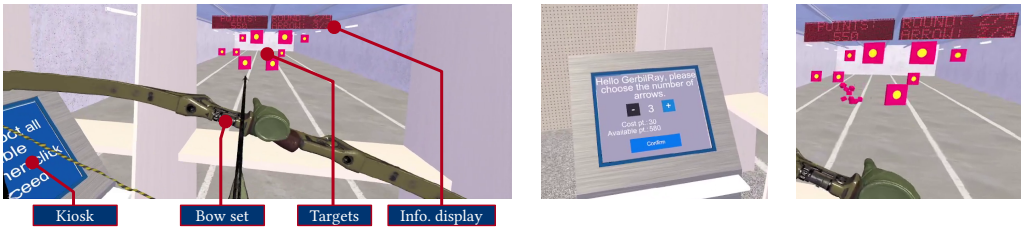
designs only package the essential frontend interaction instead of the full backend of authentication mechanisms, e.g., verifying an encrypted token or comparing signatures. Instead, our study adopts the idea in "Wizard-of-Oz" studies [197] in creating an impression that the system has full functionality through a mock registration process. Also, our probes also support logging of participants' interaction with the interface, including the fine-grained timing, to supplement the analysis.

Figure 5.1: Our four probes for VR authentication.

Study Context: VR Archery Game

We design a VR game – an archery contest – where participants trade in-game credits, using the probes to authenticate their payment. Participants earn credits by shooting virtual targets and consume these points when they refill arrows. To make the game and payment realistic to participants, we match their in-game credits with a physical compensation: the participant who scores the highest wins a "grand prize" (a small gift). Such compensation, which is added to our base compensation for participating in the study, is standard in prior studies that include gamification [38, 195, 278]. We determined the value of the prize (a 90 USD fitness tracker) based on our local wages. Below, we describe the main components of this game (see Figure 5.2).

Environment. We situate the contest in an indoor archery range. We assign the participant a chamber. In the chamber, the participant can



(a) Environment overview (b) Add arrows (c) Shoot target

Figure 5.2: The environment of our archery game with key components marked and key scenes illustrated.

find the bow set and a kiosk instructing them to shoot, refill arrows, and authenticate their payment. We place the targets towards the other end of the room at a distance. The participant can also find the information display that shows their current credits and remaining progress.

Bow set. The participant can interact with the bow using the two hand-held VR controllers. They need one hand to hold the bow and the other hand to draw. The participant shoots the arrow toward the target by releasing the drawing hand. The participant will need to refill after they use up three arrows. Each arrow consumes ten credits.

Targets. We place eight targets in the range at two different distances. Taking down each closer target rewards 20 credits to the participants while 30 for the farther targets.

Kiosk. The kiosk displays game instructions and payment interfaces. To refill arrows, participants interact with the kiosk to select how many arrows to load. After the participant confirms the selection, the kiosk will display the authentication interface among one of the four probes. Once the participant completes authentication, the kiosk will display a waiting page that emulates the running backend processes. After that, the kiosk will accept the credit payment and refill arrows.

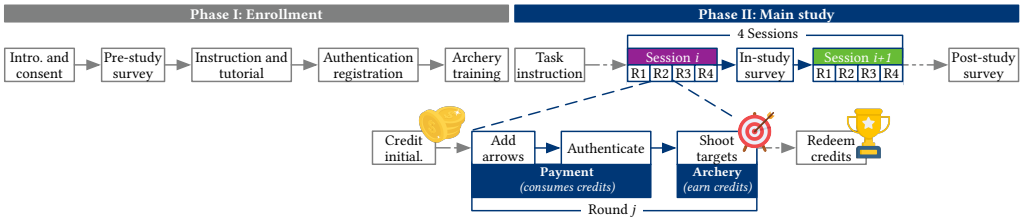


Figure 5.3: Our study procedure.

Information display. The display shows the current credit, the remaining arrows, and how many rounds the participant has completed.

Implementation. We implemented our VR game using Unity, a mainstream VR engine, and C#. In the study, we ran the game on a commodity PC (CPU: Intel i7-12700F, GPU: Nvidia 3060Ti), which is connected to an Oculus Quest 2 VR headset. Participants interacted using the headset and its hand-held controllers.

Study Design

Here, we explain our study procedure, the instruments we used to collect participants' responses, and our recruitment.

Study Procedure

We designed a with-in-subject study to evaluate the four probes. The study consists of two phases–an enrollment phase and the main study phase. The purpose of enrollment is to familiarize participants with our VR and authentication setups. In the main study, participants played the VR game and interacted with our probes during authentication. Figure 5.3 illustrates our study procedure.

Enrollment. The experimenter first obtained consent and asked the participant to complete a pre-study survey. After that, the experimenter explained the game context to the participants. We assigned participants a "nickname" for the game. The experimenter then introduced the PIN, the virtual card, and the signature setup to the participant (we assigned each participant the same credential information for the statistics of authentication time). We described to the participants that they would enroll the information, including the signature, for the use of authentication in the second phase. The experimenter then walked participants through how to use our VR setup. After the tutorial, the experimenter let the participants practice with the registration kiosk, similar to the one they will use for authentication, and present their credentials to it. After registration, the participant will practice archery. The enrollment phase took about 20 minutes.

Main study. The one-hour main study took place on another day to reduce participants' fatigue. We first reminded the participants of the study procedure, game context, and the authentication interaction.

Then, the participants entered the game with 400 points. They proceeded to finish four sessions; each included the game with an in-study survey. In each session, the participants paid using one of the probes (in a randomized order) to authenticate. Each session consisted of three rounds. Participants authenticated to pay for the arrows (min: 1, max: 3), except for the first round of every session where three arrows were given for free. At the end of each session, the participants completed an in-study survey, and following the second session, they were given a brief break outside of VR. Once all four sessions were completed, participants were instructed to fill out a post-study survey. After the completion of the survey, participants redeemed their credits to compete for the grand prize.

After the study, we disclosed our full study purpose to the participants.

We clarified that our focus is to evaluate the interaction experiences and perceptions from the frontend, and we did not process their information, i.e., signature, in the backend.

Instruments

Our study mainly relies on surveys that elicit participants' responses for our qualitative analysis. We chose to use surveys instead of more active methods such as think-aloud studies [49] to minimize the interference with participants' game experiences.

Pre-study survey. In the pre-study survey, we collected participants' demographic backgrounds, including their gender, age, education, and experience with VR. We also used the standard affinity for technology interaction (ATI) score (computed from 6-point Likert ratings) to understand participants' tech-savviness [89].

In-study survey. The in-study survey assessed (1) participants' general experience in the VR context and (2) the perceived usability of authentication probes. For the former, we used the IPQ VR presence questionnaire, a standard measure of participants' sense of presence in VR. The sense of presence is a commonly adopted measure of VR experience. It is defined as participants' subjective perception of being and acting in the virtual world (though their body resides in the physical world) [252]. This measure consists of four sub-scales (1) sense of being here (PRES), (2) spatial presence (SP), (3) involvement (INV), and (4) experienced realism (REAL). For the perceived usability, we collected participants' open-ended comments on the usability issues of each authentication probe when participants exited VR. Moreover, participants completed the system usability scale (SUS), a standard measure to assess the overall perceived usability, for each authentication probe [39].

Game log. In addition, we logged participants' behaviors and timings in the game, including their archery performance as well as the time spent on authentication. We used these objective behavior logs to support our qualitative findings.

Post-study survey. The post-study survey consists of three major components. First, to understand participants' engagement in the routine payment of the game, we asked them to explain how they decided on the number of arrows to pay. Second, we wanted to understand participants' security perception of payment authentication. We designed questions to elicit participants' responses from different angles. From prior research, we identified the five aspects related to the security of payment authentication, namely *consent* [108, 186], *security* [137], *privacy* [333], *being alerted* [239, 134, 306], and *in control* [209]. We asked the participants to evaluate and elaborate on their agreement on statements related to the five aspects. One example is: when the participant used `TAP` to pay for arrows, "I [the participant] felt that my [the participant's] payment was secured". We collected both their Likert-scale rating and open-ended responses to explain perceptions. However, rather than relying on the quantitative ratings, we mainly study the relation between interaction experiences and perception from their open-ended responses. Third, the post-study survey asked participants about their preferences among the four probes, their quality expectations that affect their preferences, and their suggestions to improve these probes. We use these questions to further understand participants' expectations of VR authentication.

Participants and Recruitment

Consistent with prior work in VR authentication [192], we recruited 24 participants from our organization. We stopped recruiting when we ob-

served data saturation from our qualitative coding [96]. Each participant received compensation (a 30 USD gift) after they completed the study.

The demographics of participants are as follows. 18 out of 24 identified themselves as man (6 women). The average age of our participants is 29.6 years old (std: 4.8). 20 of them completed or were studying for a graduate degree. Most participants have a background in computer science. Participants' ATI score (mean: 4.35, std: 1.1) shows a high affinity for technology interaction) [89]. 20 participants had used VR before (mainly for gaming). But none of the 20 participants frequently used it.

Ethics and participant safety

Our study and recruitment were approved by the IRB-equivalent body of our organization. While the study poses a low risk to participants, we took the following steps to protect their physical safety and data privacy. First, we communicated potential risks and their right to withdraw through our informed consent and disclosure processes. Second, we spread our study into two days to minimize fatigue. In addition, we regularly checked in with the participants and made sure they were comfortable to proceed during the study. Third, we ensured that our physical space was clear and safe for the use of VR. The VR application notified our participants when they got close to physical boundaries. Last, our data collection and analysis do not include sensitive personal information, and all data were anonymized and stored in a secure server in our organization.

Data Analysis

We analyzed survey and log data as described in Section 5.1. To analyze such qualitative data, the first author started open coding and took memos while recruiting participants. Meanwhile, another researcher coded data independently. We coded each response the participant made correspond-

ing to a question. The whole team discussed the memos, reconciled the codes, and refined the codebook iteratively. The two coders reached high inter-rater reliability (Cohen's Kappa $\kappa = 0.81$) using responses from 6 randomly sampled participants for each question (78 out of 312 responses in total), then we converged on a codebook to code the rest of the data. Using Grounded Theory [296], high-level themes emerge from our coding. When we observed data saturation from our coding [251], we stopped recruiting. We make our codebook along with the surveys available in Appendix A.2.

In Figure 5.4, we show our analysis framework along with themes we identified from the analysis. In addition, we report the quantitative data to support our qualitative analysis, including the IPQ questionnaire, SUS scales, and participants' Likert-scale rating on security and privacy perception.

Limitations

Our study has the following limitations. First, our participant population has a demographic bias, e.g., most participants are tech-savvy. Our population and demographics are comparable with prior work applying similar methods [212, 315, 170, 192]. Using a technology probe, our study's objective is not to generalize but present a set of findings and recommendations that guide future design. Nevertheless, our analysis reveals that even these tech-savvy participants faced challenges in interacting with VR and assessing security, let alone other users. Future work may study the probes with a more diverse population, generalize the findings, and quantitatively measure users' experiences and perceptions. Second, our study does not investigate how users use VR authentication longitudinally in the wild where users' experience and perception of security may change over time. In addition, participants' self-reported responses may be biased due to social desirability [221]. Last, the purpose of our study is to understand

users with probes in an early design phase. We focus on the novel experience of fundamental authentication concepts in VR payment. As such, we did not explore all alternative implementations of the VR interaction and interface. Despite these limitations, we believe that our work still presents a significant contribution. To the best of our knowledge, our exploratory study is the first to investigate the interplay between interaction experience and perception of security for VR authentication.

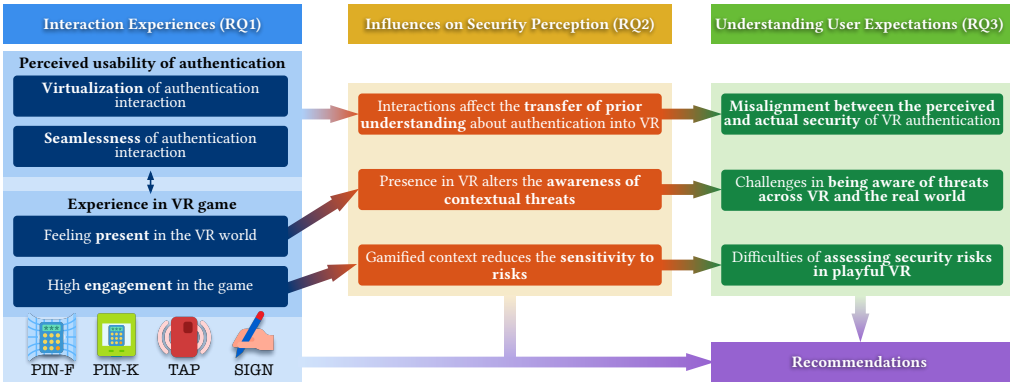


Figure 5.4: Our analysis framework and summary of results. We connect themes that are most related across research questions.

5.2 RQ1: Interaction Experiences

We observed two overarching themes of participants' experiences that influence their security perception of authentication. The first is the **perceived usability of authentication**, which consists of two sub-themes: *virtualization* and *seamlessness* of the authentication interaction. The second theme is **experience in VR game context**, which is influenced by the two sub-themes: participant *presence* and *engagement*. In the following, we elaborate on these themes and discuss how our probes uncovered these experiences.

Perceived Usability of Authentication

The virtualization and seamlessness of the authentication interaction affect the perceived usability. Our probes entail additional aspects regarding *authentication interface*, *process of authentication*, and *motion control* that are not typically present in real-world authentication interaction. These aspects contribute to the virtualization and seamlessness of the VR authentication, leading to different perceptions of usability.

Virtualization of Authentication Interaction

Participants expressed positive feedback regarding the usability of *intuitive* VR interactions and interfaces that resemble their real-life experiences. Virtualizing *familiar* interfaces from the real world made them feel more *immersed*. However, complex VR interactions introduced an additional learning curve, particularly with `SIGN`, which involved more friction compared to PIN methods and `TAP`. For instance, one participant felt *"signing in the virtual world was very different as compared to the real world"* (P13). Nevertheless, another participant found `SIGN` was easy after practicing.

> *"It's very easy to learn and use, and the functionality can be easily picked up."* (P11)

An intuitive and realistic *authentication interface*, in terms of presentation and feedback, can enhance the virtualization of interaction, even though the authentication concept is new to participants. The token-based authentication `TAP` was favorable to participants, primarily due to its interface presentation:

> *"since it closely mimics the way I use card payments in my day-to-day life."* (P8)

Participants exhibited different preferences over the interaction modality, often by comparing it with real-world counterparts. One participant

mentioned, when describing the virtual PIN pad, that the *"laser pointer is less similar to using keyboards in real world"* (P2). Moreover, participants expected more realistic interaction feedback like what they receive in the real world, especially for the highly interactive `SIGN` approach. For example, they demanded a better sense of writing on a "paper" in VR.

> *"It is still usable and the pen writes like real writing but still did not get the feeling of writing on a paper."* (P1)

In addition, participants perceived virtualization differently due to their prior experience in the real world, i.e., how they performed payment. For example, one participant was not used to a shuffled PIN pad as *"most terminals have a fixed layout* (*I am* [*they are*] *thinking of ATMs and gas stations*)*"* (P5).

Generally, the virtualization of real-world authentication interactions helped improve perceived usability. However, some participants expected VR interfaces to save them more effort compared to authentication interactions in the physical world, e.g., moving *"to get a better look at the numpad"* on the kiosk (P13).

Seamlessness of Authentication Interaction

The seamlessness of authentication interaction also contributed to better-perceived usability. Interestingly, participants associated seamlessness with specific aspects such as the *comfort* of interacting in VR, the *physical, mental, and time efforts*, and the *smoothness* of the interaction process.

As mentioned earlier, our VR authentication probes exhibit familiar properties of real-world authentication. They also inherit some of the difficulties from previous authentication experiences, including the transitions in the workflow, the memorability of PIN, and the additional usability challenge to use a shuffled PIN pad where *"the numbers shifted locations between attempts"* (P13).

The *motion control* in VR introduced unique challenges that made authentication interaction less seamless. This aspect had a more noticeable impact on probes that required complex interactions (`SIGN`, `PIN-F`, `PIN-K`) compared to `TAP`. The challenges with motion control included inconsistency in action control when interacting with digital components. This inconsistency affected the accuracy of translating participants' actions, such as how their virtual gestures aligned with their intentions. For example, one participant preferred `PIN-K` over `PIN-F` due to the observation that *"stable kiosk was easier to use as it allowed for better calibration"* in entering PINs (P7).

Moreover, spatial awareness in VR, such as coordinating the movements of a virtual object and the avatar, contributed to the challenge of motion control. Some participants indicated that lacking spatial awareness hindered their ability to control a virtual object, e.g., stretching their arm to sign a signature:

> *"It takes a while to get used to the proper distance between the pen and the kiosk screen."* (P7)

These aspects became evident when participants commented on the seamlessness of the `SIGN` probe. As mentioned before, some participants thought `SIGN` was easy, while others did not. One participant felt that performing the VR gesture was easy conceptually but hard in practice.

> *"Signature: it was conceptually easy, but the execution was somewhat cumbersome and it required a complex gesture."* (P6)

Quantitative Usability Analysis

The above interaction experiences manifest in different usability ratings for the four probes. By comparing participants' SUS rating (Figure 5.5), a standard usability metric, we find that `TAP` received the highest SUS score (mean: 82.1, std: 14.6) and cost the least time (mean: 3.3s, std: 2.0s),

indicating "excellent" usability [39]. It benefits from seamless and intuitive interaction. PIN-F (mean: 70.6, std: 19.9) and PIN-K (mean: 76.1, std: 11.7) closely follow TAP, demonstrating a "good" usability. They consumed comparable time to complete: PIN-F (mean: 10.8s, std: 11.1s) and PIN-K (mean: 10.4s, std: 6.6s). As discussed, participants appreciated the usability of these methods due to their familiarity and relative ease, despite the effort required to memorize and enter the PINs. Moreover, the differences in scores between PIN-F and PIN-K are small. Participants commented on different usability issues for them, such as the inconvenience of walking towards the kiosk for PIN-K and the distraction caused by entering PINs on a moving panel for PIN-F. SIGN (mean: 60.7, std: 14.7), which is also the most time-consuming probe (mean: 18.3s, std: 9.8s), received the lowest score among the four probes but still achieved "OK" usability. Although its interface appeared intuitive, signing with the virtual pen proved to be challenging for multiple participants in the VR environment.

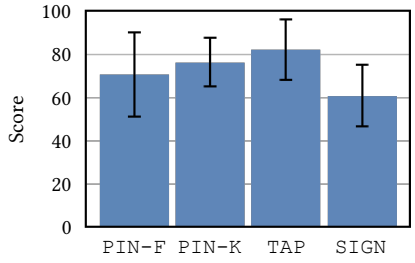


Figure 5.5: SUS scores (average and standard deviation) for each authentication probe.

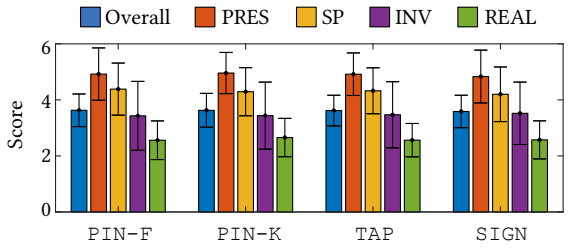


Figure 5.6: IPQ sense of presence scores for game contexts with the four authentication probes. A higher score indicates a higher level of presence. From 0 to 6, a score higher than 3 stands for neutral. All subscales have a positive mean score, except REAL. The IPQ scores of our setups are consistent with prior implementations for VR authentication [192].

Experience in the VR Game Context

In addition to the interaction experience with the authentication probe, participants' experience with the VR game (as a proxy for the VR context) appeared to influence their perception of security. We observed two sub-themes related to how participants felt a sense of *presence* and *engagement* in the game.

Feeling Present in the VR World

Using the IPQ presence questionnaire (Figure 5.6), we observed that participants rated their presence in the VR world positively. However, we did not observe a significant difference in the IPQ scores between sessions with different authentication probes, which is consistent with previous findings about authentication is often perceived as a secondary task [69]. When participants felt present in this world, some of them expressed dissatisfaction when authentication disrupted their immersive experience through an "unreal" interface, e.g., `PIN-F`:

> *"the floated pad makes it so unreal that I know it is in VR rather than real life. I do not like the experience."* (P24)

High Engagement in the Game

We observed the participants to be highly engaged in the archery game and the routine payment. When deciding which arrows to pay for, participants mentioned various factors, including *their strategy to compete* and *enjoyment of the game*. Among the participants, 19 explicitly mentioned that they would like to shoot as many arrows as they could, and six people said it was fun to play, despite differences in their archery performances (highest score: 1470, lowest score: 190). This high level of engagement also led some participants to expect less effort during authentication compared to the time spent actively playing the game, as expressed by one participant:

"the time I am [they are] actually playing the game" (P18).

Takeaway-RQ1. Participants' interaction experiences included their perceived usability of authentication and their experience in the VR game context. Our findings were consistent with prior observations that PIN authentication has acceptable usability [94] and that user authentication is perceived as a secondary task compared to the overall context [192]. We further observed that participants associated the usability of authentication with better virtualization and more seamless interactions. When fully engaged in the VR game, participants preferred less disruptive experiences to keep their immersion. Moreover, we found that unique usability challenges of VR interactions, such as motion control, influenced participants' authentication experience, particularly with behavior-based authentication.

5.3 RQ2: Influences on Security Perception

Previously, we listed the interaction themes that influenced user perception of the security of authentication. In the following, we discuss *how* these interaction themes influenced participants' security risk assessment while authentication. Our qualitative analysis of participant responses revealed three themes of influences in Figure 5.4. These themes include forming security perceptions, being aware of threats, and assigning risk to threats.

Interactions Affect the Transfer of Prior Understanding About Authentication into VR

Facing a limited understanding of novel VR authentication concepts, participants transferred their prior experience with real-world authentication to VR. This transfer of knowledge, however, did not result in a consistent

security perception. The virtualization and seamlessness of authentication interactions, while enhancing usability, affected how participants formed their security perceptions of VR.

Impact of Prior Knowledge and Experience

Participants *associated the virtual probes to authentication processes from the physical world with which they are familiar*. Examples of such processes include possessing secret knowledge (PIN) that is *"only known by me* [*the participant*]*"* and using a shuffled PIN pad that *"gave me* [*the participant*] *some sense of security"* (P20, P17).

However, participants' *prior understanding and real-life experiences varied*, leading to differences in their perceptions. For example, the participants held different views on how consent works in VR payments, as shown in Figure 5.7. For example, one user viewed `SIGN` as less familiar as they *"rarely signed to pay* (*only recently in the US*...)*"* (P6). Meanwhile, others *"naturally perceive it as giving my* [*their*] *consent"* as they would in the physical world (P12). Some participants associated consent with privacy implications. For instance, P20 raised a possible privacy concern with `SIGN`, which *"requires more information* [*signature*]*"* (P20) than PINs.

Meanwhile, participants *expressed varying levels of confidence about the security of authentication*. Some participants were not confident due to lacking information, for example, *"without more information on how the payment actually works"* or without knowing *"how the backend works"* (P13, P3). P3 also suspected that `TAP` would need additional verification or certification steps for the virtual card. On the contrary, some participants more confidently assumed that *"the technology behind tapping makes me* [*them*] *believe it is safe"* (P14). This finding contributes to the polarized responses, e.g., security and privacy of `TAP` (Figure 5.7(a, b)).

Impact of Usability

We observed that, because of a limited understanding of VR authentication, the usability properties of the probe affected the participants' security perception. This was the case for the SIGN and TAP probes. In particular, as VR interactions do not mimic real-world sensations very well, participants felt a loss of control over authentication. For example, some participants were not confident they were providing consent when using SIGN and TAP, compared to PIN-F and PIN-K (Figure 5.7(a)). When signing in VR, some participants did not feel that the VR signature belonged to them. One participant expressed:

> *"The sign-to-pay method was a bit hard to use, so I think I just tried to write something, and I felt less like providing my signature."* (P10)

Similarly, the same participant felt that TAP was not secure as they still thought *"it's not my [their] real card but a card-like object"*. In addition, the lack of feedback associated with the seamlessness of the VR interaction contributed to the loss of control, particularly for TAP. For instance, P2 expressed concern that TAP appeared too *" no-brainer "* without any warnings (P2).

The *usability characteristics of the interaction appear to undermine participants' confidence in security*. Some participants expressed concerns about the security of SIGN, the authentication method with the lowest usability. They believed that this method compromised security by accepting inconsistent signatures, which could potentially make them more susceptible to impersonation.

> *"The sign-to-pay felt the most insecure as people easily have access to my cheques and can probably fake in the VR world since the VR signatures were clearly less accurate than the real-world."* (P5)

The participants expressed greater confidence in the security of traditional PIN methods, which also offered more acceptable usability. In such cases, a higher level of interaction provided participants with a sense of control (Figure 5.7(e)). In a similar vein, a subgroup of participants expressed a willingness to accept additional interaction to prioritize security and transparency. This included incorporating extra warnings and utilizing a shuffled PIN pad to avoid *"making accident payment"* (P15).

Presence in VR Alters the Awareness of Contextual Threats

Participants engage in threat modeling as they analyze the security properties of the VR authentication. During this process, they identified several entities based on their understanding of real-world payment and the VR environment. These entities included payment and authentication service providers, physical and virtual bystanders, third-party apps, and malware. Participants also *recognized software vulnerabilities* that could lead to compromising personal information, such as PINs and signatures:

> *"Similar reason that pins and signatures are just exposed to the game or malware in the gaming system."* (P24)

Another example is participants' *awareness of virtual and physical bystanders*. For example, P6 considered `PIN-K` less secure than `PIN-F` if bystanders in VR were able to see the kiosk.

Participants expressed concerns that their presence in the VR environment reduced their awareness of both the virtual and physical worlds. This lack of awareness potentially exposed their information (PIN, signature) and assets (cards and tokens) to adversaries. They worried that *attacks in VR could be more imperceptible than in the physical world*, such as being deceived by an invisible terminal for phishing purposes:

> *"Tap-to-pay is the easiest, but I feel it not safe because I can easily touch it to an invisible system in the VR world."* (P19)

Noticing attacks depended on the type of authentication interaction in VR. For example, it might be easy for a person to notice an attacker *"stealing a card,"* but harder to observe a physical or virtual shoulder surfing attack when signing or entering a PIN (P6). Participants were also concerned that malicious users could leverage prior knowledge about the victim in the physical world to launch attacks in VR:

> *"signing can be also copied by anyone who knows my signature in the real world."* (P12)

Last, some participants had *a general lack of trust due to their ambiguity surrounding the VR technology and authentication*. P12, for instance, expressed skepticism about the security of `TAP`, which stems from their distrust of the VR technology:

> *"Tap-to-pay was the simplest, but it was way too simple to believe that the entire payment process behind the scene was dealt with as I wanted. I'm not sure if this is due to my distrust to the specific payment system, or just to the VR world, or both."* (P12)

Gamified Context Reduces the Sensitivity to Risks

Our context for user authentication revolves around participants' routine payment in a VR game. During these activities, participants were highly engaged in the game and the payment process. Some participants noted that their *sensitivity to authentication security related to the relevance of virtualization to payment*. For instance, P9 felt a greater sense of control when using `TAP`, as it simulated an actual payment experience by physically grabbing the card and initiating the payment. On the other hand, `PIN-K`, `PIN-F`, and `SIGN` were perceived as less specific to payment.

However, the *gamified context may reduce participants' sensitivity to security and privacy risks*. For instance, P7 argued that they did not feel a loss of privacy in a VR game compared to real-life transactions.

> *"In the context of the game I didn't feel like giving away privacy. However if I were to imagine this with real transactions, then I'd feel like giving some of my privacy away, similar to every time I pay with something else than cash."* (P7)

Furthermore, the gamified interactions and interfaces of certain authentication probes made some participants feel less attentive to security. When signing a signature in VR, the gaming aspect overshadowed P7's sense of giving consent.

> *"The sign to pay wasn't completely obvious you were actually paying for something, it could have been part of the game to have to write your name."* (P7)

Similarly, P10 thought `PIN-F` *"felt a bit too much like being in a game as well"* compared to `PIN-K` (P10).

We also observed a dichotomy in whether participants prefer to be alerted to the fact that a payment is taking place (Figure 5.7(d)). Some participants felt like they need to be alerted because the lack of feedback and their limited understanding make them feel less secure (as we explained in Section 5.3). Other participants felt like the payment authentication was secure, and they did not see a need to be alerted if not prompted. For example, the more personal interface of `PIN-F` captured a participant's attention because it *"clearly wouldn't let me* [*them*] *proceed through the game"* (P16).

Takeaway-RQ2. Our findings confirmed our hypothesis that authentication interactions have an impact on the security perception of VR authentication, similar to previous studies on applications like implicit authentication [134, 304]. Our findings in VR authentication further complement previous work as follows. Stephenson et al.'s survey [262] mainly discovered that the usability of knowledge-based authentication in VR is

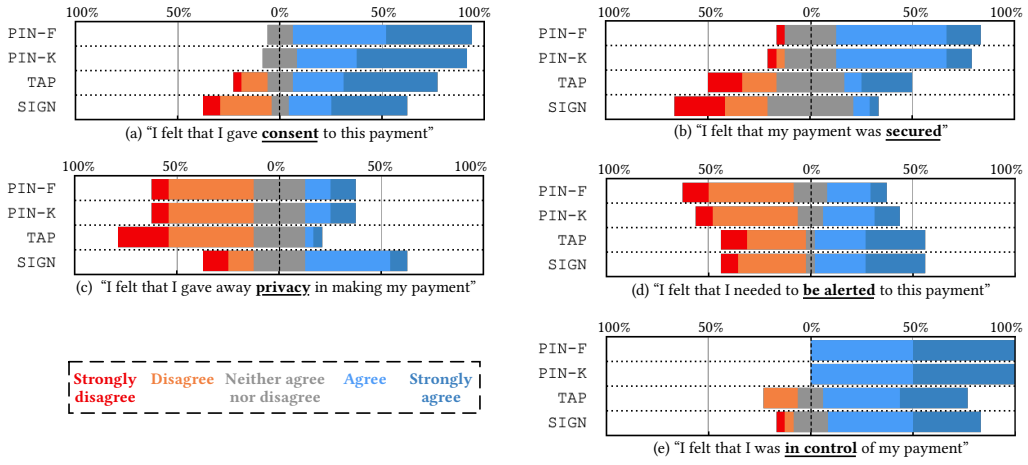


Figure 5.7: The overall security perception of the four probes. We color-coded the bars that represent the percentages of participants (red: strongly disagree, orange: disagree, grey: neither agree nor disagree, light blue: agree, dark blue: strongly agree).

perceived as a tradeoff between security and usability. Participants in our study generalized this tradeoff to other forms of VR authentication, including behavioral biometrics. Moreover, we observed that the virtualization of authentication interactions in VR may bias participants when transferring prior understanding of authentication to VR, potentially leading to a false yet confident sense of security. Our participants also expressed uncertainties regarding threats in both the virtual and physical realms due to the immersive nature of VR. Additionally, the gamified context of VR had the potential to reduce participants' sensitivity to security risks.

5.4 RQ3: Understanding User Expectations

After identifying how interaction experiences shape the security perception of VR authentication users, we turn our attention to participants' expectations of an "ideal" authentication experience. Our subsequent

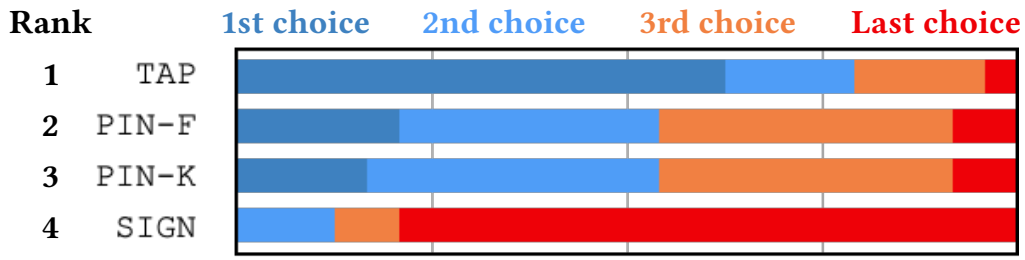


Figure 5.8: Overall payment preference of the four probes. The bars are color-coded dark blue, light blue, orange, and red to indicate the fraction of participants that selected each probe respectively to be their 1st, 2nd, 3rd, and last choice.

discussion builds on our findings in **RQ1** and **RQ2** as well as qualitatively analyzing participant responses to the surveys.

Misalignment between the Perceived and Actual Security of VR Authentication

User authentication in VR is a new concept for users, where *a gap still exists between the perceived and actual security.* The technical properties of VR authentication mechanisms may differ between the real world and VR, despite the similar virtualization metaphors. For example, the technology to secure a VR token would be different than the EMV chip in a physical card even though they share a card metaphor. We observed that some participants, even those who were technologically savvy, tended to overestimate the security of VR. On the other hand, we found that making the authentication more oblivious and seamless to users may make them underestimate the actual security in VR (Section 5.3). The observations are associated with users' lack of understanding of these novel technologies in VR to assess the actual security. For example, one participant expected the transparency for the VR authentication protocol to be *"well-defined and open-sourced"* for properly evaluating its security (P3).

Another reason behind this gap is that participants prioritize interaction experience over security, resulting in an incorrect perception of the security of the authentication method. Figure 5.7 and 5.8 show participants selecting what they considered the less secure, yet the more usable, TAP as their preferred authentication method. Participants attempted to bridge this gap by envisioning new alternatives of interaction modalities for all different concepts of authentication, such as new ways to interact with a card and drawing 3D signature *"instead of a flat 2D traditional signature"* (P5). Along the same lines, researchers and developers have been actively improving both the usability and security of VR authentication [193, 301, 213, 330, 171].

Challenges in Being Aware of Threats Across VR and the Real World

Participants enjoyed being present and immersed in the VR environment. Meanwhile, they desired awareness of security threats in VR and the physical world. For example, when immersed in VR, participants wanted the VR application *"to reflect the surrounding environment"* both virtually and physically for them to notice bystanders (P21). Another instance is one participant's need to know about suspicious activities in VR by *"sending me [them] an SMS/email/etc."* even when they are back in the physical world (P8).

Unfortunately, offering such cross-contextual awareness with an enjoyable experience is still an open challenge, even when the users develop a proper understanding of VR authentication. There are several reasons. First, VR creates an immersion effect, and users have limited perceptual capacities [281], which is more challenging compared to more conventional digital mediums, such as websites and mobile apps. Furthermore, it is challenging to interpret suspicious activities when users are present in VR. Because the VR threats can be made more imperceptible, e.g., invisible

card-skimming attacks (Section 5.3), and users miss the physical context to better understand bystander activities in VR.

Difficulties of Assessing Security Risks in Playful VR

Many users have high expectations for VR technology and applications, seeking creative and enjoyable experiences [214, 274]. In our study, we found that participants also desired a playful authentication experience and proposed various ideas to achieve it. These ideas focused on improving usability and implementation, such as the interface to manage multiple virtual cards where users could *"select the card from a pop up UI"* (P20) and informing payment success with *"more fancy effects"* (P24). Participants also anticipated emotional appeals (e.g., enjoyment and playfulness) from the authentication interfaces, especially in the game context.

> *"I want the payment process to be cool and make me feel good. some visual effects could help and make me happy to make the payment."* (P24)

However, it is important to consider the trade-off associated with a playful authentication experience. As discussed in Section 5.3, gamifying the VR experience can reduce participants' sensitivity to security risks and feedback. This reduction in awareness may have significant consequences and pose risks in security-sensitive scenarios, such as high-risk payments. This concern aligns with previous research conducted in the domain of digital games [53].

Takeaway-RQ3. While participants prioritized interaction experience, the security of authentication is important to them. Participants expected both usability and security, but their expectations appeared conflicting. Our findings highlight three tensions. First, we noticed a gap between the perceived and actual security in VR authentication as some other contexts [134], and the VR interactions further exacerbate the discrepancy.

Second, participants liked the immersion of VR while expressing concerns about threats in both the virtual and physical worlds. Third, although participants envisioned playful experiences, such experiences can diminish their sensitivity to security risks.

5.5 Discussion

Our findings yield recommendations to guide the design of future VR authentication to calibrate users' security perceptions, enhance VR systems' awareness of threats, and provide flexible feedback on security risks.

Exploring virtualized metaphors for calibrating security perceptions. As we discussed in Section 5.3, users have a limited and varying understanding of VR's inherent security properties, and how authentication interaction is virtualized may lead to misalignment between users' perceptions and the actual security properties. Calibrating such perceptions is challenging in the two following aspects, for which we provide corresponding suggestions. In general, we can actively leverage interaction metaphors [226, 19] in helping VR users align their security perceptions in a usable way.

First, the security model of VR authentication can be more involved than the traditional platforms, including the use of unconventional and multi-modal modalities, e.g., eye tracking input and biometrics [193, 171]. As such, the designers of VR authentication services could convey the security properties by ***integrating metaphors apt to VR authentication modalities***. For example, when using eye tracking for two-factor authentication (PIN and biometrics), symbolizing eye interaction may help inform users of the use of biometrics when comfortably engaging them [161].

On the other hand, the designer could ***better connect the context with authentication metaphors***. We observed the positive effect that `TAP` re-

minded some participants of the payment context (Section 5.3). We can further strengthen this association with the context by naturally embedding metaphors in the VR interactions with the primary task [200]. For example, to indicate enhanced security, one VR game can award users persistent credits after users make an effort to complete the multi-factor authentication. Another opportunity is to communicate security by storytelling in VR [148].

System support for enhancing awareness of threats across VR and the real world As we found in Section 5.3, VR users lack awareness and face difficulties in interpreting threats in both VR and the real world. These difficulties further prevent users from adequately assessing security. Though solutions exist to improve VR users' contextual awareness for safety reasons [144], it is not practical to rely on users only to comprehend all the security threats. Thus, we propose multiple improvements for the VR system to automatically detect threats across virtual and physical contexts and adapt security measures in authentication accordingly.

First, we can design ***intelligent systems to recognize and comprehend threats related to VR***. Our observations in Section 5.3 and prior studies identified threats in users' virtual world (dark patterns [246], bystanders [262], etc.) and physical contexts (physical imposter [193], unauthorized users [68], etc.). The VR system may utilize machine learning to enhance their cross-contextual awareness [144]. It may comprehend the security implication of a virtual or physical entity based on the VR scene, physical environment, and multi-sensory inputs. For instance, the VR system can tell whether a bystander is actively observing the users during payment.

Next, the VR system could employ ***access control to safeguard users' virtual assets actively***. Based on the contextual understanding, the VR platform could control other entities' access to one user's scene and asset,

e.g., payment token, when they perform authentication. We noticed our participants' need for personalizing such access control for multi-user settings, e.g., to avoid *"accident payment from the kids"* in a family with children (P20).

Moreover, the VR system may ***automatically adapt authentication methods according to threats in the context*** while balancing usability requirements. Though users prioritize usability over security, they still consider the security of authentication crucial (Section 5.4). The VR system can adaptively enhance security measures to defend against perceived threats. For example, when detecting an active observer, the VR authentication may shuffle the PIN pad from the default.

Multiplexing communication of security risk in gamified VR. Current VR applications are predominately games, and users expect a playful experience (Section 5.4). Section 5.3 explained how a gamified VR context might reduce people's sensitivity to risks and security feedback. Our findings in Section 5.4 echo prior work that users still demand transparency and control for payment authentication, e.g., properly understanding suspicious behavior. This complements the above recommendation to reduce users' effort in making security decisions. For security-sensitive applications, especially payment, risk communication is often necessary for users' assessment of security. Here we discuss the potential to improve the efficacy of risk communication by multiplexing the feedback in VR.

VR applications deliver feedback over multiple modalities (e.g., visual, audio, and haptic) to their users. The payment authentication service may choose to ***utilize orthogonal feedback modalities*** to the primary application context. As such, we avoid overwhelming users and causing security fatigue [103]. In addition, the application can interleave the timing of authentication feedback with the primary activities in VR. The choice of feedback modality also stands on an ***understanding about VR users'***

perceptual capacity and sensitivity, which may also vary across different demographics, e.g., people with accessibility challenges [262].

Future research direction. Our work opens up several directions for future research, which stem from our findings, methodology, and limitations. First, our recommendations primarily focus on improving the perceived and actual security of VR authentication through effective security communication and infrastructural support. Furthermore, the VR ecosystem involves multiple stakeholders, such as platform providers, application developers, authentication service providers, and banks, each with distinct roles and responsibilities in payment authentication. Coordinating with these stakeholders to systematically enhance security poses an ongoing challenge.

Building upon the recommendations, future research could explore effective ways to assist users in making informed security decisions and implementing security measures, for example, making trade-offs between security and usability of authentication. More specifically, it would be valuable to investigate the optimal level of control to give users versus automating decision-making on their behalf [87].

In our user study's population, the participants primarily consisted of young and tech-savvy individuals who are often early adopters of VR technology. To generalize our findings, it would be valuable for researchers to explore how users from diverse backgrounds, including different age groups and levels of technology expertise, perceive VR authentication [228].

In addition, future work may explore more possibilities of VR interaction for authentication with users involved in the co-design process [316]. Through participatory design studies, experts and users can collaborate to design both the frontend interaction and backend infrastructure of VR authentication systems.

Next, we recommend conducting longitudinal research, such as diary studies [105], to examine users' long-term adoption and usage of VR authentication. Users' security attitudes and behaviors may evolve over time as they interact with the system [166]. These longitudinal studies will provide valuable insights into usability and security issues in real-world scenarios, including how users respond to suspicious activities and threats [76].

5.6 Related Work

User authentication in virtual/augmented reality (VR/AR). Prevailing user authentication methods on smart devices, such as smartphones, use one or more of the following factors: (1) unique knowledge (e.g., PIN or unlock patterns [293]), (2) tokens (e.g., a device with coded ID data [208]), and (3) behavioral and physical biometrics (e.g., gestures and iris [176, 146]). In the context of VR/AR, authentication schemes build upon these methods but offer unique security properties compared to real-world authentication. For knowledge-based authentication in VR/AR, virtual PINs displayed in the 3D space can make shoulder-surfing more difficult [193]. Meanwhile, multiple input modalities can be used to select PINs or draw unlock patterns, such as eye gaze, head pose, controller tapping, and foot movements [301, 213]. Biometric authentication, particularly behavioral biometrics, is a prominent area of research in VR/AR. It leverages the multi-modal input modalities to capture users' biometric traits, such as motion trajectory [150], electromyography [56], eye tracking [330]. Behavioral biometrics are often associated with particular tasks users perform in VR/AR, e.g., throwing a ball [171]. There are also preliminary efforts in exploring token-based authentication mainly for AR, e.g., QR codes. User authentication in VR/AR often requires active interaction with the VR/AR interfaces. This raises usability issues and presents a

tradeoff between security and privacy [262].

Security and privacy perception. Prior research investigated how users perceive the security and privacy of user authentication mechanisms, with a focus on established methods like FIDO2 authentication. Lyastani et al. [186] discovered that users express concerns about security issues related to the loss of authentication tokens. Lassak et al.'s study [156] identified misconceptions among users regarding the storage of biometric data in FIDO2 biometric authentication. These studies highlight the disconnect between users' security perception and the actual security provided by authentication methods.

Recent research has also examined users' security and privacy perception in VR. VR developers and users felt the lack of privacy due to opaque data collection policies [12]. Many users center their concern around the threats from other users, e.g., as a bystander [68]. Additionally, users are worried about potential deception by digital content in VR [158]. Users' security and privacy perception of VR authentication received more attention recently. Stephenson et al. discovered that users often thought VR/AR authentication was as secure as other platforms in their online survey [262].

Users' perceptions of security and privacy are associated with multiple factors, including their interaction with the system. For instance, Distler et al. [72] studied how the user interface (UI) designs impact users' perceived security of mobile e-voting apps. They discovered that inadequate UI feedback and contextual information reduce users' sense of security. Users' security and privacy perceptions also depend on other factors, such as personal experience. Jeong and Chiasson [125] found that children and adults have different interpretations and perceptions of security warnings, e.g., the symbolism of a police officer icon. Differing preconceptions are challenging for establishing trust with the system, even with visual security

clues [264].

User authentication for payment. User authentication plays a crucial role in payment services by preventing fraud and minimizing financial risks. The authentication requirements vary depending on the context, such as using chip cards for in-store transactions or requiring one-time passwords (OTPs) for online shopping [5]. Users' perception of authentication security significantly influences their adoption and use of payment services. Mobile payments have gained popularity due to users associating perceived control and security with user authentication on their device [322]. Similarly, some users also desire enhanced security using biometrics in cryptocurrency wallets [294]. Trust is impacted by users' understanding of how payment services ensure authentication security, such as password confidentiality [329]. Different authentication processes in various geographies can lead to differing security perceptions [94, 43]. The payment environment also affects security perception, with users considering ATM authentication riskier than payments in a restaurant due to their unawareness of attacks [292]. In addition to security, factors such as usability in using a user authentication method also impact the use of associated payment service [145].

Contributions to the literature. To the best of our knowledge, our work is the first to systematically study the interplay between users' interaction experiences and security perception of VR authentication. Prior studies indicated such interplay in other contexts. For example, Khan et al. studied how interruptions of implicit authentication affect users' sense of security [134]. As an emerging interaction technology, VR brings novel interaction experiences and security issues to its users [12, 68]. Thus, it is natural for us to hypothesize that interactions with user authentication in VR play a distinct role in shaping users' security perceptions. However, prior work on user authentication in VR has primarily focused

on different research questions, such as enhancing its security properties [193, 150, 56, 330, 171] or studying the usability and security perception [94, 262, 2] of authentication respectively. Prior work in VR showed preliminary findings, e.g., users' security behaviors changed regarding their virtual surroundings [192]. These findings motivate our in-depth investigation of this interplay. Unlike prior studies based solely on surveys or interviews [262], our study method helped us attain in-context insights by employing "technology probes" [120] for authentication in a realistic VR payment use case. We also discuss how our findings supplement prior studies.

5.7 Conclusion

We presented a technology probe study to investigate how interaction experiences in VR authentication affect users' security perception. We designed four probes, using variants of PIN authentication, a virtual card, and a signature, that represent the paradigms of user authentication. We embedded these probes in the routine payments of a VR archery game. In our user study, we collected participants' responses using surveys regarding interaction experiences, security perceptions, and expectations for authentication. We revealed how participants benefited from the virtualization of authentication in VR and faced unique challenges in interactions, e.g., motion control. Participants encountered difficulties and ambiguity due to VR interactions when transferring their prior authentication knowledge to the VR context. Participants' expectations centered around improving interaction factors with security remained a crucial but secondary factor. However, their expectations were conflicting. We identified tensions in their expectations, which drive our recommendations for future work.

6 CONCLUSION

6.1 Dissertation Summary

Smart devices, such as smart homes and AR/VR, have transformed our living experience. They automate our daily routines and interact with us seamlessly and immersively. However, such convenience comes at the price of security and privacy. The data used for interaction, especially the massive amount of biosignals, carry huge security and privacy implications. Yet, smart device users are still living under security and privacy threats despite the rights granted by recent regulations. This dissertation aims to make security and privacy a first-class citizen in our interactions with smart devices. It contributes in two aspects: (1) designing solutions to enhancing security and privacy control in the systems of smart devices and (2) investigating how smart device users perceive, assess, and react to security and privacy issues. We overcame several challenges presented in prior work. In particular, we enabled usable security and privacy controls for applications, making them compatible with the functionalities and user experience. On the other hand, we advanced the understanding of smart device users' security and privacy considerations that further inform solutions and practices to protect users' security and privacy, which were not readily visible in prior studies.

This dissertation demonstrates my representative contributions. Our work on the privacy of eye-tracking systems in Chapter 2 exemplifies how to empower users to control privacy when preserving real-time utilities. Chapter 3 designs a challenge-response biometric authentication, demonstrating our approach to making user authentication more secure and resilient by mitigating privacy exposure of biometric templates. Beyond designing more secure and private systems, Chapter 4 and 5 advance the knowledge of users' security and privacy considerations. Chapter 4 con-

tributes a systematic framework to analyze how smart home users develop security and privacy considerations and attitudes in real life. Chapter 5 explores how the novel interaction experiences of VR authentication affect users' security perceptions.

6.2 Reflections on My Research Methodology

Through my research, I have developed an interdisciplinary perspective that inspires and guides my endeavors in security and privacy. Firstly, I view security and privacy from a technological standpoint. However, security and privacy are more than just technological problems. Resolving these challenges requires technical solutions and concerted societal and institutional efforts. Recognizing this gap, I have expanded my research horizon beyond designing technical solutions for individual users to guiding security and privacy practices for broader stakeholders.

Addressing the above challenges drives me to adopt diverse methods and skills for system design and user research. As a security and privacy researcher, one core capability shared in both aspects is building proper abstraction and models for security and privacy problems. I explore the problem scope and design space across various system layers, including identifying potential threat vectors, creating appropriate abstractions, and designing solutions for security and privacy. For example, after exploring the security and privacy semantics of biosignal modalities, Chapter 2 designs a privacy control that fits in the software processing stack of the data, and Chapter 3 redesigns the hardware sensing frontend for biometrics. The modeling capabilities extend to my research on understanding users. My approach involves identifying the most suitable platform for measuring users' experiences and reactions, designing measurement probes, and creating a model to contextualize and explain the findings. For instance, Chapter 4 studies smart home users from their online discussion as it is a

non-intrusive medium that yields rich real-life data; Chapter 5 employs technology probes for exploring the open design space of VR authentication in its user study. Qualitative analyses unveil nuanced insights from both studies, effectively shedding light on the underlying patterns of user experiences and behaviors.

6.3 Future Work

We are witnessing a technological revolution in various domains of intelligent, pervasive, and immersive computing. This revolution will not only reform our living experience but also influence society and ourselves. As technology advances, new barriers to security and privacy are emerging. I identify several challenges as examples. First, users embrace an overwhelmingly growing number of intelligent, interconnected digital services that ceaselessly collect private data and automate work and life routines. Recent advances in generative artificial intelligence and "metaverse" are driving this change, yet there is little understanding of how they impact our security and privacy. Second, many digital services are deployed on heterogeneous software and hardware systems without efficient security and privacy solutions built in. Moreover, these platforms will fall short of meeting the security and privacy requirements of emerging applications. Finally, the general public lacks a proper understanding of the security and privacy issues associated with new technologies, and our prior experiences may not directly apply to them. What is more, when spreading among underrepresented populations, digital technologies often slack in accommodating their unique privacy needs. My research agenda will build on my knowledge in the areas of security and privacy, human-centered design, and power-efficient systems (my contributions in research collaborations across disciplines during my Ph.D. study are discussed in Appendix A.3.) to tackle the above challenges as described

below.

Developing autonomous agents for security and privacy. The key to making security and privacy solutions effective is empowering users to recognize security and privacy risks and proper mitigations. I will develop security and privacy agents that intuitively communicate security and privacy to users and aid them in mediating security and privacy conflicts across domains. An example is an AR assistant that informs users to turn off smart speakers during a private conversation. The challenge of accommodating varying security and privacy semantics of digital services (e.g., smart devices and online social platforms) and individuals' security and privacy requirements will be investigated and addressed. I will leverage machine learning to establish security and privacy awareness for autonomous agents from the complex dynamics of user interactions. I will also investigate how these agents should communicate security and privacy naturally by embodying personified interactions (e.g., visual, verbal, and gestural). For broader impact, my research will utilize and optimize these agents to promote security and privacy awareness for people more vulnerable to security and privacy risks, such as children and immigrants.

Building software/hardware systems with security and privacy by design. Security and privacy-by-design systems deepen the root of trust in sensing and computation for numerous critical healthcare, industrial, and transportation services. I will build security and privacy-by-design systems that are efficient, resilient, transparent, and reconfigurable through software and hardware improvements. This will be achieved by designing power-efficient systems for pervasive sensing and computing, improving hardware resiliency to attacks with cross-stack verifiability, and developing hardware-reconfigurable security and privacy protection that reacts to emerging threats quickly. My research will investigate balancing multiple design objectives, including security and privacy, energy efficiency, and connectivity, based on my prior cross-domain, interdisciplinary research.

This direction will also create synergy with the autonomous security and privacy agent, which can map users' requirements to low-level system functions and operations. I will further expand this direction through collaboration with domain experts in building, fabricating, and deploying security and privacy-by-design systems in the real world.

Making security and privacy accessible for everyone. Let alone the lack of security and privacy support for emerging technologies, prior security and privacy research underrepresents vulnerable populations and users with other cultural and social backgrounds, such as immigrants and low-income populations. My future research will investigate the security and privacy needs regarding emerging technologies and further seek solutions tailored to underrepresented demographics. I will first develop instruments to measure people's security and privacy requirements and attitudes, for example, from social media in different cultures. Second, I will build infrastructure to model user behaviors, e.g., using non-intrusive instruments such as eye tracking for people with accessibility issues, and quantify the influences, including security and privacy norms, geopolitics, and regulations, on underrepresented users' security and privacy attitudes at a scale. With such understanding, I will design security and privacy-enhancing tools and tech-support platforms for connecting multiple stakeholders, including underrepresented users, advocates, and government agencies.

A.1 Appendix for Chapter 4

Codebook

Contextual factors

- **Adoption phases**
  - **Consideration of product acquisition:** acquisition may include purchase or non-purchase acquisitions such as sharing others' devices
  - **Acquisition of the product but not in use:** the user acquires or buys the product but not yet in use
  - **(Active) use of a product:** the user has a relevant product in use, including personalization of a routine
  - **Abandonment or transfer of product ownership:** the user transfers product to others or abandons a product
- **Product factors**
  - **Quality requirements**
    - **Compatibility:** compatibility or interoperability of the product
    - **Customization:** evaluation of customization or customizability
    - **Ease of use:** How easy, convenient, or usable the product is to use
    - **Price-performance:** evaluation of price-performance of the product

    - **Reliability or functional flaws:** evaluation of a product's reliability or detection of flaws in the product

- **Technology features**
    - **Dependency on cloud:** product's dependency on cloud to function
    - **Dependency on network:** product's dependency on certain networking/connectivity
    - **Expectation of primary functionality:** user's expectations of the primary functionality of the device
    - **Firmware/app or software integration:** product's integration with a particular app, software, or firmware, including open-source software
    - **General technology:** user's general view of smart home or home automation technology
    - **Hardware design:** hardware design, e.g., encapsulation, of the device
    - **Intelligent virtual assistant:** smart assistant, e.g., Siri/Alexa, features
    - **Level of user control:** evaluation of the level of user control of the product
    - **Notification and status reporting:** Notification, e.g., of a status, features
    - **Product certification:** product is certified to operate within a specified standard
    - **Product or software update:** update of product or the product's software
    - **Remote control:** the ability to remotely control the product
    - **Storage and backup:** evaluation of the storage and backup functions of the product

- **Auxiliary information**
  - **Public information channels**
    - **Customer review:** others' customer reviews of the product or company
    - **"Conspiracy theory":** information perceived as conspiracy
    - **News or reports:** news media or public reports about security and privacy
    - **Privacy policy:** privacy policy or terms of the product
    - **Public relation:** stakeholders', e.g., company, public communications about security and privacy
    - **Social media:** referencing reviews and complaints posted on social media platforms, including Reddit
    - **Threat exploit:** referencing a public report of a threat or exploit related to the device
    - **Regulation and legal protection:** the regulation or legal protection about security and privacy
    - **Unidentified public sources:** referencing an unidentified public information source, e.g., anecdotes
  - **Evidence in real-life interaction**
    - **Suspicious product activity:** unusual activity involving their product, e.g., warning and abnormal status
    - **Customer support:** communication with customer service provided by the company or manufacturer
- **Relevant stakeholders**
  - **Company or brand:** company or brand which manufactures, sells, or provides services for the product

  - **Government or other countries:** governmental bodies or foregin entities that are relevant to product security and privacy
  - **Third parties:** third parties ,e.g., service provider, organizations, or unknown malicious entities., involved
  - **End-users**
    * **Children:** concern about children
    * **Device ownership:** who owns the device
    * **Elderly:** concern about elderly people
    * **Shared device usage:** sharing device usage with multiple users spatially or temporally

- **Security and privacy features**
  - **Account access**
    * **Account management:** the process of managing and/or setting up accounts and the information for the product
    * **Resource authorization:** the capability or privilege that the user has to control and determine the access to their product or product functions
  - **Privacy options**
    * **Opt-in/opt-out:** users' right to opt in or opt out
    * **Remote data retention:** remote retention, e.g., in cloud, of the user's data
    * **Sensitive information:** perceived sensitive personal information with the product
    * **User right to view/edit/delete data:** users' option to view/edit/delete their own information
    * **User right to be informed:** be notified and informed transparently about the practices

- **Safety measures**
  - **Data backup:** critical data is backed up
  - **Fallback options:** fallback options in case of tampering
- **Security features**
  - **Device authentication:** device authentication is required to enable or use a product, function, or software
  - **Device pairing:** pairing two products in the network to establish communication or linkage
  - **Encryption:** communication, data, storage, etc. encrypted for security purposes
  - **Intrusion detection and notification:** monitoring the status and detecting any attempts to invade the system security
  - **Password and strength:** using password and the security level of password
  - **Security compliance test:** security or privacy compliance test enforced on products
  - **Security setting and enforcement:** the function and control with the product to enforce its security, e.g., firewall, VPN, etc.
  - **Security update or patch:** patch of a product for security
  - **User authentication, verification, and identification:** the product or service requires authentication, verification, or identification of users' identity
- **System integration**
  - **Dependency resiliency and vulnerability:** the resiliency or vulnerability is imposed by the product's integration with other products

    * **Device exposure to network:** how the product information, function, or control is exposed to network or other entities connected to the network
    * **Secure system layout and integration:** how the structure or layout of integration of different products or services is set up to ensure security
    * **Server security:** how the server or cloud that hosts the services is secured

Security and privacy concerns

- **Threats**
    - **Backdoor:** Backdoor can be implemented by hackers or malware or implemented by manufacturers to bypass security measures
    - **Data mishandling:** measures to handle data is not proper and leads to concerns
    - **Data misuse:** data is misused and leads to concern
    - **Device hacking:** the threat of hacking smart home products, including malware and botnet
    - **Exploit tracking:** threat involves exploitation of tracking capacity of a product, e.g., location tracking
    - **Level of attacker sophistication:** how sophisticated the attacker is when performing the attack
    - **Permission abuse:** permission or access or the product is abused by others for unauthorized use or purposes
    - **Physical tampering:** the threat on the security of products enabled by physical access and tampering

  - **Ransomware:** ransomware that compromises the product
  - **Scam:** users receive scam that breaches critical information or leads to product compromise
  - **Selling user data:** selling user information leading to privacy concern
  - **Surveillance:** the threat from surveillance, e.g., from companies and governments
  - **(Unwanted) information collection and sharing:** users' unwanted sharing or collection of information that leads to concern
  - **Wireless or remote attack:** the attack is performed remotely and wirelessly, specifically on the network or communication protocol

- **Risks**

  - **Damage to physical properties:** the threat causes a risk in the physical space, like fire hazard, breakin, etc.
  - **Financial risk:** the threat causes an impact financially, like financial loss
  - **Health or human risk:** the threat causes harm on people, such as the health concern or injury
  - **Loss and breach of digital properties:** the threat leads to a loss or breach of digital properties such as data or control of a product/service

Security and privacy protective measures

- **Mitigating an option:** avoid acquiring or using a less secure/private product/option in adoption (purchase, abandonment, etc.) decision-making

- **Adopt fallback:** use fallback or redundancy against tampering
- **Resource authorization:** authorize and manage resource and permission as protection
- **Configure security or privacy setting:** configure a function for security and privacy
- **Secure system layout and integration:** adopt secure system layout and integration to mitigate concern
- **Monitor intrusion:** monitor intrusion as protection, including general advice
- **Disable internet connection:** disabling internet to mitigate concern
- **Contact customer support:** ask customer support for help
- **Unplug power:** unplug power to mitigate use of a product
- **Reflashing system or data:** reflashing system or data to mitigate privacy concern or security risk, e.g., malware

Security and privacy considerations (themes and subthemes for RQ1)

- **Developing security and privacy concerns**
  - **S&P awareness**
    - Contextual factors contribute to awareness
    - Awareness evolves based on changing contextual factors
  - **Threat identification**
    - Technical expertise affects vulnerability assessment
    - Preconceptions shape users' views toward products and stakeholders

        * Assumptions regarding stakeholders' behaviors result in different role assignments
    - **Risk assessment**
        * Users' assumptions about adversaries drive their likelihood assessment
        * Users' valuation of users and assets influences their severity assessment

- **Incorporating protective strategies**
    - **Strategy identification**
        * Users leverage information sources to identify protective strategies
        * Contextual factors constrain the scope of strategies
    - **Tradeoff recognition**
        * Price and technical effort inform cost awareness
        * Benefits of strategies stand on the security and perceived improvements in use cases
    - **Strategy assessment**

Security and privacy attitudes (themes for RQ2)

- **Dismissiveness:** lack of concern, reluctance to incorporate further strategies
- **Exploration:** proactively developing considerations
- **Resignation:** security and privacy gave up in tradeoff
- **Positive pragmatics:** security and privacy satisfied in tradeoff
- **Devotion:** highly devoted to incorporate, usually technically competent

Interaction in discussion

- **Seeking advice or information:** the user is seeking info/input/advice/feedback/opinion
- **Showing expectation and needs:** the user shows specific demand, expectation, or need on a certain product, product feature or functionality
- **Providing information or advice:** the user is sharing info/input/advice/feedback/opinion
- **Adopting opinion or advice:** the user adopts or accepts the recommendation or advice from other users
- **Rejecting opinion or advice:** the user rejects the recommendation or advice from other users
- **Debating on opinion or advice:** the user raises questions or doubts in debate or argumentation with others
- **Self-recognition of experience level:** the user acknowledges their experience or lack of experience (technically)

Influence of online discourse (themes and subthemes for RQ3)

- **Strategies to resolve ambiguity**
  - Collaborative exploration through elaboration
  - Transfer of personal experience to a new context
  - Supplementary information as evidence
- **Contribution to attitude development**
  - The discourse affects S&P concerns

  - The discourse informs alternative strategies

- **Influence on the discourse environment**

  - Opposing attitudes result in topic incoherence
  - Empathy resonates across attitudes
  - Opposing attitudes create social pressure

Smart home product type (built on Amazon's smart home department [15])

- Smart lighting
- Smart locks and openers
- Smart security cameras and systems
- Smart power plugs
- Smart thermostats and climate
- Smart detectors and sensors
- Home entertainment
- Smart voice assistants and hubs
- Smart kitchen tools
- Robotic vacuums
- Smart garden device
- WiFi and networking
- Smart pet products

A.2 Appendix for Chapter 5

Pre-Study Survey

1. What is your age:
2. What is your gender?

- Woman
- Man
- Non-binary
- Prefer not to say
- Other

3. What is your highest education level (on-going or completed)?

- High school diploma
- Bachelor's degree
- Graduate or professional degree
- PhD
- Other

4. What is your major field of study or work:
5. Do you own any virtual reality or augmented reality headsets (e.g., Oculus, HTC Vive, Hololens, Google Glasses, etc.)?

- Yes, virtual reality
- Yes, augmented reality

- No

6. How often do you use or have used VR headsets/devices?

- Very frequently
- Frequently
- Occasionally
- Rarely
- Very rarely
- Never

7. How often do you use or have used AR headsets/devices?

- Very frequently
- Frequently
- Occasionally
- Rarely
- Very rarely
- Never

8. Have you ever developed any applications for AR/VR devices?

- Yes
- No
- Other

9. What types of applications have you experienced in AR/VR?

- Gaming/entertainment
- Socializing/communication
- Education
- Work-related (training, productivity, etc.)
- Finance
- Other

10. How often do you pay via the following methods? (Choices for each method: Very frequently, Frequently, Occasionally, Rarely, Very rarely, Never)

- Cash
- Card present (physically)
- Card on file (card information stored)
- Check
- Smart phone wallet (e.g., Apple pay or Google pay)
- Cash app and wallet (e.g., Venmo or Paypal)
- Card online
- Online banking
- Cryptocurrency wallet
- Game currency/credit system

11. Do you have other payment method you use frequently:
12. What kinds of login do you have experience with for payment, if any? (more than one answer allowed)

- Password
- Unlock pattern
- Signature
- Biometrics (face, iris, fingerprint, etc.)
- Token (e.g., credit card)
- Paired device (e.g., smartphone)
- Paired account (e.g., Chase Bank with Venmo)
- N/A
- Other

13. What kinds of login do you have experience within AR/VR, if any? (more than one answer allowed)

- Password
- Unlock pattern
- Signature
- Biometrics (face, iris, fingerprint, etc.)
- Token (e.g., credit card)
- Paired device (e.g., smartphone)
- Paired account (e.g., Chase Bank with Venmo)
- N/A
- Other

14. ATI scale [24] (Choices for each statement: Completely disagree, Largely disagree, Slightly disagree, Slightly agree, Largely agree, Completely agree)

- I like to occupy myself in greater detail with technical systems.
- I like testing the functions of new technical systems.
- I predominantly deal with technical systems because I have to.
- When I have a new technical system in front of me, I try it out intensively.
- I enjoy spending time becoming acquainted with a new technical system.
- It is enough for me that a technical system works; I don't care how or why.
- I try to understand how a technical system exactly works.
- It is enough for me to know the basic functions of a technical system.
- I try to make full use of the capabilities of a technical system.

In-Study Survey

Note that we refer to PIN-F/PIN-K/TAP/SIGN as pin-to-pay (floating)/ pin-to-pay (on kiosk)/ tap-to-pay/ sign-to-pay in the survey. Participants are asked to respond to the survey for each probe/setup they experience.
1. IPQ scale [121] (statements following by choices for each statement; choices in a seven-point Likert scale)

- In the computer generated world I had a sense of "being there." (Not at all–Very much)

- Somehow I felt that the virtual world surrounded me. (Fully disagree–Fully agree)
- I felt like I was just perceiving pictures. (Fully disagree–Fully agree)
- I did not feel present in the virtual space. (Did not feel–Felt present)
- I had a sense of acting in the virtual space, rather than operating something from outside. (Fully disagree–Fully agree)
- I felt present in the virtual space. (Fully disagree–Fully agree)
- How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)? (Extremely aware-Moderately aware-Not aware at all)
- I was not aware of my real environment. (Fully disagree–Fully agree)
- I still paid attention to the real environment. (Fully disagree–Fully agree)
- I was completely captivated by the virtual world. (Fully disagree–Fully agree)
- How real did the virtual world seem to you? (Completely real–Not real at all)
- How much did your experience in the virtual environment seem consistent with your real world experience? (Not consistent-Moderately consistent-Very consistent)
- How real did the virtual world seem to you? (About as real as an imagined world–Indistinguishable from the real world)
- The virtual world seemed more realistic than the real world. (Fully disagree–Fully agree)

2. SUS [286] (Choice for each statement: Strongly agree to Strongly disagree in a five-point Likert scale)

- I think that I would like to use this system frequently.
- I found the system unnecessarily complex.
- I thought the system was easy to use.
- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.
- I thought there was too much inconsistency in this system.
- I would imagine that most people would learn to use this system very quickly.
- I found the system very cumbersome to use.
- I felt very confident using the system.
- I needed to learn a lot of things before I could get going with this system.

3. I am satisfied with my performance of arrow shooting.

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree

4. Please recall and describe how it was challenging or easy when using tap-to-pay/sign-to-pay/pin-to-pay (floating)/pin-to-pay (on kiosk) here to load the arrows:

Post-Study Survey

1. How did you decide (what motivated you to decide) on how many arrows to load and pay for:
2. Please rate how much you agree with the following statements when you used the four login methods (tap-to-pay/sign-to-pay/pin-to-pay (floating)/pin-to-pay (on kiosk)) to pay for your arrows. Then, please explain why and justify your selections, e.g., by ranking and comparing among the methods. (Choices for each statement: Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree) (*repeat this question for each probe/setup*)

- I felt that I gave consent to this payment
- I felt that I needed to be alerted to this payment
- I felt that my payment was secured
- I felt that I was in control of my payment
- I felt that I gave away privacy in making my payment

3. Please rank the four methods you would prefer to adopt to make payments in virtual reality. (Put the one you prefer the most on the top "1")

- Tap-to-pay
- Sign-to-pay

- Pin-to-pay (floating)
- Pin-to-pay (on kiosk)

Please comment, explain and justify your above ranking, including any ties.
4. What are the qualities, for example, easiness, you would prefer when logging in to pay in virtual reality, and why? Please name top three qualities and explain.
5. How would you improve the four presented methods tap-to-pay, sign-to-pay, pin-to-pay (floating), pin-to-pay (on kiosk), and can you imagine or suggest any other forms of payment in virtual reality? Please explain.

Codebook

- **Seamlessness of interaction**
  - **Easiness and smoothness:** how easy and smooth the interaction is
  - **Comfortability:** interactions make users feel comfortable
  - **Physical, mental, and time efforts:** for example, when users feel there are a lot of actions to make
- **Virtualization of interaction**
  - **Realistic:** how realistic the interaction seems to users
  - **Familiar:** feeling familiar with the interaction
  - **Intuitiveness:** e.g., the interface seems straightforward
- **Interaction components**
  - **Motion control**

    - **Spatial awareness:** people feeling digital objects and surroundings in relation to their avatar in the VR space
    - **Action control and consistency:** people controlling their motion and action in VR, and evaluation of consistency
    - **Interaction modality:** what modality is used to interact, e.g., keyboard, grab, and tap
  - **Authentication interface**
    - **Interface presentation:** how the interface is visualized and displayed
    - **Interaction feedback:** what feedback users receive in response to their actions
  - **Process of authentication**
    - **Learning curve:** how easy users get used to the interaction
    - **Transition in workflow:** how many steps/transitions are there
    - **Knowledge to memorize:** what users should remember for their use of a method

- **Context of payment and VR game**
  - **Presence/immersion:** users feel being present/immersed in VR
  - **Entities in making the payment:** who gets involved in the payment, e.g., service providers
  - **Importance/sensitivity of the payment:** users' sensitivity to security risks

- **Engagement in the game**

  - **Strategy to complete:** users' calculation, e.g., based on confidence, in the game
  - **Enjoyment of the game:** e.g., users having fun in shooting arrows

- **Understanding of the authentication system**
  - **Security and privacy properties of the system**: what measures are implemented for security or privacy, e.g., anonymity
  - **Sense of ownership:** e.g., secret knowledge they own
  - **Security and privacy feedback:** what feedback users can get and how
  - **Trust and confidence:** if users trust or are confident about the system

- **Threat model**
  - **Privacy vulnerability and attacks:** users feel privacy being threatened, for example, data breach
  - **Security vulnerability and attacks**
    * **unauthorized behavior:** e.g., other users access the assets maliciously
    * **malicious objects:** a malicious object in the digital world
    * **Shoulder surfing:** people shoulder surfing others' behavior
  - **Other (malicious) entities**
    * **Threats from the VR surroundings:** for example, an invisible digital object
    * **Other (malicious) users:** e.g., physical and virtual bystanders

    - **Other (malicious) apps/software:** e.g., malware

- **Security and privacy requirements:**
  - **Leveled security:** security can be enhanced depending on the context
  - **Robustness:** being resilient to unexpected behaviors
  - **Cancelability:** having the cancelability to revoke behaviors
- **Transparency and informing:**
  - **Open-sourced:** the method that can be open-sourced
  - **Informing channels:** cues being used to inform users
- **Engagement and creativity:**
  - **Novelty:** how novel the interaction is
  - **Fun, attractiveness, and playfulness:** how fun, attractive, or playful the interaction is

A.3 Contributions in Research Collaborations

I would like to acknowledge my research collaborations during my Ph.D. study and explain my contributions. These collaborations span security and privacy, human-centered and power-efficient systems.

- **Developing usable privacy control for real-time eye-tracking systems [164]:** I am the lead author of this paper. I proposed the idea, designed the system, built and evaluated software prototypes with user studies and trace-based analysis. Amrita Roy Chowdhury helped formulate the theoretical framework of the method.

- **Securing user authentication using nonlinear vibration challenge-responses [165]:** I am the lead author of this paper. I proposed the method, designed and built the hardware/software system, and conducted user experiments.
- **Understanding how smart home users develop security and privacy considerations and attitudes [166]:** I am the lead author of this paper. I proposed the research questions, designed the analysis framework, and led the analysis team. Kaiwen Sun, Brittany Skye Huff, and Anna Marie Bierley contributed to the qualitative coding and thematic analysis.
- **Understanding how interaction experiences influence security perceptions of VR authentication [163]:** I am the lead author of this paper. The work is my internship project at Visa Research. I proposed the idea, led a team to design the experiment and study, developed software prototypes, conducted user studies, and analyzed the data.
- **Acoustic sensor fingerprinting for spoof-resistant mobile device authentication [162]:** Yongwoo Lee is the lead author of this paper. I contributed to the study design, evaluation, and writing.
- **Power-efficient unary computing and its applications [310, 309, 311]:** Di Wu is the lead author of the papers. I contributed to the problem scoping, system and application evaluations, and writing of the papers.
- **Virtualizing and unifying nonlinear operations for emerging neural networks [308]:** Di Wu is the lead author of the paper. I contributed to problem scoping, idea formulation, application evaluation, and writing of the paper.
- **Quality-configurable approximate dividers [32, 198]:** Setarah Behroozi is the lead author of the conference paper [32], and Jackson Melchert

is the lead author of the journal publication [198]. I contributed to the application analysis and paper writing.

- **Monitoring the core body temperature of dairy cows using implantable biosensors and wearable systems [61, 60]:** Hanwook Chung is the lead author of the papers. I designed, built, and tested the electronic and networking systems for the experiments.

BIBLIOGRAPHY

[1] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In *ACM CCS*, 2016.

[2] Y. Abdelrahman, F. Mathis, P. Knierim, A. Kettler, F. Alt, and M. Khamis. Cuevr: Studying the usability of cue-based authentication for virtual reality. In *AVI*, 2022.

[3] N. Abdi, K. M Ramokapane, and J. M. Such. More than smart speakers: Security and privacy perceptions of smart home personal assistants. In *SOUPS*, 2019.

[4] J. M. Abowd and I. M. Schmutte. An economic analysis of privacy protection and statistical accuracy as social choices. *American Economic Review*, 109(1):171–202, 2019.

[5] S. Acharya, A. Polawar, and P. Y. Pawar. Two factor authentication using smartphone generated one time password. *IOSR Journal of Computer Engineering*, 11(2):85–90, 2013.

[6] A. Açık, A. Sarwary, R. Schultze-Kraft, S. Onat, and P. König. Developmental changes in natural viewing behavior: bottom-up and top-down differences between children, young adults and older adults. *Frontiers in Psychology*, 1:207, 2010.

[7] A. Acquisti. Privacy in electronic commerce and the economics of immediate gratification. In *ACM EC*, 2004.

[8] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, et al. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM CSUR*, 50(3):1–41, 2017.

[9] A. Acquisti, L. Brandimarte, and G. Loewenstein. Privacy and human behavior in the age of information. *Science*, 347(6221):509–514, 2015.

[10] A. Acquisti, L. Brandimarte, and G. Loewenstein. Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology*, 30(4):736–758, 2020.

[11] A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *PETS*, 2006.

[12] D. Adams, A. Bah, C. Barwulor, N. Musaby, K. Pitkin, and E. M. Redmiles. Ethics emerging: The story of privacy and security perceptions in virtual reality. In *SOUPS*, 2018.

[13] S. Adewusi, S. Rakheja, P. Marcotte, and J. Boutine. Vibration transmissibility characteristics of the human hand–arm system under different postures, hand forces and excitation levels. *Journal of Sound and Vibration*, 329(14):2953–2971, 2010.

[14] I. Agtzidis, M. Startsev, and M. Dorr. 360-degree video gaze behaviour: A ground-truth data set and a classification algorithm for eye movements. In *ACM MM*, 2019.

[15] Amazon. *Smart Home Devices and Systems*, 2021. `https://www.amazon.com/Smart-Home/b?ie=UTF8&node=6563140011`.

[16] T. Ammari, J. Kaye, J. Y. Tsai, and F. Bentley. Music, search, and IoT: How people (really) use voice assistants. *ACM TOCHI*, 26(3):17–1, 2019.

[17] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems. In *ACM CCS*, 2013.

[18] J. Angulo, E. Wästlund, and J. Högberg. What would it take for you to tell your secrets to a cloud? In *NordSec*, 2014.

[19] I. G. Angus and H. A. Sowizral. Embedding the 2d interaction metaphor in a real 3d virtual environment. In *Stereoscopic Displays and Virtual Reality Systems II*, volume 2409, pages 282–293, 1995.

[20] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, et al. Understanding the Mirai botnet. In *USENIX Security*, 2017.

[21] E. Arabadzhiyska, O. T. Tursun, K. Myszkowski, H. Seidel, and P. Didyk. Saccade landing position prediction for gaze-contingent rendering. *ACM TOG*, 36(4):1–12, 2017.

[22] F. Argelaguet and C. Andujar. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013.

[23] C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein. Low-cost electroencephalogram (EEG) based authentication. In *IEEE/EMBS NER*, 2011.

[24] Ati-scale.org. *ATI Scale — ati-scale.org*, 2023. `https://ati-scale.org/`.

[25] K. Bannier, E. Jain, and O. Le Meur. Deepcomics: Saliency estimation for comics. In *ACM ETRA*, 2018.

[26] S. Barra, M. De Marsico, M. Nappi, F. Narducci, and D. Riccio. A hand-based biometric system in visible light for mobile environments. *Information Sciences*, 479:472–485, 2019.

[27] S. Barth and M. D. T. De Jong. The privacy paradox–investigating discrepancies between expressed privacy concerns and actual online behavior–a systematic literature review. *Telematics and Informatics*, 34(7):1038–1058, 2017.

[28] J. Baumgartner. Pushshift's update on removal requests. `https://twitter.com/jasonbaumgartne/status/1431845831984943104?s=21`, 2021.

[29] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. The Pushshift Reddit dataset. In *ICWSM*, 2020.

[30] W. Becker and A. F. Fuchs. Further properties of the human saccadic system: eye movements and correction saccades with and without visual fixation points. *Vision Research*, 9(10):1247–1258, 1969.

[31] S. A. Beedie, D. M. St. Clair, and P. J. Benson. Atypical scanpaths in schizophrenia: evidence of a trait-or state-dependent phenomenon? *Journal of Psychiatry & Neuroscience*, 36(3):150, 2011.

[32] S. Behroozi, J. Li, J. Melchert, and Y. Kim. Saadi: A scalable accuracy approximate divider for dynamic energy-quality scaling. In *ASP-DAC*, 2019.

[33] R. Bellini, E. Tseng, N. McDonald, R. Greenstadt, D. McCoy, T. Ristenpart, and N. Dell. "So-called privacy breeds evil" narrative justifications for intimate partner surveillance in online forums. *PACM HCI*, 4(CSCW3):1–27, 2021.

[34] S. Berkovsky, R. Taib, I. Koprinska, E. Wang, Y. Zeng, J. Li, and S. Kleitman. Detecting personality traits using eye-tracking data. In *ACM CHI*, 2019.

[35] A. Borji, D. N. Sihite, and L. Itti. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE TIP*, 22(1):55–69, 2012.

[36] E. Bozkir, O. Günlü, W. Fuhl, R. F. Schaefer, and E. Kasneci. Differential privacy for eye tracking with temporal correlations. *arXiv:2002.08972*, 2020.

[37] E. Bozkir, A. B. Ünal, M. Akgün, E. Kasneci, and N. Pfeifer. Privacy preserving gaze estimation using synthetic images via a randomized encoding based framework. *arXiv:1911.07936*, 2019.

[38] R. Brewer, L. Anthony, Q. Brown, G. Irwin, J. Nias, and B. Tate. Using gamification to motivate children to complete empirical studies in lab environments. In *ACM IDC*, 2013.

[39] J. Brooke. SUS: A retrospective. *Journal of Usability Studies*, 8(2):29–40, 2013.

[40] F. Broz, H. Lehmann, B. Mutlu, and Y. Nakano. *Gaze in Human-Robot Communication*, volume 81. John Benjamins Publishing Company, 2015.

[41] P. Bukaty. *The California Consumer Privacy Act (CCPA): An Implementation Guide*. 2019.

[42] G. C. Burdea and P. Coiffet. *Virtual reality technology*. 2003.

[43] K. Busse, M. Tahaei, K. Krombholz, E. von Zezschwitz, M. Smith, J. Tian, and W. Xu. Cash, cards or cryptocurrencies? A study of payment culture in four countries. In *EuroS&PW*, 2020.

[44] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand. What do different evaluation metrics tell us about saliency models? *IEEE TPAMI*, 41(3):740–757, 2018.

[45] S. Castagnos, N. Jones, and P. Pu. Eye-tracking product recommenders' usage. In *ACM RecSys*, 2010.

[46] G. Chalhoub, M. J. Kraemer, N. Nthala, and I. Flechais. "It did not give me an option to decline": A longitudinal analysis of the user experience of security and privacy in smart home products. In *ACM CHI*, 2021.

[47] V. Chandrasekaran, S. Banerjee, B. Mutlu, and K. Fawaz. PowerCut and obfuscator: An exploration of the design space for privacy-preserving interventions for smart speakers. In *SOUPS*, 2021.

[48] G. Charness, U. Gneezy, and M. A. Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1):1–8, 2012.

[49] E. Charters. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal*, 12(2), 2003.

[50] K. Chatzikokolakis, E. Elsalamouny, and C. Palamidessi. Efficient utility improvement for location privacy. In *PETS*, 2017.

[51] K. Chatzikokolakis, C. Palamidessi, and M. Stronati. A predictive differentially-private mechanism for mobility traces. In *PETS*, 2014.

[52] J. Chauhan, Y. Hu, S. Seneviratne, A. Misra, A. Seneviratne, and Y. Lee. Breathprint: Breathing acoustics-based user authentication. In *ACM MobiSys*, 2017.

[53] L. S.-L. Chen. The impact of perceived risk, intangibility and consumer characteristics on online game playing. *Computers in Human Behavior*, 26(6):1607–1613, 2010.

[54] W. Chen, L. Chen, Y. Huang, X. Zhang, L. Wang, R. Ruby, and K. Wu. Taprint: Secure text input for commodity smart wearables. In *MobiCom*, 2019.

[55] Y. Chen, J. Sun, R. Zhang, and Y. Zhang. Your song your way: Rhythm-based two-factor authentication for multi-touch mobile devices. In *IEEE INFOCOM*, 2015.

[56] Y. Chen, Z. Yang, R. Abbou, P. Lopes, B. Y. Zhao, and H. Zheng. User authentication via electrical muscle stimulation. In *ACM CHI*, 2021.

[57] E. K. Choe, S. Consolvo, J. Jung, B. Harrison, and J. A. Kientz. Living in a glass house: A survey of private moments in the home. In *ACM UbiComp*, 2011.

[58] E. K. Choe, S. Consolvo, J. Jung, B. Harrison, S. N. Patel, and J. A. Kientz. Investigating receptiveness to sensing and inference in the home using sensor proxies. In *ACM UbiComp*, 2012.

[59] D. Choi, J. Han, T. Chung, Y. Y. Ahn, B. G. Chun, and T. T. Kwon. Characterizing conversation patterns in Reddit: From the perspectives of content properties and user participation behaviors. In *ACM COSN*, 2015.

[60] H. Chung, J. Li, Y. Kim, and C. Y. Choi. Continuous and wireless skin contact and ear implant temperature measurements and relations to the core body temperature of heat stressed dairy cows. In *ILES*, 2018.

[61] H. Chung, J. Li, Y. Kim, J. M. C. Van Os, S. H. Brounts, and C. Y. Choi. Using implantable biosensors and wearable scanners to monitor dairy cattle's core body temperature in real-time. *Computers and electronics in agriculture*, 174:105453, 2020.

[62] J. Clawson, J. A. Pater, A. D. Miller, E. D. Mynatt, and L. Mamykina. No longer wearing: Investigating the abandonment of personal health-tracking technologies on Craigslist. In *ACM UbiComp*, 2015.

[63] C. Cobb, S. Bhagavatula, K. A. Garrett, A. Hoffman, V. Rao, and L. Bauer. "I would have to evaluate their objections": Privacy tensions between smart home device owners and incidental users. In *PETS*, 2021.

[64] A. Coutrot, J. H. Hsiao, and A. B. Chan. Scanpath modeling and classification with hidden markov models. *Behavior Research Methods*, 50(1):362–379, 2018.

[65] Nora A. D. and Joseph T. The corporate cultivation of digital resignation. *New Media & Society*, 21(8):1824–1839, 2019.

[66] A. Daffalla, L. Simko, T. Kohno, and A. G. Bardas. Defensive technology use by political activists during the Sudanese revolution. In *IEEE S&P*, 2021.

[67] S. Das, L. A. Dabbish, and J. I. Hong. A typology of perceived triggers for end-user security and privacy behaviors. In *SOUPS*, 2019.

[68] J. A. De Guzman, K. Thilakarathna, and A. Seneviratne. Security and privacy approaches in mixed reality: A literature survey. *ACM CSUR*, 52(6):1–37, 2019.

[69] A. De Luca, M. Langheinrich, and He. Hussmann. Towards understanding ATM security: A field study of real world ATM use. In *SOUPS*, 2010.

[70] M. O. Derawi, C. Nickel, P. Bours, and C. Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *IEEE IIH-MSP*, 2010.

[71] R. Dewhurst, M. Nyström, H. Jarodzka, T. Foulsham, R. Johansson, and K. Holmqvist. It depends on how you look at it: Scanpath comparison in multiple dimensions with multimatch, a vector-based approach. *Behavior Research Methods*, 44(4):1079–1100, 2012.

[72] V. Distler, M.-L. Zollinger, C. Lallemand, P. B. Roenne, P. Y. A. Ryan, and V. Koenig. Security-visible, yet unseen? In *ACM CHI*, 2019.

[73] R. G. Dong, Aaron W. Schopper, T. McDowell, D. E. Welcome, J. Wu, W. P. Smutz, C. M. Warren, and S. Rakheja. Vibration energy absorption (VEA) in human fingers-hand-arm system. *Medical engineering & physics*, 26(6):483–492, 2004.

[74] R. G. Dong, J. Wu, and D. E. Welcome. Recent advances in biodynamics of human hand-arm system. *Industrial health*, 43(3):449–471, 2005.

[75] P. Dourish and K. Anderson. Collective information practice: Exploring privacy and security as social and cultural phenomena. *Human-Computer Interaction*, 21(3):319–342, 2006.

[76] J. S. Downs, M. Holbrook, and L. F. Cranor. Behavioral response to phishing risk. In *eCrime*, 2007.

[77] J. L. Dupree, R. Devries, D. M. Berry, and E. Lank. Privacy personas: Clustering users via attitudes and behaviors toward security practices. In *ACM CHI*, 2016.

[78] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(Nos. 3-4):211–407, 2014.

[79] S. Eberz, G. Lovisotto, A. Patane, M. Kwiatkowska, V. Lenders, and I. Martinovic. When your fitness tracker betrays you: Quantifying the predictability of biometric features across contexts. In *IEEE S&P*, 2018.

[80] S. Eberz, G. Lovisotto, K. B. Rasmussen, V. Lenders, and I. Martinovic. 28 blinks later: Tackling practical challenges of eye movement biometrics. In *ACM CCS*, 2019.

[81] P. Emami-Naeini, Y. Agarwal, L. F. Cranor, and H. Hibshi. Ask the experts: What should be on an IoT privacy and security label? In *IEEE S&P*, 2020.

[82] P. Emami-Naeini, M. Degeling, L. Bauer, R. Chow, L. F. Cranor, M. R. Haghighat, and H. Patterson. The influence of friends and experts on privacy decision making in IoT scenarios. *PACM HCI*, 2(CSCW):1–26, 2018.

[83] P. Emami-Naeini, J. Dheenadhayalan, Y. Agarwal, and L. Faith. Cranor. Which privacy and security attributes most impact consumers' risk perception and willingness to purchase IoT devices? In *IEEE S&P*, 2021.

[84] P. Emami-Naeini, H. Dixon, Y. Agarwal, and L. F. Cranor. Exploring how privacy and security factor into IoT device purchase behavior. In *ACM CHI*, 2019.

[85] J. Engel, C. Resnick, A. Roberts, S. Dieleman, M. Norouzi, D. Eck, and K. Simonyan. Neural audio synthesis of musical notes with wavenet autoencoders. In *ICML*, 2017.

[86] S. Eraslan, Y. Yesilada, and S. Harper. Scanpath trend analysis on web pages: Clustering eye tracking scanpaths. *ACM TWEB*, 10(4):1–35, 2016.

[87] D. Filipczuk, T. Baarslag, E. H. Gerding, and M. C. Schraefel. Automated privacy negotiations with preference uncertainty. *Autonomous Agents and Multi-Agent Systems*, 36(2):49, 2022.

[88] C. Flender and G. Müller. Type indeterminacy in privacy decisions: The privacy paradox revisited. In *QI*, 2012.

[89] T. Franke, C. Attig, and D. Wessel. A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ATI) scale. *IJHCI*, 35(6):456–467, 2019.

[90] G. Freeman and D. Maloney. Body, avatar, and me: The presentation and perception of self in social virtual reality. *PACM HCI*, 4(CSCW3):1–27, 2021.

[91] R. Garg and S. Sengupta. He is just like me: A study of the long-term use of smart speakers by parents and children. *ACM IMWUT*, 4(1):1–24, 2020.

[92] GazeRecorder. *GazeRecorder–webcam eye tracking*, 2020. `https://gazerecorder.com`.

[93] C. Geeng and F. Roesner. Who's in control? Interactions in multi-user smart homes. In *ACM CHI*, 2019.

[94] C. George, M. Khamis, E. von Zezschwitz, M. Burger, H. Schmidt, F. Alt, and H. Hussmann. Seamless and secure vr: Adapting and evaluating established authentication systems for virtual reality. In *USEC*, 2017.

[95] A. Gibaldi, M. Vanegas, P. J. Bex, and G. Maiello. Evaluation of the tobii eyex eye tracking controller and matlab toolkit for research. *Behavior Research Methods*, 49(3):923–946, 2017.

[96] G. Guest, A. Bunce, and L. Johnson. How many interviews are enough? an experiment with data saturation and variability. *Field methods*, 18(1):59–82, 2006.

[97] S. R. Gulliver and G. Ghinea. The perceptual and attentive impact of delay and jitter in multimedia delivery. *IEEE Transactions on Broadcasting*, 53(2):449–458, 2007.

[98] M. J. Haass, L. E. Matzen, K. M. Butler, and M. Armenta. A new method for categorizing scanpaths from eye tracking data. In *ACM ETRA*, 2016.

[99] Hackread. *Hackers access 150,000+ security cameras in massive Verkada hack — hackread.com*, 2021. `https://www.hackread.com/hackers-access-security-camera-footage-verkada-hack/`.

[100] I. Hagestedt, M. Backes, and A. Bulling. Adversarial attacks on classifiers for eye-based user modelling. In *ACM ETRA*, 2020.

[101] J. M. Haney, Y. Acar, and S. M. Furman. "It's the company, the government, you and I": User perceptions of responsibility for smart home privacy and security. In *USENIX Security*, 2021.

[102] J. M. Haney, S. M. Furman, and Y. Acar. User perceptions of smart home privacy and security. *NIST Interagency/Internal Report*, 2020.

[103] M. Harbach, M. Hettig, S. Weber, and M. Smith. Using personal examples to improve risk communication for security & privacy decisions. In *ACM CHI*, 2014.

[104] H. Harkous, K. Fawaz, R. Lebret, F. Schaub, K. G. Shin, and K. Aberer. Polisis: Automated analysis and presentation of privacy policies using deep learning. In *USENIX Security*, 2018.

[105] E. Hayashi and J. Hong. A diary study of password usage in daily life. In *ACM CHI*, 2011.

[106] P. He, X. Liu, J. Gao, and W. Chen. DeBERTa: Decoding-enhanced BERT with disentangled attention. In *ICLR*, 2020.

[107] W. He, M. Golla, R. Padhi, J. Ofek, M. Dürmuth, E. Fernandes, and B. Ur. Rethinking access control and authentication for the home Internet of Things (IoT). In *USENIX Security*, 2018.

[108] A. Herzberg. Payments and banking with mobile personal devices. *Communications of the ACM*, 46(5):53–58, 2003.

[109] R. S. Hessels, C. Kemner, C. van den Boomen, and I. T. C. Hooge. The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behavior Research Methods*, 48(4):1694–1712, 2016.

[110] R. S. Hessels, D. C. Niehorster, C. Kemner, and I. T. C. Hooge. Noise-robust fixation detection in eye movement data: Identification by two-means clustering (i2mc). *Behavior Research Methods*, 49(5):1802–1823, 2017.

[111] P. Hock, S. Benedikter, J. Gugenheimer, and E. Rukzio. Carvr: Enabling in-car virtual reality entertainment. In *ACM CHI*, 2017.

[112] C. Holland, A. Garza, E. Kurtova, J. Cruz, and O. Komogortsev. Usability evaluation of eye tracking on an unmodified common tablet. In *ACM CHI EA*, 2013.

[113] C. Holland and O. V. Komogortsev. Biometric identification via eye movement scanpaths in reading. In *IEEE IJCB*, 2011.

[114] J. Holvast. History of privacy. *The History of Information Security*, pages 737–769, 2007.

[115] I. TH. C. Hooge and C. J. Erkelens. Adjustment of fixation duration in visual search. *Vision Research*, 38(9):1295–IN4, 1998.

[116] J. Hsu, M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayan, B. C. Pierce, and A. Roth. Differential privacy: An economic method for choosing epsilon. In *IEEE CSF*, 2014.

[117] J. Hua, W. Tong, F. Xu, and S. Zhong. A geo-indistinguishable location perturbation mechanism for location-based services supporting frequent queries. *IEEE TIFS*, 13(5):1155–1168, 2017.

[118] Y. Huang, B. Obada-Obieh, and K. Beznosov. Amazon vs. my brother: How users of shared smart speakers perceive and cope with privacy risks. In *ACM CHI*, 2020.

[119] H. Hutchinson and et al. Technology probes: Inspiring design for and with families. In *ACM CHI*, 2003.

[120] H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, et al. Technology probes: inspiring design for and with families. In *ACM CHI*, 2003.

[121] Igroup.org. *igroup presence questionnaire (IPQ) overview*, 2023. `http://www.igroup.org/pq/ipq/index.php`.

[122] A. Jain, L. Hong, and S. Pankanti. Biometric identification. *ACM Communications*, 43(2):90–98, 2000.

[123] S. Jalaliniya, D. Mardanbegi, I. Sintos, and D. G. Garcia. Eyedroid: an open source mobile gaze tracker on android for eyewear computers. In *ACM UbiComp*, 2015.

[124] S. Jana, A. Narayanan, and V. Shmatikov. A scanner darkly: Protecting user privacy from perceptual applications. In *IEEE S&P*, 2013.

[125] R. Jeong and S. Chiasson. 'lime','open lock', and'blocked' children's perception of colors, symbols, and words in cybersecurity warnings. In *ACM CHI*, 2020.

[126] M. Jiang, S. Huang, J. Duan, and Q. Zhao. Salicon: Saliency in context. In *IEEE CVPR*, 2015.

[127] B. John, P. Raiturkar, O. Le Meur, and E. Jain. A benchmark of four methods for generating 360° saliency maps from eye tracking data. *IJSC*, 13(03):329–341, 2019.

[128] A. N. Joinson, U. D. Reips, T. Buchanan, and C. B. P. Schofield. Privacy, trust, and self-disclosure online. *Human-Computer Interaction*, 25(1):1–24, 2010.

[129] N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic. Image quality assessment for iris biometric. In *Biometric technology for human identification III*, volume 6202, page 62020D, 2006.

[130] Z. Kapoula, Q. Yang, J. Otero-Millan, S. Xiao, S. L. Macknik, A. Lang, M. Verny, and S. Martinez-Conde. Distinctive features of microsaccades in alzheimer's disease and in mild cognitive impairment. *Age*, 36(2):535–543, 2014.

[131] E. Karapanos, J. Zimmerman, J. Forlizzi, and J. B. Martens. User experience over time: An initial framework. In *ACM CHI*, 2009.

[132] G. Kellaris, S. Papadopoulos, X. Xiao, and D. Papadias. Differentially private event sequences over infinite streams. In *VLDB*, 2014.

[133] M. Khamis, F. Alt, M. Hassib, E. von Zezschwitz, R. Hasholzner, and A. Bulling. Gazetouchpass: Multimodal authentication using gaze and touch on mobile devices. In *ACM CHI EA*, 2016.

[134] H. Khan, U. Hengartner, and D. Vogel. Usability and security perceptions of implicit authentication: convenient, secure, sometimes annoying. In *SOUPS*, 2015.

[135] M. K. Khan, J. Zhang, and X. Wang. Chaotic hash-based fingerprint biometric remote user authentication scheme on mobile devices. *Chaos, Solitons & Fractals*, 35(3):519–524, 2008.

[136] M. Khavkin and M. Last. Preserving differential privacy and utility of non-stationary data streams. In *IEEE ICDMW*, 2018.

[137] C. Kim, W. Tao, N. Shin, and K.-S. Kim. An empirical study of customers' perceptions of security and trust in e-payment systems. *Electronic commerce research and applications*, 9(1):84–95, 2010.

[138] T. Kinnunen, F. Sedlak, and R. Bednarik. Towards task-independent person authentication using eye movement signals. In *ACM ETRA*, 2010.

[139] W. Klippel. Tutorial: Loudspeaker nonlinearities–causes, parameters, symptoms. *Journal of the Audio Engineering Society*, 54(10):907–939, 2006.

[140] S. D. König and E. A. Buffalo. A nonparametric method for detecting fixations and saccades using cluster analysis: Removing the need for arbitrary thresholds. *Journal of Neuroscience Methods*, 227:121–131, 2014.

[141] V. Koshy, J. S. Park, T. C. Cheng, and K. Karahalios. "We just use what they give us": Understanding passenger user perspectives in smart homes. In *ACM CHI*, 2021.

[142] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba. Eye tracking for everyone. In *IEEE CVPR*, 2016.

[143] J. Kropczynski, R. Ghaiumy Anaraky, M. Akter, A. J. Godfrey, H. Lipford, and P. J. Wisniewski. Examining collaborative support for privacy and security in the broader context of tech caregiving. *PACM HCI*, 5(CSCW2):1–23, 2021.

[144] Y. Kudo, A. Tang, K. Fujita, I. Endo, K. Takashima, and Y. Kitamura. Towards balancing vr immersion and bystander awareness. *PACM HCI*, 5(ISS):1–22, 2021.

[145] S. Kujala, R. Mugge, and T. Miron-Shatz. The role of expectations in service evaluation: A longitudinal study of a proximity mobile payment service. *International Journal of Human-Computer Studies*, 98:51–61, 2017.

[146] A. Kumar and A. Passi. Comparison and combination of iris matchers for reliable personal authentication. *Pattern recognition*, 43(3):1016–1026, 2010.

[147] M. Kumar, T. Winograd, and A. Paepcke. Gaze-enhanced scrolling techniques. In *ACM CHI EA*, 2007.

[148] P. Kumar, J. Vitak, M. Chetty, T. L. Clegg, J. Yang, B. McNally, and E. Bonsignore. Co-designing online privacy-related games and stories with children. In *ACM IDC*, 2018.

[149] P. Kumaraguru and L. F. Cranor. *Privacy indexes: A survey of Westin's studies*. CMU-ISRI, 2005.

[150] A. Kupin, B. Moeller, Y. Jiang, N. K. Banerjee, and S. Banerjee. Task-driven biometric authentication of users in virtual reality (vr) environments. In *MMM*, pages 55–67, 2019.

[151] Pupil Labs. *Gaze Datum Format*, 2020. `https://docs.pupil-labs.com/developer/core/overview/#gaze-datum-format`.

[152] E. Lafontaine, A. Sabir, and A. Das. Understanding people's attitude and concerns towards adopting IoT devices. In *ACM CHI EA*, 2021.

[153] D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, and P. Zaphiris. *Human-Computer Interaction–INTERACT 2019*, volume 11748. Springer, 2019.

[154] M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25-26):3559–3565, 2001.

[155] G. Laput, R. Xiao, and C. Harrison. Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. In *UIST*, 2016.

[156] L. Lassak, A. Hildebrandt, M. Golla, and B. Ur. "It's stored, hopefully, on an encrypted server": Mitigating users' misconceptions about {FIDO2} biometric {WebAuthn}. In *USENIX Security*, 2021.

[157] Y. Lau. *JPMorgan bets the metaverse is a $1 trillion yearly opportunity and opens the first virtual bank*, 2022. `https://fortune.com/2022/02/16/jpmorgan-first-bank-join-metaverse/`.

[158] K. Lebeck, K. Ruth, T. Kohno, and F. Roesner. Towards security and privacy for multi-user augmented reality: Foundations with end users. In *IEEE S&P*, 2018.

[159] D. Lee, R. Larose, and N. Rifon. Keeping our network safe: A model of online protection behaviour. *Behaviour & Information Technology*, 27(5):445–454, 2008.

[160] J. Lee and C. Clifton. How much is enough? choosing ϵ for differential privacy. In *ISC*, 2011.

[161] S. Lee, A. El Ali, M. Wijntjes, and P. Cesar. Understanding and designing avatar biosignal visualizations for social virtual reality entertainment. In *ACM CHI*, 2022.

[162] Y. Lee, J. Li, and Y. Kim. Micprint: acoustic sensor fingerprinting for spoof-resistant mobile device authentication. In *EAI MobiQuitous*, 2019.

[163] J. Li, S. S. Arora, K. Fawaz, Y. Kim, C. Liu, S. Meiser, M. Minaei, M. Shirvanian, and K. Wagner. How interactions influence users' security perception of virtual reality authentication? *arXiv:2303.11575*, 2023.

[164] J. Li, A. R. Chowdhury, K. Fawaz, and Y. Kim. Kalεido: Real-time privacy control for eye-tracking systems. In *USENIX Security*, 2021.

[165] J. Li, K. Fawaz, and Y. Kim. Velody: Nonlinear vibration challenge-response for resilient user authentication. In *ACM CCS*, 2019.

[166] J. Li, K. Sun, B. S. Huff, A. M. Bierley, Y. Kim, F. Schaub, and K. Fawaz. "it's up to the consumer to be smart": Understanding the security and privacy attitudes of smart home users on reddit. In *IEEE S&P*, 2023.

[167] T. Li, E. Louie, L. Dabbish, and J. I. Hong. How developers talk about personal data and what it means for user privacy: A case study of a developer forum on Reddit. *PACM HCI*, 4(CSCW3):1–28, 2021.

[168] Y. Li, Z. Cao, and J. Wang. Gazture: Design and implementation of a gaze based gesture control system on tablets. *ACM IMWUT*, 1(3):1–17, 2017.

[169] Z. Li, Z. Yang, C. Song, C. Li, Z. Peng, and W. Xu. E-eye: Hidden electronics recognition through mmwave nonlinear effects. In *ACM SenSys*, 2018.

[170] Z. Liang and B. Ploderer. How does fitbit measure brainwaves: a qualitative study into the credibility of sleep-tracking technologies. *ACM IMWUT*, 4(1):1–29, 2020.

[171] J. Liebers, M. Abdelaziz, L. Mecke, A. Saad, J. Auda, U. Gruenefeld, F. Alt, and S. Schneegass. Understanding user identification in virtual reality through behavioral biometrics and the effect of body normalization. In *ACM CHI*, 2021.

[172] F. Lin, K. W. Cho, C. Song, W. Xu, and Z. Jin. Brain password: A secure and truly cancelable brain biometrics for smart headwear. In *ACM MobiSys*, 2018.

[173] F. Lin, C. Song, Y. Zhuang, W. Xu, C. Li, and K. Ren. Cardiac scan: A non-contact and continuous heart-based user authentication system. In *ACM MobiCom*, 2017.

[174] TY. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014.

[175] A. Liu, L. Xia, A. Duchowski, R. Bailey, K. Holmqvist, and E. Jain. Differential privacy for eye-tracking data. In *ACM ETRA*, 2019.

[176] C. Liu, G. D. Clark, and J. Lindqvist. Where usability and security go hand-in-hand: Robust gesture-based authentication for mobile systems. In *ACM CHI*, 2017.

[177] J. Liu, . Wang, Y. Chen, and N. Saxena. Vibwrite: Towards finger-input authentication on ubiquitous surfaces via physical vibration. In *ACM CCS*, 2017.

[178] L. Liu, H. Li, and M. Gruteser. Edge assisted real-time object detection for mobile augmented reality. In *ACM MobiCom*, 2019.

[179] P. Liu and H. Sohn. Development of nonlinear spectral correlation between ultrasonic modulation components. *NDT & E International*, 91:120–128, 2017.

[180] R. Liu, C. Cornelius, R. Rawassizadeh, R. Peterson, and D. Kotz. Vocal resonance: Using internal body voice for wearable authentication. *ACM IMWUT*, 2(1):19, 2018.

[181] Y.-l. Liu, L. Huang, W. Yan, X. Wang, and R. Zhang. Privacy in AI and the IoT: The privacy concerns of smart speaker users and the personal information protection law in China. *Telecommunications Policy*, 46(7):102334, 2022.

[182] K. Logan. Why isn't everyone doing it? A comparison of antecedents to following brands on Twitter and Facebook. *Journal of Interactive Advertising*, 14(2):60–72, 2014.

[183] T. Lokot and N. Diakopoulos. News bots: Automating news and information dissemination on Twitter. *Digital Journalism*, 4(6):682–699, 2016.

[184] J. Lorenzo-Trueba, F. Fang, X. Wang, I. Echizen, J. Yamagishi, and T. Kinnunen. Can we steal your vocal identity from the internet?: Initial investigation of cloning obama's voice using gan, wavenet and low-quality found data. *arXiv:1803.00860*, 2018.

[185] Z. Lu and H. Shen. Differentially private k-means clustering with guaranteed convergence. *arXiv:2002.01043*, 2020.

[186] S. G. Lyastani, M. Schilling, M. Neumayr, M. Backes, and S. Bugiel. Is FIDO2 the kingslayer of user authentication? a comparative usability study of FIDO2 passwordless authentication. In *IEEE S&P*, 2020.

[187] P. Majaranta and A. Bulling. Eye tracking and eye-based human–computer interaction. In *Advances in Physiological Computing*, pages 39–65. Springer, 2014.

[188] N. Malazizi, H. Alipour, and H. Olya. Risk perceptions of Airbnb hosts: Evidence from a Mediterranean island. *Sustainability*, 10(5):1349, 2018.

[189] S. Malmasi and M. Zampieri. Detecting hate speech in social media. In *RANLP*, 2017.

[190] S. Mare, M. Baker, and J. Gummeson. A study of authentication in daily life. In *SOUPS*, 2016.

[191] S. Marwecki, A. D. Wilson, E. Ofek, M. Gonzalez Franco, and C. Holz. Mise-unseen: Using eye tracking to hide virtual reality scene changes in plain sight. In *ACM UIST*, 2019.

[192] F. Mathis, K. Vaniea, and M. Khamis. Can I borrow your ATM? Using virtual reality for (simulated) in situ authentication research. In *IEEE VR*, pages 301–310, 2022.

[193] F. Mathis, J. H. Williamson, K. Vaniea, and M. Khamis. Fast and secure authentication in virtual reality using coordinated 3d manipulation and pointing. *ACM TOCHI*, 28(1):1–44, 2021.

[194] P. Mayer, Y. Zou, F. Schaub, and A. J. Aviv. "Now I'm a bit angry:" Individuals' awareness, perception, and responses to data breaches that affected them. In *USENIX Security*, 2021.

[195] S. I. McCoy, R. Buzdugan, R. Grimball, L. Natoli, C. M. Mejia, J. D. Klausner, and M. R. McGrath. Stick to it: pilot study results of an intervention using gamification to increase hiv screening among young men who have sex with men in california. *Mhealth*, 4, 2018.

[196] S. A. McMains and S. Kastner. Visual attention. *Encyclopedia of Neuroscience*, 1:4296–4302, 2009.

[197] L. Mecke, K. Pfeuffer, S. Prange, and F. Alt. Open sesame! user perception of physical, biometric, and behavioural authentication concepts to open doors. In *MUM*, 2018.

[198] J. Melchert, S. Behroozi, J. Li, and Y. Kim. Saadi-ec: A quality-configurable approximate divider for energy efficiency. *IEEE TVLSI*, 27(11):2680–2692, 2019.

[199] N. Meng, D. Keküllüoğlu, and K. Vaniea. Owning and sharing: Privacy perceptions of smart speaker users. *PACM HCI*, 5(CSCW1):1–29, 2021.

[200] N. Micallef and N. A. G. Arachchilage. A gamified approach to improve users' memorability of fall-back authentication. In *SOUPS*, 2017.

[201] Microsoft. *Scene understanding SDK overview*, 2020. `https://docs.microsoft.com/en-us/windows/mixed-reality/develop/platform-capabilities-and-apis/scene-understanding-SDK`.

[202] A. E. Millen and P. J. B. Hancock. Eye see through you! eye tracking unmasks concealed face recognition despite countermeasures. *Cognitive Research: Principles and Implications*, 4(1):23, 2019.

[203] E. Miluzzo, T. Wang, and A. T. Campbell. Eyephone: activating mobile phones with your eyes. In *ACM MobiHeld*, 2010.

[204] C. H. Morimoto and M. R. M. Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1):4–24, 2005.

[205] S. Mukherjee and P. K. Bala. Detecting sarcasm in customer tweets: an NLP based approach. *Industrial Management & Data Systems*, 2017.

[206] C. Müller, W. Stoll, and F. Schmäl. The effect of optical devices and repeated trials on the velocity of saccadic eye movements. *Acta Oto-Laryngologica*, 123(4):471–476, 2003.

[207] P. Negi, P. Sharma, V. Jain, and B. Bahmani. K-means++ vs. behavioral biometrics: One loop to rule them all. In *NDSS*, 2018.

[208] P. Nguyen, U. Muncuk, A. Ashok, K. R. Chowdhury, M. Gruteser, and T. Vu. Battery-free identification token for touch sensing devices. In *ACM SenSys*, 2016.

[209] M. Nilsson, A. Adams, and S. Herd. Building security and trust in online banking. In *ACM CHI EA*, 2005.

[210] K. Nissim, S. Raskhodnikova, and A. Smith. Smooth sensitivity and sampling in private data analysis. In *ACM SOTC*, 2007.

[211] P. Norman, H. Boer, E. R. Seydel, and B. Mullan. Protection motivation theory. *Predicting and Changing Health Behavior*, pages 70–106, 2015.

[212] J. O'Hagan, J. R. Williamson, F. Mathis, M. Khamis, and M. McGill. Re-evaluating vr user awareness needs during bystander interactions. In *ACM CHI*, 2023.

[213] I. Olade, H.-N. Liang, C. Fleming, and C. Champion. Exploring the vulnerabilities and advantages of swipe or pattern authentication in virtual reality (VR). In *ICVARS*, 2020.

[214] T. Olsson, E. Lagerstam, T. Kärkkäinen, and K. Väänänen-Vainio-Mattila. Expected user experience of mobile augmented reality services: a user study in the context of shopping centres. *Personal and ubiquitous computing*, 17:287–304, 2013.

[215] S. Oya, C. Troncoso, and F. Pérez-González. Is geo-indistinguishability what you are looking for? In *ACM WPES*, 2017.

[216] S. Panjwani and A. Prakash. Crowdsourcing attacks on biometric systems. In *SOUPS*, 2014.

[217] A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays. Webgazer: Scalable webcam eye tracking using user interactions. In *IJCAI*, 2016.

[218] J. Park, N. Pažin, J. Friedman, V. M. Zatsiorsky, and M. L. Latash. Mechanical properties of the human hand digits: Age-related differences. *Clinical Biomechanics*, 29(2):129–137, 2014.

[219] S. Parkin, E. M. Redmiles, L. Coventry, and M. A. Sasse. Security when it is welcome: Exploring device purchase as an opportune moment for security behavior change. In *USEC*, 2019.

[220] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn. Towards foveated rendering for gaze-tracked virtual reality. *ACM TOG*, 35(6):179, 2016.

[221] D. L. Phillips and K. J. Clancy. Some effects of "Social desirability" in survey studies. *American journal of sociology*, 77(5):921–940, 1972.

[222] R. Pieters, E. Rosbergen, and M. Wedel. Visual attention to repeated print advertising: A test of scanpath theory. *Journal of Marketing Research*, 36(4):424–438, 1999.

[223] A. Presas, D. Valentin, E. Egusquiza, C. Valero, M. Egusquiza, and M. Bossio. Accurate determination of the frequency response function of submerged and confined structures by using pzt-patches. *Sensors*, 17(3):660, 2017.

[224] N. Proferes, N. Jones, S. Gilbert, C. Fiesler, and M. Zimmer. Studying Reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media+ Society*, 7(2):20563051211019004, 2021.

[225] Pushshift.io. `https://pushshift.io/`, 2022.

[226] F. Raja, K. Hawkey, S. Hsu, K.-L. C. Wang, and K. Beznosov. A brick wall, a locked door, and a bandit: a physical security metaphor for firewall warnings. In *SOUPS*, 2011.

[227] P. Rajivan and J. Camp. Influence of privacy attitude and privacy cue framing on android app choices. In *SOUPS*, 2016.

[228] D. K. Ratakonda, T. French, and J. A. Fails. My name is my password: Understanding children's authentication practices. In *ACM IDC*, 2019.

[229] N. Raval, A. Machanavajjhala, and J. Pan. Olympus: Sensor privacy through utility aware obfuscation. In *PETS*, 2019.

[230] N. Raval, A. Srivastava, K. Lebeck, L. Cox, and A. Machanavajjhala. Markit: Privacy markers for protecting visual secrets. In *ACM UbiComp*, 2014.

[231] K. Rayner, M. S. Castelhano, and J. Yang. Eye movements when looking at unusual/weird scenes: Are there cultural differences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1):254, 2009.

[232] M. Rébillat, R. Hennequin, E. Corteel, and B. F. G. Katz. Identification of cascade of hammerstein models for the description of nonlinearities in vibrating devices. *Journal of Sound and Vibration*, 330(5):1018–1038, 2011.

[233] Reddit. Reddit Content Policy. `https://www.redditinc.com/policies/content-policy`, 2022.

[234] Reddit. Subreddit Search. https://www.reddit.com/subreddits/search?q=smart+home, 2022.

[235] Reddit. /r/homeautomation. https://www.reddit.com/r/HomeAutomation, 2023.

[236] Reddit. /r/homeautomation wiki. https://www.reddit.com/r/HomeAutomation/wiki/index, 2023.

[237] E. M. Redmiles, N. Warford, A. Jayanti, A. Koneru, S. Kross, M. Morales, R. Stevens, and M. L. Mazurek. A comprehensive quality evaluation of security and privacy advice on the web. In *USENIX Security*, 2020.

[238] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv:1804.02767*, 2018.

[239] K. Reese, T. Smith, J. Dutson, J. Armknecht, J. Cameron, and K. Seamons. A usability study of five two-factor authentication methods. In *SOUPS 2019*, 2019.

[240] P. M. Regan. Privacy as a common good in the digital world. *Information, Communication & Society*, 5(3):382–405, 2002.

[241] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NIPS*, 2015.

[242] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10(1–3):19–41, 2000.

[243] A. Rizzo, T. D. Parsons, B. Lange, P. Kenny, J. G. Buckwalter, B. Rothbaum, J. Difede, J. Frazier, B. Newman, J. Williams, et al. Virtual reality goes to war: A brief review of the future of military behavioral healthcare. *Journal of clinical psychology in medical settings*, 18(2):176–187, 2011.

[244] F. Roesner, D. Molnar, A. Moshchuk, T. Kohno, and H. J. Wang. World-driven access control for continuous sensing. In *ACM CCS*, 2014.

[245] R. Ross, V. Pillitteri, K. Dempsey, M. Riddle, and G. Guissanie. Protecting controlled unclassified information in nonfederal systems and organizations. Technical report, National Institute of Standards and Technology, 2019.

[246] K. Ruth, T. Kohno, and F. Roesner. Secure {Multi-User} content sharing for augmented reality applications. In *USENIX Security*, 2019.

[247] M. Saeed, S. Ali, J. Blackburn, E. De Cristofaro, S. Zannettou, and G. Stringhini. Trollmagnifier: Detecting state-sponsored troll accounts on Reddit. In *IEEE S&P*, 2022.

[248] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *ACM ETRA*, 2000.

[249] A. Sanchez, C. Vazquez, C. Marker, J. LeMoult, and J. Joormann. Attentional disengagement predicts stress recovery in depression: An eye-tracking study. *Journal of Abnormal Psychology*, 122(2):303, 2013.

[250] B. Saunders, J. Sim, T. Kingstone, S. Baker, J. Waterfield, B. Bartlam, H. Burroughs, and C. Jinks. Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity*, 52(4):1893–1907, 2018.

[251] B. Saunders, J. Sim, T. Kingstone, S. Baker, J. Waterfield, B. Bartlam, H. Burroughs, and C. Jinks. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity*, 52(4):1893–1907, 2018.

[252] V. Schwind, P. Knierim, N. Haas, and N. Henze. Using presence questionnaires in virtual reality. In *ACM CHI*, 2019.

[253] H. Shen, C. Faklaris, H. Jin, L. Dabbish, and J. I. Hong. 'I can't even buy apples if i don't use mobile pay?' When mobile payments become infrastructural in China. *PACM HCI*, 4(CSCW2):1–26, 2020.

[254] M. Sherman, G. Clark, Y. Yang, S. Sugrim, A. Modig, J. Lindqvist, A. Oulasvirta, and T. Roos. User-generated free-form gestures for authentication: Security and memorability. In *ACM MobiSys*, 2014.

[255] J. S. Silk, L. R. Stroud, G. J. Siegle, R. E. Dahl, K. H. Lee, and E. E. Nelson. Peer acceptance and rejection through the eyes of youth: pupillary, eyetracking and ecological data from the chatroom interact task. *Social Cognitive and Affective Neuroscience*, 7(1):93–105, 2012.

[256] M. Siqueiros Sanchez, E. Pettersson, D. P. Kennedy, S. Bölte, P. Lichtenstein, B. M. D'Onofrio, and T. Falck-Ytter. Visual disengagement: Genetic architecture and relation to autistic traits in the general population. *Journal of Autism and Developmental Disorders*, 2019.

[257] I. Sluganovic, M. Roeschlin, K. B. Rasmussen, and I. Martinovic. Using reflexive eye movements for fast challenge-response authentication. In *ACM CCS*, 2016.

[258] Y. Song, Z. Cai, and Z.-L. Zhang. Multi-touch authentication using hand geometry and behavioral information. In *IEEE S&P*, 2017.

[259] M. Speicher, S. Cucerca, and A. Krüger. Vrshop: a mobile interactive virtual reality shopping environment combining the benefits of on- and offline shopping. *ACM IMWUT*, 1(3):1–31, 2017.

[260] P. C. Stacey, S. Walker, and J. D. M. Underwood. Face processing and familiarity: Evidence from eye-movement data. *British Journal of Psychology*, 96(4):407–422, 2005.

[261] J. Steil, I. Hagestedt, M. X. Huang, and A. Bulling. Privacy-aware eye tracking using differential privacy. In *ACM ETRA*, 2019.

[262] S. Stephenson, B. Pal, S. Fan, E. Fernandes, Y. Zhao, and R. Chatterjee. Sok: Authentication in augmented and virtual reality. In *IEEE S&P*, 2022.

[263] W. Steptoe, R. Wolff, A. Murgia, E. Guimaraes, J. Rae, P. Sharkey, D. Roberts, and A. Steed. Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments. In *ACM CSCW*, 2008.

[264] C. Stransky, D. Wermke, J. Schrader, N. Huaman, Y. Acar, A. L. Fehlhaber, M. Wei, B. Ur, and S. Fahl. On the limited impact of visualizing encryption: Perceptions of E2E messaging security. In *SOUPS 2021*, 2021.

[265] A. Strauss and J. Corbin. *Basics of Qualitative Research Techniques*. Sage, 1998.

[266] G. E. Suh and S. Devadas. Physical unclonable functions for device authentication and secret key generation. In *DAC*, 2007.

[267] K. Sun, Y. Zou, J. Radesky, C. Brooks, and F. Schaub. Child safety in the smart home: Parents' perceptions, needs, and mitigation strategies. *PACM HCI*, 5(CSCW2):1–41, 2021.

[268] M. Tabassum, T. Kosinski, and H. R. Lipford. "I don't own the data": End user perceptions of smart home device data practices and risks. In *SOUPS*, 2019.

[269] M. Tahaei, A. Frik, and K. Vaniea. Privacy champions in software teams: Understanding their motivations, strategies, and challenges. In *ACM CHI*, 2021.

[270] M. Tahaei, T. Li, and K. Vaniea. Understanding privacy-related advice on Stack Overflow. *PETS*, 2022.

[271] M. Tahaei, K. Vaniea, and N. Saphra. Understanding privacy-related questions on Stack Overflow. In *ACM CHI*, 2020.

[272] A. Talarico, M. Malvezzi, and D. Prattichizzo. Modeling the human touch: A fem model of the human hand fingertips for haptic application. In *COMSOL Conference*, 2014.

[273] K. Tang, K. Gerling, and L. Geurts. Virtual feed: A simulated breastfeeding experience in virtual reality. In *ACM CHI EA*, 2021.

[274] K. Tang, K. Gerling, V. Vanden Abeele, L. Geurts, and M. Aufheimer. Playful reflection: Impact of gamification on a virtual reality simulation of breastfeeding. In *ACM CHI*, 2023.

[275] P. Termsarasab, T. Thammongkolchai, J. C. Rucker, and S. J. Frucht. The diagnostic value of saccades in movement disorder patients: a practical guide and review. *Journal of Clinical Movement Disorders*, 2(1):14, 2015.

[276] M. Teyssier, M. Koelle, P. Strohmeier, B. Fruchard, and J. Steimle. Eyecam: Revealing relations between humans and sensing devices through an anthropomorphic webcam. In *ACM CHI*, 2021.

[277] Tobii. *Scripting API of Tobii Unity SDK*, 2020. `https://developer.tobii.com/pc-gaming/unity-sdk/scripting-api/`.

[278] G. F. Tondello, A. Mora, A. Marczewski, and L. E. Nacke. Empirical validation of the gamification user types hexad scale in english and spanish. *International Journal of Human-Computer Studies*, 127:95–111, 2019.

[279] T. Toyama, D. Sonntag, A. Dengel, T. Matsuda, M. Iwamura, and K. Kise. A mixed reality head-mounted text translation system using eye gaze input. In *ACM IUI*, 2014.

[280] E. Tseng, R. Bellini, N. McDonald, M. Danos, R. Greenstadt, D. McCoy, N. Dell, and T. Ristenpart. The tools and tactics used in intimate partner surveillance: An analysis of online infidelity forums. In *USENIX Security*, 2020.

[281] W.-J. Tseng, E. Bonnail, M. Mcgill, M. Khamis, E. Lecolinet, S. Huron, and J. Gugenheimer. The dark side of perceptual manipulations in virtual reality. In *ACM CHI*, 2022.

[282] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders. Selective search for object recognition. *IJCV*, 104(2):154–171, 2013.

[283] Unity. *Scripting Reference of XR.Eyes*, 2020. `https://docs.unity3d.com/ScriptReference/XR.Eyes.html`.

[284] Unity. *Survival shooter tutorial*, 2020. `https://learn.unity.com/project/survival-shooter-tutorial/?tab=overview`.

[285] Unity. *Unity Scripting API: Keyframe*, 2020. `https://docs.unity3d.com/ScriptReference/Keyframe.html`.

[286] Usability.gov. *System Usability Scale (SUS)*, 2023. `https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html`.

[287] A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv:1609.03499*, 2016.

[288] J. Varona, C. Manresa-Yee, and F. J. Perales. Hands-free vision-based interface for computer accessibility. *Journal of Network and Computer Applications*, 31(4):357–374, 2008.

[289] S. Venugopalan, F. Juefei-Xu, B. Cowley, and M. Savvides. Electromyograph and keystroke dynamics for spoof-resistant biometric authentication. In *IEEE CVPR Workshops*, 2015.

[290] P. Voigt and A. Von dem Bussche. The EU general data protection regulation (GDPR). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 10(3152676):10–5555, 2017.

[291] A. Voishvillo, A. Terekhov, E. Czerwinski, and S. Alexandrov. Graphing, interpretation, and comparison of results of loudspeaker nonlinear distortion measurements. *Journal of the Audio Engineering Society*, 52(4):332–357, 2004.

[292] M. Volkamer, A. Gutmann, K. Renaud, P. Gerber, and P. Mayer. Replication study: A {Cross-Country} field observation study of real world {PIN} usage at {ATMs} and in various electronic payment scenarios. In *SOUPS*, 2018.

[293] E. Von Zezschwitz, P. Dunphy, and A. De Luca. Patterns in the wild: a field study of the usability of pattern and pin-based authentication on mobile devices. In *MobileHCI*, 2013.

[294] A. Voskobojnikov, O. Wiese, M. Mehrabi Koushki, V. Roth, and K. Beznosov. The u in crypto stands for usable: An empirical study of user experience with mobile cryptocurrency wallets. In *ACM CHI*, 2021.

[295] D. Votipka, M. N. Punzalan, S. M. Rabin, Y. Tausczik, and M. L. Mazurek. An investigation of online reverse engineering community discussions in the context of ghidra. In *IEEE EuroS&P*, 2021.

[296] D. Walker and F. Myrick. Grounded theory: An exploration of process and procedure. *Qualitative health research*, 16(4):547–559, 2006.

[297] L. Wang, K. Huang, K. Sun, W. Wang, C. Tian, L. Xie, and Q. Gu. Unlock with your heart: Heartbeat-based authentication on commercial mobile phones. *ACM IMWUT*, 2(3):140, 2018.

[298] R. J. Wang, X. Li, and C. X. Ling. Pelee: A real-time object detection system on mobile devices. In *NIPS*, 2018.

[299] H. Watson, E. Moju-Igbene, A. Kumari, and S. Das. "We hold each other accountable": Unpacking how social groups approach cybersecurity and privacy together. In *ACM CHI*, 2020.

[300] J. Watson, H. R. Lipford, and A. Besmer. Mapping user preference to privacy default settings. *ACM TOCHI*, 22(6):1–20, 2015.

[301] K. Watson, R. Bretin, M. Khamis, and F. Mathis. The feet in human-centred security: Investigating foot-based user authentication for public displays. In *ACM CHI EA*, 2022.

[302] J. B. Whiting, R. D. Olufuwote, J. D. Cravens-Pickens, and A. Banford Witting. Online blaming and intimate partner violence: A content analysis of social media comments. *The Qualitative Report*, 2019.

[303] A. Whitten and J. D. Tygar. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *USENIX Security*, 1999.

[304] S. Wiefling, M. Dürmuth, and L. Lo Iacono. More than just good passwords? a study on usability and security perceptions of risk-based authentication. In *ACSAC*, 2020.

[305] N. Wilming, S. Onat, J. P. Ossandón, A. Açık, T. C. Kietzmann, K. Kaspar, R. R. Gameiro, A. Vormberg, and P. König. An extensive dataset of eye movements during viewing of complex images. *Scientific Data*, 4(1):1–11, 2017.

[306] F. Wolf, R. Kuber, and A. J. Aviv. "Pretty close to a must-have:" Balancing usability desire and security concern in biometric adoption. In *ACM CHI*, 2019.

[307] A. Woodruff, V. Pihur, S. Consolvo, L. Brandimarte, and A. Acquisti. Would a privacy fundamentalist sell their DNA for $1000... if nothing bad happened as a result? the Westin categories, behavioral intentions, and consequences. In *SOUPS*, 2014.

[308] D. Wu, J. Li, S. Behroozi, Y. Kim, and J. San Miguel. UNO: Virtualizing and unifying nonlinear operations for emerging neural networks. In *IEEE/ACM ISLPED*, 2021.

[309] D. Wu, J. Li, Z. Pan, Y. Kim, and J. San Miguel. uBrain: A unary brain computer interface. In *ACM/IEEE ISCA*, 2022.

[310] D. Wu, J. Li, R. Yin, H. Hsiao, Y. Kim, and J. San Miguel. uGEMM: Unary computing architecture for gemm applications. In *ACM/IEEE ISCA*, 2020.

[311] D. Wu, J. Li, R. Yin, H. Hsiao, Y. Kim, and J. San Miguel. uGEMM: Unary computing for gemm applications. *IEEE Micro*, 41(3):50–56, 2021.

[312] Z. Wu, J. Yamagishi, T. Kinnunen, C. Hanilçi, M. Sahidullah, A. Sizov, N. Evans, and M. Todisco. ASVspoof: the automatic speaker verification spoofing and countermeasures challenge. *IEEE Journal of Selected Topics in Signal Processing*, 11(4):588–604, 2017.

[313] W. Xie, A. Fowler-Dawson, and A. Tvauri. Revealing the relationship between rational fatalism and the online privacy paradox. *Behaviour & Information Technology*, 38(7):742–759, 2019.

[314] S. Xu, H. Jiang, and F. C. M. Lau. Personalized online document, image and video recommendation via commodity eye-tracking. In *ACM RecSys*, 2008.

[315] X. Yan, S. Raj, B. Huang, S. Y. Park, and M. W. Newman. Toward lightweight in-situ self-reporting: An exploratory study of alternative smartwatch interface designs in context. *ACM IMWUT*, 4(4):1–22, 2020.

[316] Y. Yao, J. R. Basdeo, S. Kaushik, and Y. Wang. Defending my castle: A co-design study of privacy mechanisms for smart homes. In *ACM CHI*, 2019.

[317] Z. Ye, Y. Li, A. Fathi, Y. Han, A. Rozga, G. D. Abowd, and J. M. Rehg. Detecting eye contact using wearable eye-tracking glasses. In *ACM Ubicomp*, 2012.

[318] X. Yi and N. Ling. Fast pixel-based video scene change detection. In *IEEE ISCAS*, 2005.

[319] V. Y. Zaitsev, L. A. Matveev, and A. Matveyev. Elastic-wave modulation approach to crack detection: Comparison of conventional modulation and higher-order interactions. *NDT & E International*, 44(1):21–31, 2011.

[320] E. Zeng, S. Mare, and F. Roesner. End user security and privacy concerns with smart homes. In *SOUPS*, 2017.

[321] E. Zeng and F. Roesner. Understanding and improving security and privacy in multi-user smart homes: A design exploration and in-home user study. In *USENIX Security*, 2019.

[322] J. Zhang, Y. Luximon, and Y. Song. The role of consumers' perceived security, perceived control, interface design features, and conscientiousness in continuous use of mobile payment services. *Sustainability*, 11(23):6843, 2019.

[323] L. Zhang, S. Tan, and J. Yang. Hearing your voice is not enough: An articulatory gesture based liveness detection for voice authentication. In *ACM CCS*, 2017.

[324] L. Zhang, S. Tan, J. Yang, and Y. Chen. Voicelive: A phoneme localization based liveness detection for voice authentication on smartphones. In *ACM CCS*, 2016.

[325] Z. Zhang, H. Wang, A. V. Vasilakos, and H. Fang. ECG-cryptography and authentication in body area networks. *IEEE Trans. Inf. Technol. Biomed.*, 16(6):1070–1078, 2012.

[326] J. C. Zhao, R. C. Davis, P. S. Foong, and S. Zhao. Cofaçade: A customizable assistive approach for elders and their helpers. In *ACM CHI*, 2015.

[327] Y. Zhao, D. Yuan, J. T. Du, and J. Chen. Geo-ellipse-indistinguishability: community-aware location privacy protection for directional distribution. *IEEE TKDE*, 35(7):6957–6967, 2022.

[328] S. Zheng, N. Apthorpe, M. Chetty, and N. Feamster. User perceptions of smart home IoT privacy. *PACM HCI*, 2(CSCW):1–20, 2018.

[329] T. Zhou. The effect of initial trust on user adoption of mobile payment. *Information development*, 27(4):290–300, 2011.

[330] H. Zhu, W. Jin, M. Xiao, S. Murali, and M. Li. Blinkey: A two-factor user authentication method for virtual reality devices. *ACM IMWUT*, 4(4):1–29, 2020.

[331] J. J. Zhu, Y. C. Chang, C. H. Ku, S. Y. Li, and C. J. Chen. Online critical review classification in response strategy and service provider rating: Algorithms from heuristic processing, sentiment analysis to deep learning. *Journal of Business Research*, 129:860–877, 2021.

[332] V. Zimmermann, M. Bennighof, M. Edel, O. Hofmann, J. Jung, and M. von Wick. 'Home, smart home'–Exploring end users' mental models of smart homes. In *Mensch und Computer*, 2018.

[333] V. Zimmermann and N. Gerber. The password is dead, long live the password–a laboratory study on user perceptions of authentication schemes. *International Journal of Human-Computer Studies*, 133:26–44, 2020.

[334] Y. Zou, M. Zhao, Z. Zhou, J. Lin, M. Li, and K. Wu. Bilock: User authentication via dental occlusion biometrics. *ACM IMWUT*, 2(3):152, 2018.