The Unprecedented Risks and Opportunities of Extended Reality Motion Data

By

Vivek Chinar Nair

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Dawn Song, Chair
Professor James F. O'Brien
Associate Professor Björn Hartmann
Dr. Louis B. Rosenberg

Fall 2023

The Unprecedented Risks and Opportunities of Extended Reality Motion Data

Abstract

The Unprecedented Risks and Opportunities of Extended Reality Motion Data

by

Vivek Chinar Nair

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Dawn Song, Chair

The adoption of virtual reality (VR) technologies has rapidly gained momentum in recent years as companies around the world begin to position the "metaverse" as the next major medium of human-computer interaction. The latest generation of VR devices, including the Apple Vision Pro and Meta Quest 3, blur the lines between virtual and augmented reality (AR), resulting in extended reality (XR) systems that are expected to be more deeply and seamlessly integrated with our daily lives than ever before. As companies with a clouded reputation for respecting user privacy become increasingly involved in XR development, the attention of researchers and the general public is rightly shifting toward the unique security and privacy threats that these platforms may pose.

Motion tracking "telemetry" data lies at the core of nearly all modern XR and metaverse experiences. While it has long been known that people reveal information about themselves via their motion, the extent to which these findings apply to XR platforms has, until recently, not been widely understood, with most users perceiving motion to be amongst the more innocuous categories of data in XR. Contrary to these perceptions, this dissertation explores the unprecedented risks and opportunities of XR motion data. We present both a series of attacks that illustrate the severity of the XR privacy threat and a set of defensive countermeasures to protect user privacy in XR while maintaining a positive user experience.

We first present a detailed systematization of the landscape of VR privacy attacks and defenses by proposing a comprehensive taxonomy of data attributes, information flow, adversaries, and countermeasures based an analysis of over 60 prior studies. We then identify and describe a novel dataset of over 4.7 million motion capture recordings, voluntarily submitted by over 105,000 XR device users from over 50 countries. In addition to being over 200 times larger than the largest prior motion capture research dataset, this data is critical to enabling several major contributions throughout this dissertation.

First, using our new dataset, we show that a large number of real VR users (N=55,541) can be uniquely and reliably identified across multiple sessions using just their head and hand motion relative to virtual objects. After training a classification model on 5 minutes of data per person, a user can be uniquely identified amongst the entire pool of 55,541 with 94.33% accuracy from 100 seconds of motion, and with 73.20% accuracy from just 10 seconds of motion. Then, we go a step further, showing that a variety of private user information can be inferred just by analyzing motion data recorded from VR devices. After conducting a large-scale survey of VR users (N=1,006) with dozens of questions ranging from background and demographics to behavioral patterns and health information, we demonstrate that simple machine learning models can accurately and consistently infer over 40 personal attributes from VR motion data alone. In a third study, we show that adversarially designed VR games can infer an even wider range of attributes than can be observed by passive observation alone. After inviting 50 study participants to play an innocent-looking "escape room" game in VR, we show that an adversarial program could accurately infer over 25 of their data attributes, from anthropometrics like height and wingspan to demographics like age and gender.

While users have, to some extent, grown accustomed to privacy attacks on the web, metaverse platforms carry many of the privacy risks of the conventional internet (and more) while at present offering few of the defensive utilities that users are accustomed to having access to. To remedy this, we present the first known method of implementing an "incognito mode" for VR. Our technique leverages local ε-differential privacy to quantifiably obscure sensitive user data attributes, with a focus on intelligently adding noise when and where it is needed most to maximize privacy while minimizing usability impact. However, we then demonstrate a state-of-the-art VR identification model architecture that can convincingly bypass this anonymization technique when trained on a sufficiently large dataset. Therefore, we ultimately propose a "deep motion masking" approach that scalably and effectively facilitates the real-time anonymization of VR telemetry data. Through a large-scale user study ($N = 182$), we demonstrate that our method is effective at achieving both cross-session unlinkability and indistinguishability of anonymized motion data.

This dissertation represents a comprehensive tour of the unique set of privacy risks presented by XR technologies. In doing so, our aim is not to discourage the use of XR devices but rather to provide users with an enhanced understanding of the associated hazards and arm them with the tools necessary to mitigate those risks.

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation, and the original research supporting this work, was made possible by the generous advice, support, and contributions of Allen Yang, Ananya Kharche, Atticus Cull, Beni Issler, Björn Hartmann, Brandon Huang, Charles Dove, Chris He, Christian Rack, Christopher Harth-Kitzerow, Conor Gilsenan, Dawn Song, Dziugas Ramonas, Eric Paulos, Eric Wallace, Gonzalo Munilla Garrido, Ines Bouissou, James F. O'Brien, James Smith, Jason Sun, Julien Piet, Justus Mattern, Louis Rosenberg, Lun Wang, Mani Malek, Mark Roman Miller, Nikita Samarin, Rui Wang, Shuixian Li, Sriram Sridhar, Syomantak Chaudhuri, Viktor Radulov, Wenbo Guo, Xiaoyuan Liu, Yu Gai, and Zade Lobo.

This research would not have been possible without the guidance of an excellent dissertation committee: Dawn Song, James F. O'Brien, Björn Hartmann, and Louis Rosenberg.

The research constituting this dissertation took place at Berkeley's Center for Responsible, Decentralized Intelligence (RDI). This work was funded in part by Meta Platforms, the Minderoo Foundation, the National Science Foundation (NSF), the National Physical Science Consortium (NPSC), and the Fannie and John Hertz Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of their employers or the supporting entities.

We sincerely thank all of our study participants for making this work possible.

Chapter 1

Introduction: Truth in Motion

1.1 Introduction

While virtual reality (VR) has been around in some form since well before the modern internet, the recent introduction of affordable standalone VR devices, such as the Meta Quest 2, has marked a turning point in the accessibility of VR to average consumers. In 2022 alone, more than 10 million VR headsets were sold, showing that the technology has begun to reach mass-market adoption. While augmented reality (AR) devices, such as the Microsoft HoloLens, Meta Quest 3, and Apple Vision Pro, are currently less popular than VR, AR is already being used in a growing number of industries and professional applications.

VR and AR, collectively known as "extended reality" (XR), are envisioned by their proponents as a step towards the eventual creation of a massively connected "metaverse": an immersive virtual world where users meet to work, learn, and socialize. Indeed, future iterations of these devices, particularly those that support AR, are well-positioned to become a major medium of human-computer interaction in the near future.

While modern XR devices contain a wide variety of sensors, at the core of nearly all XR experiences is a stream of motion capture "telemetry" data that records the position and orientation of tracked locations on the user's body in 3D space. Metaverse platforms, by their very nature, turn every movement of a user into a stream of data broadcast to other users anywhere in the world in order to facilitate real-time interaction.

Today's XR platforms and experiences have been built under the assumption that this telemetry data is relatively innocuous: useful for rendering an avatar representing one user on another's device, but not much more. However, this dissertation challenges that notion.

In this dissertation, we present studies that paint a very different picture of motion data. What appears at first to be random variations in movement may perhaps be more akin to a DNA sequence, revealing the identity, biometrics, demographics, and even health information of XR users to anyone else in the same virtual world.

The privacy consequences of XR motion data are more striking still in light of how these devices are actually used in practice. While proponents emphasize brand-friendly work meetings and social gatherings, XR usage today often includes rowdy gaming sessions or adult experiences. The ability to link user identities across applications, and perhaps even to their real-world identity, could entail severe consequences for ordinary XR users and tarnish the reputation of metaverse technologies as a whole.

The news is not entirely negative. We are still in the early days of XR adoption, and have the opportunity to learn from decades of security and privacy research on the conventional internet. In addition to describing the security and privacy challenges presented by XR motion data, we describe several clear approaches for counteracting these threats. If researchers act quickly to test and implement privacy-preserving mechanisms for the metaverse, security and privacy can be at the foundation of future metaverse systems.

> "Cassius: 'Tis Cinna; I do know him by his gait; He is a friend."
>
> William Shakespeare
> in *Julius Caesar*, 1627

1.2 Truth in Motion

Most people have an intuitive understanding that the way we move around in our daily lives is as much an expression of our individuality as is the way we speak. Because movement patterns are a product of each individual's unique physiology, muscle memory, and even personality, we all learn, without really trying, to recognize people based on their motion, and to make subconscious assumptions about people based on the way they move. Actors in film and television are well aware of this, and are often instructed to adopt specific mannerisms to convey subtle cues about the character they wish to portray.

The phenomenon of persons being characterizable by their motion patterns first became the subject of rigorous academic interest in the 1970s, with a series of studies demonstrating the extent to which individuals unknowingly reveal identifying information about themselves via their movements. Most famously, in a 1977 study of six participants, Cutting and Kozlowski demonstrated that individuals can identify their friends just by viewing motion-tracked objects affixed to the body [48]. At a time well before the advent of modern computer graphics, the authors creatively resorted to taping highly reflective objects to a number of points on the participants' bodies. The scientists then streamed a camera feed of the subjects through a television monitor, and increased the contrast until the participants' silhouettes disappeared and only the individual points of light could be seen, as shown in Figure 1.1.

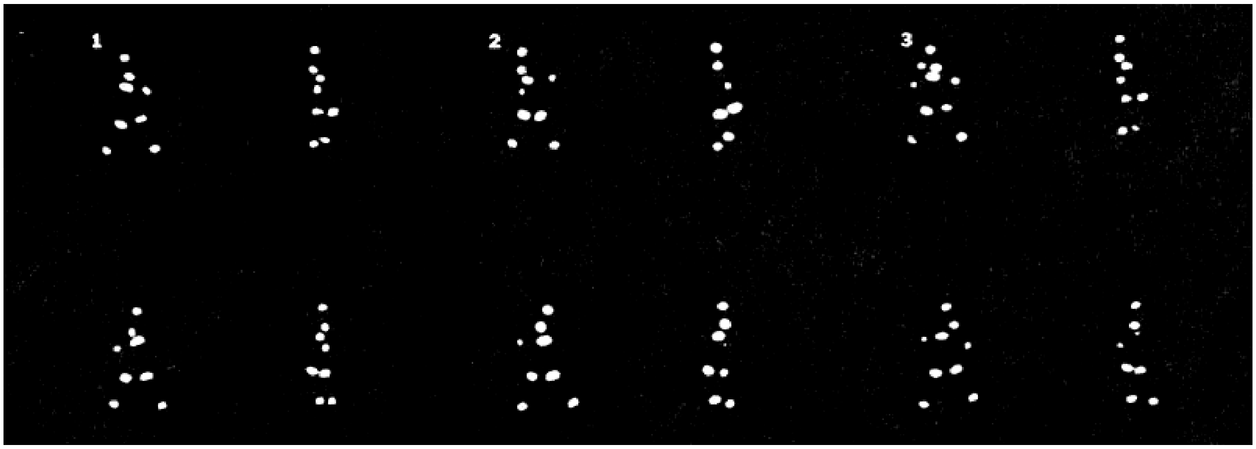


Figure 1.1: Three subjects are shown walking around a laboratory with point-light markers affixed to their bodies. (Adapted from 'Recognizing friends by their walk: Gait perception without familiarity cues' [48], with permission.)

After recording the motions of six participants, their friends were asked to come into the lab and identify the name of each subject based only on the movement of the points of light, which they were able to do with 38% accuracy ($p < .005$). In a later study, the same recordings were shown to new participants, who were able to infer the gender of the original subjects with 79% accuracy ($p < .05$) [146]. More recently, researchers have also shown that the motion of children can be differentiated from that of adults with 66% accuracy [123].

These results tell us something fundamental about human motion: it is a *biometric* that belongs in the same category as blood type or an iris scan. While we have clearly known that this is the case for quite some time, it is becoming particularly relevant today as we potentially enter a new era of extended reality proliferation.

1.3 Moving about the Metaverse

To those who are acquainted with XR, the motion data illustrated in Figure 1.1 may seem quite familiar. Fundamentally, an XR device uses an array of sensors to generate a stream of motion data from its user. As in the Cutting and Kozlowski study, XR devices function by tracking the location of individual parts of the body in 3D space. However, instead of using visible points of light, these measurements are typically generated using a combination of inertial measurement units (IMUs) and either onboard cameras (also known as "inside-out" tracking) or external tracking stations (known as "outside-in" tracking). At a minimum, the location and orientation of the user's head and hands are tracked, though eye tracking and full-body tracking systems are becoming increasingly common.

In a standard consumer-grade XR system, the points of interest on the user's body are measured by the XR hardware between 60 and 144 times per second. This data is then passed to the software application running on the device, which uses it to render stimuli for the user, thereby creating an immersive experience. In the case of multi-user or "metaverse" applications, the motion data is also streamed from the device to a remote game server, which in turn may forward it to other users around the world so that a virtual "avatar" of the first user can be rendered on their devices. In Chapter 2, we systematize this information flow and present a threat model that characterizes the capabilities of each involved entity.

Despite changing hands several times, the information contained within this data stream is fundamentally unchanged: individual points, representing specific body parts of the XR user, moving around in 3D space. In other words, each of the involved entities (the hardware, the application, the server, and the other users) are receiving the same data that we have known for decades can be used to identify and profile individuals.

We are not the first to make this observation. Researchers have for some time been studying the ability to uniquely identify users based on their motions in XR. However, it is only with the recent widespread adoption of XR that sufficiently large datasets have become available to truly understand the true scale and implications of this threat. In Chapter 3, we identify and describe a novel dataset of XR motion data, which we then use throughout this dissertation to further our understanding of XR security and privacy risks.

1.4 Motion as Identity

In 2020, a team of scientists at Stanford University's Virtual Human Interaction Lab (VHIL) performed an experiment to investigate whether ordinary people can be identified in VR based solely on their movement patterns. The researchers set up an interactive VR exhibit at The Tech Interactive, a science and technology museum in San Jose, California. Visitors to the exhibit were asked for permission to have their motion data recorded while they interacted with the VR devices being displayed. Later, the researchers anonymized a portion of the data from each visitor to see if they could re-identify them based on their motions. The results show that 95% of users were correctly re-identified by simple "random forest" machine learning models trained on less than five minutes of tracking data per person [178].

This finding is noteworthy in light of the fact that the users weren't doing anything particularly identifiable; in fact, participants were just asked to passively observe 360° videos while their movements were recorded. Still, in doing so, most users subconsciously revealed enough information about themselves to consistently stand out from the other 510 participants.

While this study was the first to genuinely establish the possibility of telemetry-based identification in VR, it does not tell the full story about the extent of the resulting privacy threat. For instance, identification of 511 users does not preclude the possibility that basic static measurements like height and wingspan were enough to tell each of the users apart, nor does it necessarily prove that users can be linked from one usage session to the next.

The findings of the VHIL study motivated us to scale up the prior efforts to a size more representative of future metaverse environments. In Chapter 4, we describe a VR identification study we performed with data from over 55,000 users. Using over 2.5 million motion capture recordings from “Beat Saber,” a popular VR rhythm game, we analyzed the possibility of training machine learning models based on the gameplay recordings of each user. We used LightGBM, a tree-based machine learning framework, to train a hierarchical classification model on summary statistics derived from five minutes of motion data per user. Using this model, we were able to identify the same users from their motions on different in-game “maps” and on a completely different day.

Our results, presented at *USENIX Security ’23*, demonstrate that users can be uniquely identified with 94.33% accuracy from 100 seconds of data, and with 73.20% accuracy from just 10 seconds of motion data [196]. In other words, by observing the movements of an anonymous VR user, we can usually determine exactly which of the 55,000 known users they are within 10 seconds, and almost always within 100 seconds.

Our research in this area indicates that movement patterns, as measured by VR devices, are a much stronger biometric signal than previously imagined. It’s only with the recent explosion in the popularity of VR gaming that a study of this scale has become possible. As larger datasets emerge, we may soon find that motion data can in fact identify users at an even greater rate, perhaps 1 in 100,000 or more.

1.5 Motion as a Fingerprint

To contextualize the strength of VR motion data as an identifying signal, it is helpful to compare the biometric uniqueness of motion to more traditional biometrics like iris, fingerprint, or facial scans, as illustrated below in Figure 1.2.

To date, the most comprehensive analysis of biometric identification is a 2003 study from the National Institutes of Standards and Technology (NIST), which analyzed dozens of commercially available biometric sensors using real data from over 100,000 users [301]. The results indicate that high-end fingerprint sensors could, at the time, identify users within a population of 10,000 with 90% accuracy. The best-performing facial recognition systems could only identify 1 in 500 users with the same accuracy. Voice recognition was shown to be even worse, with no system achieving greater than 85% accuracy regardless of the population size [161]. We already know that XR motion data can be used to identify at least 55,000 users, with over 90% accuracy. In fact, of the technologies evaluated by NIST, only iris scans out-performed motion, with an identification rate better than 1:150,000 [101].

Of course, biometric technology has greatly improved since 2003, but an equally comprehensive analysis has not since been performed. Still, the comparison remains informative; we are in the early days of motion-based identification and should expect to see similar improvements in motion biometrics over time. Overall, the ability to identify users via their motion is at least comparable to other biometrics at a similar stage in their development.

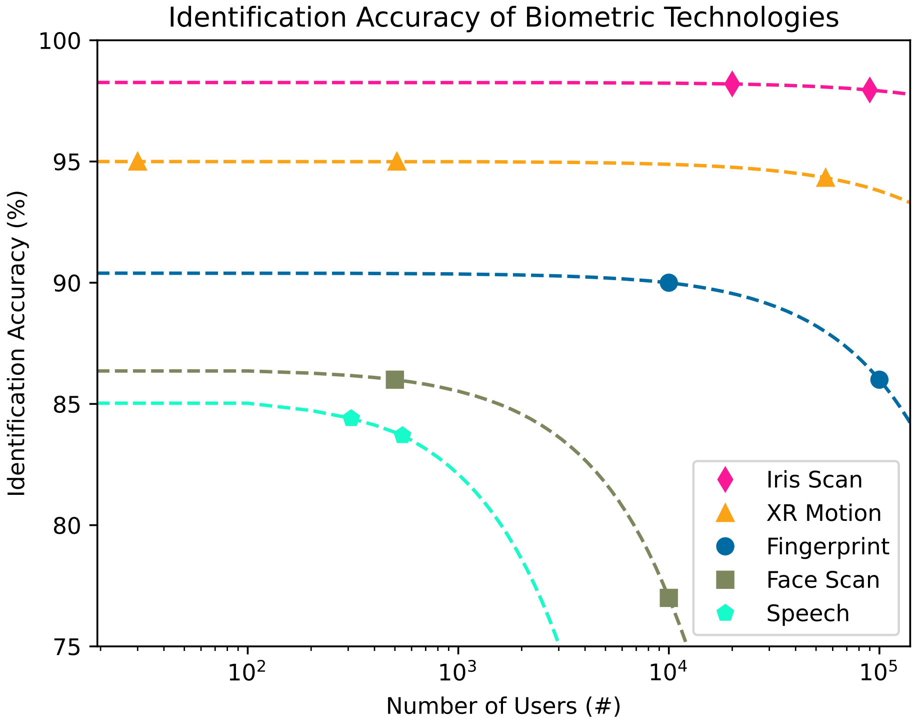


Figure 1.2: Graph of user count vs. identification accuracy for various biometric technologies; log-log scale with log-linear projection on top-1 error rate.

Because the way we move evolves over time, it may seem different in kind to other forms of biometrics. Yet here too, we find clear analogs to widely accepted biometric technologies like facial recognition. Just as one might walk differently from day to day depending on their clothing and footwear, one's face might appear different to recognition software from one day to the next depending on changes to their makeup or facial hair. Similarly, one's mood is as likely to affect their movement style as it is their facial expression. Our motion changes over time, just as our faces gradually change with age. Overall, the relative novelty of XR motion data as a biometric has left many unanswered questions about its temporal and situational robustness, but comparable biometric technologies have learned to ignore daily fluctuations and develop a consistent, long-term signal.

Still, there is at least one critical difference between conventional biometrics and the motion data captured in XR. Sharing fingerprints, and similar biometrics, is not strictly required to browse the web, but motion data is a fundamental part of how XR devices work, and must be shared in real time with a variety of parties to enable metaverse experiences. The equivalent would be if logging into a social media website entailed sending a scan of your fingerprints not only to the platform but also to every other user you interact with.

1.6 A Moving Threat

Consider a public figure who regularly uses a VR system with their corporate credentials to hold meetings and do professional work. In the evening, they log on with a different account to play multiplayer VR games (where they might not behave in the most professional way), and later in the evening, they use a third account for adult VR experiences. Most people in this situation would reasonably prefer that the service providers not be able to tie these accounts together. As it stands, the user's unique motion patterns would allow any observer (or group of colluding observers) to quickly link all of these accounts together.

On the web, "browser fingerprinting," which uses subtle differences between browser configurations to link people across web services, is a highly analogous attack that is generally regarded as a significant privacy concern. However, while one can replace their browser, they cannot easily change the physiology and muscle memory that dictates their movements.

Users have, for better or for worse, grown accustomed to privacy risks like browser fingerprinting being a part of daily life in the digital era. On the other hand, motion-based privacy risks are so seldom discussed that they are poorly understood even by experienced XR users. For example, in one recent study, researchers from Carnegie Mellon University asked a number of test subjects to rank various types of data collected by an XR device from least to most concerning [87]. Participants with all levels of XR experience consistently rated body movement data as amongst the least concerning XR data streams. This disconnect between the known privacy consequences of a technology and users' understanding of the same mirrors attitudes observed in the early days of the web. We may therefore find that users' perceptions of privacy in the metaverse follow a similar trajectory, eventually coming to treat the metaverse as a broadly public space with a reduced expectation of privacy.

Unlike browser fingerprinting, motion-based attacks are by no means limited to virtual spaces. In fact, motion patterns are so intrinsically tied to our physical selves that they may soon be able to follow us out of the metaverse and into the real world. Machine learning models designed to extract 3D motion data from monocular video feeds are rapidly improving; we can reasonably extrapolate that it will eventually be possible to match a person's VR movements to surveillance video. Unlike your face, which can be covered with a mask, no reasonable countermeasure can obscure all of your movements from public view.

On the flip side, the relatively consistent nature of identifiable motion patterns could provide an unparalleled opportunity for passive authentication in future metaverse applications. XR users could benefit from the convenience of having their motion data also be used to verify their identity rather than needing to authenticate explicitly. Unfortunately, the laissez-faire nature with which VR motion data is currently broadcasted and uploaded to the internet undermines its future use in authentication. The equivalent would be using fingerprint login on your accounts if pictures of your fingerprints were already uploaded to the internet. In a sense, today's VR users are paying a heavy early adoption penalty by sharing their motion data with the world before comprehensive defenses are in place.

Finally, like a fingerprint, one may be inclined to believe that motion identification is the virtue of random but ultimately meaningless variations. In reality, our movement style is crafted over time as the result of our background and experiences, and can later be “decoded” to not only identify us, but also to infer a variety of attributes that we may prefer remain private. These risks also extend to children, who will increasingly use XR devices in the coming years, not only for gaming but also for school and other educational contexts.

1.7 Motion as DNA

Thus far, we have explored the analogy of motion to a fingerprint that follows users throughout the metaverse, allowing them to be tracked across devices and applications. This analogy is true, but incomplete. Recall, for example, that point-light motion data has long been known to reveal not only the identity of participants, but also their age and gender. Perhaps a more appropriate analogy is DNA, which is not only unique to an individual, but also encodes information about their personal characteristics.

In Chapter 5, we present a second study of the same Beat Saber users. In this study, we surveyed over 1,000 Beat Saber players to ask them a variety of questions about their background, biometrics, demographics, health information, behavioral patterns, and technical device specifications. Later, we trained a series of machine learning models to see which, if any, of these responses could be accurately inferred just by examining the motion patterns of these users [194]. The models utilized a transformer architecture (similar to that found in language models like GPT), and were trained using up to 30 minutes of motion data per user. Then, to avoid simply re-identifying those participants, we evaluated the trained models on a completely different set of users than those used for training.

The results go far beyond inferring the expected anthropometrics like height and wingspan, or even demographics like age and gender. We found that even behavioral attributes, such as substance use, could be inferred from the telemetry data with statistically significant accuracy. Everything from the country that a user is from to the clothes that they are wearing can be determined using features derived from their motions alone. Perhaps most strikingly, the presence of mental and physical disabilities could clearly be discerned from the motion data; all from recordings of users playing an otherwise innocuous VR rhythm game.

With open access to device APIs, XR developers are not limited to creating legitimate applications. Malicious developers can go further, creating games and applications that are deliberately designed to covertly harvest user data. In Chapter 6, we present a third study in which we recruited 50 participants to play an innocent-looking VR game called “MetaData,” shown in Figure 1.3. The game appears to be a harmless “escape room” in which players complete a series of challenges to progress through the game. In reality, we had carefully constructed each puzzle to covertly reveal more information about the players based on their interactions with the virtual world than would be possible via passive observation alone.

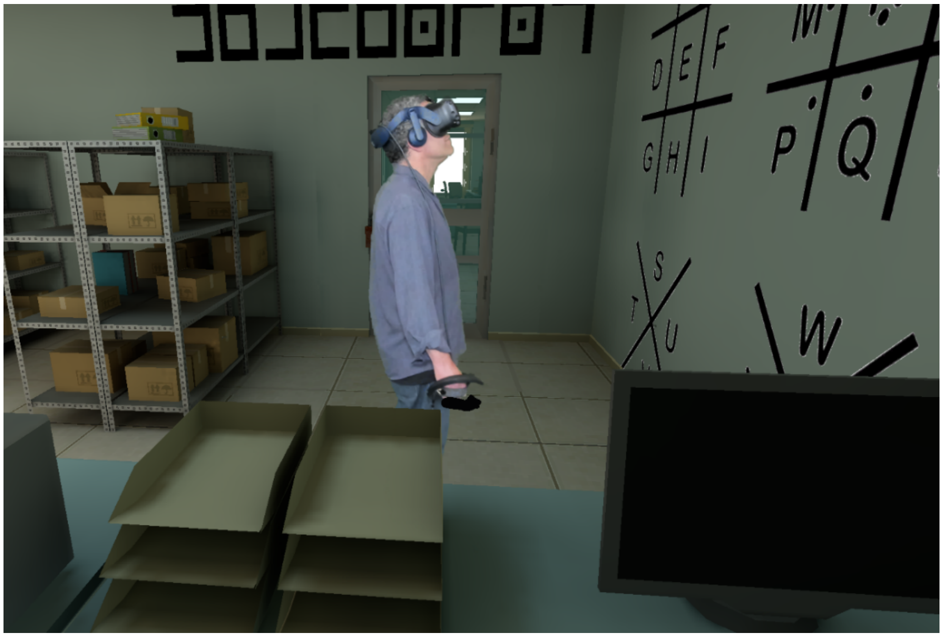


Figure 1.3: Mixed reality image of Louis Rosenberg playing “MetaData,” an adversarially-designed VR “escape room” game that harvests private user information.

Our results, recently presented at *PETS ’23*, show that over 25 personal data attributes, from anthropometrics like height and wingspan to demographics like age and gender, can be inferred from these users within just a few minutes of gameplay [193]. If “Beat Saber” is like a typical website, passively recording interactions in an otherwise normal application, then “MetaData” prototypes a concept more akin to the online quizzes deployed by Cambridge Analytica and others to actively harvest user data while being disguised as a harmless activity. By incorporating the identification methods discussed earlier, adversaries can attempt to aggregate detailed user profiles from data across many such applications.

These findings highlight that the privacy risks of XR devices stem not only from their sensors but also from the immersive nature of their displays, which can be used to totally control a user’s virtual environment to influence the information they reveal.

1.8 A Fast-Moving Field

All of the attacks we have described thus far utilize just three tracked locations: one on the user’s head, and one on each hand. While that’s already enough to identify and profile a large number of users, future XR headsets will likely feature full-body tracking systems, in which at least six to ten body parts are tracked.

Any risk to privacy in XR is further exacerbated by other modalities found on many devices. For example, Apple's new "Vision Pro" device is known to feature a LIDAR array, eye tracking, microphones, and no less than 14 cameras, in addition to full-body tracking.

As seen in Apple's upcoming device, the industry is rapidly transitioning from traditional VR headsets that are used in fully simulated worlds to lighter-weight mixed and augmented reality devices that enable immersive content to be integrated into a user's view of their physical surroundings. The upcoming headset from Apple uses "passthrough cameras" to augment the real world with virtual content and is intended for regular use within a user's home or office. This means users will be able to perform many of their common daily activities while wearing these mixed reality headsets, from sitting on their living room couch and opening their refrigerator to grabbing coffee mugs off the shelf or climbing into bed.

Given that Beat Saber data, which demands a relatively narrow range of human motions, can be used to distinguish 1 user among 55,000, we can reasonably predict that as XR devices are integrated into common daily activities, a wider range of motion patterns will be captured, which could be used to identify and profile users with even greater precision. Consider, for example, the ubiquitous task of grabbing a doorknob and opening a door. Each of us has performed this motion so many times that it's likely to be deeply ingrained in our muscle memory and likely at least as uniquely repeatable as the saber swings in Beat Saber.

Major technology companies are already developing AR and MR eyewear that they hope will be so lightweight and fashionable that users will be comfortable wearing them outside the home or office as they go about their normal daily routines: walking down city streets, shopping in retail establishments, and visiting restaurants and bars. Google, Samsung, and Qualcomm have publicly announced a partnership to develop XR devices built on the Android operating system with the goal of enabling similar usage patterns as mobile phones. In fact, many experts believe that lightweight XR eyewear will largely replace the handheld mobile phone market in the near future: "The phone is already dead," claims Alex Kipman, inventor of Microsoft's HoloLens AR glasses. "People just haven't realized."

These mobile XR devices are likely to also include an array of sensors that measure our interactions with the physical world, including actions as routine as grabbing products off of store shelves. Like turning a doorknob, or shaking a hand, reaching for a product is likely ingrained in our muscle memory and uniquely identifiable. However, tracking this action is particularly interesting because mobile XR devices are capable of displaying promotional content to users based on where they are, what they're looking at, and even what they reach for [236]. So, when picking up a can of soup, a mobile XR device could deploy targeted promotional content that links to data about that user's personal preferences and shopping habits. Because we know XR users can be uniquely identified via their motions, it may be difficult for platform providers to maintain the privacy of users who do not wish to be individually targeted by real-time marketing materials.

1.9 Safeguarding Motion

Data privacy issues are obviously not unique to the metaverse. In fact, nearly every major communications technology advancement of the past century has been accompanied by corresponding privacy risks, from the wiretapping of landlines beginning in the 1890s through to emerging privacy concerns with smart home, mobile, and wearable devices today. As is the case with XR motion data, information that exists to provide necessary, legitimate functionality can often also be leveraged for adversarial purposes.

On the web, tracking cookies are a quintessential example of this phenomenon. While cookies serve an important, legitimate purpose in enabling persistent sessions, adversaries can leverage them to track users across websites. But unlike in VR, the maturation of web technologies has brought a suite of countermeasures to such attacks. Technologies like VPNs, proxies, Tor, and incognito mode in browsers, have provided users with vital defensive tools for reclaiming their privacy in the face of such attacks. Until recently, no equivalent comprehensive defensive utilities had been developed for extended reality devices.

We thus find ourselves in the dangerous situation of facing unprecedented privacy threats in VR while lacking the defensive resources we have become accustomed to on the web. This is not necessarily due to a lack of interest in XR privacy, though research in this area is certainly far less common today than in web security and privacy. Rather, it is due to a fundamental challenge with XR motion data: the same telemetry data that is necessary to provide legitimate multi-user functionality can also be used for adversarial purposes.

Consider, by contrast, the permission-based model used by a typical smartphone application. The data sources accessible to each application are segmented into discrete permissions, which must be granted to the application by the end-user on an individual basis. If a navigation app requests access to view a user's GPS location, they might approve the request based on the understanding that the application needs this information to function. However, if it instead asks to see their contact list or listen to their microphone, this would reasonably raise a red flag in the context of what the application actually needs. Researchers have indeed tried to implement similar fine-grained access control systems in XR, with systems like Erebus providing least-privilege access to XR sensor data contingent on conditions like time and location [140]. Motion data, however, has resisted attempts at granular restriction, and as such largely remains an all-or-nothing proposition. As it stands, there is fundamentally just a single stream of motion telemetry data that is used by all XR applications for a variety of purposes. Once sent off to a remote game server, there is no easy way to audit whether the data is being used for benign or nefarious reasons.

Further complicating attempts to protect the privacy of XR motion data is the need for any resulting defensive system to be real-time and almost instantaneous. In many XR devices, even a slight added delay can cause a disconnect between what the user sees and what their inner ear senses, resulting in severe motion sickness. Most standardization authorities place an upper bound on "motion to photon latency" of just 20 milliseconds before users experience significant negative effects. By contrast, VR attackers can be slow and non-causal, waiting for an entire session of motion data to be captured before beginning their attack.

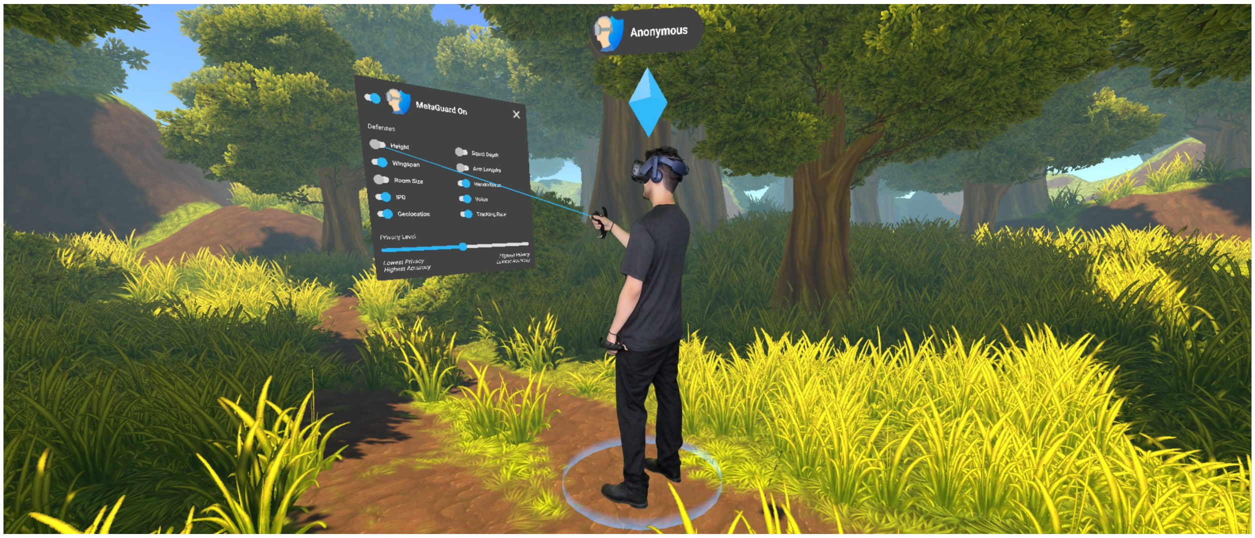


Figure 1.4: Mixed reality photo of Vivek Nair using "MetaGuard."

There is a silver lining, in that we have the opportunity to learn from the most effective privacy-preserving technologies on the web to implement metaverse architectures with security and privacy at their core. There are a few potential paths forward in this respect.

The first and most obvious approach would be to leverage "local epsilon-differential privacy," a statistical measure of information leakage that is known as the "gold standard of data privacy." We have already had moderate success in utilizing this technique. In Chapter 7, we present "MetaGuard," an open-source plugin for the Unity game engine that we think of as a proof of concept for an "incognito mode of the metaverse." MetaGuard, a "Best Paper" winner at UIST '23, works by identifying privacy-sensitive dimensions present in an XR telemetry data stream, such as those corresponding to a user's height or wingspan. These axes are then passed through a "Laplacian noise distribution," a type of differentially-private transformation function, before being transmitted to the server and on to other users. The plugin can easily be installed by end users into a variety of existing VR applications just by placing the extension files in a particular directory on their device, and can be customized to suit the specific needs and risks of each application, as shown in Figure 1.4.

To evaluate the efficacy of MetaGuard at protecting VR users, we replayed the motion recordings of users from the MetaData study, as well as of the 55,000 Beat Saber users, within a virtual environment to simulate what their data would have looked like had they been using MetaGuard. We found that MetaGuard is reasonably effective at mitigating both identification and inference attacks. MetaGuard reduced the accuracy of identifying users across sessions from nearly 95% to less than 5% when using the same identification models and techniques described previously, trained on a single recording per user. Attacks targeting private user data were also hindered, with the ability to infer demographics like age and gender dropping below the threshold of statistical significance [190].

These countermeasures do come at the cost of usability, however; by changing the user's motions to protect their privacy, users may experience a discrepancy between their true and apparent joint locations. For example, when reaching out to shake the hand of another virtual user, a person may find that the other user perceives their hand to be in a different location than expected, in order to hide their true wingspan from the other user. However, the level of "error" experienced by users is distributed according to a theoretical optimality that minimizes the error experienced by users for any given level of privacy.

Machine learning provides an alternative approach to differential privacy for removing sensitive data from XR telemetry. In Chapter 8, we describe "deep motion masking," a machine learning architecture designed to transform, or "corrupt," XR telemetry streams in order to remove user data embedded in the motion while minimally impacting legitimate application functionality. At a high level, our deep motion masking works by decomposing the plausible variance of human motion sequences into action-related variance and user-related variance. It then anonymizes telemetry sequences by modifying their user-related component without changing the action-related component of the motion.

In using this method, we lose the mathematically provable properties of a differential privacy approach, as formal verification on complex machine learning models remains a known difficult problem. On the flip side, the model is actually more effective at protecting user privacy in practice, due to its ability to detect and obscure not only primary sensitive attributes but also hidden correlations to these variables. In our evaluation, we found that deep motion masking presents a 2.7× improvement over MetaGuard in the indistinguishability of anonymized motion data, and an over 20× improvement in cross-session unlinkability.

While deep motion masking in highly suitable for motion data intended for consumption by human observers, there will always be VR applications in which very high precision is required, such as telemedicine, competitive e-sports, or remote operation of equipment. If anonymity is still desired in such an application, an alternative solution, a defense worth exploring in future work is the use of trusted execution environments (TEEs) or secure multi-party computation (MPC) to provide transparency into how metaverse servers actually utilize the telemetry data shared by users. TEEs like Intel's SGX or Amazon's Nitro provide a hardware-based attestation mechanism that allows users to verify the software running on a remote machine before sending their data to that server, ensuring that only legitimate operations are being performed. For a subset of the operations offered by TEEs, MPC can also provide a purely cryptographic mechanism for achieving the same verifiable computations, regardless of the underlying hardware. These solutions are also not without their fair share of concerns. Most forms of MPC are currently far too inefficient to facilitate the high-throughput and low-latency data streams required for XR. TEEs, on the other hand, are fast enough, but researchers constantly demonstrate new security vulnerabilities that undermine their fundamental security properties. Still, technologies that enable users to audit exactly how their data is being used by metaverse entities may ultimately prove more resilient than motion transformation methods that cannot provide strong guarantees against an adaptive adversary that develops new ways to attack XR data streams over time.

1.10 Statement of Claimed Contributions

- In Chapter 2, "Data Privacy in Virtual Reality," our claimed contributions are as follows:
  - We propose a holistic information flow and threat model for VR privacy studies (§2.5).
  - We build a taxonomy of data attributes observable in virtual reality (§2.6).
  - We categorize about prior 30 attacks (§2.7) and 35 defenses (§2.8) related to VR privacy.
- In Chapter 3, "4,700,000 Motion Recordings from 105,000 VR Users," our claimed contributions are as follows:
  - We describe and publish a novel VR motion capture dataset containing 4,717,215 motion capture recordings uploaded by 105,852 XR device users from over 50 countries (§3.3).
  - We present a new lossless "Extended Reality Open Recording" (XROR) file format that is about 30% more space efficient than the original motion capture file formats (§3.6).
  - We present the results of a large-scale survey ($N = 1,006$) of the users contained in our dataset we conducted to better understand their demographics (§3.10).
- In Chapter 4, "Unique Identification of Over 50,000 Virtual Reality Users from Head and Hand Motion Data," our claimed contributions are as follows:
  - We describe a novel motion featurization technique that incorporates VR application context information to enhance VR user identification (§4.3).
  - We present a hierarchical classification approach that allows us to build a scalable identification model with over 50,000 distinct classes (§4.4).
  - We achieve 94.33% identification accuracy of 55,541 VR users from head and hand motion data (§4.6) and provide detailed explainability results (§4.8).
- In Chapter 5, "Inferring Private Personal Attributes of Virtual Reality Users from Head and Hand Motion Data," our claimed contributions are as follows:
  - We surveyed over 1,000 VR users to generate a comprehensive dataset of diverse user data attributes with corresponding VR motion recordings (§5.3).
  - We present a general-purpose transformer-based machine learning technique for inferring user data attributes from head and hand motion streams (§5.5).
  - We demonstrate over that 40 binary classes relating to personal user data attributes can be inferred from motion data in standard non-adversarial VR games (§5.7).

- In Chapter 6, "Exploring the Privacy Risks of Adversarial VR Game Design," our claimed contributions are as follows:
  - Through a series of examples, we demonstrate how active VR adversaries can harvest further user information that is visible through passive observation alone (§6.3).
  - We present "MetaData," an open-source VR game that illustrates how malicious game developers can design adversarial yet seemingly innocuous VR environments (§6.4).
  - Through a user study of 50 participants, we experimentally demonstrate that an attacker can covertly harvest over 25 unique data attributes from VR users (§6.6).
- In Chapter 7, "Going Incognito in the Metaverse: Achieving Theoretically Optimal Privacy-Usability Tradeoffs in VR," our claimed contributions are as follows:
  - We provide the first ε-differentially private framework for protecting a range of sensitive data attributes in VR motion telemetry streams (§7.3).
  - We describe MetaGuard, a concrete implementation of a modular "incognito mode for VR," realized as an open-source plugin for the Unity game engine (§7.4).
  - We show that our approach is effective at defeating specific VR privacy attacks (§7.5).
- In Chapter 8, "Deep Motion Masking for Secure, Usable, and Scalable Real-Time Anonymization of Virtual Reality Motion Data," our claimed contributions are as follows:
  - We present a new, state-of-the-art VR identification model that can convincingly bypass existing VR anonymity systems such as MetaGuard (§8.3).
  - We propose a new "deep motion masking" technique that facilitates the scalable, real-time anonymization of VR telemetry data (§8.4).
  - Using new and existing VR identification models, our evaluation shows at least a 20× improvement in anonymity over prior VR privacy approaches (§8.6).
  - Our large-scale usability study ($N = 182$ participants) demonstrates a nearly 3× improvement in the indistinguishability of resulting anonymized motion data (§8.6).
  - In realistic simulations, we show that our anonymizer has minimal impact on perceived interactions between users and virtual objects (§8.6).

1.11 Statement of Multiple Authorship and Prior Publication

Portions of the research presented in this dissertation have previously been published as papers or preprints with multiple authors other than the principal author of this dissertation:

- Chapter 1 (this chapter) is derived from "Truth in Motion: The Unprecedented Risks and Opportunities of Extended Reality Motion Data" [195], published in *IEEE Security & Privacy*, and co-authored by Louis Rosenberg, James F. O'Brien, and Dawn Song.

- Chapter 2 is derived from "SoK: Data Privacy in Virtual Reality" [94], published in *Privacy Enhancing Technologies Symposium (PETS) '24*, and co-authored by Gonzalo Munilla Garrido and Dawn Song.

- Chapter 3 is derived from "BOXRR-23: 4.7 Million Motion Capture Recordings from 105,852 Extended Reality Device Users" [192], published as a preprint, and co-authored by Wenbo Guo, Rui Wang, James F. O'Brien, Louis Rosenberg, and Dawn Song.

- Chapter 4 is derived from "Unique Identification of 50,000+ VR Users from Head & Hand Motion Data" [196], published in *USENIX Security '23*, and co-authored by Wenbo Guo, Justus Mattern, Rui Wang, James F. O'Brien, Louis Rosenberg, and Dawn Song.

- Chapter 5 is derived from "Inferring Private Personal Attributes of Virtual Reality Users from Head and Hand Motion Data" [194], published as a preprint, and co-authored by Christian Rack, Wenbo Guo, Rui Wang, Shuixian Li, Brandon Huang, Atticus Cull, James F. O'Brien, Marc Latoschik, Louis Rosenberg, and Dawn Song.

- Chapter 6 is derived from "Exploring the Privacy Risks of Adversarial VR Game Design" [193], published in *Privacy Enhancing Technologies Symposium (PETS) '23*, and co-authored by Gonzalo Munilla Garrido, Dawn Song, and James F. O'Brien.

- Chapter 7 is derived from "Going Incognito in the Metaverse: Achieving Theoretically Optimal Privacy-Usability Tradeoffs in VR," published in *User Interface Software and Technology (UIST) '23*, and co-authored by Gonzalo Munilla Garrido and Dawn Song.

- Chapter 8 is derived from "Deep Motion Masking for Secure, Usable, and Scalable Real-Time Anonymization of Virtual Reality Motion Data," co-authored by Wenbo Guo, James F. O'Brien, Louis Rosenberg, and Dawn Song.

In cases where co-authored material is incorporated in this dissertation, all major contributors were informed of their inclusion herein in addition to being credited above.

Chapter 2

Background: Data Privacy in Virtual Reality

2.1 Introduction

Major players in the extended reality (XR) industry are racing to create the “metaverse,” a paradigm shift in human-computer interaction that represents the internet as an immersive 3D virtual world [233]. Motion tracking devices are set to be a fundamental part of this “new internet,” with hand-held controllers or other body tracking systems being used to digitize and relay an individual’s movement patterns to other users around the world for immersive real-time interaction. While the idea of a “metaverse” promises to offer a richer social experience with more lifelike interactions than today’s internet, as with many advancements in communication technologies, it also accompanies an unprecedented set of privacy risks.

Recent studies have demonstrated that the exact same XR device “telemetry” data that is fundamental to the operation of nearly all existing XR applications can also be used to identify [178, 36, 207] and profile [274, 251, 193] users with or without their knowledge. These risks are exacerbated by VR’s unparalleled immersiveness, which can make users more susceptible to self-disclosure [164, 266], and social engineering [9, 62].

Unlike current internet platforms, where users now have access to a suite of defensive privacy tools (such as Tor, VPNs, proxies, and “incognito mode” in browsers) to mitigate privacy threats, there is currently no mature suite of defenses for combating the equivalent risks in VR. Extant literature offers a scattered set of privacy defenses at an early prototype stage, with no significant knowledge transfer to commercial-grade applications.

In this chapter, we lay the groundwork for tackling this impending challenge by providing a new, comprehensive VR information flow and threat model, taxonomy of data attributes observable in VR, and systematization of over 60 existing VR privacy attacks and defenses. In subsequent chapters, we refer back to this systematization to position our contributions within the broader landscape of VR privacy research.

2.2 Related Work

The “reality and virtuality continuum,” originally proposed by Milgram and Kishino [176], is a continuous scale ranging between a completely virtual environment (virtuality), and a completely real environment (reality). The majority of the research discussed in this chapter, and in this dissertation as a whole, is positioned toward the virtuality end of this continuum, with attacks specific to mixed reality (MR) and augmented reality (AR) receiving less emphasis than attacks on virtual reality (VR). However, many of the risks associated with VR devices are equally applicable to MR and AR, and should be interpreted throughout this dissertation as potentially relevant to the entire reality and virtuality continuum.

VR Devices

Since around 2016, the general public has had the opportunity to experience immersive VR devices like never before. In their most basic form, VR devices incorporate a head-mounted display (HMD), typically with integrated microphones and speakers, and two handheld controllers, typically with a variety of buttons, haptics, and other interface components [170]. Some HMDs tether directly to a PC [289], while others, like the Meta Quest 2, can operate as a standalone device [170]. The VR system tracks the HMD and the controllers by outside-in tracking (using stationary external sensors [285]) or inside-out tracking (employing built-in optical sensors [170]). Front cameras for inside-out tracking also enable the user to observe their real-world surroundings by using a “pass-through” mode. This basic setup generates realistic 3D graphics, spatial audio, verbal interaction, and six degrees of head and hand tracking (X, Y, and Z positions, and yaw, pitch, and roll rotations). Today’s VR devices are typically intended for short-term use in a “controlled” environment (e.g., a home, backyard, or office), with future iterations eventually targeting comfortable all-day use.

In addition to the basic setup described above, many VR devices incorporate additional devices and sensors that make VR experiences more immersive yet present further opportunities for data harvesting. Optical sensors for eye-tracking enable foveated rendering [290], increasing the quality of the visual output [7] and lengthening HMD battery life by reducing GPU load. Moreover, eye-tracking can be combined with additional optical sensors that register facial features [286] to improve telepresence applications by enabling more realistic and expressive avatars [39]. In addition to basic buttons and haptic feedback, handheld controllers can have touch sensors for detecting gestures [288], and even outward-facing cameras for improved tracking [171]. Taking controllers to their extreme, force feedback gloves can provide even more ergonomic and realistic virtual interactions [109]. The latest generation of VR devices support full-body tracking [253], enabling more expressive experiences with other users in virtual worlds. Advanced users can even don haptic vests [23] that deliver positional haptic feedback, or masks that aim to reproduce specific scents [209]. Some VR applications, particularly for healthcare, also include sensors that measure galvanic skin response [122], electrodermal activity [10], heart rate [268], skin temperature [208] and superficial brain waves (e.g., EEGs built into HMDs for brain-computer interfacing) [200, 306, 22].

While the plethora of input and output devices associated with VR systems enable users to become deeply immersed in digital environments, it is critical to analyze devices that track users from a privacy perspective, as they can directly expose users' sensitive biometrics, behavior, identity, and real-world surroundings [71, 193, 275, 178]. While some of these data points are also measured by other mobile devices, the unprecedented nature of the VR privacy threat stems partly from the ability to simultaneously obtain a wide range of attributes that would previously have required the combined data of several devices. This confluence of attributes heightens the threat of fingerprinting and inferring demographics to profile, identify, and track users across applications in unrivaled new ways [193].

VR Attacks & Defenses

Tangentially related to this chapter are a number of reviews and survey papers on extended reality displays (1994) [176], classifications (1996) [20], early challenges (1997) [15], integrity and ownership (2000) [85], and enabling technologies (2001) [14]. In the 2010s, researchers began studying the ethical considerations of XR (2014, 2018) [111, 4], presented newfound challenges (2016) [220], discussed the threats of combining VR with social networks (2016) [203], and investigated VR safety (2018) [16]. Recently, practitioners have continued researching VR attacks (2021, 2022) [34, 279] and VR user authentication (2022) [263, 70], and have supported new regulations for upcoming metaverse platforms (2022) [237].

With respect to VR privacy specifically, we identified 10 relevant literature reviews [54, 71, 61, 136, 62, 254, 294, 148, 97, 205], three of which are the closest to the analysis set forth herein. First, Shrestha and Saxena (2017) [254] provided an offensive and defensive overview of XR devices with a focus on optical cameras with respect to privacy, security, and safety. Next, De Guzman et al. (2019) [54] expanded the AR privacy and security defense classification of Roesner et al. [235] to MR without an in-depth analysis of data attributes and attacks. Finally, Odeleye et al. (2022) [205] provided a taxonomy of cybersecurity VR attacks related to authentication and privacy, comprising 5 privacy defenses and 10 attack-focused studies, which are also included in §2.5, §2.7, and §2.8.

Among the rest of the selected literature reviews, Katsini et al. [136] and Kröger et al. [148] specifically studied the privacy implications and research directions of eye-tracking, which we incorporate into this chapter. Additionally, we included the relevant privacy-related insights and VR application taxonomies of two comprehensive reviews that covered general metaverse topics as varied as data management, privacy, legal issues, and economic threats [71, 294]. Lastly, we included key information from narrower surveys of VR security and privacy [61, 97], and VR data attributes and privacy considerations [62].

2.3 Data Collection Method

Our data collection method was inspired by noteworthy systemization of knowledge (SoK) papers and literature reviews in the field of security and privacy [70, 95, 54]. For this review, we sought literature presenting at least one of the following artifacts in the context of VR privacy: (i) a privacy threat, (ii) a privacy defense, or (iii) a taxonomy of data attributes. For papers targeting MR or AR, we included the work only if the presented artifacts overlapped, at least partially, with VR. Before the search, we knew of 12 relevant studies containing the target artifacts (the "base literature"). We then curated the following search string by studying the base literature and conducting a manual preliminary search in Google Scholar for papers containing the targeted artifacts:

Search string: ("virtual reality" OR "virtual telepresence" OR "head-mounted displays" OR "head-worn display" OR "metaverse") AND "data" AND "privacy" AND ("attack" OR "offense" OR "defense" OR "protection")

We used seven of the most prominent digital libraries focused on computer science and software engineering, in combination with Google Scholar, to perform an exhaustive search of the extant literature. Specifically, with this search string, we queried the seven most relevant digital libraries: IEEE Xplore [118], ACM Digital Library [3], ScienceDirect [246], SpringerLink [257], Scopus [77], Wiley InterScience [300], and Web of Science [41]. We included work published between 2010 and 2022, and excluded books, resulting in 1700 hits. We then sequentially filtered the publications by title (47 selected from 1700), abstract (35 selected from 47), and full text (16 selected from 35). Combining the base literature (12) with the filtered studies (16) resulted in 23 selected studies after deduplication.

To further ensure we collected as many relevant publications as possible, we conducted a backward search of the references of the 23 selected studies under the same criteria, and contacted the authors of the 23 works to obtain further relevant publications. The backward search revealed 26 studies, and from the corpus signaled by the scholars, we included another 7 after deduplication. Lastly, we collected another 12 publications thereafter throughout the research and writing of this manuscript following the same criteria.

2.4 Data Collection Results

Table 2.1 shows our final list of 68 publications obtained using the method described above. The rest of this chapter analyzes these 68 publications to provide a comprehensive information flow, threat model, and attribute taxonomy to guide VR privacy research.

Study focus	**Studies** (68)
Primary Studies *(*58*)*	
Defenses (35)	[262], [53], [298], [157], [180]†, [190], [52]‡, [80], [151]‡, [261], [50], [125], [104], [105], [26], [152]‡, [36]‡, [270], [124], [83], [312], [267], [293], [66], [30], [308], [129], [130], [106]‡, [154]†, [214]†, [250]†, [207]†, [229], [244]
Attacks (19)	[107], [314]†, [9]‡, [193], [178], [251], [302]†, [291]‡, [274], [268]†, [167]†, [11]†, [238]†, [208]†, [22]†, [306]†, [122]†, [156], [12]
Surveys (2)	[164], [266]†
Evaluations (2)	[201], [275]
Secondary Studies *(*10*)*	
Literature Reviews (10)	[54], [71], [61], [136], [62], [254], [294], [148], [97], [205]

Table 2.1: The 68 collected studies.

[†] An attacker can leverage the associated defense/mechanism for adversarial purposes.
[‡] The study is defense/attack focused but there is an adversarial/defensive component.

2.5 VR Information Flow & Threat Model

From the 68 selected studies, we identified 5 studies that proposed a VR information flow [71, 193, 52, 294, 97], and 21 that discuss VR threat models [54, 262, 53, 71, 193, 151, 125, 104, 105, 152, 124, 270, 83, 130, 251, 106, 291, 250, 207, 156, 12]. We extracted and combined the associated artifacts to produce a holistic VR information flow and threat model that satisfactorily encapsulates all of the surveyed research.

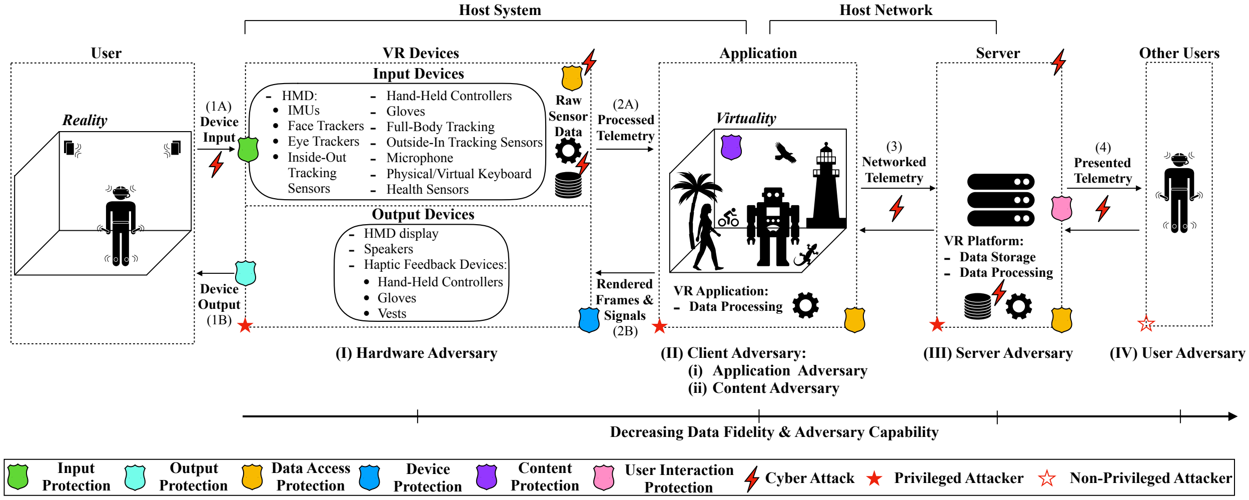


Figure 2.1: Virtual reality information flow and threat model. (cf. [193, 190, 52, 294]).

VR Information Flow

VR device manufacturers or vendors provide app stores where users can download VR applications and games (e.g., the Oculus Store or Steam). Fig. 2.1 illustrates the information flow after installing such an application. These applications typically run in the host VR system, which ingests user input: *spatial & inertial motion data*, *audio*, *text*, *video*, and *physiological signals* (1A). The VR device firmware processes raw sensor data and other input types into useful telemetry, which the application accesses via a standard API (e.g., OpenVR) (2A). The VR application then uses this data to generate stimuli, such as visuals, audio, and haptics, via a rendering pipeline (2B). The output devices present this processed information to the user as an immersive, interactive virtual world (1B). For multi-user online experiences, the client-side application exchanges processed telemetry with an external server through a network, which can reveal *system* and *network* information (3). Finally, the server updates the global state of the virtual world and relays the telemetry data to other users (4). As the information flows from steps (1A) to (4), intermediate data processing steps like compression and downsampling may degrade the quality of the underlying signals.

VR Threats

Within the frame of this study, we consider a state of *privacy* as the lack of a breach of any individual's sensitive data attributes [307]. In our threat model, attackers breach user privacy by collecting and inferring enough information to reliably *identify* and comprehensively *profile* a user across VR applications over multiple usage sessions (tracking). Attackers (i) *identify* an individual when they can uniquely distinguish the user from others, and (ii) *profile* users when they unwarrantedly attach information related to the user's characteristics (e.g., demographics, preferences, browsing history, etc.) [59, 134, 274].

The collected studies discussing or proposing threat models consider application developers [53, 193, 151, 125, 104, 105, 152, 124, 270, 83, 251, 106, 156], servers [106, 193, 12], content creators [71, 193], device manufacturers [262, 193], other users [250, 291, 207], and hackers[1] [71, 270, 130] as the attackers in VR, or rely on general privacy threat models like Lindunn [59, 134, 54, 151]. Based on these studies, we adopt a more comprehensive and pervasive privacy-centered attacker classification specific to VR that encompasses the privacy repercussions of the above threat models. The adversary types of Fig. 2.1 correspond to four distinct entities associated with data processing in VR applications at different privilege levels. These adversaries might coalesce, e.g., a developer of a VR application can also run the server providing multi-user functionality. Table 2.2 shows these attackers' capabilities.

	Observable Attribute Classes							
Adversary	**Spat. & Iner. Telemetry**	**Text**	**Audio**	**Video**	**Phy. Signals**	**System**	**Network**	**Behavior**
(I) Hardware	✓	✓	✓	✓	✓	✓	✗	✓
(II) Client	✓*	✓	✓	✓*	✓	✓	✓	✓
(III) Server	✓*	✓	✓*	✗	✓	✓	✓	✓
(IV) User	✓*	✓	✓*	✗	✗	✗	✗	✓

Table 2.2: VR attacker capabilities (cf. [193]).

*Observable with deteriorated data quality or in abstracted form.
Legend: Spat. = Spatial, Iner. = Intertial, Phy. = Physiological.

[1] Extensive security literature covers how hackers can abuse VR devices as well as metaverse servers, networks, and databases [211, 234, 212, 212, 103].

(I) Hardware Adversaries control the hardware and firmware of the VR device and, thus, can access raw user inputs and arbitrarily manipulate the information provided to the application (2A) and presented to the user (1B).

(II) Client Adversaries represent the developers of the client-side VR application (*Application Adversary* [193]) and the content creators (*Content Adversary* [9]). Content adversaries can create *immersive falsehoods*, i.e., designing immersive experiences with misinformative, manipulative, and deceptive content [9]. Application adversaries can access the input data via system APIs, and arbitrarily manipulate the rendered frames and signals output to the VR devices (2B) and the information streamed to external servers (3).

(III) Server Adversaries control the external server facilitating multi-user functionality and, therefore, can arbitrarily process received networked data before streaming such information to other users' devices (4).

(IV) User Adversaries represent other users of the same VR application. They receive user data streams from a server and can interact with the target user.

VR Defenses

We highlight in Fig. 2.1 where the defenses can counter potential attacks and classify them based on five adapted categories. They consist of the two categories that De Guzman et al. [54] added to the primary three proposed by Roesner et al. [235], which are present in other privacy literature [95, 273, 297]. Given that many researchers highlighted the potential harm of deceptive immersive content [62, 180, 111, 4, 34, 279, 203, 27], we add a category for virtual content protection. Note that not all of these protections are related to *privacy* (§2.5), but also to *security* (i.e., measures to impede unauthorized data access [25]) and *safety* (i.e., measures to preserve the physical and mental well-being of users [62]). We highlight the following literature for guidance in security and safety attacks and protections: [54, 254, 205, 61, 294, 97, 70, 263]. We frame our SoK around attacks and defenses related to the *privacy* aspects of these defenses, mainly to input protection.

(I) Input Protection (Security & Privacy). Software that, e.g., perturbs [190] or abstracts [83] active (user) and passive (user's surrounding environment) sensitive input information to prevent user privacy breaches. Additionally, systems should be secured against adversarial inputs that deceive detection algorithms (cyberattack in Fig. 2.1: 1A).

(II) Data Access Protection (Security & Privacy). Active and passive user inputs are stored, relayed and accessed to deliver user-consumable output. The corresponding privacy and security measures extensively overlap with other systems, which existing literature covers comprehensively [95, 273, 297, 211, 110].

(III) Output Protection (Security & Safety). Detecting and censoring [204] malicious manipulation of outputs can prevent security breaches like "clickjacking" [235] or physical harm (e.g., inducing collisions with objects [279, 34], VR sickness [34], or epilepsy [16]).

(IV) User Interaction Protection (Privacy & Safety). Privacy protections can enhance confidentiality (i.e., data is only revealed to selected entities [95]) in physical or virtual spaces shared by multiple interacting users, e.g., a private virtual enclave that other users cannot enter [80]. We add to this category safety measures such as invisible avatar barriers to avoid psychological harm from virtual harassment [27] or buylling [205].

(V) Device Protection (Security & Safety). Device security measures can implicitly protect users and data in the above defensive aspects (e.g., authentication prevents impersonation [160]), and defend against cyberattacks targeting devices [205] and networks [102], and VR tracking system jamming [225], which could lead to physical harm.

(VI) Content Protection (Privacy & Safety). Safety measures such as age verification and content moderation can protect users against immersive falsehoods, and inappropriate, unsolicited, and harmful content that may lead to mental harm, disinformation, or manipulation of views and opinions [27, 180]. The privacy aspect relates to detecting virtual content and environments designed so that users are more likely to reveal sensitive information, e.g., prompting users to solve puzzles that reveal health data inconspicuously [193].

2.6 Taxonomy of VR Data & Applications

Thanks to the sensor-generated data and the applications processing this information, users can experience VR. However, applications are also the gateway for adversaries to harvest sensitive user data and use such information against them. The following classifies and discusses the data attributes and the applications subject to our threat model.

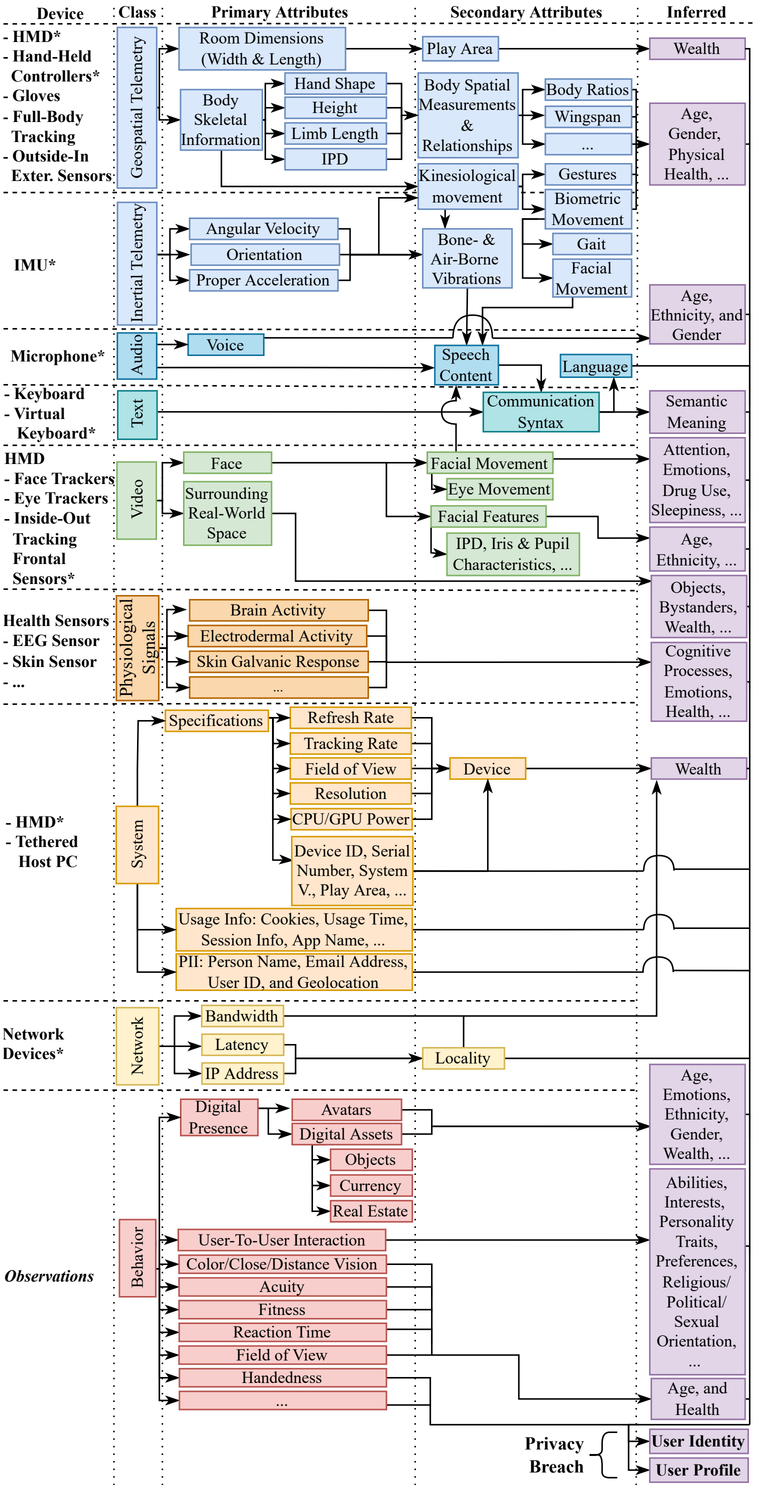


Figure 2.2: Taxonomy of VR data attributes.

VR Attributes

Using the same 68 publications discussed above, we now present a taxonomy of the data attributes that originate from using the input devices of § 2.2. Would-be attackers can collect these attributes at different stages of the VR information flow of § 2.5. Fig. 2.2 presents the resulting taxonomy of VR-derived data. We base our categorization on observable attribute classes and indicate which attributes or observations an attacker can directly capture from a data source (*primary*), deterministically derive from primary attributes (*secondary*), and infer from primary and secondary attributes employing ML or other learning procedures (*inferred*). Furthermore, we use the 68 publications to draw the connections between attributes, thus, there might be other connections outside VR and new ones might arise in future work, e.g., deriving ethnicity or personality traits from VR inertial telemetry.

Spatial & Inertial Telemetry. The position, orientation, and acceleration of body tracking devices over time reveal anthropometric measurements. Such measurements can be *direct* (body skeletal information such as arm-length and height [178]), *combined* to obtain further biometrics (e.g., wingspan [193]), or *compared* to draw relationships (ratios may reveal a user's body asymmetries [190]). An attacker may also record kinesiological movements, which can reveal unique gestures [254, 83], or biometric movements [207] such as gait [250]. Additionally, the devices' coordinates can map the play area's boundries, revealing its surface [193]. Even without full-body tracking devices, Winkler et al. [302] showed that reinforcement learning techniques could infer a full-body pose with telemetry from only an HMD, its IMUs, and hand-held controllers. Furthermore, Chen et al. [251] derived speech from the bone- and air-borne vibrations registered by an HMD's IMU telemetry data. Note that hardware and client adversaries have a privileged position to observe device telemetry. In contrast, server and user adversaries will experience degraded precision in their attribute estimations due to intermediate data processing, e.g., filtering and compression.

Audio & Text. Users can verbally interact with other users in virtual telepresence applications or give voice commands to their VR devices through a microphone [162]. Attackers can listen to vocalizations to fingerprint users based on vocal characteristics (e.g., frequency or accent) [193, 254] and profile them with communication semantics [71]. While voice biometrics may degrade along the data flow, speech semantics are more robust and could remain vulnerable to user adversaries. Additionally, the messaging functionality enabled by physical or virtual keyboards operated with hand-held controllers or gloves increases the attack surface [156, 12, 244].

Video. HMD's face optical sensors can register and track eye and facial movements and features to render expressive photorealistic avatars [39]. However, the facial video feed can also serve to identify an individual (e.g., using IPD, or Iris, and pupil characteristics [36, 129]) or infer emotions [314]. Notably, Kröger et al. [148] provided a comprehensive overview of the plethora of attributes that privileged adversaries can infer from eye tracking. Moreover, with expressive avatars, server and user adversaries could also learn other users' mental state. Additionally, while more prevalent in AR applications, the inside-out tracking frontal cameras of a VR HMD [170] also expose the real-world environment surrounding users, which can reveal sensitive information to hardware and client adversaries, such as personal objects [151, 312], the surrounding space type [107, 52], or bystanders [125, 124].

Physiological Signals. As health sensors like EEGs make their way into commercial-grade HMDs [200], the possibilities of VR (and privileged adversaries) expand dramatically. With these sensors, applications can adjust immersive experiences based on physiological signals that meet users' particular needs in real-time [62, 11, 208, 306] and can help users with rehabilitation treatments [254, 10, 238]. Such improvements, however, will also expose critically sensitive user information, such as physical and mental health conditions [62, 306, 238], behavior [268, 11, 22], language semantics [55, 269], and other sensitive PII like credit cards, PINs, and locations or persons known to the user [166].

System & Network. Adversaries can determine a user's VR device, host PC, network characteristics, and related internet session information [275]. Specifically, hardware and client adversaries can query system APIs to collect system specifications (e.g., tracking rate, resolution, etc.), and less privileged adversaries may devise attacks to gauge a target user's refresh rate without access to system APIs or user agents [193]. Notably, Trimananda et al. [275] captured the plethora of system information relayed to servers, which included all the above, in addition to PII like a person's name and usage information such as cookies or app names. While not specific to VR, as virtual telepresence applications rely on multiple servers to reduce perceived latency [292], attackers can observe network traffic to determine users' geolocation without an IP address. Altogether, these additional data points help adversaries fingerprint users to track them across internet VR sessions.

Behaviour. Observing users' avatar likeness, expressed emotions, interactions and reactions to virtual stimuli from other avatars or virtual content can reveal various sensitive human characteristics [149, 266]. In practice, malicious developers may carefully and inconspicuously deliver stimuli in a virtual experience to prompt the user to unconsciously reveal their reaction time, handedness, fitness level, visual and mental acuity, etc. [193]. Additionally, how a user chooses to represent their likeness as avatars, together with the digital assets they own, can reveal information such as their demographics or wealth [62, 130]. Lastly, user-to-user interaction in social VR can lead to attackers directly spying on or engaging with the target user [80, 291]. The information required to meaningfully observe user interactions is typically enough at each stage of the information flow for any attacker to extract such sensitive behavioral data.

Inferred Attributes. By deploying the appropriate machine learning algorithm [79, 213, 175], the attributes discussed above can reveal demographics [9] and other related sensitive attributes such as emotions [314], physical and mental health [148, 157], wealth, and political or sexual orientation or preferences over different users or products [80, 127], among others [62]. Users may also unintentionally or voluntarily self-disclose such information or additional biographical data (e.g., age, home address, education, social status, work history, etc.) [262, 266], or be deceived by the application or other users to reveal inferable attributes [9]. Ultimately, using known techniques, adversaries can leverage the breadth of harvested information to identify and profile users across VR applications.

VR Applications

For decades, the gaming industry has advanced 3D graphics hardware and low-latency content delivery to create immersive, time-intensive online user experiences. Their expertise has pushed gaming to become the current dominant application in VR [17]. However, VR promises applications beyond entertainment: social life, education, healthcare, fitness, military training, architecture, retail, business, productivity (virtual offices), engineering, and manufacturing [201, 40]. Specifically, social VR has recently increased in popularity with titles such as *VRChat*, whereby users worldwide interact with each other in real-time [266].

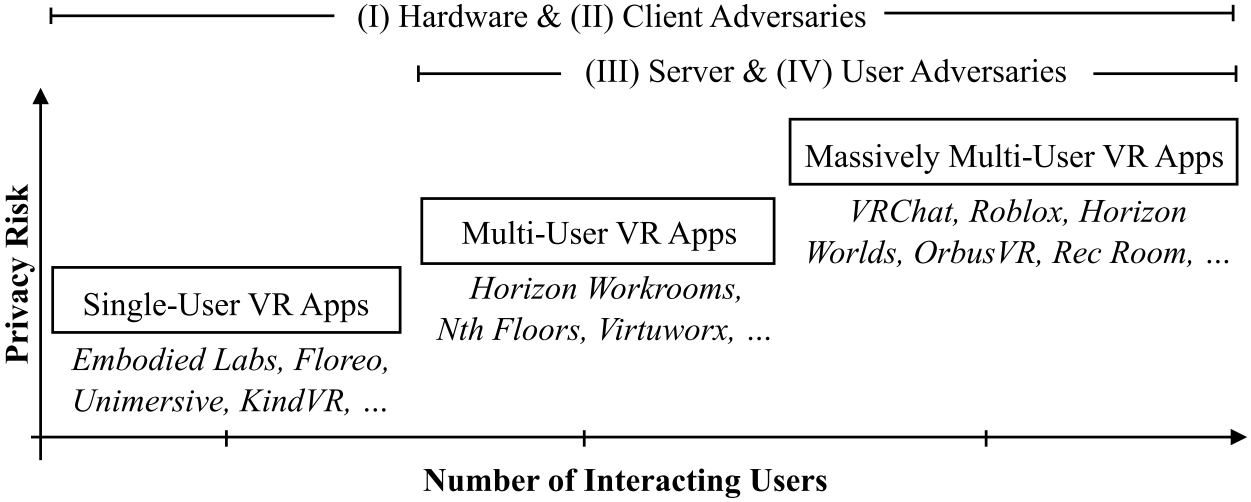

(I) Hardware & (II) Client Adversaries
(III) Server & (IV) User Adversaries
Privacy Risk
Single-User VR Apps
Embodied Labs, Floreo, Unimersive, KindVR, ...
Multi-User VR Apps
Horizon Workrooms, Nth Floors, Virtuworx, ...
Massively Multi-User VR Apps
VRChat, Roblox, Horizon Worlds, OrbusVR, Rec Room, ...
Number of Interacting Users

Figure 2.3: Privacy risk of VR applications as adversary exposure increases.

Only two of the 68 collected studies classified VR applications based on the target industry [71, 294]. We provide an orthogonal categorization from a privacy standpoint based on our threat model and taxonomy of attributes. Accordingly, we contemplate privacy risks in VR from three perspectives: (i) *adversarial*, (ii) *defensive*, and (iii) *data sensitivity*. Accordingly, VR application developers may consider answering three questions:

(i) *How much adversarial exposure could the application suffer?* Fig. 2.3 shows the prevalence of hardware and client adversaries across all applications and the rise in privacy risks as users require servers to interact with others. While massively multi-user VR applications such as social VR are the most privacy-hostile environments, single-user applications are at least vulnerable to the VR firmware itself, as it may have direct network access to exfiltrate collected data from an application (e.g., Oculus Quest 2).

(ii) *How much privacy is the user willing to forgo?* Some users are willing to expose any information necessary to experience VR at its full immersive potential, while others are more reserved [61]. If protecting or opting out of specific data inputs is possible, the privacy risks an application entails may vary from user to user [190]. Ideally, developers should offer these choices and design applications such that the privacy preferences are customizable.

(iii) *How sensitive is the data handled by the application?* Most VR applications ingest *spatial* and *inertial telemetry*, and require a *system* and a *network* to join interactive experiences, where adversaries can extract *behavioral information*. These attribute classes form a privacy risk baseline. The application context raises the risks above this baseline, e.g., virtual health clinics, classrooms, and offices handle more PII and critically sensitive data than a game, such as *physiological signals*, *text* in homework or emails, and context-specific behavioral information such as attention to the lecturer or emotions during a meeting.

2.7 VR Privacy Attacks

Of the 68 collected studies, we found 30 attacks introducing explicit, offensive mechanisms (15) or methods that an attacker could leverage for adversarial purposes (15). For example, an attacker can misuse motion-based authentication models to perform identification attacks across VR sessions. We systematically classified these 30 attacks in Table 2.3 (labeled with IDs A2 to A31) based on the threat model of §2.5 and attribute classification of §2.6. We categorized the attacks according to the information presented in the associated papers and included the most distinct or prevalent metrics to benchmark the attacks. Where information was lacking (e.g., not all attacks had an explicit adversary model), we used our best judgment supported by the publications artifacts (e.g., the client was the most common adversary, and studies such as A9, A11, and A12 developed an application). We then present a set of research questions that use the findings of prior studies to motivate the rest of this dissertation.

ID	Name	Devices	Observable Attribute Classes									PB		Adversary				Metric
			Spat. Tele.	Iner. Tele.	Text	Audio	Video	Phys. Sig.	System	Network	Behavior	Profiling	Ident.	Hardware	Client	Server	User	
A2	Malicious Design [9]	N/A	-	-	-	-	-	-	-	-	●	●	-	-	●	-	-	N/A
A3	*MitR* [291]	N/A	-	-	-	●	-	-	-	-	●	●	●	-	-	-	●	N/A
A4	*QuestSim*† [302]	🥽 🎮 🧭 ◎	●	●	-	-	-	-	-	-	-	-	●	-	●	-	-	Geo. Errors
A5	*Face-Mic* [251]	🧭 ◎	-	●	-	-	-	-	-	-	-	●	●	-	●	-	-	Accuracy
A6	*GaitLock*† [250]	🧭 ◎	-	●	-	-	-	-	-	-	-	-	●	●	-	-	-	Accuracy
A7	Movement Biometrics† [180]	🥽 🎮 🧭	●	●	-	-	-	-	-	-	-	-	●	●	-	-	-	EER
A8	Movement Biometrics [178]	🥽 🎮 🧭	●	●	-	-	-	-	-	-	-	-	●	-	●	-	-	Accuracy
A9	Movement Biometrics† [154]	🥽 🎮 🧭	●	●	-	-	-	-	-	-	-	-	●	-	●	-	-	Accuracy
A10	Movement Biometrics [274]	🥽 🎮 🧭 👁	●	●	-	-	●	-	-	-	-	●	●	-	●	-	-	F1-Score
A11	Movement Biometrics† [214]	🥽 🎮 🧭 ◎ 👁	●	●	-	-	●	-	-	-	-	-	●	-	●	-	-	Accuracy
A12	*BioMove*† [207]	🥽 🎮 🧭 ◎ 👁	●	●	-	-	●	-	-	-	-	-	●	-	●	-	-	Accuracy
A13	Eye Tracking‡ [52]	👁	-	-	-	-	●	-	-	-	-	-	●	-	●	-	-	Accuracy
A14	Iris Identification ‡ [36]	👁	-	-	-	-	●	-	-	-	-	-	●	-	●	-	-	Accuracy
A15	*Kalεido*‡ [152]	👁	-	-	-	-	●	-	-	-	-	●	●	-	●	-	-	F1-Score
A16	*EMOShip*† [314]	👁	-	-	-	-	●	-	-	-	●	●	-	-	●	-	-	F1-Score
A17	Spatial Recognition [107]	📹	-	-	-	-	●	-	-	-	-	●	-	-	●	-	-	MER
A18	Spatial Recognition‡ [151]	📹	-	-	-	-	●	-	-	-	-	●	-	-	●	-	-	F1-Score
A19	*SafeMR*‡ [106]	📹	-	-	-	-	●	-	-	-	-	●	-	-	●	●	-	Accuracy
A20	*Vreed*† [268]	👁 💓	-	-	-	-	●	●	-	-	-	●	-	-	●	-	-	Signal Statistics
A21	*Galea*† [22]	👁 💓	-	-	-	-	●	●	-	-	●	●	-	●	-	-	-	Signal Statistics
A22	Signal Processing† [167]	🧭 ◎ 💓	-	●	-	-	-	●	-	-	-	●	-	-	●	-	-	EER
A23	Signal Processing† [11]	◎ 💓	-	●	-	-	-	●	-	-	●	●	-	-	●	-	-	Signal Peaks
A24	Signal Processing† [208]	💓	-	-	-	-	-	●	-	-	-	●	-	●	-	-	-	Signal Stability
A25	Signal Processing† [238]	💓	-	-	-	-	-	●	-	-	-	●	-	-	●	-	-	Signal Statistics
A26	Signal Processing† [306]	💓	-	-	-	-	-	●	-	-	-	●	-	-	●	-	-	Accuracy
A27	Signal Processing† [122]	💓	-	-	-	-	-	●	-	-	-	●	-	-	●	-	-	Signal Statistics
A28	Virtual Typing [156]	🥽 🎮 🧭 ◎ 📹	●	●	●	-	●	-	-	-	-	●	●	●	●	-	-	Accuracy
A29	*VR-Spy* [12]	📶	-	-	●	-	-	-	-	-	-	●	●	-	-	●	-	Accuracy
A30	Self-Disclosure† [266]	N/A	-	-	-	●	-	-	-	-	●	●	●	-	-	-	●	N/A
A31	Movement Biometrics [196]	🥽 🎮	●	-	-	-	-	-	●	-	●	-	●	-	-	●	●	Accuracy

Table 2.3: Systematization of VR attacks from collected papers.

In interpreting Table 2.3:

†An attacker can leverage the defense/mechanism for adversarial purposes.

‡Although the study is defense-focused, there is an adversarial component.

Names: Names in italics correspond to the preferred title; other names are descriptive.

VR Device: 🕶 = HMD, 👁 = Eye Trackers, 📹 = Inside-Out Tracking Optical Sensors, ▶■ = Outside-In Tracking Optical Sensors, ◔ = IMU Orientation, ◉ = IMU Velocity, 🎮 = Hand-Held Controllers, 🎤 = Microphone, 🖴 = Tethered PC, 📶 = Network Devices, 💓 = Health Sensor, N/A = Not Applicable.

Abbreviations: Spat. = Spatial, Iner. = Inertial, Tele. = Telemetry, Phys. = Physiological, PB = Privacy breach, MER = Mean error rate, EER = Equal error rates.

(RQ1) How do VR devices enable unprecedented attack opportunities? The relevant literature demonstrates that the unique privacy risks of VR devices stem mostly from their vast array of sensors and inputs, which generally capture far more user data than other mobile computing platforms. For example, most identification attacks rely on *HMDs* and *hand-held controllers* to capture kinesiological movements (A6-12), while eye trackers mainly have a supportive role (A10-12). Profiling attacks that predict sensitive information, such as emotions (e.g., arousal and stress levels), often rely on built-in *health sensors* (A20-27). These attacks use devices such as EEGs (A26), EMGs (A23), and ECGs (A20, A26), but also blood pressure (A25), galvanic (A20, A25-27), thermal (A24-25), respiratory (A26-27), and photoplethysmographic (A9, A22) sensors. Studies show that accelerometer and EMG data are an effective combination for identifying users' reactions to virtual stimuli (A23), and EEGs are especially suitable for predicting emotions (A20). Thus, while some VR devices and applications have specific security vulnerabilities [201], the vast majority of security and privacy threats in VR stem from misusing sensor data intended to facilitate legitimate application functionality. Therefore, throughout this dissertation, we are motivated to research attacks that misuse VR sensor data rather than leveraging application vulnerabilities.

(RQ2) Which VR data attributes expose attack opportunities? The *spatial telemetry* of HMDs and hand-held controllers clearly represent low-hanging fruit for adversaries. Attackers can easily measure anthropometrics like height and wingspan to uniquely identify a small set of users (A1), as well as to register uniquely identifiable motions and gestures such as pointing (A11). Combined with *inertial telemetry*, an attacker can infer a user's full-body pose, even with avatars of different scales (A4), perform highly accurate identification attacks (A7, A8, A12), and infer age (A10). In addition to passive attacks, there is always the danger of intentional or unintentional self-disclosure through movement (A30). Accordingly, throughout this dissertation, we are motivated to focus on VR motion data, particularly head and hand motion, as a near-universal privacy risk in VR.

(RQ3) How invasive are VR attacks? The malicious accumulation of user data through profiling and tracking across internet sessions can lead to surveillance advertisement [46], price discrimination [93], cyber abuse [27], personal autonomy curtailment [62], and pushing political agendas [206], among other risks [111, 4]. These threats are exacerbated when adversaries have access to users' emotional states and reactions to stimuli [62], which are more easily observable in VR (A20). We are motivated to better understand the set of attributes inferable in VR, which we explore in Chapter 5 through a large-scale user study.

(RQ4) What is the true scale of the VR privacy threat? Most VR privacy attack studies remain relatively small, with the largest prior study containing about 500 users (A8). However, most VR applications are several orders of magnitude larger than this, with future metaverse environments potentially hosting millions of users. Researchers currently lack sufficient data to understand the scale of the VR privacy threat in comparison with conventional biometrics. Therefore, we are motivated to identify larger datasets for VR privacy research, which we discuss in Chapter 3, and to massively scale up VR privacy attack studies, which we do in Chapter 4.

(RQ5) How practical and effective are privacy attacks? Based on the literature, *user* adversarial attacks are easy to execute (A3, A30), as users can, at a minimum, join a VR session and social engineer information from users. Furthermore, *identification* attacks targeting kinesiological movements are highly accurate, with the most effective *identification* attacks targeting achieving an accuracy of 98% using only IMUs (A6). However, all of these attacks rely on passive observation, and what is less understood is whether an active adversary can access additional capabilities. Accordingly, we are motivated to research the privacy risks of adversarial VR game design, which we discuss in Chapter 6.

(RQ6) Where do VR attacks lie on a comprehensive threat model? Finally, it is likely combining several of the discussed attacks could further enhance adversarial capabilities. For example, combining identification and profiling attacks may allow an adversary to track users across sessions, curating an increasingly detailed profile of the user over time. Throughout this dissertation, we refer back to the threat model established above to illustrate how several attacks and defenses may interact with each other.

2.8 VR Privacy Defenses

Following an equivalent method to VR attacks, we now turn our attention to the 35 identified defenses (labeled with IDs D2 to D35) according to the threat model of §2.5 and attribute classification of §2.6. Table 2.4 systematically categorizes the defenses based on the corresponding papers. As before, we designed a set of research questions based on the existing literature to motivate the later chapters of this dissertation.

			Defended Attribute Classes									BP		Protection						
ID	Name	Devices	Spat. Tele.	Iner. Tele.	Text	Audio	Video	Phys. Sig.	System	Network	Behavior	Profiling	Ident.	Input	Data Access	Output*	User Inter.	Device	Content	Metric
D2	Eye Tracking [52]	👁	-	-	-	-	●	-	-	-	-	-	●	●	-	-	-	-	-	F1-Score
D3	Iris De-Identification [36]	👁	-	-	-	-	●	-	-	-	-	-	●	●	-	-	-	-	-	Accuracy
D4	Kalεido [152]	👁	-	-	-	-	●	-	-	-	-	-	●	●	-	-	-	-	-	Accuracy
D5	Eye Tracking [262]	👁	-	-	-	-	●	-	-	-	-	●	●	●	-	-	-	-	-	Accuracy
D6	Eye Tracking [298]	👁	-	-	-	-	●	-	-	-	-	●	●	●	-	-	-	-	-	QoE
D7	Eye Tracking [53]	👁	-	-	-	-	●	-	-	-	-	-	●	-	●	-	-	-	-	Accuracy
D8	Eye Tracking [157]	👁	-	-	-	-	●	-	-	-	-	●	-	-	●	-	-	-	-	CC, MSE
D9	Eye Tracking [130]	👁	-	-	-	-	●	-	-	-	-	-	●	●	-	-	-	-	-	CRR
D10	Eye Tracking [26]	👁	-	-	-	-	●	-	-	-	-	-	●	-	●	-	-	-	-	CC, NMSE
D11	*EyeVEIL* [129]	👁	-	-	-	-	●	-	-	-	-	-	●	●	-	-	-	-	-	Accuracy
D12	*PrivacEye* [261]	👁 📹	-	-	-	-	●	-	-	-	-	●	-	●	-	-	-	-	-	Accuracy
D13	Spatial Recognition [151]	📹	-	-	-	-	●	-	-	-	-	●	-	-	●	-	-	-	-	F1-Score
D14	*SafeMR* [106]	📹	-	-	-	-	●	-	-	-	-	●	-	-	●	-	-	-	-	Accuracy
D15	OS Support [50, 125]	📹	-	-	-	-	●	-	-	-	-	●	-	●	-	-	-	-	-	FN, FP
D16	Spatial Recognition [104, 105]	📹	-	-	-	-	●	-	-	-	-	●	-	●	-	-	-	-	-	Accuracy
D17	*PlaceAvoider* [270]	📹	-	-	-	-	●	-	-	-	-	●	-	●	-	-	-	-	-	Accuracy
D18	*Darkly* [124]	📹	-	-	-	-	●	-	-	-	-	●	●	●	-	-	-	-	-	# Breaches
D19	Spatial Recognition [312]	📹	-	-	-	-	●	-	-	-	-	●	-	●	-	-	-	-	-	Accuracy
D20	Spatial Recognition [267]	📹	-	-	-	-	●	-	-	-	-	-	●	●	-	-	-	-	-	N/A
D21	*OpenFace* [293]	📹	-	-	-	-	●	-	-	-	-	-	●	●	-	-	-	-	-	Accuracy
D22	*GARP-Face* [66]	📹	-	-	-	-	●	-	-	-	-	-	●	●	-	-	-	-	-	Accuracy
D23	GAN-Based Defense [30]	📹	-	-	-	-	●	-	-	-	-	-	●	●	-	-	-	-	-	Natruralness
D24	GAN-Based Defense [308]	📹	-	-	-	-	●	-	-	-	-	-	●	●	-	-	-	-	-	Accuracy
D25	*Prepose* [83]	📹	●	-	-	-	●	-	-	-	-	-	●	●	-	-	-	-	-	Expression
D26	Movement Biometrics [154]	🕶 🎮 🎯	●	●	-	-	-	-	-	-	-	-	†	-	-	-	-	●	-	Accuracy
D27	Movement Biometrics [214]	🕶 🎮 🎯 ◎ 👁	●	●	-	-	●	-	-	-	-	-	†	-	-	-	-	●	-	Accuracy
D28	*GaitLock* [250]	🎯 ◎	-	●	-	-	-	-	-	-	-	-	†	-	-	-	-	●	-	Accuracy
D29	*BioMove* [207]	🕶 🎮 🎯 ◎ 👁	●	●	-	-	●	-	-	-	-	-	†	-	-	-	-	●	-	Accuracy
D30	Digital Presence [80]	N/A	-	-	-	-	-	-	-	-	●	●	●	-	-	-	●	-	●	N/A
D31	*SecSpace* [229]	N/A	-	-	-	-	-	-	-	-	●	-	●	-	-	-	●	-	-	N/A
D32	Design Defense [9]‡	N/A	-	-	-	-	-	-	-	-	●	●	-	-	-	-	●	-	●	N/A
D33	*MitR Defense* [291]‡	N/A	-	-	-	●	-	-	-	-	●	●	●	-	-	-	●	-	-	FN, FP
D34	Self-Disclosure Defense [164]	N/A	-	-	-	●	-	-	-	-	●	●	●	-	-	-	●	-	-	N/A
D35	*ReconViguRation* [244]	⌨	-	-	●	-	-	-	-	-	-	●	●	●	-	-	-	-	-	Error Rate

Table 2.4: Systematization of VR defenses from collected papers.

In interpreting Table 2.4:

‡Although the study is attack-focused, there is a defensive component.

†Authentication protection (as opposed to identification protection).

*Output safety and security attacks and defenses are covered in dedicated literature [235, 279, 34, 204, 16].

Names: Names in italics correspond to the preferred title; other names are descriptive.

VR Device: 🕶 = HMD, 👁 = Eye Trackers, 📹 = Inside-Out Tracking Optical Sensors, 🎥 = Outside-In Tracking Optical Sensors, ⊙ = IMU Orientation, ◉ = IMU Velocity, 🎮 = Hand-Held Controllers, 🎤 = Microphone, 📶 = Network Devices, ⌨ = Physical keyboard; N/A = Not Applicable/Available.

Abbreviations: Spat. = Spatial, Iner. = Inertial, Tele. = Telemetry, Phys. = Physiological, Inter. = Interaction, BP = Breach prevention, QoE = Quality of experience, CC = Correlation coefficient, NMSE = Normalized mean squared error, CRR = Correct recognition rate, FN = False negative, FP = False positive.

(RQ7) What are the general categories of proposed VR privacy defenses?

(i) *Perturbation.* Some studies provide privacy guarantees by adding noise to spatial or eye tracking data (e.g., D4), while others blur (e.g., D11, D18) or mask (e.g., D3, D17) regions of a video like facial features, sensitive objects, and bystanders. Perturbation is the primary privacy-preserving technique explored in the later parts of this dissertation.

(ii) *Information abstraction.* Alternatively, some studies suggest using software that extracts exclusively the relevant features from the surrounding space (e.g., surfaces, D16) or shares only the events triggered by sensitive inputs (e.g., unique gestures, D25).

(iii) *Recognizers.* Studies have proposed automated deepfake detection (D32) or middleware that detects and warns the user of sensitive surrounding objects and bystanders (D15).

(iv) *Static & dynamic analyzers.* Research from the application security domain can be repurposed for the detection of VR application vulnerabilities, e.g., unauthorized access to a virtual room (D33), or malware that exfiltrates sensitive surrounding objects (D13).

(v) *Platform features.* Some studies suggest platform-level defenses, primarily to protect user interactions. Examples include virtual private enclaves that only authorized users can access (D31). Other defenses focus on confusing adversaries, e.g., with avatar clones dispersed across multiple VR applications, teleportation to new virtual locations, private copies of the virtual public environment, and platform-generated non-identifiable or invisible avatars (D30). Furthermore, platforms could include embedded voice modulators and social media privacy settings, whereby, for example, only friends can see one's avatar (D32).

(vi) *Authentication.* Finally, motion-based identification can also be deployed defensively, such as for logging into a VR device or defeating sybil attacks (D26-29).

(RQ8) How do defenses balance usability and privacy? Practical VR privacy defenses must maintain the usability and immersion of VR applications; thus, researchers have designed utility metrics to assess the loss of usability when enabling VR privacy protections. Aspects that impact usability are *energy consumption* (D14, D17), *latency* (D2, D4), and *playability* (D1, D4, D9, D14), i.e., how enjoyable or productive a VR experience is. Approaches that help to minimize *energy consumption* are use of a tethered PC, offloading computation to the cloud (though bandwidth may be a challenge, D17), and sharing processing resources like object detection with other applications, which also reduces *latency* (D14). VR protections can decrease *playability* if the defense perturbs data, which is measurable via metrics such as game scores (D4), subjective enjoyment (D4), attentiveness, comfort (D9), naturalness (D23), and accuracy loss (D4, D16). In Chapter 7, we use "theoretical optimality" as our usability goal, defined as minimizing mean-squared error at any privacy level, while in Chapter 8, we use "indistinguishability" as our usability goal.

(RQ9) What are the common limitations of VR privacy defenses? One common pitfall in the design of privacy-preserving systems for VR is not respecting causality; i.e., "looking into the future" when anonymizing sequential telemetry data. Mechanisms that do so may display impressive evaluation metrics when evaluated asynchronously but are unsuitable for real-world deployment in a streaming setting. This motivates us to emphasize real-time, low-latency, causal defenses in Chapters 7 and 8. Another common limitation in proposed defenses is the lack of customizability, with different VR applications having vastly different considerations in the balance of privacy and usability. Accordingly, the defenses in Chapters 7 and 8 are designed to offer tunable privacy parameters.

(RQ10) How comprehensive are VR privacy defenses? Researchers typically implement defenses as middleware (D15-16, D18) that pre-processes data before a potentially malicious application ingests it, or as an easy-to-install plugin within the VR application. However, the latter would only defend against *server* and *user* adversaries while the former would also defend against *application* adversaries. It is particularly challenging to defend against *hardware* adversaries without being able to directly audit VR device firmware. On the other hand, it is at least somewhat feasible to implement and enforce policy-based solutions to VR *hardware* and *application* threats, while VR *servers* and *users* may be scattered around the world, presenting jurisdictional issues. Therefore, the defenses in Chapters 7 and 8 primarily focus on mitigating the threats presented by *server* and *user* adversaries.

2.9 VR Privacy Opportunities

Coverage of Attacks & Defenses

Table 2.5 shows the most comprehensive attacks per attribute class and the most fitting applicable defenses, if any. As illustrated below, while many of the known attacks have at least one associated defense, there are no comprehensive defenses protecting spatial and inertial telemetry data for identification and profiling. As such, protecting VR motion data is the primary focus of the defenses we present in Chapters 7 and 8.

Class	**Privacy Attack**	**Privacy Defense**
Identification		
Spatial Telemetry	(A12) *BioMove*†	⋆
Inertial Telemetry	(A6) *GaitLock*†	⋆
	(A5) *Face-Mic*	⋆
Text	(A28) Virtual Typing	(D35) *ReconViguRation*
	(A29) *VR-Spy*	(D35) *ReconViguRation*⋆, VPN, Tor, Proxies, etc.
Audio	(A3) *MitR*	(D33) *MitR Defense*‡
	Speech Recognition	Voice Modulation [198, 164]
Video	(A13) Eye Tracking‡	(D2) *Kalεido*⋆
	(A14) Iris Identifi.‡	(D11) *EyeVEIL*
Physio. Signals	⋆	⋆
System	⋆	⋆
Network	⋆	⋆
Profiling		
Spatial Telemetry	(A10) Movement Bio.	⋆
Inertial Telemetry	(A5) *Face-Mic*	⋆
	(A10) Movement Bio.	⋆
Text	(A5) *Face-Mic*	⋆
	(A10) Movement Bio.	⋆
Audio	(A3) *MitR*	(D33) *MitR Defense*‡⋆
Video	(A16) *EMOShip*†	(D12) *Kalεido*⋆
	(A17) Spatial Recog.	(D16) Spatial Re.
	(A18) Spatial Recog.‡	(D13) Spatial Re.
Physio. Signals	(A20) *Vreed*†	⋆
	(A22) Signal Proces.†	⋆
Behavior	(A3) *MitR*	(D33) *MitR Defense*‡⋆
	(A2) *Malicious Design*	(D32) Design Defense‡⋆

Table 2.5: Coverage of attacks and defenses.

†An attacker can leverage the associated defense/mechanism for adversarial purposes.

‡While the study is attack/defense focused, there is a(n) defensive/adversarial component.

⋆Privacy opportunity.

2.10 Discussion

VR Defenses in Practice

Of the 54 studies focused on attacks and defenses, we found that less than 20% had a functional open-source repository: 3 defenses (D16, D25, D33) and 6 attacks (A7-8, A16-17, A21, A27). In addition to increasing the difficulty of building on prior work, this may be a factor limiting the transfer of academic privacy research into the VR industry. By contrast, producing transparent and reproducible results is a major focus of this dissertation, with Chapter 9.2 providing open-source code for every experiment contained herein.

Furthermore, most defensive tools have remained entirely in the academic domain, with limited industry collaboration. Of the 54 studies focused on attacks and defenses, *Prepose* (D25) had an official affiliation with Microsoft, with no evidence of its use in production, and *EMOShip* (A16) forms part of the technology stack of Pupil Labs. The Bigscreen company used the recommendations from *MitR Defense* (D33) to patch their privacy vulnerabilities. Beyond these three examples, we scarcely observe VR privacy research deployed in practice. For this reason, we developed the technique of Chapter 8 in direct collaboration with the VR industry, with the goal of producing a system that is suitable for real-world deployment.

Key Findings

A few key findings have emerged from studying the 68 selected publications and results. First, there is a lack of understanding about the true scale of the VR privacy threat. There is a pressing need for large-scale datasets and studies that approach the scale of conventional biometric technologies. Second, there is a fundamental imbalance between the capabilities of VR attackers and defenders. While VR attacks can be slow, asynchronous, and non-causal, VR defenses must be fast, usable, and low-latency. This has resulted in many attacks lacking corresponding defenses, as shown in Table 2.5 above. Finally, security and usability must be kept in careful balance, with usability and performance metrics being incorporated into all defensive research in this area. As emphasized in the underlined portions of this chapter, these findings have informed the approach taken in the remainder of this dissertation.

2.11 Conclusion

In this chapter, we have drawn the landscape of data privacy in VR by proposing a comprehensive VR information flow, threat model, and attribute taxonomy, as well as outlining the open privacy opportunities in this field. In the following chapter, we describe a novel dataset that can be used to address many of the unanswered questions described above. Then, in the rest of this dissertation, we use that dataset, along with the frameworks presented here, to explore new VR privacy attacks and defenses with novel insights for the field as a whole.

Chapter 3

Dataset: 4,700,000 Motion Recordings from 105,000 Virtual Reality Users

3.1 Introduction

For decades, human motion capture (MoCap) recordings have been an important resource in a variety of fields, ranging from animation and computer-generated imagery (CGI) to authentication and human-computer interaction (HCI). As discussed in the previous chapters, the proliferation of extended reality (XR) devices has created a prominent new application for this data, with motion data being central to almost all XR and "metaverse" experiences. Since 2002, at least 25 motion capture datasets have been created based on laboratory studies of up to a few hundred users to facilitate research in this important domain.

A significant area of interest in this dissertation is the passive identification and authentication of XR users based on their movement patterns. However, as demonstrated by our literature survey in Chapter 2, XR identification and authentication studies have, until now, been limited to a few hundred users due to the lack of large-scale human motion datasets. By contrast, studies involving traditional biometrics, such as fingerprints or facial recognition, often use datasets with 100,000 or more subjects [301].

In this chapter, we introduce the BOXRR-23 dataset, which contains 4,717,215 motion capture recordings uploaded by 105,852 XR device users from over 50 countries. Our data is derived from two popular VR games, "Beat Saber" and "Tilt Brush." In addition to being more diverse and ecologically valid than laboratory studies, BOXRR-23 is over 200 times larger than the largest known public motion capture dataset. This dataset is used in the following chapters to enhance our understanding of VR security and privacy; for example, we use this dataset in Chapter 4 to demonstrate that XR motion data provides a biometric signal on par with fingerprints. However, the potential uses of this data could go far beyond XR security and privacy to include areas such as motion synthesis, human-computer interaction, and theoretical machine learning research.

In addition to assembling this dataset from three public sources and enriching it with additional metadata, we developed a new lossless “Extended Reality Open Recording” (XROR) file format due to the lack of an existing format suitable for this use case. The XROR format is about 30% more space efficient than the original motion capture file formats.

To help interested researchers evaluate this dataset, and to clarify the results in the remainder of this dissertation, we provide documentation pursuant to the Dataset Nutrition Label [113] standard. Furthermore, we conduct and analyze a large-scale survey ($N = 1,006$) of the users contained in this dataset to better understand their demographics.

3.2 Background

Since the 1990s, computerized motion tracking systems have been used for animation and CGI in a large number of popular movies, television series, and video games. A typical commercial motion capture solution uses optical tracking or inertial measurement units (IMUs) to measure the location of various parts of the body, with prices ranging from $10,000 to over $250,000 for a full-body tracking system. Conventional motion capture datasets have involved expensive laboratory studies with up to 300 subjects paid to perform a variety of tasks while wearing a professional motion capture setup. However, as discussed above, motion capture data is also central to the operation of extended reality (XR) systems, which use either external or onboard sensors to measure the position and orientation of the user’s head and hands in 3D space. In essence, XR devices have recently become an affordable and widely adopted form of motion tracking system. The motion data generated by an XR device is used by a client-side application, such as “Beat Saber” or “Tilt Brush,” to render auditory, visual, and haptic stimuli, creating an immersive 3D experience. In some cases, users capture and share recordings of the motion data generated during an XR usage session to allow other users to “replay” the same virtual experience.

Beat Saber

“Beat Saber” [88], shown in Figure 3.1, is a VR rhythm game where players slice blocks representing musical beats with a pair of sabers they hold in each hand. It is the primary data source for the BOXRR-23 dataset. With over 6 million copies sold, Beat Saber is the most popular VR application of all time [303]. The game contains a number of “maps,” which consist of an audio track (typically a song) and a series of objects presented to the user in time with the audio. These objects include “blocks,” which the player must hit at the correct angle with the correct saber, “bombs,” which the player must avoid hitting with their sabers, and “walls,” which the player must avoid with their head. The player is given a score based on their accuracy in completing these tasks. Reacting to these events typically requires skilled users to deploy high-speed “ballistic” movements [284, 65].

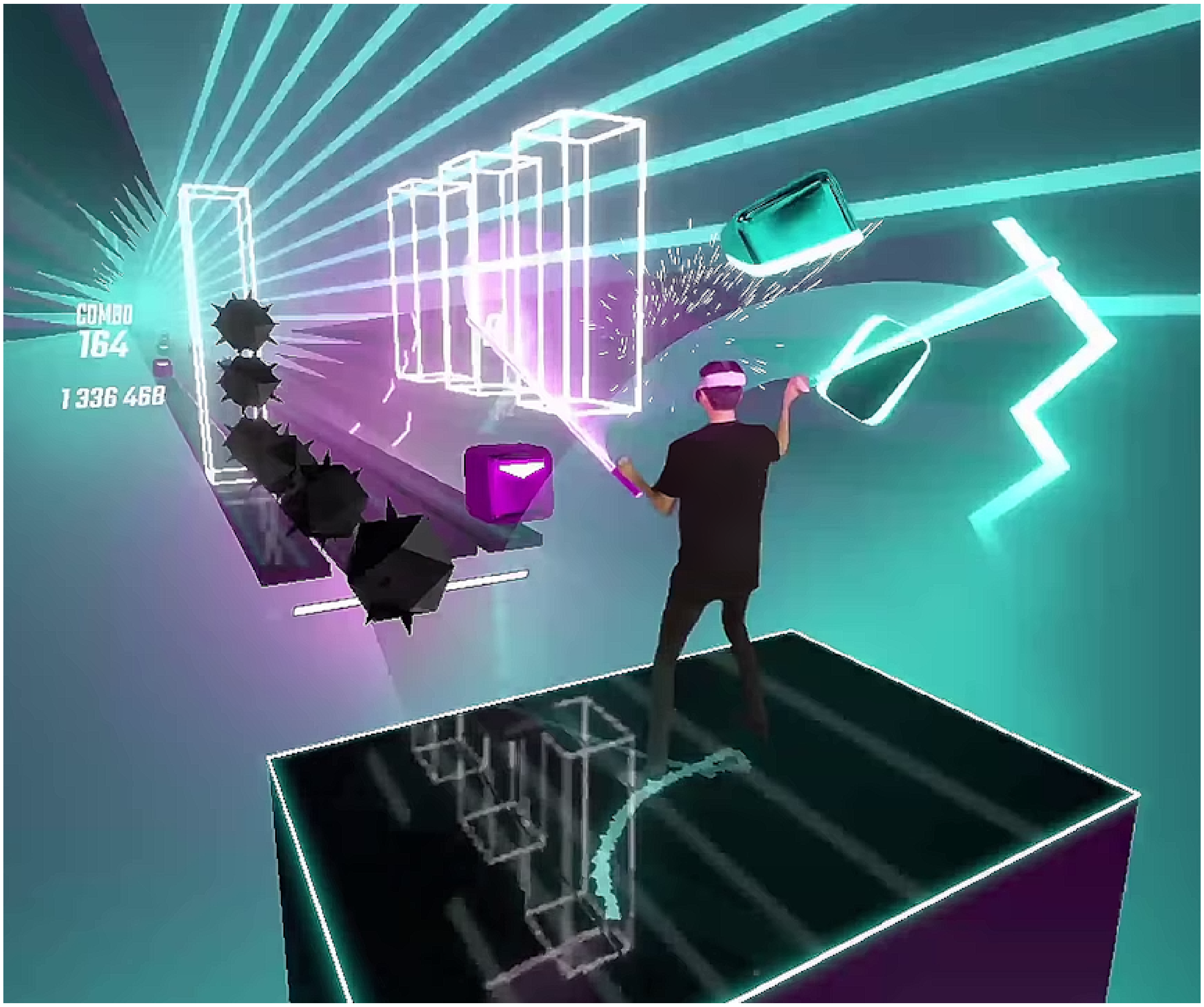


Figure 3.1: “Beat Saber,” a popular VR rhythm game.

While hundreds of maps are included in the base game, over 100,000 user-created maps can be played by installing open-source game modifications. Beat Saber enthusiasts may choose to install open-source leaderboard extensions in order to compete with other players to achieve a higher “rank” on the leaderboards for popular maps. Two of the most popular Beat Saber leaderboard services are “BeatLeader” [222] and “ScoreSaber” [247], with a combined 4 million scores being submitted to the platforms to date. When submitting a score to either of these services, users attach a motion capture recording of them playing the corresponding Beat Saber map, which is then made publicly available on the BeatLeader or ScoreSaber website to allow others to audit the legitimacy of the claimed score.

Figure 3.2: “Tilt Brush,” a popular VR painting app.

Tilt Brush

“Tilt Brush” [271], shown in Figure 3.2, is a VR painting game created by Google that allows users to create 3D virtual objects using a variety of brushes and tools. Users can then export their drawings in various file formats, along with a motion capture recording of them creating the object, allowing other users to re-watch the original painting process. From 2017 to 2021, Google hosted “Google Poly,” a free service for sharing virtual creations (and accompanying motion capture recordings) from Tilt Brush. After the shutdown of Google Poly in 2021, the “PolyGone” project [216] was created to host a free archive of over 50,000 user-submitted creations from Google Poly under a CC-BY license. Contrary to Beat Saber, Tilt Brush motion consists primarily of precise fine motor movements.

3.3 Data Collection

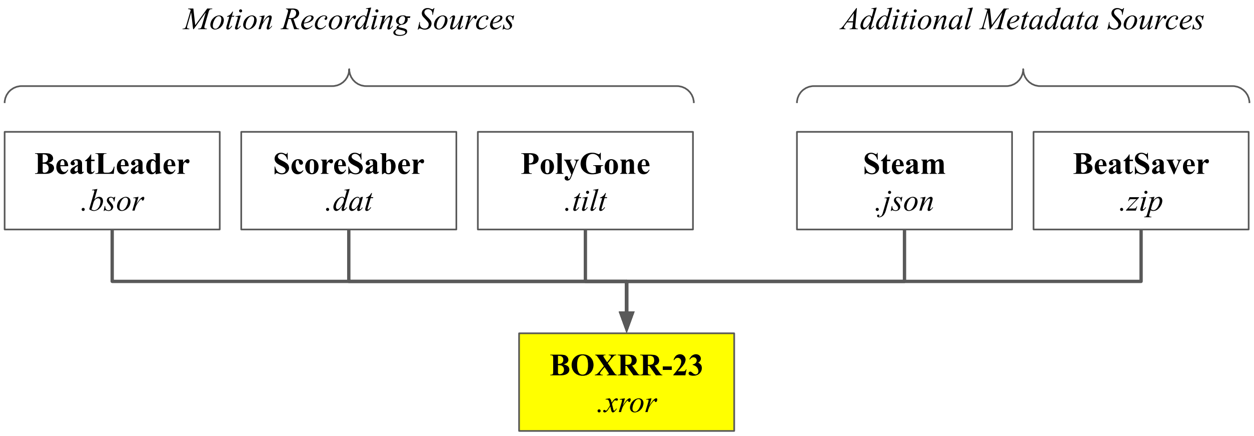


Figure 3.3: Data collection/processing pipeline for BOXRR-23.

Figure 3.3 shows the data collection process used to produce the BOXRR-23 dataset. We downloaded over 4.7 million publicly available motion capture recordings stored on the BeatLeader, ScoreSaber, and PolyGone websites, and obtained additional metadata information, such as player experience levels and in-game events, from the public web APIs of Steam [259] and BeatSaver [18]. We then removed identifiable details like player IDs and pseudonyms to protect the identity of each user. Finally, we converted all recordings from their original formats into our purpose-built XROR format, described in §3.6. The sizes of each of the sources, and of the dataset, are summarized in Table 3.1. We performed this data collection process in April 2023 and have included all valid, non-corrupt recordings submitted to all three platforms between November 1st, 2017 and April 15th, 2023.

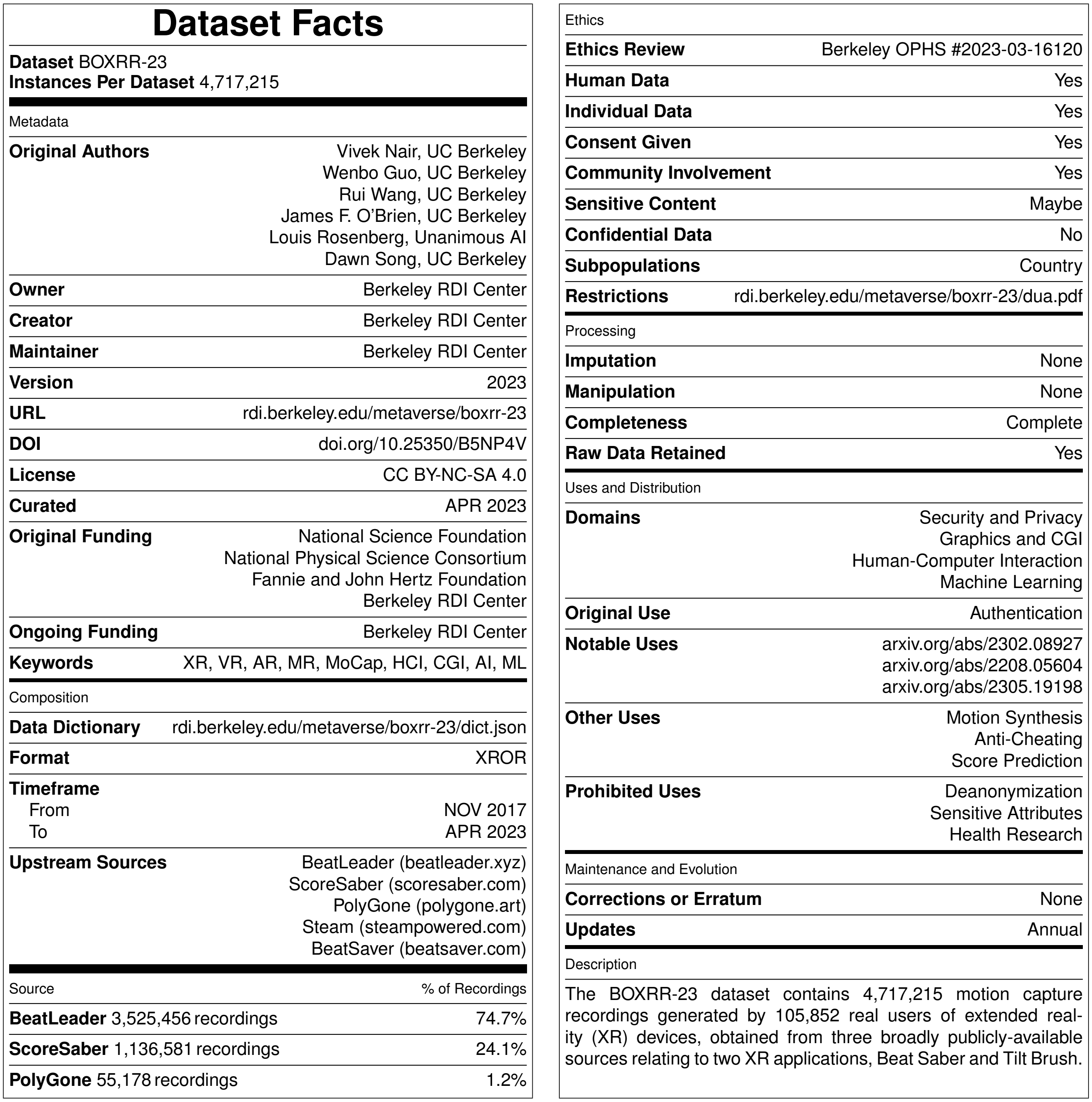
Table 3.1(A): Sources for data in BOXRR-23 dataset.

Source	**Application**	**Users**	**Recordings**	**Format**	**Size**
BeatLeader	Beat Saber	95,192	3,525,456	.bsor	6.25 TB
ScoreSaber	Beat Saber	55,331	1,136,581	.dat	1.44 TB
PolyGone	Tilt Brush	27,693	55,178	.tilt	1.87 TB

Table 3.1(B): Output characteristics of BOXRR-23 dataset.

Dataset	**Users**	**Recordings**	**Format**	**Size**
BOXRR-23 Dataset	105,852	4,717,215	.xror	4.71 TB

3.4 Dataset Nutrition Label



Dataset Facts

Dataset BOXRR-23
Instances Per Dataset 4,717,215

Metadata

Original Authors	Vivek Nair, UC Berkeley Wenbo Guo, UC Berkeley Rui Wang, UC Berkeley James F. O'Brien, UC Berkeley Louis Rosenberg, Unanimous AI Dawn Song, UC Berkeley
Owner	Berkeley RDI Center
Creator	Berkeley RDI Center
Maintainer	Berkeley RDI Center
Version	2023
URL	rdi.berkeley.edu/metaverse/boxrr-23
DOI	doi.org/10.25350/B5NP4V
License	CC BY-NC-SA 4.0
Curated	APR 2023
Original Funding	National Science Foundation National Physical Science Consortium Fannie and John Hertz Foundation Berkeley RDI Center
Ongoing Funding	Berkeley RDI Center
Keywords	XR, VR, AR, MR, MoCap, HCI, CGI, AI, ML

Composition

Data Dictionary	rdi.berkeley.edu/metaverse/boxrr-23/dict.json
Format	XROR
Timeframe	
From	NOV 2017
To	APR 2023
Upstream Sources	BeatLeader (beatleader.xyz) ScoreSaber (scoresaber.com) PolyGone (polygone.art) Steam (steampowered.com) BeatSaver (beatsaver.com)

Source	% of Recordings
BeatLeader 3,525,456 recordings	74.7%
ScoreSaber 1,136,581 recordings	24.1%
PolyGone 55,178 recordings	1.2%

Ethics

Ethics Review	Berkeley OPHS #2023-03-16120
Human Data	Yes
Individual Data	Yes
Consent Given	Yes
Community Involvement	Yes
Sensitive Content	Maybe
Confidential Data	No
Subpopulations	Country
Restrictions	rdi.berkeley.edu/metaverse/boxrr-23/dua.pdf

Processing

Imputation	None
Manipulation	None
Completeness	Complete
Raw Data Retained	Yes

Uses and Distribution

Domains	Security and Privacy Graphics and CGI Human-Computer Interaction Machine Learning
Original Use	Authentication
Notable Uses	arxiv.org/abs/2302.08927 arxiv.org/abs/2208.05604 arxiv.org/abs/2305.19198
Other Uses	Motion Synthesis Anti-Cheating Score Prediction
Prohibited Uses	Deanonymization Sensitive Attributes Health Research

Maintenance and Evolution

Corrections or Erratum	None
Updates	Annual

Description

The BOXRR-23 dataset contains 4,717,215 motion capture recordings generated by 105,852 real users of extended reality (XR) devices, obtained from three broadly publicly-available sources relating to two XR applications, Beat Saber and Tilt Brush.

Figure 3.4: Dataset label according to the dataset nutrition [113] standard.

3.5 Related Work

We searched for existing datasets relating to “motion capture,” “telemetry,” “VR motion,” “XR motion,” etc., on dataset hosting platforms like Kaggle, Zenodo, and Dryad, as well as for academic papers relating to motion capture data and experiments. We found over 25 existing datasets containing human motion recordings. The majority of these datasets come from conventional non-XR motion tracking systems, as listed in Table 3.2(A), while several originate from XR-based laboratory studies, listed in Table 3.2(B). The largest existing study contained 511 subjects [178], with a single session captured from each subject. By contrast, our dataset, summarized in Table 3.2(C), contains over 105,000 subjects and 4.7 million recordings from the three sources described in §3.3.

Table 3.2(A): Current motion capture datasets outside XR.

Dataset	**Organization**	**Year**	**Subjects**	**Recordings**	**Markers**
BMLrub [276]	Ruhr Univ. Bochum	2002	111	3,061	41, 3DoF
HDM05 [187]	Max Planck Society	2007	4	215	41, 3DoF
CMU-MMAC [272]	Carnegie Mellon Univ.	2008	5	5	41, 3DoF
EYES Japan [185]	EYES Japan	2009	12	750	37, 3DoF
HumanEva [255]	Univ. of Toronto	2010	3	28	39, 3DoF
SFU MoCap [249]	Simon Fraser Univ.	2012	7	44	53, 3DoF
ACCAD [2]	Ohio State Univ.	2012	20	252	82, 3DoF
Sleight of Hand [116]	Trinity College Dublin	2012	1	62	91, 3DoF
Human3.6m [120]	Romanian Academy	2013	11	44	24, 3DoF
MoSh [158]	Max Planck Society	2014	19	77	87, 3DoF
MPI Limits [6]	Max Planck Society	2015	3	35	53, 3DoF
KIT MoCap [165]	Karlsruhe Inst. of Tech.	2016	232	2,925	50, 3DoF
Total Capture [277]	Univ. of Surrey	2017	5	37	53, 3DoF
AMASS [163]	Max Planck Society	2019	344	11,265	37, 3DoF
CMU MoCap [43]	Carnegie Mellon Univ.	2019	144	2,605	41, 3DoF
MoVi [96]	Queen’s Univ.	2021	90	1,890	12, 3DoF

Table 3.2(B): Current motion capture datasets inside XR.

Dataset	**Organization**	**Year**	**Subjects**	**Recordings**	**Trackers**
Behavioural Biometrics [214]	Bundeswehr Univ. Munich	2019	22	88	3, 6DoF
TTI [178]	Stanford Univ.	2020	511	511	3, 6DoF
Body Normalization [154]	Univ. of Duisburg-Essen	2021	16	48	3, 6DoF
Obfuscation [186]	Univ. of Central Florida	2021	60	120	3, 6DoF
Body Sway [60]	Purdue Univ.	2021	28	336	3, 6DoF
You Can’t Hide [274]	Univ. of Padova	2022	35	69	3, 6DoF
Motion Matching [217]	Univ. of Catalonia	2022	1	12	3, 6DoF
Personal Identifiability [177]	Stanford Univ.	2023	232	1856	3, 6DoF
Who is Alyx [242]	Univ. of Würzburg	2023	71	142	3, 6DoF

Table 3.2(C): Our new XR motion capture dataset.

Dataset	**Organization**	**Year**	**Subjects**	**Recordings**	**Trackers**
BOXRR-23	Anonymized	2023	105,852	4,717,215	3, 6DoF

In addition to being over 200 times larger than the largest existing dataset, we found that all of the existing datasets come from a laboratory study in which participants used a small number of homogeneous devices and were generally physically present in a narrow geographical area. Thus, the BOXRR-23 dataset is more useful for obtaining a representative sample of XR users, as it originates from real XR users using their own devices in their own homes. As a result, it contains diverse data from over 40 types of XR devices, and includes users from over 50 countries around the world.

As evidenced by Table 3.2, BOXRR-23 is more comparable to existing XR datasets with a small number of 6DoF trackers than non-XR datasets with a large number of 3DoF markers. In applications where detailed full-body tracking is required, a conventional MoCap dataset may be more appropriate.

3.6 XROR Format

As detailed in §3.3, the data included in the BOXRR-23 dataset was scraped from three separate sources (BeatLeader, ScoreSaber, and PolyGone), each using three separate custom file formats designed specifically for those platforms (.BSOR, .DAT, and .TILT, respectively, summarized in Table 3.3(A)). We felt that the experience of future consumers of this dataset would be improved if the recordings were all converted to a single file format that could be analyzed and ingested via a unified pipeline.

We began by evaluating open-source motion capture file formats such as .BVA, .BVH, and .MVNX. Unfortunately, we found that the existing formats were unsuitable for this database for a variety of reasons. Some formats, such as .BVA and .BVH, only have support for motion data, and did not allow us to embed the rich metadata and event data streams we wished to include in the dataset. Others, like .MVNX, did support the inclusion of arbitrary metadata and event data streams, but used an inefficient underlying text-based file format (.XML) that would have caused the dataset to balloon to over 300 TB in size. Finally, some proprietary formats did contain all of the necessary features in an efficient binary format, but were not open-source and required paid tools or licenses to utilize them. Overall, we found that none of the existing open-source file formats were unsuitable for this dataset.

A formal specification of the XROR format, using the BSON version of the JSON Schema notation, is provided on our website.

To address the issues with existing open-source file formats, we introduce the new "Extened Reality Open Recording (XROR)" file format. XROR files contain metadata as well as rich event and motion data streams, and are based internally on BSON (Binary JSON), a flexible, widely-supported format with libraries in dozens of languages. Metadata is stored as JSON key-value pairs, while event data and motion data streams are converted to 2D floating-point arrays and compressed using fpzip, a lossless compressor of multidimensional floating-point arrays designed by Lawrence Livermore National Laboratory specifically for the efficient storage and transmission of scientific datasets.

To evaluate the relative efficiency of our new format, we converted a portion of our dataset into a variety of existing open formats, summarized in Table 3.3(B), as well as our proposed XROR format, as shown in Table 3.3(C). Even compared to the original source formats shown in Table 3.3(A), XROR achieves lossless space savings of at least 30%.

Table 3(A): Source file formats for motion data.

Format	**Metadata**	**Motion Data**	**Event Data**	**Compression**	**Avg. Size**
.tilt	✓	✓	✓		33.89 MB
.bsor	✓	✓	✓		1.77 MB
.dat	✓	✓	✓		1.27 MB

Table 3(B): Existing general file formats for motion data.

Format	**Metadata**	**Motion Data**	**Event Data**	**Compression**	**Avg. Size**
.mvnx	✓	✓	✓		61.90 MB
.bvh		✓			25.79 MB
.bva		✓			13.98 MB

Table 3(C): Proposed new open file format for motion data.

Format	**Metadata**	**Motion Data**	**Event Data**	**Compression**	**Avg. Size**
.xror	✓	✓	✓	✓	0.99 MB

Due to the advantages of our new XROR format over the existing alternatives, the entire BOXRR-23 dataset is offered exclusively as XROR files. To help researchers process this format, we have provided open-source tools to parse XROR files, and convert them to and from a variety of formats (e.g., TILT, BSOR, DAT, and JSON).

3.7 Recording Contents

Figures 3.5–3.8 illustrate the typical contents of each recording in the BOXRR-23 dataset. Specifically, the following data is included in each recording:

1. **Metadata**. A variety of metadata is included with each entry, including anonymized user IDs, hardware and software information, and virtual environment and activity details.

2. **Motion data**. Recordings consist of motion data captured in 6DoF at between 60 Hz and 144 Hz. Beat Saber recordings include head and hand motion data (see Fig. 3.5), while Tilt Brush recordings include brush motion and pressure data (see Fig. 3.6).

3. **Event data**. Motion data is accompanied by rich contextual information about events occurring in the virtual world. This includes information about the in-game objects and obstacles in the case of Beat Saber (see Fig. 3.7), and about each brush stroke in the case of Tilt Brush (see Fig. 3.8).

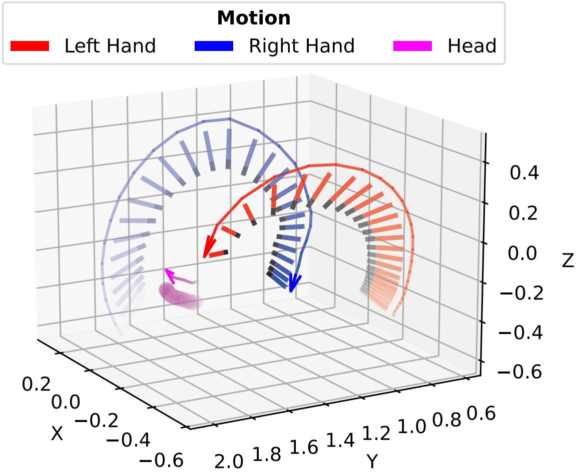


Figure 3.5: “Beat Saber” motion data.

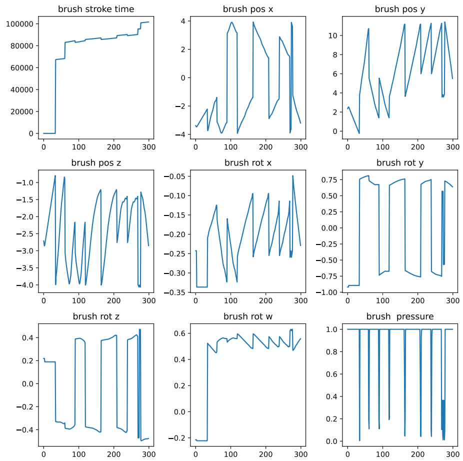


Figure 3.6: “Tilt Brush” motion data.

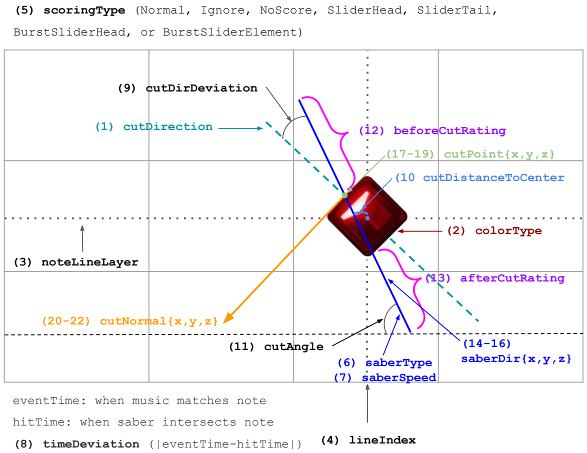


Figure 3.7: “Beat Saber” event data.

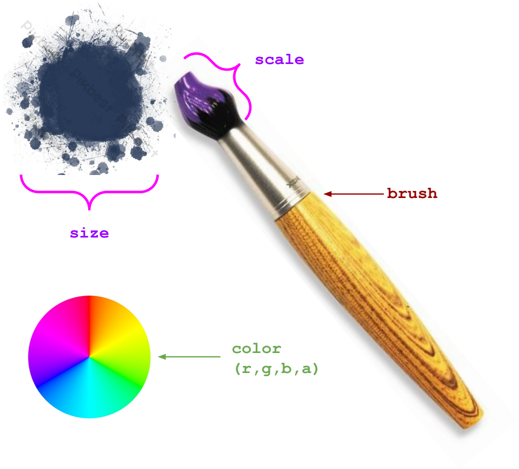


Figure 3.8: “Tilt Brush” event data.

3.8 Access Instructions

Researchers interested in using the BOXRR-23 dataset are invited to visit our website: `https://rdi.berkeley.edu/metaverse/boxrr-23`. The DOI is 10.25350/B5NP4V. For ease of access, the dataset has been split into 106 .zip files, each containing up to 1,000 users. Each user is represented by a folder containing .xror recordings from that user.

We developed the licensing terms for this dataset in conjunction with the IRB and IP office at our institution, with the chief goal of protecting the human subjects contained in this dataset. The dataset is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license, and is additionally subject to an ethical data use agreement (DUA) that prohibits unethical uses of the data, such as attempts to deanonymize the subjects. Access to the dataset is automatically granted upon agreeing to the CC BY-NC-SA 4.0 license and DUA.

3.9 Intended Use Cases

As discussed above, known uses of this dataset are primarily in the authentication and biometrics domain. Most of the research in this dissertation uses this dataset to advance knowledge of XR security and privacy. However, there are a number of interesting envisioned uses for this dataset within the VR community, beyond security and privacy research.

Future Directions

Outside the security and privacy domain, we can envision a number of additional interesting applications for this data. Historically, motion capture data has primarily been used for computer graphics, animation, and CGI, and our data could also be used in this domain. For example, it could be used to train large-scale generative machine learning models for natural human motion synthesis tasks. It may also be of interest to researchers studying human-computer interaction in XR. For example, researchers could use the data to investigate interaction patterns likely to cause discomfort or injury.

One area of active research that is relevant to our dataset is the inference of full-body pose information from sparse tracking inputs. Researchers have demonstrated the ability to recover full-body motion data from the motion of a few tracked points [128, 67]. Using these techniques, the sparse tracking data offered by our dataset could be used to recover inferred full-body motion for various uses.

Furthermore, the dataset contains numerous labels, including anonymized user IDs, hardware and software descriptions, and virtual environment and activity descriptions, that can be used to construct novel classification and regression tasks. For example, a very interesting use of the Tilt Brush portion of the dataset could be to use the brushstroke motion data to infer the title or description of the drawing, which are provided in the metadata.

Finally, this dataset presents a challenging and unique opportunity for theoretical machine learning research, because it consists of long, sequential data, with sequence lengths often in excess of 100,000. Most existing deep learning algorithms are not well equipped to handle sequential data of this size. Currently, our dataset is a rare instance of a task in which classical ML algorithms seem to outperform deep learning methods [196]. Developing models that can accurately and efficiently ingest the data contained in this dataset may require theoretical advances in machine learning techniques.

3.10 Population Survey

To shed additional light on the demographics of the users within our dataset, we conducted a large-scale online survey of VR users. The survey contained about 50 questions and received 1,006 responses, of which 830 users were present in the BOXRR-23 dataset.

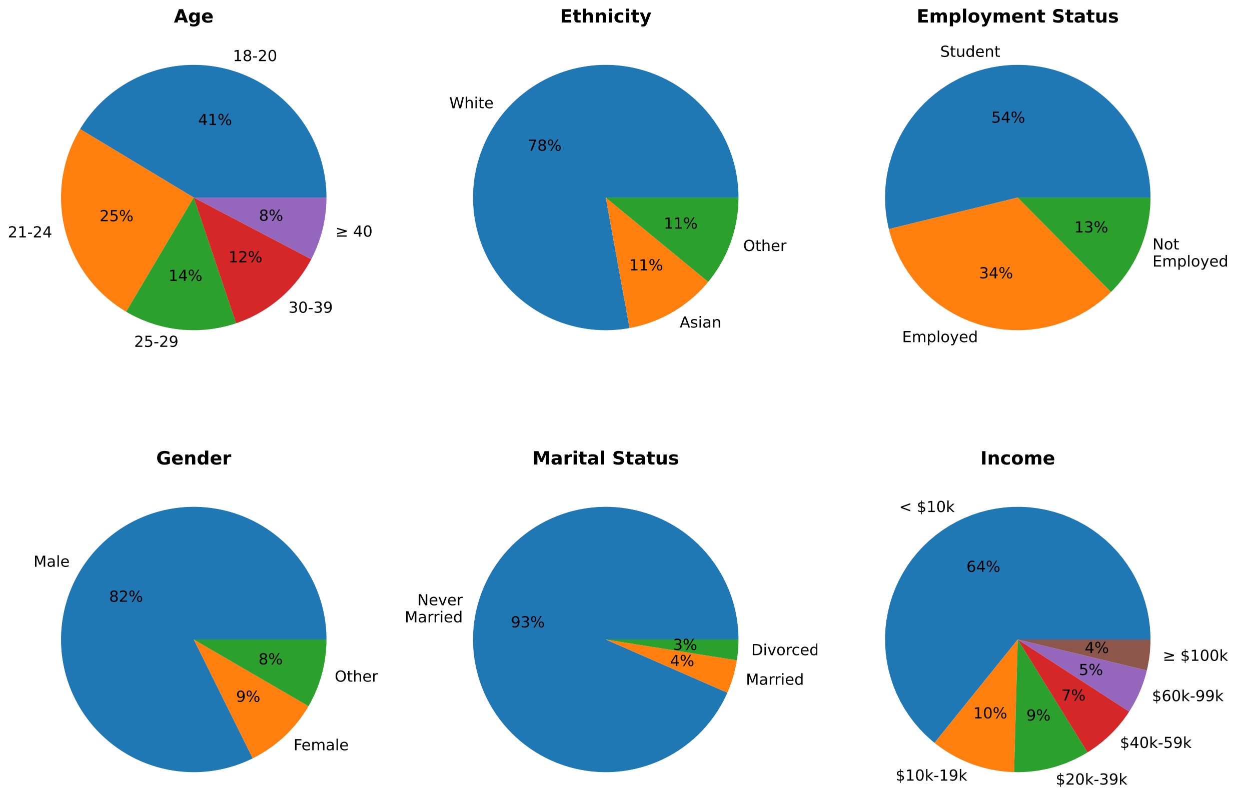


Figure 3.9: Survey results from 830 users in the BOXRR-23 dataset.

Our survey was conducted in coordination with BeatLeader and other Beat Saber organizations, and thus did not reach the 1% of BOXRR-23 users from Tilt Brush. The full results of this survey are available online [191], and are summarized in Figure 3.9 above.

3.11 Limitations

As may be evident by the survey results provided in §3.10, the users included in our dataset are not necessarily representative of a general population. For example, the dataset consists primarily of white and male subjects. While the subjects are demographically similar to the overall population of VR device users [230], they consist entirely of users who chose to upload a BeatSaber performance or TiltBrush drawing to a public platform. As such, we believe enthusiast or expert-level users are likely to be overrepresented in the dataset. However, for the same reason, the dataset likely contains far more geographic diversity than existing laboratory-based datasets. Furthermore, the data is derived from just two VR applications, Beat Saber and Tilt Brush, with almost 75% of the users and 99% of the recordings being from Beat Saber alone. Overall, researchers should be cautious when attempting to use this dataset to draw conclusions about larger populations than the ones directly included. When attempting to use BOXRR-23 to draw conclusions about broader populations, researchers are advised to follow known best practices for accounting for sampling bias in datasets [210, 143]. Additionally, there are some risks associated with the dataset being derived from ordinary XR users. Some metadata values, such as Beat Saber song titles or Tilt Brush drawing descriptions, may contain objectionable content due to their user-submitted nature. Metadata constituting user-configured settings like height and handedness should be considered self-reported, and are subject to the typical response biases associated with self-reported values. Finally, because the data is from "the wild" rather than a laboratory study, it originates from a wide variety of heterogeneous XR devices and physical environments, and may include more noise and tracking errors than a lab-created dataset.

3.12 Ethical Considerations

Because our dataset consists entirely of motion capture recordings from human subjects, significant attention was given to ethics throughout the process of designing and collecting the dataset. Our collection of this dataset for research purposes was approved by UC Berkeley's Committee for Protection of Human Subjects (CPHS), an OHRP-certified Institutional Review Board (IRB), as protocol #2023-03-16120.

We note that in producing this dataset, the authors had no direct contact with human subjects. Instead, our data is derived from three public sources. All data utilized in this study was already broadly, publicly available, to any person in the world with an internet connection, without the need for permissions, credentials, authentication, or any special tools or applications, via the websites of ScoreSaber, BeatLeader, and PolyGone. No new data is being made accessible to the public in the publication of this dataset; our contribution is in finding, scraping, aggregating, reprocessing, enriching, and distributing this existing data, and in surveying the underlying population.

Despite the public nature of the data and the IRB approval, we chose to obtain written permission from ScoreSaber, BeatLeader, and PolyGone before proceeding out of an abundance of caution and respect for the communities from which this data originates. We did not begin collecting data until authorized to do so by these communities, and sought their input throughout the collection process.

Users of the ScoreSaber, BeatLeader, and PolyGone platforms must voluntary install custom software to share their motion recording data with these platforms. They are fully aware of the nature of the data being shared, as uploading and publicly sharing XR data is the explicit purpose of these platforms. They also consent to their recordings being made publicly available in the privacy policies of these platforms. For example, the BeatLeader Privacy Policy, which can be found at `https://www.beatleader.xyz/privacy`, states that "Replays may contain personally identifiable information... Your data, including associated personally identifiable information, will be broadly publicly available to anyone with an internet connection via the BeatLeader website." Users of Google Poly (and PolyGone) consent to making their data publicly available under a CC-BY license.

Beyond consenting to the publication of their data in privacy policies and license agreements, we made further attempts to notify users of their involvement in academic research. Because users authenticate with these platforms via OAuth, their contact information is not known to the platforms, making direct consultation infeasible. However, we worked in collaboration with the BeatLeader team to inform users of their inclusion in academic research via their website and the official social media channels of the platform.

Although users knowingly consented to the public availability of their motion data, we took two additional steps to protect the privacy of data subjects. First, all known explicit identifiers, such as usernames and user IDs, have been removed from the dataset. No potentially sensitive information, such as protected health information, is included in the data or metadata. Second, the dataset is offered under a data use agreement (DUA) that prohibits researchers from attempting to deanonymize or contact the users, or to infer private attributes of the users that may be deemed sensitive. We voluntarily followed the strictest PII data handling standards and guidelines offered by our institution throughout the dataset collection process to preclude the accidental release of non-anonymized data.

Participants originally submitted their motion data to the ScoreSaber, BeatLeader, and PolyGone platforms for purposes other than academic research. Namely, they chose to make their data freely publicly available for reasons such as competitive e-sports or collaborative artwork; as such, users were not compensated for their original submissions, nor for their inclusion in the dataset. Moreover, any participant risks associated with the use of an extended reality device would have been realized by the users regardless of the later inclusion of the resultant motion recordings in this dataset. The scraping and redistribution of publicly available online data is a highly common and widely accepted practice within the computer security and machine learning communities [226, 81].

While it is impossible to entirely eliminate the risks associated with a new dataset, we believe the additional risk posed by our dataset is minimal in light of the fact that all of the included data was already public. On the other hand, the data has the potential to facilitate significant advances in fields like graphics, HCI, XR, AI/ML, and computer security and privacy. We have taken significant steps to mitigate the potential harms of this dataset while maximizing its utility for beneficial research. Overall, we believe this research constitutes a net benefit to the subjects whose data was included by shedding light on the implications of the motion capture data which they have already, independently chosen to publish. For instance, security and privacy research using this dataset benefits society by highlighting the magnitude of the VR privacy threat and motivating future work on countermeasures.

3.13 Conclusion

In this chapter, we have presented the BOXRR-23 dataset, a 4.7 TB collection of extended reality motion capture recordings from users around the world. Unlike existing motion capture datasets, BOXRR-23 is derived from recordings submitted by participants using their own XR devices, rather than a laboratory setup. As a result, it contains over 200 times more users, and over 400 times more recordings, than all known comparable datasets, while simultaneously being more diverse and ecologically valid. The two XR applications included in BOXRR-23, Beat Saber and Tilt Brush, provide highly complementary motion data. Beat Saber consists almost entirely of fast ballistic movements while Tilt Brush consists almost entirely of fine motor movements, each controlled by a separate part of the brain [86]. By combining these sources, BOXRR-23 provides a diverse collection of motion patterns.

In addition to identifying three new sources of motion data not previously widely known to academic researchers, we contributed a new XROR format to enable the efficient storage and transmission of this data. Our XROR format is approximately 30% more efficient than the three original data formats, without any loss in precision, while also being more versatile than most existing open-source formats. We also conducted a large survey of over 800 users present in the dataset to help researchers understand its demographic constituency.

In the next chapter, we will begin to explore this dataset by performing a large-scale study of motion-based identification in VR. For the first time, BOXRR-23 will allow the identifiability of human motion data to be directly compared with biometrics like fingerprints and facial recognition, revealing the stunning strength of VR motion biometrics.

Chapter 4

Unique Identification of Over 55,000 Virtual Reality Users from Head and Hand Motion Data

4.1 Introduction

Identification (i.e., deanonymization) is one of the most basic privacy threats relevant to VR motion data. While it has long been known that individuals exhibit distinct biomechanical motion patterns that can be used to identify them or infer their personal attributes [202, 142, 48, 146, 215, 123], the extent to which the subset of this information that is observable in VR can be used to uniquely identify users is less well understood.

Although prior research has been conducted on the personal identifiability of VR tracking data [214, 214, 154, 274, 193], existing works have utilized data from small lab studies with 16 to 511 participants. By contrast, the dataset described in Chapter 3 is not only more than 100 times larger than the largest prior result, but is also far more representative of a realistic use case. Gaming has thus far been the predominant driver of VR adoption, with 91 of the 100 most popular VR applications being games as of early 2023 [260]. In this chapter, we examine the extent to which spatial telemetry captured during VR gaming sessions can be used to uniquely identify an otherwise anonymous player. Using the dataset from Chapter 3 with a combination of context-aware featurization and hierarchical machine learning, we show that players can be identified out of a pool of over 50,000 candidates with 94.33% accuracy from 100 seconds of head and hand motion data.

Despite the difficulty of identification growing in proportion to the number of users, we achieve comparable identification accuracy to prior works. We show that while identifying users in smaller sets (≤ 511) can be accomplished just by learning static attributes like height, actual behavioral differences in movement patterns must be utilized to identify users within our substantially larger dataset. As such, this study is the first to truly demonstrate the extent to which motion can be an identifying feature in VR.

4.2 Method

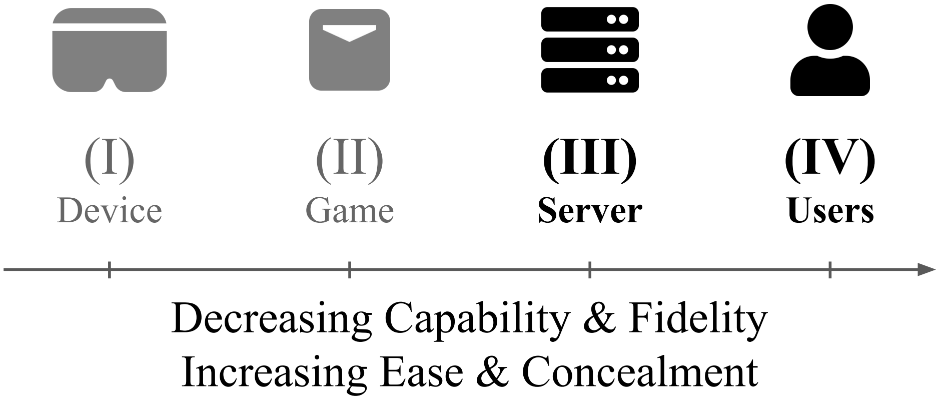

(I)
Device
(II)
Game
(III)
Server
(IV)
Users
Decreasing Capability & Fidelity
Increasing Ease & Concealment

Figure 4.1: Selected VR threat actors relevant to this work (see Chapter 2).

VR Adversaries

Referring back to the VR information flow and threat model of Chapter 2, recall that each entity in the VR information flow that can view the VR device telemetry of a target user is considered a potential adversary. Specifically, the attackers generally considered in VR privacy research are (I) VR hardware, (II) VR applications, (III) external servers, and (IV) external users. Each of these adversaries receives a view of the telemetry stream, which it could use to make adversarial inferences of private VR user information instead of (or in addition to) its intended purpose of facilitating application functionality. However, because the data can be reduced and compressed at each stage of the information flow, adversaries in higher tiers are considered "weaker" in this model.

Fig. 4.1 illustrates the information flow and threat actors discussed in 2. In this chapter, we are particularly interested in the game server (III) and other users (IV) as potential adversaries. These parties receive data processed by and filtered through the prior entities, meaning that attacks available to them can often be performed by other entities with even greater precision. They are also amongst the hardest attacks to detect due to their remote nature. This study exclusively analyzes data sent from a popular VR game to a remote server or other users, meaning that the attacks analyzed in this chapter represent the hardest and most pernicious realistic threats in VR.

VR Threat Scenarios

Why does motion-based identification in VR represent a compelling privacy threat? Consider a public figure who frequently uses a VR system with their corporate credentials to do professional work. In the evening, they log on with a different account for multiplayer VR gaming (where they might not behave in the most professional way), and later in the evening, they use a third account for adult VR experiences. Most individuals in this situation would reasonably prefer that the adversaries outlined above not be able to tie these accounts together. However, if a user can be uniquely identified by their VR motion patterns, any observer (or potentially even a group of colluding adversaries) could quickly link all of these accounts to them simply by observing their movement in each context.

On the web, "browser fingerprinting," which uses subtle differences between browsers to link people across web services, is highly analogous and is regarded as a significant privacy concern [76]. However, while one can replace their browser, they cannot easily change the distinct physiology and muscle memory that dictates their apparent movements, making motion identification a particularly challenging privacy threat.

Dataset

The BOXRR-23 dataset, described in detail in Chapter 3, is the primary source of data for this chapter. While the dataset contains data from BeatLeader, ScoreSaber, and PolyGone, only the BeatLeader portion of the data is used in this chapter.

This 3.96 TB subset of the dataset dataset consists of 2,669,886 replays from 55,541 users across 713,013 separate play sessions. The dataset has between 1 and 4,509 replays per user, with a median of 14. The replays range in length from 5 seconds to over an hour,[1] with a median length of 2 minutes and 56 seconds.

Because the data used in this chapter originates from a single, popular VR game (Beat Saber), we cannot yet demonstrate the ability to track users across applications, which we hope to see attempted in future work.

Ethical Considerations

Because our work involves data derived from human subjects, significant attention was given to ethics throughout the study. We note that no original data collection was performed by the authors; we used an existing dataset from an external source. All data utilized in this study was already broadly, publicly available, to any person in the world with an internet connection, without the need for permissions, credentials, authentication, or any special tools or applications. See §3.12 for more details on the ethics of the BOXRR dataset.

[1] Some maps are longer than a single song; e.g., an entire film soundtrack.

We submitted a detailed research proposal to UC Berkeley's IRB, in which we described precisely the BeatLeader telemetry data and its potentially sensitive nature, as well as our PII handling procedures, and our research goals. Since no original data was collected from human subjects, and BeatLeader data is already public, the protocol was deemed IRB-exempt under 45 C.F.R. § 46.104(d)(4)(i) and was issued a Notice of Approval.

Overall, we believe this research constitutes a net benefit to society by highlighting the magnitude of the VR privacy threat and motivating future work on defensive countermeasures. It further benefits the Beat Saber users whose data was utilized by highlighting the possible implications of the telemetry data which they had already made public, and also enabling the potential future development of anti-cheating tools.

Machine Learning

Classical ML. In Chapter 2, we describe a number of existing VR identification studies ranging from 16 to 511 participants in size. These existing VR privacy studies model user identification as a classification problem and leverage machine learning to classify users based on feature vectors of extracted data. Given that the existing studies process the telemetry data into a relatively small tabular dataset, these works usually leverage classical ML techniques (such as random forest [29] and gradient boosting [37]).

Underlying these models are decision trees, which construct a tree-based rule structure for a learning problem. A random forest ensembles multiple decision trees to improve the model's capacity, and thus is capable of handling more sophisticated learning problems. Gradient boosting takes this a step further by iteratively optimizing the set of trees rather than simply aggregating them. During the training process, gradient boosting actively updates the trees and their weights based on the current prediction results, allowing it to generally achieve a better performance than random forests alone [42]. We observe similar results in our study, with gradient boosting models providing by far the best performance.

Deep Learning. Interestingly, few of the existing studies have used deep learning for VR user identification, and their results are amongst the least accurate [154]. This is counterintuitive, as deep learning has become a mainstream technique in the machine learning community. In different domains, deep learning algorithms (e.g., Multi-layer Perceptrons) outperform traditional (e.g., tree-based) ML models in dealing with tabular data [99].

However, theses findings may not apply to our particular use case. This application has a very large number of users, which means that the classifier has to distinguish a large number of classes. It is challenging for deep learning models to train and converge under these conditions because they require a multi-class classifier to contain a large number of neurons in the output layer. In fact, most existing benchmark datasets where deep learning demonstrates a superior performance have a small number of classes. For example, the widely used image classification datasets MNIST [58] and CIFAR-10 [147] have ten classes, and some widely used text classification datasets only have 20 classes (Newsgroups [8]). The dataset with the most classes is ImageNet [56], which has 1,000 classes.

We found that deep learning empirically fails to perform well in our study, which requires more than 50,000 classes. Still, it is likely that larger and more sophisticated deep learning models could achieve strong performance in the future.

4.3 Featurization

As described above, we chose to use tabular classical ML models for this study rather than sequential deep learning models due to their empirically better performance for this dataset. However, to do so, we must first convert the streaming time-series motion data into a fixed tabular dataset that can be used by non-sequential models. In this section, we describe our method for converting the time-series replay telemetry data into a flat feature vector which can be consumed by a basic tree-based model. The featurization techniques described in this section are used in the identification models discussed later in this chapter.

We determined the best-performing model architecture and featurization method through a complex multi-parameter optimization in which we evaluated a variety of different featurization approaches together with a variety of classification model architectures and hyperparameters. In this process, more than 1,000 separate models were trained and tested using a validation set. However, we have chosen to use the single best-performing model architecture throughout this section to simplify the explanation of our feature selection.

Specifically, in this section, we use a 500-user identification model to validate our featurization choices and compare the resulting classification accuracy to the Miller et al. approach. For each proposed featurization approach, we randomly chose 500 users from our dataset and generated 150 training and 15 testing samples per user, using the train/test split discussed above. The features were then standardized using Z-score normalization before being used to train a 500-class LightGBM classification model. The identification accuracy on a per-sample and per-user basis is used to evaluate each approach.

We define a "session" as a continuously-recorded sequence of replays from a single user where no more than 10 minutes have elapsed between each replay. Our dataset contains an average of 13 such sessions per user. For each user, we reserve 70% of the sessions for training, 10% for validation, and 20% for testing, with a minimum of 1 session per set. As such, our models always perform true cross-session user identification rather than merely learning session-specific features, such as the exact position of a user within their room.

We begin with the best-performing existing method of featurizing VR telemetry data, which is that of Miller et al. [178], achieving 95% accuracy on 511 users. We describe this method in §4.3, and improve upon it in subsequent parts.

Guiding Principles

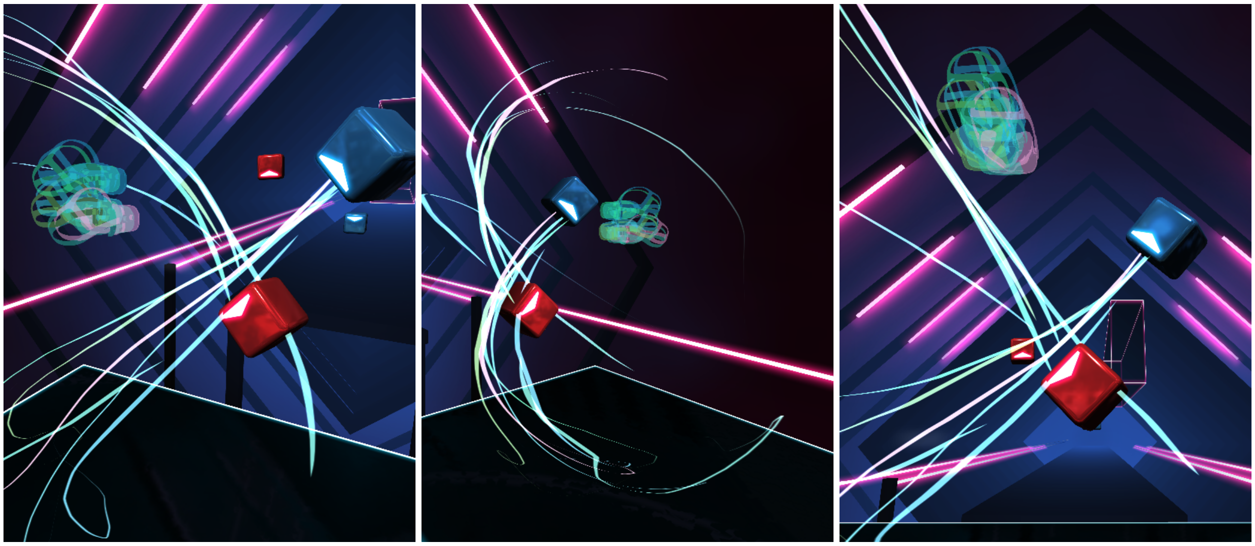


Figure 4.2: Five Beat Saber users hitting the same block pattern.

Fig. 4.2 shows, from several perspectives, the path taken by five Beat Saber users when slicing the same pair of blocks. As is clearly visible by the depictions, different users exhibit distinct motion responses even when presented with identical stimuli. These differences may be the result of physiology, learned motion patterns ("muscle memory"), random variance, or a combination thereof. The goal of the identification models presented in this chapter is to learn a set of motion characteristics that uniquely represent a user. Accordingly, the featurization techniques of this section aim to reduce the dimensionality of the telemetry stream to the extent possible while retaining the ability to differentiate between users.

Motion Features

Motion telemetry is the primary source of data for user identification and inference in VR. Fig. 4.3 shows a one-second segment of the head and hand motion of a Beat Saber user.

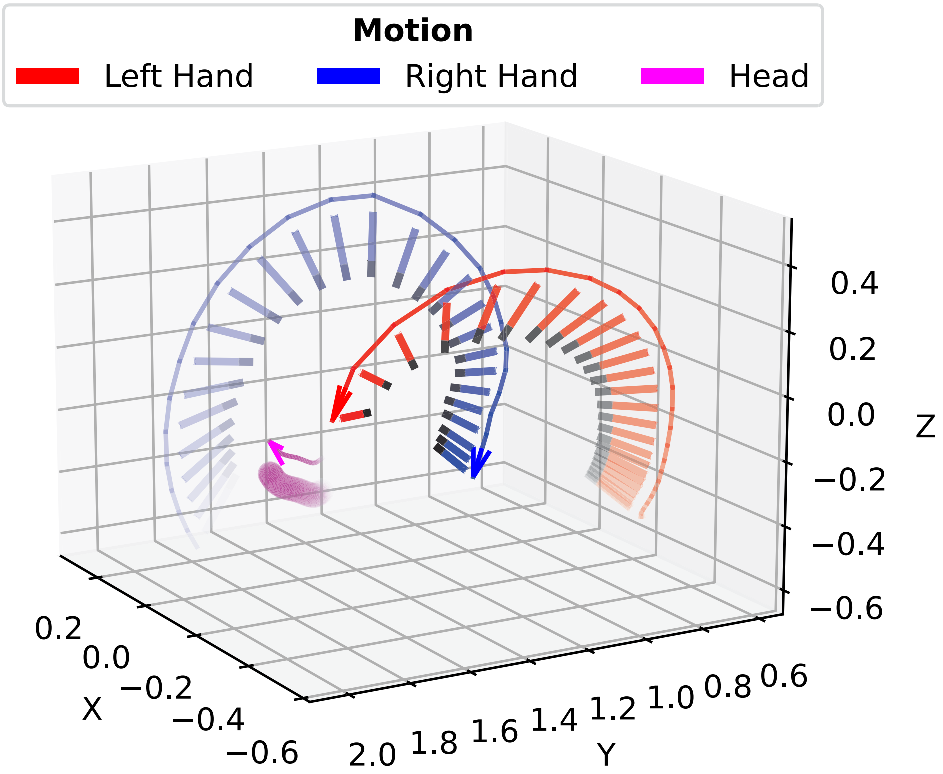


Figure 4.3: Head and hand motion from one second of telemetry.

As is visible in Fig. 4.3, each frame of telemetry data encodes 3D position and orientation coordinates across each of the three tracked objects. The Miller et al. method of motion data encoding suggests summarizing each of these 18 data streams using five summary statistics, namely the minimum, maximum, mean, median, and standard deviation, resulting in a 90-dimensional output vector. Using this approach with the Beat Saber data yields a 69.3% accurate per-sample identification and 93.4% accurate per-user identification using the evaluation method described above. This is comparable to the 95% accuracy reported by Miller et al. with their dataset.

In practice, we found that better performance is achieved by providing orientation measurements as four quaternion elements instead of three Euler angles. This modification alone resulted in an improved per-sample identification accuracy of 80.1% and per-user identification accuracy of 96.6%. Thus, our best-performing motion featurization can be represented as a 105-dimensional vector constructed as follows:

$$\begin{gathered}\{pos_x, pos_y, pos_z, rot_i, rot_j, rot_k, rot_1\}\\ \times\\ \{min, max, mean, med, stdev\}\\ \times\\ \{head, left\_hand, right\_hand\}\end{gathered}$$

Context Features

While motion alone may be sufficient to identify 500 users, additional information is needed when dealing with significantly larger datasets. In particular, models can benefit from knowing the activity-specific context in which a motion segment is captured such that different users can be compared directly when performing similar actions.

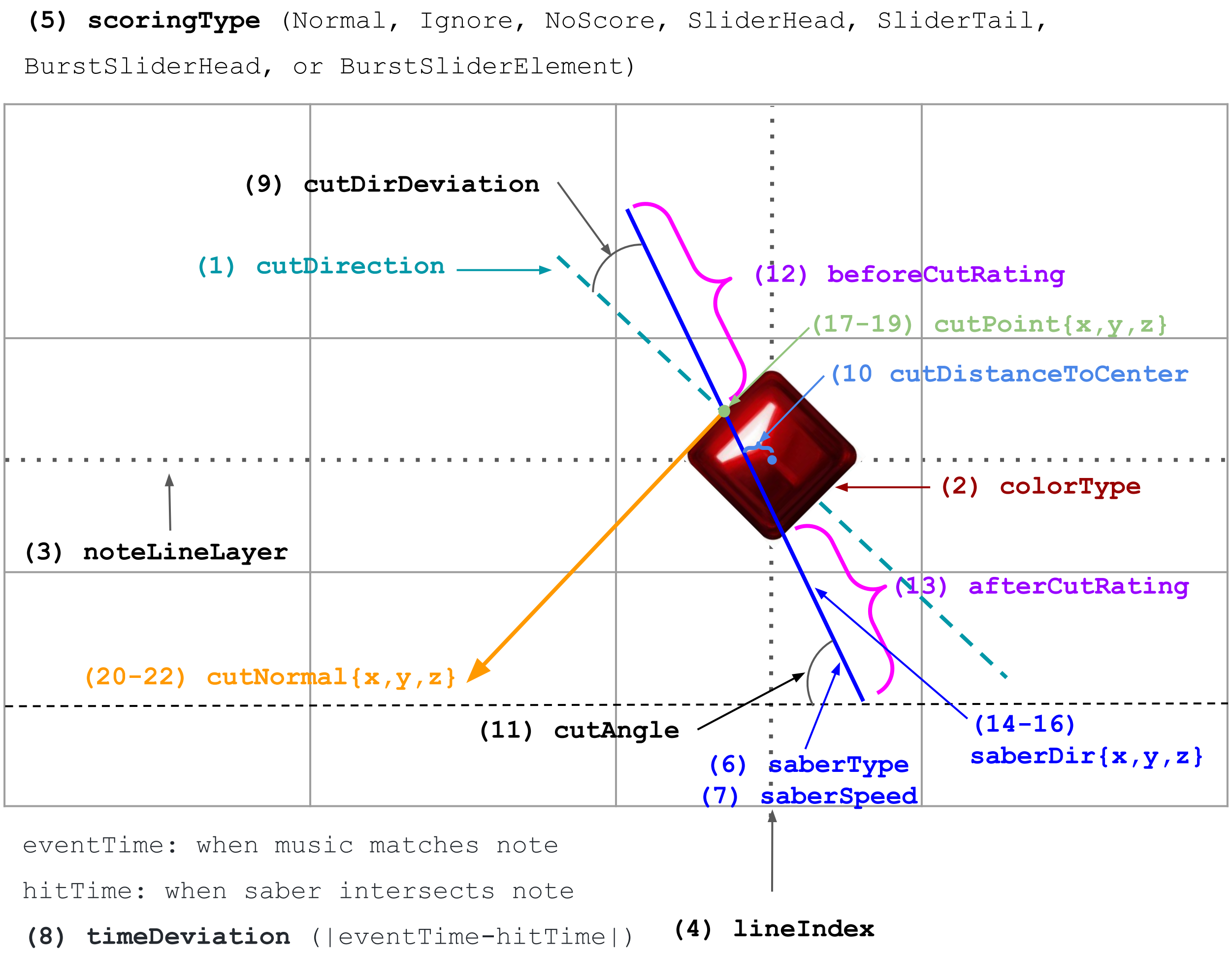


Figure 4.4: The 22 contextual features of a Beat Saber block.

In the case of Beat Saber, the activity chosen was the act of slicing an approaching block with a saber held in either hand. Specifically, we found 22 features that most accurately characterize movement relative to a single block, as shown in Fig. 4.4. These features include, for example, the position, orientation, type, and color of the block, the angle, speed, location, and accuracy of the cut, and the relative error of the cut in both space and time.

Although these 22 features provide a comprehensive yet succinct parameterization of a user's response to an individual block, they are insufficient to identify users without accompanying motion features. Using these features alone with the previously-established evaluation method yields just 14.8% accuracy per sample and 43.8% accuracy per user. While this is still highly statistically significant relative to the 0.2% accuracy one would achieve by attempting to identify one of the 500 users at random, it under-performs even the basic Miller et al. approach. Still, it demonstrates the potential to aid identification when combined with motion features.

Hybrid Featurization

Finally, we describe the inclusion of both motion and context features within a single feature vector, thus allowing models to interpret motion data specifically in relation to other users performing the same or similar actions. By combining the 22 context features of §4.3 with the 105 motion features of §4.3 corresponding to one second of motion centered on the moment of contact, a 127-dimensional hybrid feature vector can be produced. Using this feature set with our established evaluation approach yields 83.8% accurate per-note user identification, with 98.2% accurate identification per user.

While this hybrid feature set now outperforms either the motion or the contextual features alone, some useful information is still excluded. In particular, it is useful to explicitly separate the motion features from before and after a target event. For example, different information can be learned from a user's "in swing" and "out swing" relative to a block.

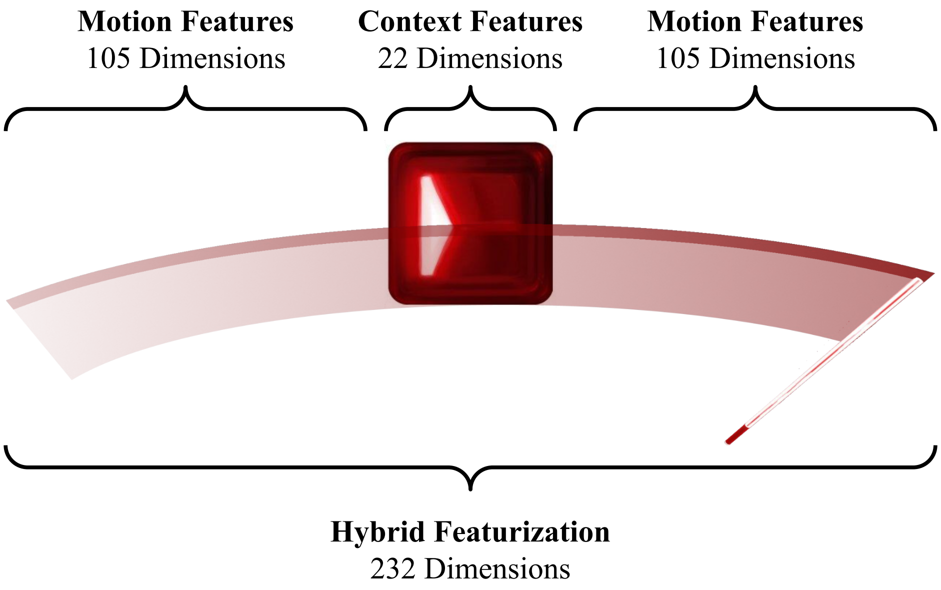


Figure 4.5: Hybrid featurization of a Beat Saber block.

Fig. 4.5 shows a full hybrid featurization of a Beat Saber block, including 22 contextual features for the block and 105 motion features corresponding to the one-second intervals before and after the block, totalling 232 dimensions. When evaluating this featurization with the same machine learning approach as before, 93.2% accurate identification is achieved per sample, with perfect (100.0% accurate) per-user identification of 500 users. The results of all approaches discussed in this section are summarized in Table 4.1.

Featurization Approach	**Features (#)**	**Accuracy (Per Sample)**	**Accuracy (Per User)**
Motion (Euler Angles)	90	69.3%	93.4%
Motion (Quaternion)	105	80.1%	96.6%
Contextual	22	14.8%	43.8%
Light Hybrid	127	83.8%	98.2%
Full Hybrid	232	93.2%	100.0%

Table 4.1: Accuracy of identifying 500 users using LightGBM with each of the discussed featurization methods.

In summary, the combination of rich contextual information about an event with separate features summarizing motion before and after said event is effective at achieving accurate identification for datasets significantly larger than 500 users. This is in part because the motion segments can be understood in the context of the corresponding stimuli, and in part because it begins to simulate a small sequential model; that is, it allows the model to ascertain which motion features are consistent and which change across two consecutive time slices. As such, we use this 232-dimension hybrid featurization method in all subsequent models for the remainder of this chapter.

4.4 Model Architecture

Having established the above featurization technique, we next describe our selected machine learning model architecture for identifying users. This remains a non-trivial problem in practice, as it requires a 50,000-class classification model, a use case that many existing machine learning algorithms are not designed to handle (see §4.2). Therefore, after selecting a performant algorithm and preprocessing method, we describe a hierarchical approach for constructing the overall classification model out of several smaller classifiers.

Algorithm Selection

Using the best-performing feature set from §4.3, we tried to construct an identification model using 6 popular classical machine learning classification algorithms with the same sample of 500 users. For each algorithm, we began by using the default hyperparameters and then ran up to 25 rounds of tuning to obtain the below results, which show the best per-sample identification performance achieved by each algorithm.

- LightGBM: **93.2%**
- XGBoost: 80.0%
- Logistic Regression: 72.2%
- Support Vector Machines: 67.13%
- Extreme Random Trees: 35.5%
- Random Forest: 32.1%
- Naive Bayes: **1.2%**

As discussed in §4.2, gradient boosting models are known to outperform other tree-based classification algorithms on tabular datasets, which matches our observations above. In particular, LightGBM [137], an industry-leading gradient boosting framework, exhibited by far the best performance. We also tried multiple sequential and non-sequential deep learning approaches with limited success. As summarized below, the deep learning attempts far underperformed the classification accuracy of the best classical ML algorithm.

- GRU: **84.0%**
- LSTM: 83.0%
- MLP: **72.0%**

Overall, we conclude that simple deep learning algorithms empirically failed to perform as well as LightGBM for the large multi-class classification task at hand. Moving forward, we use LightGBM for our identification models in view of the performance results.

Preprocessing Method

Using the hybrid featurization and LightGBM model with optimized hyperparameters, we evaluated five potential preprocessing methods, the results of which are shown below.

- StandardScaler: **93.2%**
- MinMaxScaler: 89.8%
- MaxAbsScaler: 86.4%
- SparseNormalizer: 83.5%
- TruncatedSVD: **66.5%**

The preprocessing approach with best results is standard scaling (Z-score normalization), whereby each feature is transformed by removing the mean and scaling to unit variance.

Hierarchical Approach

For smaller datasets, the above methods would be adequate. Indeed, if up to 5,500 classes are present, a single LightGBM classification model, deployed with our described featurization and preprocessing method, demonstrates strong performance in identifying users. Unfortunately, training a single LightGBM model with 50,000 classes would be infeasible with our dataset. We found that the training time and memory consumption of training a LightGBM classifier scales quadratically with the number of classes, as shown in Fig. 4.6.

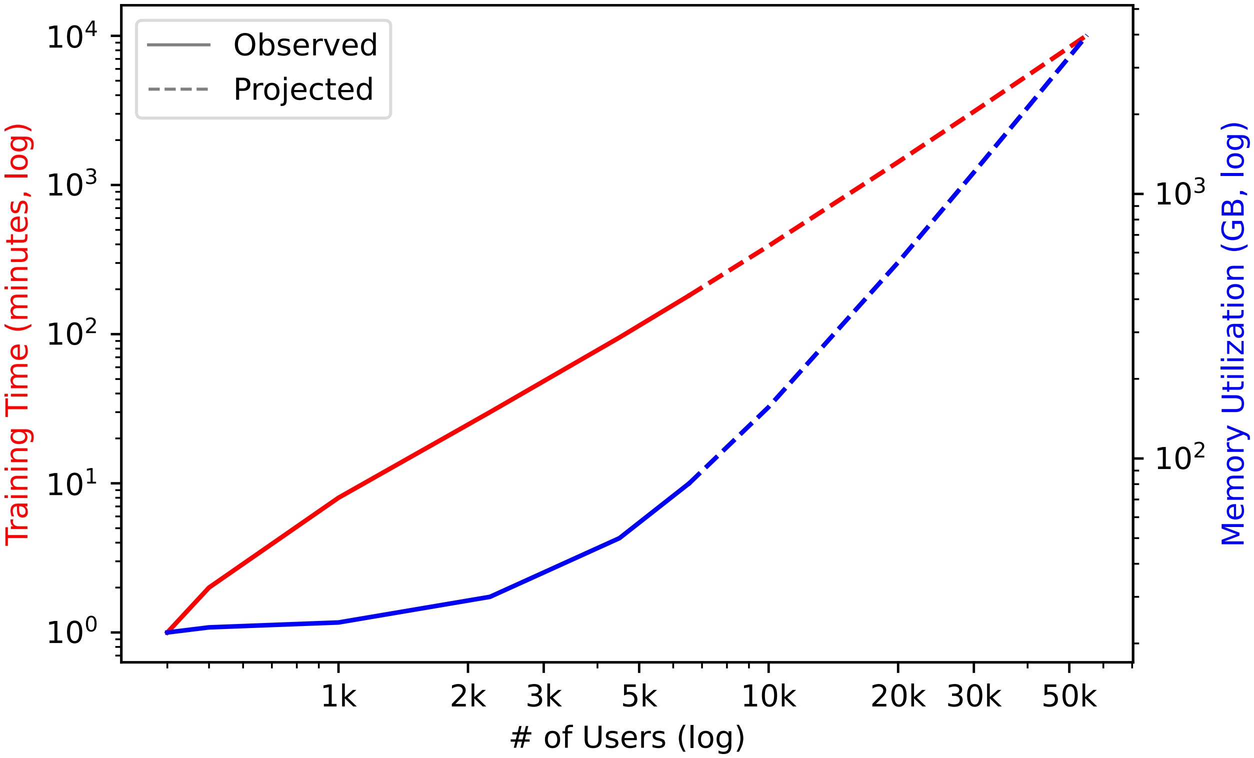


Figure 4.6: Observed and projected time and memory required to train an increasingly large LightGBM classifier.

According to a polynomial projection of our attempts to train classifiers with as many as 5,000 users, training a single classifier with all 55,000 users would take over 7 days and consume nearly 4 TB of RAM. While still within the realm of possibility when using server-grade hardware, the prospect of even larger datasets over the horizon motivates us to find a more efficient and scalable architecture.

We ultimately chose to construct a multi-layer hierarchical classifier. Our overall identification model is composed of three layers of smaller classifiers, each of which are only trained on a small set of available classes.

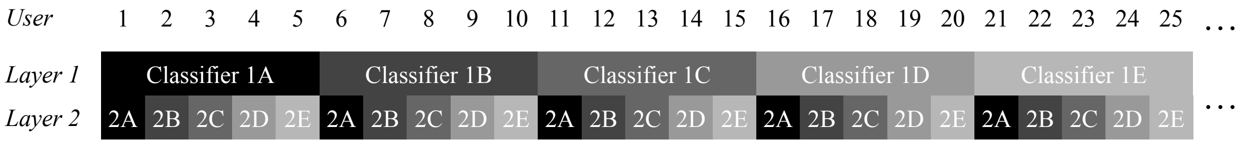


Figure 4.7: Hierarchical structure with 5 models per layer.

Fig. 4.7 illustrates the principle method by which the first two classification layers are constructed. In the first layer, N classifiers are each trained on 1/N of the available classes. In practice, we train 10 classification models with about 5,000 users each. This single layer already provides better performance than one may expect. Although each of the models will output a classification when identifying a user, regardless of whether that user is actually contained within their training set, the classification probability is usually highest in the model actually containing the target user.

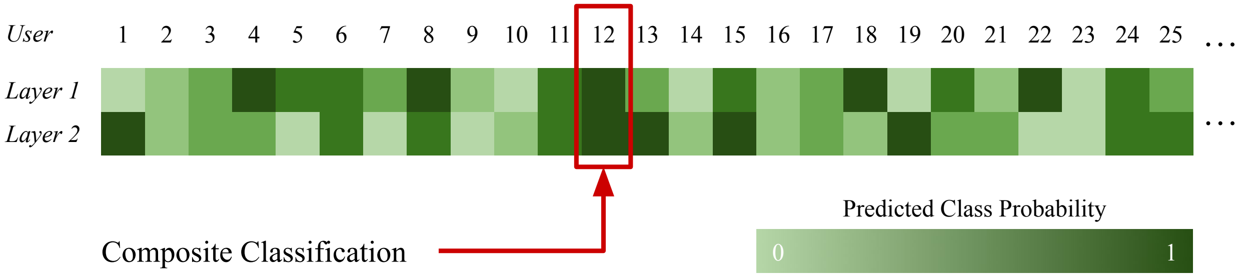


Figure 4.8: Class probabilities output by hierarchical classifier.

Further accuracy can be obtained by adding a second layer, also containing N classifiers each trained on 1/N of the available classes, with an even class redistribution from the first layer. Now, when querying each layer to identify a user, the layers are likely to agree on the correct user while disagreeing about false classifications (see Fig. 4.8). The overall classification can now be obtained by taking the highest logarithmic sum of the class probabilities output by both layers.

Adding more layers at this stage via random redistribution provides diminishing returns. Instead, a separate clustering set (independent of the train, validate, and test sets) can now be used to cluster users based on their class confusion using the existing two layers. The method for doing so using connected components in a graph is illustrated in Fig. 4.9.

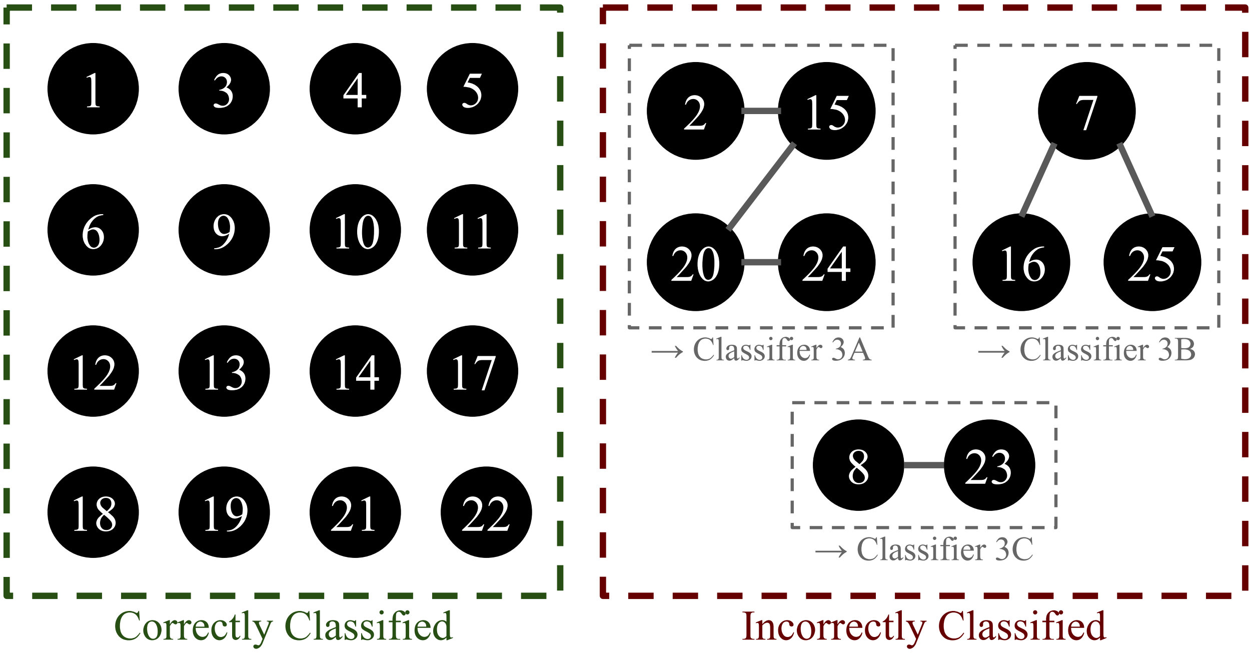


Figure 4.9: Graph-based method of selecting layer 3 groups.

As illustrated in Fig. 4.9, an undirected graph is constructed with a node for each user. Every time a user is incorrectly classified using the clustering set, an edge is added between the user and up to five apparently similar users. The connected components of this graph now represent sets of users who are likely to be misidentified as each other. In a third layer, one additional model can be trained for each component C in the graph (where $|C| > 1$), containing the users of C. When ultimately identifying a user, the logarithmic sum of the first two layers is used to obtain an initial identity. If the resulting user is present in one of the connected components, the corresponding model in the third layer is used to produce the final classification. Otherwise, the initial classification is directly returned as the predicted identity. Given limited computational resources, this approach increases the odds that similar classes are directly compared in at least one model.

Scalability

While motivated by the infeasibility of training a single multiclass classification model of insufficient size, the proposed hierarchical architecture also presents important scalability and practicality improvements over a monolithic approach. Each model in a layer can be trained in parallel, allowing for a 10-20x reduction in training time when using a cluster. Testing and inference can similarly be parallelized by evaluating each model separately.

Finally, the cost of adding a new user is significantly reduced by the hierarchical approach. When a new user is added, only one model on each layer must be retrained, rather than retraining the entire classifier. Given that most platforms where such an identification model may practically be deployed are constantly receiving new users, this alone constitutes a major improvement in the practicality of deployment.

Methodological Novelty

The primary contribution of this chapter is in presenting a VR identification result that is more than 100x larger than the next largest study in this field. Nevertheless, the unique challenges of this dataset have led us to make advances in the techniques used for identification. For instance, the hybrid featurization of §4.3 offers a significant performance advantage over the motion featurization of Miller et al., while our hierarchical model architecture in §4.4 provides a necessary improvement in scalability. To the best of our knowledge, neither of these techniques have been disclosed in prior work. We later obtained the Miller dataset (N=511), and found that these techniques improved their identification accuracy from 95.0% to 99.8%, demonstrating the significant practical improvement offered by our methods.

4.5 Hyperparameters

In this chapter, we use the following hyperparameters when training LightGBM models:

- objective='multiclass'
- boosting_type='goss'
- colsample_bytree=0.6933333333333332
- learning_rate=0.1
- max_bin=63
- max_depth=-1
- min_child_weight=7
- min_data_in_leaf=20
- min_split_gain=0.9473684210526315
- n_estimators=200
- num_leaves=33
- reg_alpha=0.7894736842105263
- reg_lambda=0.894736842105263

4.6 Evaluation

We evaluated our identification technique using a distributed machine learning cluster of 10 nodes, each with 16 vCPU cores and 128 GB of RAM. The replays of each user were separated into 4 or more distinct sessions, which were reserved for training, clustering, validation, and testing at a ratio of 70-10-10-10. For each user, 150 samples were generated from the training set using the full hybrid featurization method of §4.3. The features of all users were then z-score normalized, and used to train the hierarchical model described in §4.4.

The training process was completed in about 3 hours each for the first and second layers and about 6 hours for the third layer. The final testing process, which required over 90 million classifications to be made, took about 8 hours; an individual user identification requires less than a second.

Results

Layer	**# of Models**	**Accuracy (per Model)**	**Accuracy (per Layer)**
Layer 1	10	93.1%	90.2%
Layer 2	10	93.1%	90.2%
Layers 1 & 2	20	93.1%	91.0%
Layer 3	5	84.0%	84.0%
Layers 1, 2, & 3	25	91.3%	94.3%

Table 4.2: Accuracy of each hierarchical model layer per model (i.e., 5.5k users) and per layer (i.e., 55k users).

Table 4.2 shows the identification accuracy of each layer in the hierarchical model when evaluated using 50 test samples (100 seconds) per user. An identification accuracy of 90.1% can be achieved using a single layer, with the hierarchical architecture boosting the overall accuracy to 94.3%.

Of course, the accuracy of identification is highly dependent on the number of samples (and thus seconds of data) used to identify a user. Fig. 4.10 illustrates the identification accuracy in relation to the number of seconds used.

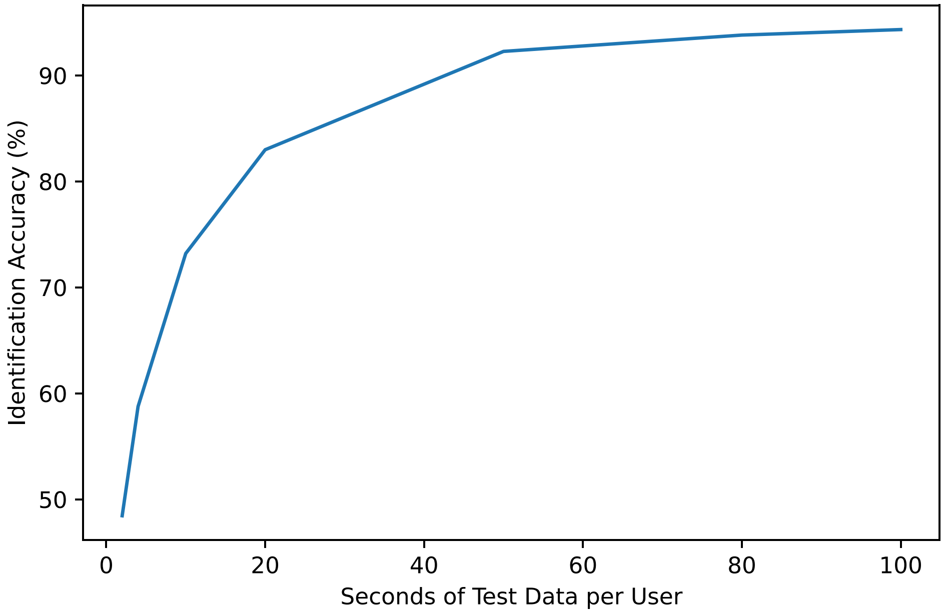


Figure 4.10: Impact of test sample size on accuracy.

Even with a single sample generated from just 2 seconds of telemetry data, the correct user out of 50,000 is identified about 48.45% of the time. Using 5 samples (10 seconds) of data increases this accuracy to 73.20%, which implies that only a short period of motion information is actually needed to uniquely characterize a user. A single minute of data yields 92.78% identification accuracy, and the full 94.33% accuracy is achieved when 50 samples (100 seconds) of data are used, with rapidly diminishing returns for each sample thereafter.

In some applications, it may be sufficient to output a small number of candidate identities rather than exactly identifying a user. In our evaluation, the correct user is amongst the top 3 candidates identified by the model in 97.25% of all instances.

Open-World Setting

Thus far, we have evaluated our models under the closed-world assumption, in which we are only concerned with classifying users that have already been seen in the training phase. However, in any realistic deployment, models will often be faced with users that have not previously been encountered. In the open-world setting, models should be able to detect the unseen classes rather than incorrectly identifying them as a previously-seen user. Ideally, the model can then be updated over time to incorporate the new users into the system.

Thankfully, it is well known that statistical techniques can be used to detect instances of concept drift in classification models. For example, Transcend [131] uses a statistical comparison of samples to identify concept drift in malware classification models. Using a similar principle, our hierarchical classification approach is already well suited to detect and reject users not previously seen during training.

To understand the performance of our models in an open-world setting, we performed a second evaluation using 10% of the existing users (5,554) and an equal number of new BeatLeader users not previously seen in training. Each of these users was classified using the first two layers of the hierarchical model. Fig. 4.11 shows the output confidence of both layers for new and existing users.

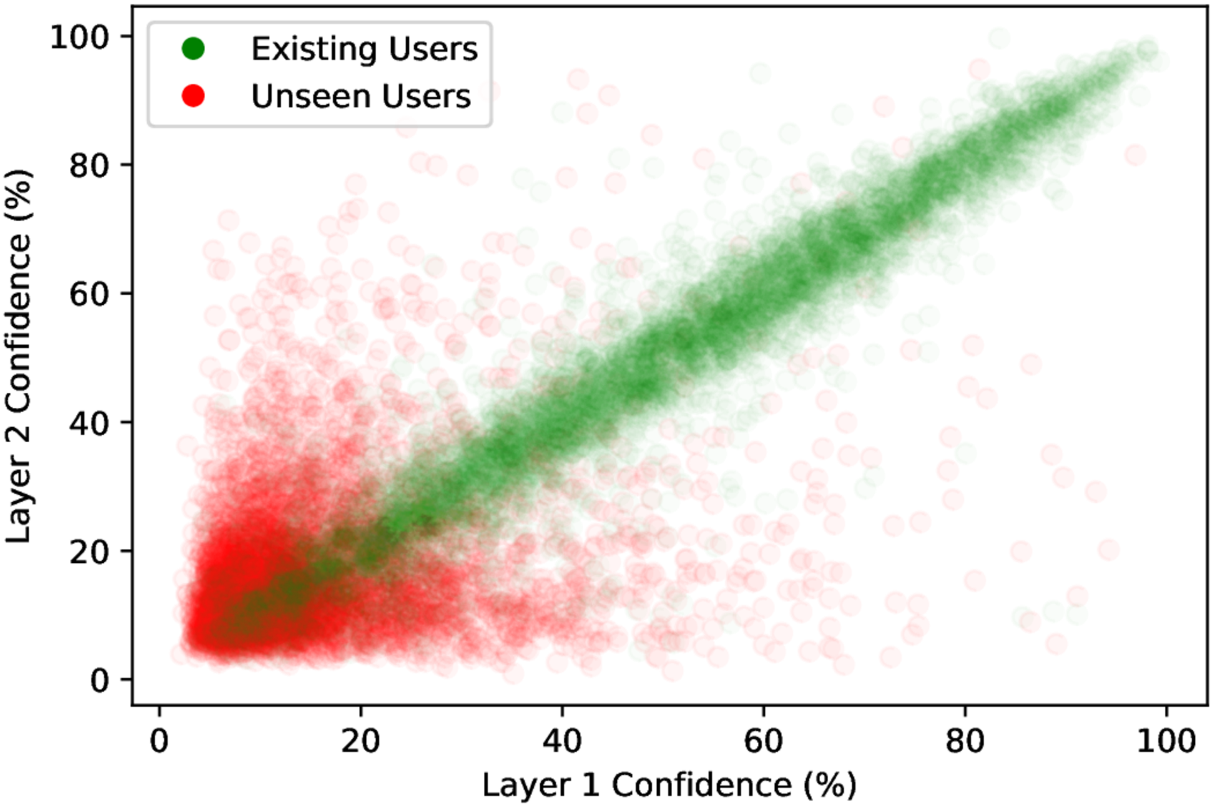


Figure 4.11: Correlation of layer 1 and layer 2 confidence values for existing and unseen users in the open-world setting.

As illustrated in Fig. 4.11, users present in the training set demonstrate a high correlation between the confidence of both layers, while previously unseen users show less correlation and have significantly lower confidence overall. Thus, a simple logistic regression model can be trained to determine whether a given user was previously seen. We chose to allocate 90% of the 5,554 new and 5,554 existing users to training, with the remaining 10% for testing. Thus, we trained the model using 4,999 existing users and 4,999 new users, and subsequently tested it using 555 existing users and 555 new users, the results of which are shown in Tab. 4.3. For each user, the inputs consisted of the max, argmax, and standard deviation of classification confidence values from each layer.

	Existing Users	**Unseen Users**
Classified as Existing	518 (93.3%)	45 (8.1%)
Classified as Unseen	37 (6.7%)	510 (91.9%)

Table 4.3: Binary classification of seen versus unseen users in the open-world setting using a simple logistic regression model on layer confidence values.

Overall, the logistic regression model was 92.6% effective at determining whether a given user had previously been seen in the training phase. This result should be interpreted in light of the fact that the accuracy of identifying and rejecting new users cannot reasonably be expected to out-perform the overall 94.3% accuracy of user identification. Thus, our approach could reasonably be deployed in the open-world setting.

4.7 Impact Factors

As explained in Chapter 3, our dataset contains labeled metadata for a number of user attributes, including device information and some basic demographics. While we avoided using this data in our identification model in order to achieve purely motion-based identification, we later used all of this information to perform a key factor analysis so as to better understand which attributes affect the identifiability of a user. The 15 most important factors are summarized in Fig. 4.12. This summary evaluates the impact of each factor on the accuracy of layers 1 and 2, as not all users are present in layer 3.

Fig. 4.12 reveals some interesting trends with respect to the factors which most impacted identification accuracy. Some devices, such as Windows Mixed Reality, are less conducive to identification, perhaps due the device's overreliance on low-quality dead reckoning for tracking. Others, like Valve Index, yield better than average user identification, which may be due to its highly precise outside-in tracking system.

Users from certain countries, particularly Japan and South Korea, are significantly easier to identify, implying there may be detectable cultural differences in play style. This result is statistically significant, with over 99% identification accuracy for users from those countries.

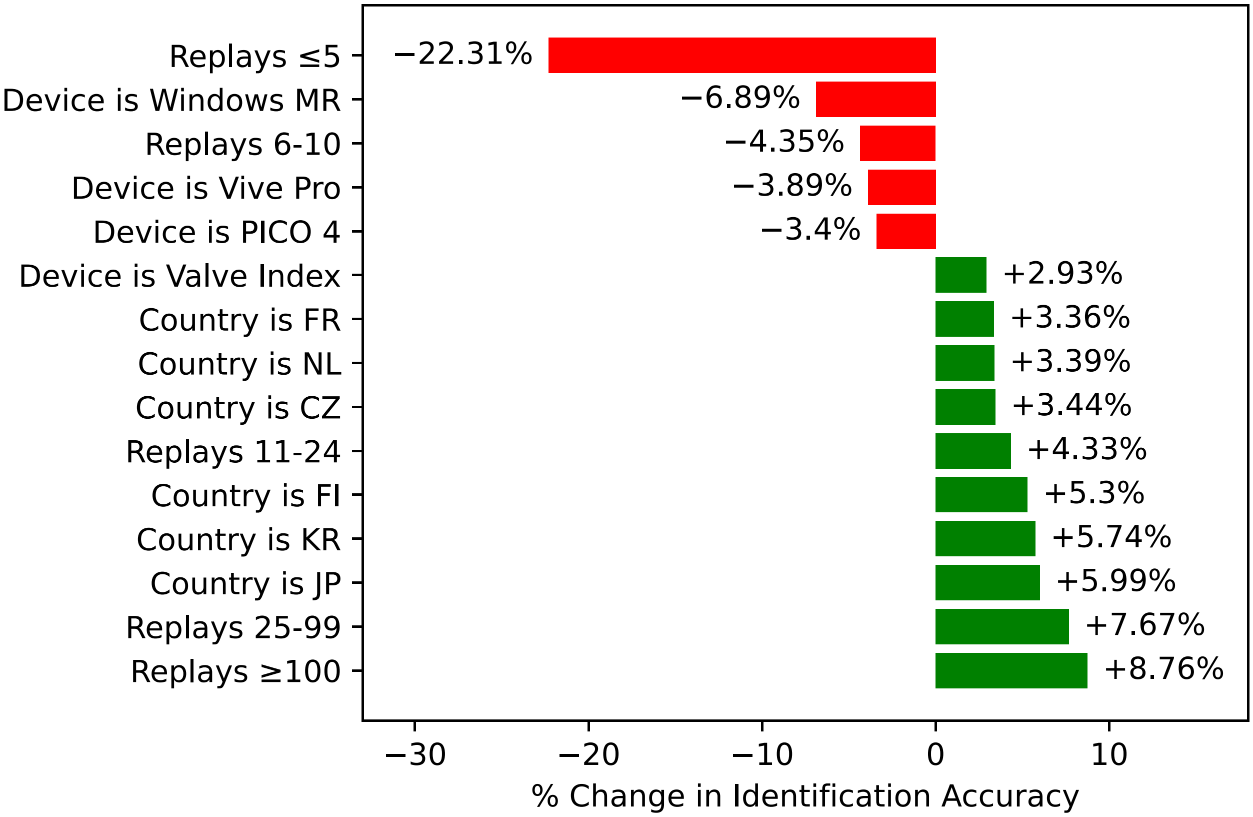


Figure 4.12: Impact of key factors on identification accuracy.

However, by far the most important factor in determining identification accuracy is the number of total replays observed from a target user, regardless of how many samples were actually used to train the model. Users with 5 or less total replays submitted were significantly harder to identify, while the 5,000 or so users with 100 or more replays could be identified with over 99.5% accuracy. The identification accuracy for users is charted against the number of replays in Fig. 4.13.

The clear trend of users with more replays (and thus more time spent in the game) being more easily identifiable is indicative of something other than more data being available, as the full 150 training features can easily be extracted from a single 5-minute session. Rather, it suggests that users with more experience are likely to develop a distinct play style (and reinforce the corresponding muscle memory) over time. Highly experienced players are thus more likely than novices to exhibit a repeatable response to the same stimulus, with veteran users becoming so consistent in their movements that they can be identified with near-perfect accuracy. This finding is a key driver of further improvements in motion identification in later chapters of this dissertation.

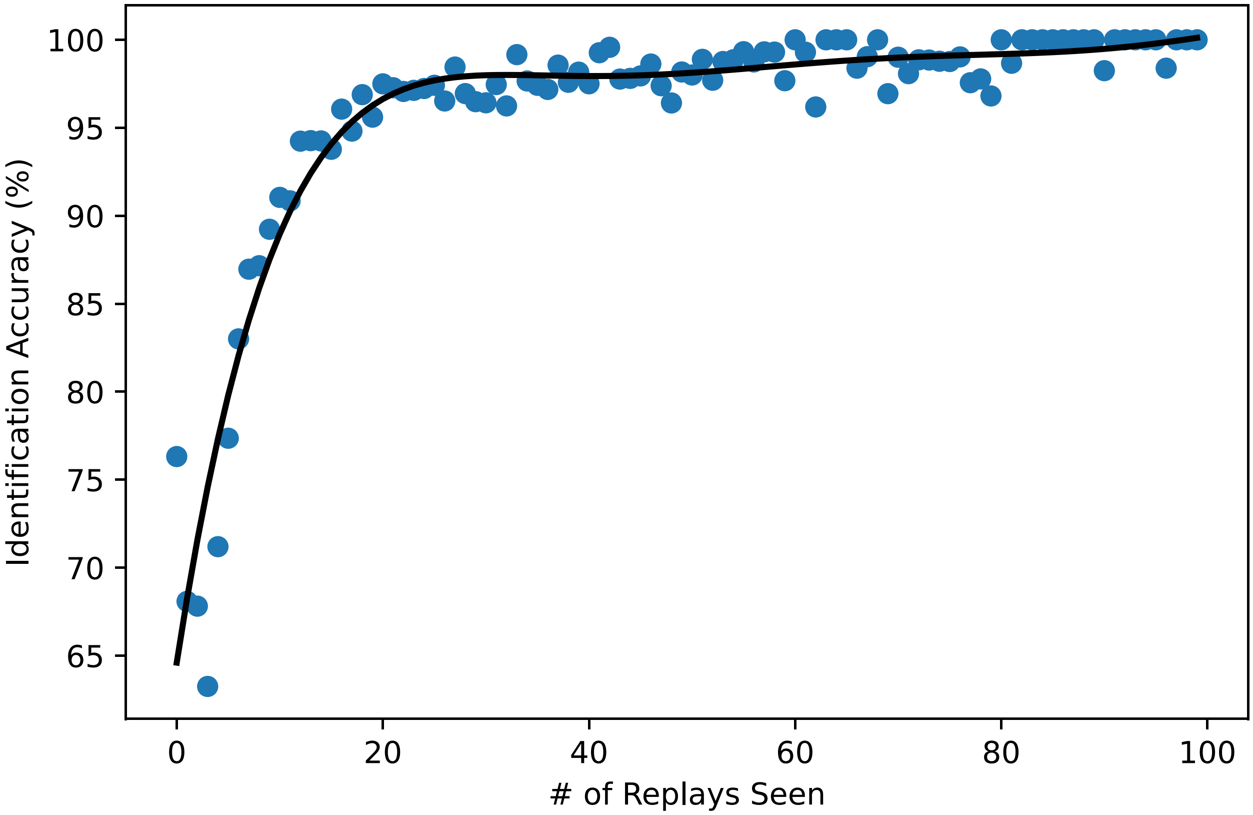


Figure 4.13: Replays per user vs. identification accuracy.

4.8 Explanations

An additional benefit of using a LightGBM model is the relative ease of explaining the importance of each feature. Fig. 4.14 shows the percentage of splits attributable to each of the 10 most important features (out of 232) in our final model.

As illustrated in Fig. 4.14, many of the most important features for identification correspond to obvious physical measurements. For example, the two most important features, which measure the maximum Y-position of the headset before and after the cut, are an obvious proxy for the user's height (and posture). Similarly, the next six most important features seemingly measure the length of the user's arms when furthest outstretched. These first eight features alone account for 6.8% of the splits and 10.2% of the gain of the identification model, providing about 12 bits of real entropy – enough information to accurately identify as many as 4,000 users.

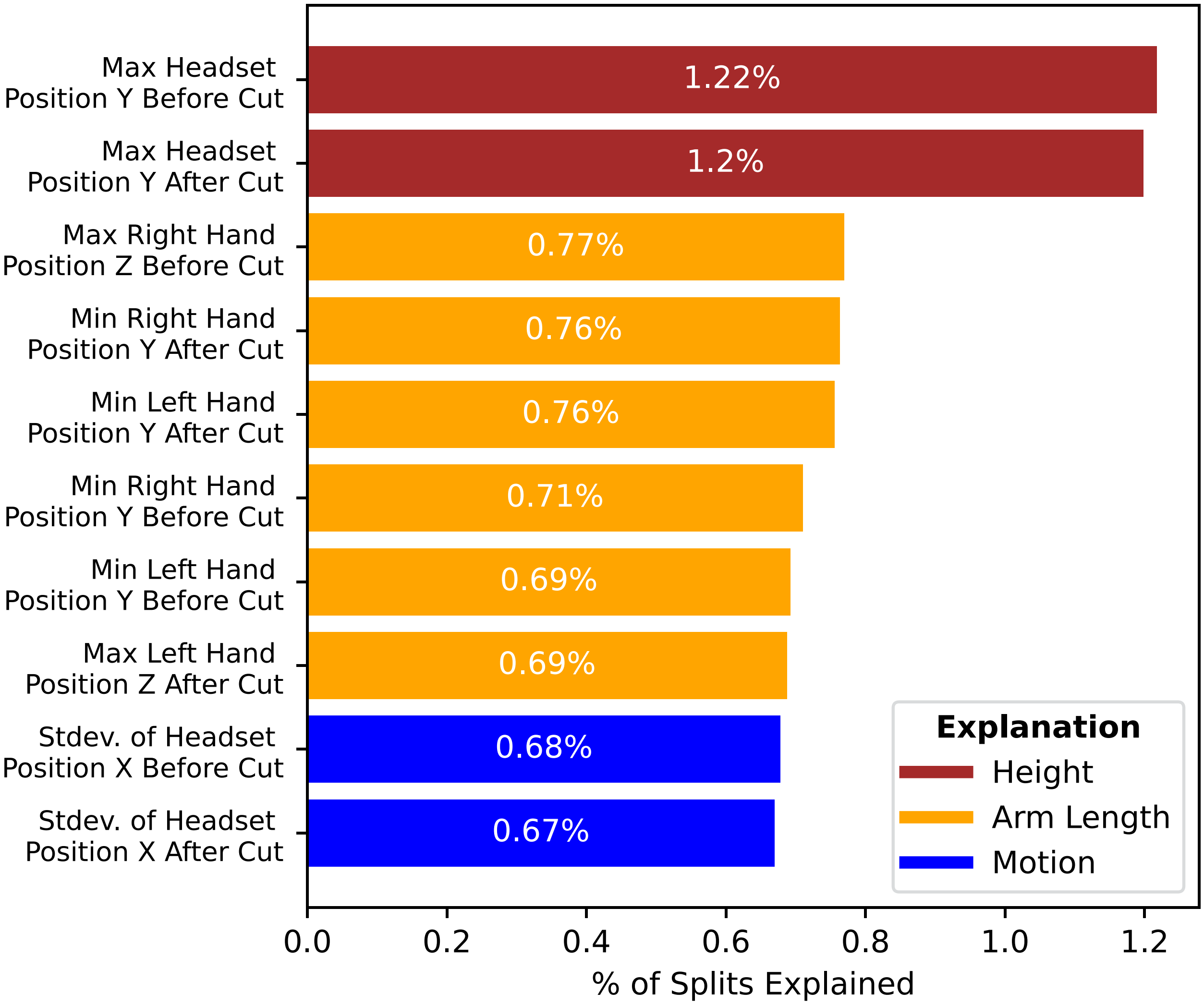


Figure 4.14: Explanation for 10 most important features.

It is no coincidence that these easily understandable features are by far the most important for identification. Unlike motion features, which are highly dependent on the specific action being taken, features that measure some static physical dimension of a user are highly consistent throughout a replay and across sessions. Thus, while the importance of any given motion feature may vary depending on the context of a sample, models can be sure to glean some information from the static features of every sample, regardless of context.

Still, these simple measurements alone hardly account for the identification of 50,000 users. A more complete picture is provided by Fig. 4.15, which shows the percentage of overall information gain explained by all 232 utilized features.

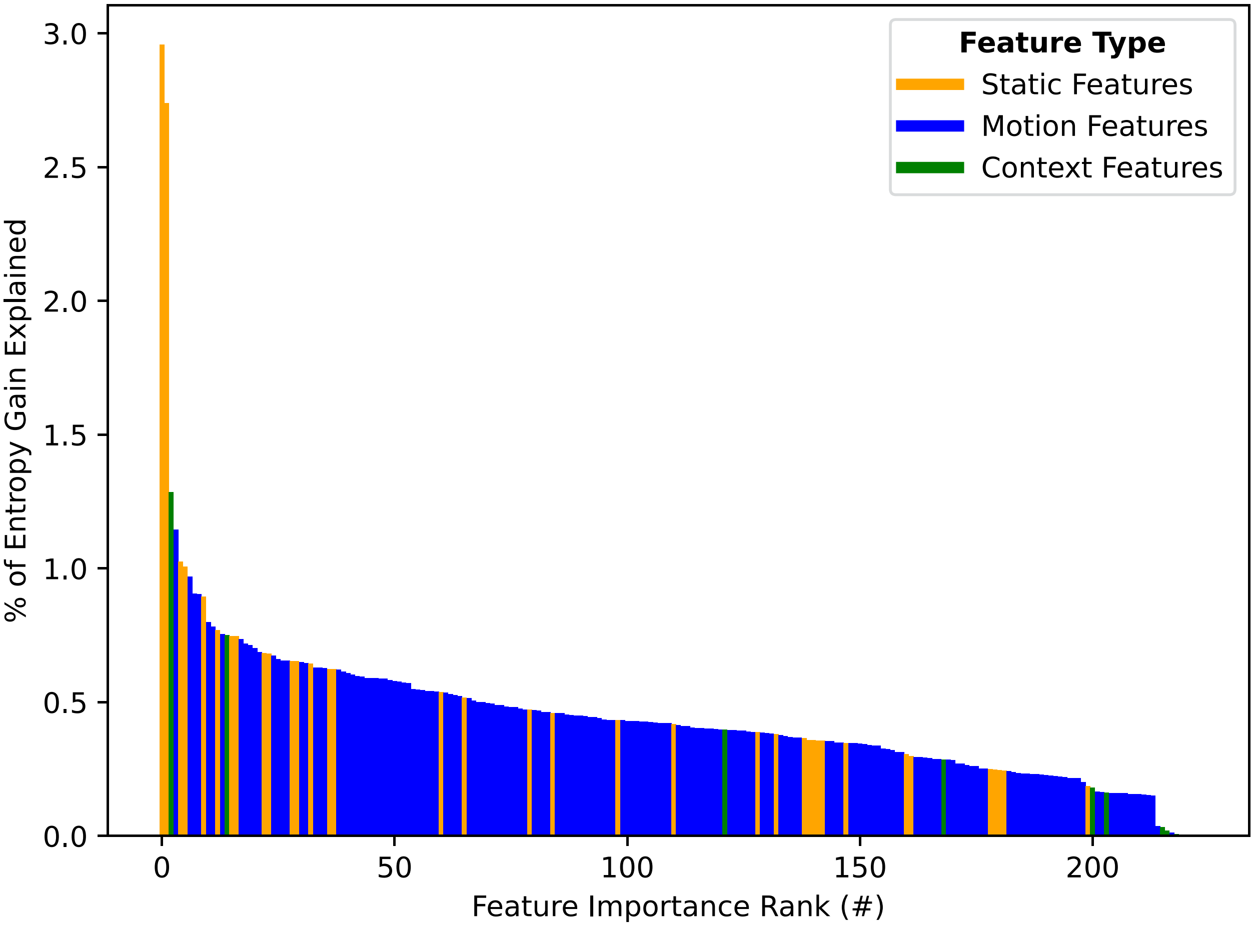


Figure 4.15: Entropy explained by all feature types.

As is evident in Fig. 4.15, motion features actually play a major role in identifying users. While static measurements comprise many of the most important features, they account for only 22.9% of the overall performance of the model. Motion features constitute 73.9% of all entropy gain, while contextual features compose the remaining 3.2%. Clearly, motion features actually represent the majority of information used by our identification model, and the task of identifying over 50,000 users would not have been possible without them.

4.9 Participant Distribution

Replays	**55,541**
≤ 5	14,945 (26.9%)
6–10	8,639 (15.6%)
11–24	12,495 (22.5%)
25–99	14,012 (25.2%)
≥ 100	5,450 (9.8%)

Platform	**55,541**
SteamVR	42,035 (75.7%)
Oculus	11,269 (20.3%)
Oculus PC	2,223 (4.0%)
Others	14 (0.0%)

Runtime	**55,541**
OpenVR	42,039 (75.7%)
Oculus	13,492 (24.3%)
Unknown	10 (0.0%)

Headset	**55,541**
Oculus Quest 2 (Standalone)	25,857 (46.6%)
Oculus Quest 2 (Quest Link)	4,124 (7.4%)
Valve Index	8,820 (15.9%)
Oculus Rift S	4,483 (8.1%)
HTC Vive	2,408 (4.3%)
Oculus Rift CV1	2,061 (3.7%)
Pico Neo 3	1,595 (2.9%)
Oculus Quest (Standalone)	1,453 (2.6%)
Oculus Quest (Quest Link)	313 (0.6%)
PICO 4	905 (1.6%)
HTC VIVE Pro	728 (1.3%)
HP Reverb G20	644 (1.2%)
HTC Vive Cosmos Elite	395 (0.7%)
HTC VIVE Pro 2	328 (0.6%)
Samsung Windows Mixed Reality	304 (0.5%)
HTC Vive Cosmos	226 (0.4%)
Others	897 (1.6%)

Controller	**55,541**
Oculus Quest Controller	16,449 (29.6%)
Oculus Touch Controller	11,240 (20.2%)
Valve Knuckles Controller	9,805 (17.7%)
Oculus Rift S Controller	3,202 (5.8%)
HTC Vive Controller	1,958 (3.5%)
Pico Neo 3 Controller	1,443 (2.6%)
Oculus Rift CV1 Controller	1,265 (2.3%)
Oculus Quest Controller	665 (1.2%)
HTC VIVE Pro Controller	602 (1.1%)
Others	8,912 (16.0%)

Handedness	**55,541**
Right	53,144 (95.7%)
Left	2,397 (4.3%)

Height	**55,541**
≤ 1.5 m	4,888 (8.8%)
1.5 m – 1.6 m	4,721 (8.5%)
1.6 m – 1.7 m	17,273 (31.1%)
1.7 m – 1.8 m	18,495 (33.3%)
1.8 m – 1.9 m	6,720 (12.1%)
≥ 1.9 m	3,444 (6.2%)

Countries	**55,541**
US	15,142 (27.3%)
DE	2,404 (4.3%)
GB	2,350 (4.2%)
CN	1,964 (3.5%)
CA	1,563 (2.8%)
JP	1,337 (2.4%)
AU	988 (1.8%)
FR	955 (1.7%)
NL	767 (1.4%)
RU	743 (1.3%)
PL	650 (1.2%)
HK	545 (1.0%)
BR	349 (0.6%)
CZ	344 (0.6%)
FI	335 (0.6%)
KR	304 (0.5%)
NO	297 (0.5%)
SE	288 (0.5%)
ES	282 (0.5%)
AT	277 (0.5%)
DK	255 (0.5%)
SG	241 (0.4%)
BE	201 (0.4%)
IT	188 (0.3%)
NZ	159 (0.3%)
TW	157 (0.3%)
MX	137 (0.2%)
CH	116 (0.2%)
HU	114 (0.2%)
CL	111 (0.2%)
IL	101 (0.2%)
TH	88 (0.2%)
AR	88 (0.2%)
IE	86 (0.2%)
Others	21389 (38.5%)

4.10 Discussion

In consideration of the differences between this study and prior work, the identification accuracy achieved in this chapter may even be stronger than it initially appears. Unlike the laboratory studies with which this work can be most directly compared, our study endures many of the pitfalls associated with utilizing a dataset from "in the wild." Chief among them is the fact that many users may actually have more than one account or play on multiple devices, resulting in the presence of multiple distinct classes which are in fact identical. Furthermore, our definition of a "session" is more rigorous than the previous work, with training and testing data for users originating from completely separate days. The largest comparable study [178] records 10 short sessions of a user on the same day. Therefore, our results represent the consistent identification of a user across wider periods of time, a task that is far more difficult than correlating motion segments recorded in close succession.

This rigorous session-based split method also provides assurances that player-map preferences are not being used for identification. One reasonable concern with the use of data from Beat Saber is that each player may have their own set of preferred maps, which could, in theory, be used by models as part of the identification process rather than motion alone. Indeed, learning a trivial relationship between a player and their favorite map would undermine the presented results. However, because our dataset consists of leaderboard high scores, we have, at most, only a single instance of a given player playing a given map. Since a replay must occur entirely within one session, our session-based split method ensures that a given player-map replay will be included in either the training or testing sets, but never both. Moreover, the hybrid featurization provides only a single note (2 seconds), from which the map cannot be inferred. Thus, it is certain, for multiple independent reasons, that player-map associations are not being used to artificially inflate identification accuracy.

Lastly, our work was the first to fully and demonstrably leverage actual movement for identification in VR. As demonstrated in §4.8, deriving simple measurements like height and arm lengths is sufficient for a model to identify tens or even hundreds of users, as is seen in Miller et al. [178]. This speculation is supported by the fact that users in that study were instructed to simply observe a number of 360-degree VR videos, a relatively static task that does not fundamentally involve much movement. By contrast, identifying 50,000 users would not have been possible without leveraging actual motion patterns, which was made possible by our featurization approach that contextualizes observed motion relative to relevant virtual objects involved in a repeatable activity. The model explainability results of §4.8 indicate that motion features played a key role in identifying users, accounting for a majority of the model's information gain. As discussed in §4.2, one cannot easily change their motion patterns, creating the potential for users to be tracked throughout the metaverse. This may, in fact, paint an incomplete picture. Motion patterns are so intrinsically tied to our physical selves that they may soon be able to follow us out of the metaverse and into the real world. Machine learning models designed to extract 3D motion data from monocular video feeds are rapidly improving [252]. We can reasonably extrapolate that it will eventually be possible to match a person's VR movements to surveillance video.

Unlike one's face, which can be covered with a mask, no physical countermeasure can reasonably obscure all of a person's movements from public view. While this threat is speculative today, the ability demonstrated in this chapter to use motion in a way comparable to other biometrics indicates that we should begin considering the realistic possibility of such scenarios in the pursuit of a future secure and private architecture for the metaverse.

On the positive side, the relatively consistent nature of identifiable motion patterns could provide an unparalleled opportunity for passive authentication in future metaverse applications. Users could benefit from the convenience of having their motion data, fundamentally required for VR functionality, also be used to verify their identity rather than needing to authenticate explicitly. Unfortunately, the laissez-faire nature with which VR motion data is currently broadcasted and shared undermines its future use in authentication; the equivalent would be using fingerprint login on your accounts if pictures of your fingerprints were already made public on the internet. Today's VR users may be paying a heavy early adoption penalty by sharing their motion data before comprehensive defenses are in place.

Limitations

There are a few notable limitations to the work presented in this chapter. Most importantly, several features were used to identify users that are arguably unique to the Beat Saber application. While Beat Saber is currently the most popular VR application in existence, it is not clear, without further investigation, whether these results will generalize to other types of VR applications. Furthermore, the "ground truth" values for some of the attributes reported in §4.9, namely height and handedness, are based on user-configurable settings, and as such, should be treated as self-reported. Indeed, many players are known to deliberately misconfigure their height setting to obtain a perceived performance advantage.

As described in §4.2 and quantified in §4.4, deep learning models, though broadly desirable, empirically underperformed tree-based models in our experiments. We found the identification performance of traditional ML models to be sufficient in light of the main focus of this chapter, which is to shed light on the sheer magnitude of the privacy concerns implicated by collecting telemetry data in VR applications. Another advantage of using LightGBM is the ability to generate rich model explanations, as shown in §4.8.

4.11 Conclusion

While perhaps not surprising to experts in biomechanics, the extent to which users can be uniquely identified by observing just a few seconds of motion of their head and hands may indeed be surprising to most. Though many don't presently think of movement patterns as a uniquely identifiable characteristic to the same extent as faces and fingerprints, results like those presented in this chapter may serve to change this assumption. The same telemetry streams which are essential to the operation of VR devices should in fact be considered highly sensitive data that may reveal a plethora of information about an end user.

Chapter 5

Inferring Private Personal Attributes of Virtual Reality Users from Head and Hand Motion Data

5.1 Introduction

As of early 2023, VR games, including “Beat Saber,” constitute 91 of the 100 most popular VR applications [260]. On conventional platforms, gaming is typically perceived as amongst the most innocuous classes of applications from a security and privacy perspective. However, in Chapter 4, we showed that motion data from Beat Saber, a non-adversarial VR rhythm game, can be used to uniquely identify over 50,000 VR users. In this study, we go a step further by exploring the extent to which popular non-adversarial VR games may inadvertently leak private information about their users by revealing their motion patterns.

To determine whether private information can be inferred from the head and hand motion data broadcast by a typical multi-player VR game, we asked over 1,000 Beat Saber players a series of about 50 questions, ranging from demographics like age and gender to personal background, behavioral patterns, and health information. We then linked each user’s responses to their corresponding motion capture recordings in the dataset described in Chapter 3. After collecting data attributes and motion samples from over 1,000 users, we designed 50 binary classification problems based on thresholding the dataset (e.g., “old” vs. “young”). We then trained and tested a deep-learning binary classifier that ingested a sequence of motion data and produced a binary classification for each attribute. We found that over 40 of the 50 attributes could consistently and reliably be inferred from user motion data alone. Thus, while these users may hold the presumption of anonymity in a VR gaming setting, this presumption is evidently flawed. Not only are their movement patterns revealing their identity, as demonstrated in Chapter 4, but our results in this chapter imply that motion patterns could actually be exposing a plethora of information about them to the device, application, server, and even other users within the same virtual environment.

The goal of this chapter is not to provide an optimal approach for inferring any particular attribute from VR motion data. Rather, we aim to demonstrate, with high statistical significance, that a wide variety of personal and privacy-sensitive variables can be inferred from head and hand motion, in order to highlight the urgent need for privacy-preserving mechanisms in multi-user VR applications.

5.2 Method

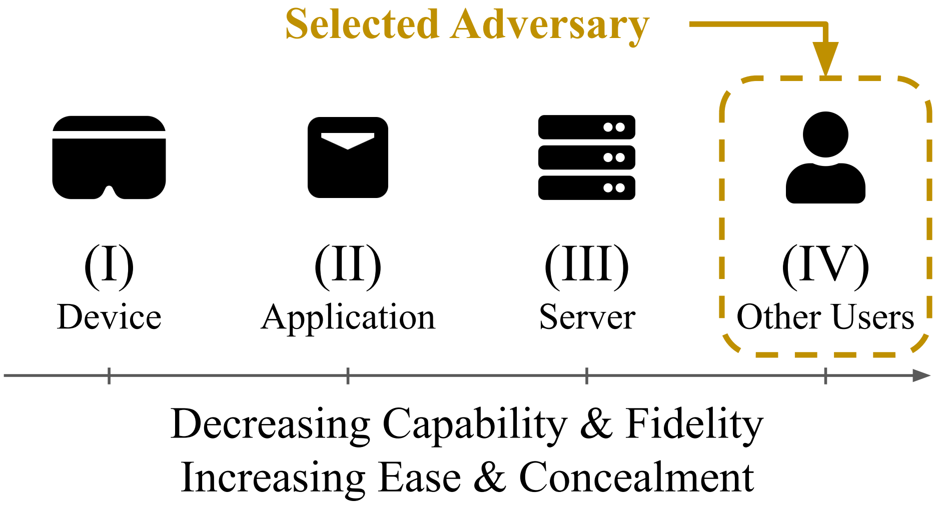


Figure 5.1: Selected VR threat actors relevant to this work (see Chapter 2.

VR Adversaries

Once again, we refer back to the information flow and threat model of Chapter 2 to contextualize this work. Recall that adversaries in this model exist on a spectrum, as shown in Figure 5.1, with adversaries becoming “weaker” from left to right due to potential interference, such as compression or transformation, at each transmission step. In this chapter, we have chosen to focus on the user adversary (IV) by using only motion data that would normally be available to ordinary users of a multi-user VR application. Because this is considered the weakest adversary in our threat model, attacks available to this user can usually also be performed by all other adversaries in the system, while also being amongst the hardest to detect due to their remote and decentralized nature.

VR Threat Scenarios

In Chapter 2, we introduced a number of existing works in the VR privacy domain in addition to proposing a standard information flow and threat model. The majority of these works are categorized as "identification," in which VR users are deanonymized or tracked across sessions based on their movement patterns. The study presented in Chapter 4 would fall into this category. A relatively smaller portion of the existing work is categorized as "profiling," whereby specific attributes, such as age or gender, are inferred from users in VR. For example, Tricomi et al. (2023) [274] accurately infer the gender and age of about 35 VR users, using eye tracking data in addition to head and hand motion. The study presented in this chapter falls into the profiling category, and aims to demonstrate that profiling is possible even by the weakest known class of adversaries, in popular benign applications like Beat Saber, and from head and hand motion data alone.

5.3 Data Collection

We partnered with the administrators of BeatLeader to conduct an official survey of BeadLeader users, consisting of about 50 personal questions across 9 categories of information. Beat Saber players were invited to take the survey via an announcement released through the official social media accounts of BeatLeader. Participation in the survey was voluntary, with all questions being optional. The categories of information collected were as follows:

1. *Participation.* Participants provided links to their BeatLeader profile from which motion capture recordings could be used.
2. *Demographics.* Participants were asked a variety of demographic questions based on the 2020 United States Census [31].
3. *Specifications.* Participants shared an automatically-generated system report containing various computer specifications.
4. *Background.* Participants were asked about their history with musical instruments, rhythm games, dancing, and athletics.
5. *Health.* Participants were asked about their mental and physical health status and disabilities as well as their visual acuity.
6. *Habits.* Participants were asked about their habits relating to Beat Saber, such as their warmup routine prior to playing.
7. *Environment.* Participants were asked about the sizes and locations of the areas in which they typically play Beat Saber.
8. *Anthropometrics.* Participants were asked to measure various physical dimensions of their body, such as height and wingspan.
9. *Clothing.* Participants were asked about the clothing and footwear they typically wear while playing Beat Saber.

The survey was conducted from April 15th, 2023 to May 1st, 2023, with 1,006 valid responses collected in that time. Participants were not monetarily compensated but were given the option to add a badge to their BeatLeader profile in recognition of their contribution. The exact questions asked in each section are given in the following section (§5.4).

Motion Recordings

Participants were required to read and agree to an informed consent document prior to beginning the survey. During the informed consent procedure for the survey, participants also gave us permission to use their publicly available motion capture recordings from the BOXRR dataset to infer the attributes contained in their survey responses. Accordingly, we cross-referenced the responses of each participant to the corresponding head and hand motion recordings in the dataset of Chapter 3. For participants with more than 100 recordings in the dataset, only the 100 most recent recordings were utilized.

Ethical Considerations

We conducted this project with significant attention to ethical considerations. Specifically, we refrained from asking questions that could be viewed as overly sensitive, and did not solicit responses from vulnerable populations, including minors under the age of 18. Participants were required to read and agree to a thorough informed consent document prior to inclusion in the study, and vulnerable subjects were excluded on a self-certification basis.

An additional source of data was the motion data in the BOXRR dataset. For further discussion of the ethical considerations of the dataset, see Chapter 3. In this study, participants explicitly agreed to our use of this motion data via the informed consent process.

Participants were not monetarily compensated or given anything of substantial value for their participation in the survey, nor penalized for non-participation. Every question in the survey was optional. Thus, participants were never unduly pressured to provide information that they were uncomfortable with disclosing.

Because the survey responses include sensitive information, such as health status, we followed the strictest data handling standards and guidelines offered by our institution throughout this study. Overall, we believe this research constitutes a net benefit to society by highlighting the magnitude of the VR privacy threat and motivating future work on defensive countermeasures. This study has been reviewed and approved by UC Berkeley's Committee for Protection of Human Subjects (CPHS), an OHRP-certified Institutional Review Board (IRB), as protocol #2023-03-16120.

5.4 Survey Questions

Participation

Mods. Have you ever played BeatSaber with the ScoreSaber and/or BeatLeader mods installed?

Secondary Accounts. Have you ever submitted a score using a BeatLeader or ScoreSaber account not listed above?

Multiple Users. Has any person other than yourself ever submitted a score to any of the BeatLeader or ScoreSaber accounts listed above?

Play Time. To the nearest hour, how many total hours have you spent playing Beat Saber?

Demographics

Sex. What is your sex?

Age. What is your age in years?

Employment Status. Which of the following options best represents your current employment status?

Marital Status. Which of the following options best represents your current marital status?

Languages. Which languages do you speak fluently? If multiple, list all languages spoken in order of proficiency.

Educational Status. What is the highest degree or level of school you have completed?

Income. Which of the following options best represents your total gross income in 2022? Convert your answer to United States Dollars (USD).

Ethnicity. What is your ethnicity?

Political Orientation. Which of the following generally best represents your political views?

Specifications

CPU Brand. According to the Steam system report, what is the vendor of the CPU in the user's PC?

Logical Cores. According to the Steam system report, how many logical CPU cores are in the user's PC?

CPU Speed. According to the Steam system report, what is the base CPU clock speed in the user's PC?

Form Factor. According to the Steam system report, is the user's PC a laptop or desktop?

Operating System. According to the Steam system report, what is the operating system of the user's PC?

System Memory. According to the Steam system report, how much RAM is in the user's PC?

Drive Space. According to the Steam system report, how much empty disk space is in the user's PC?

Base Stations. According to the Steam system report, how many lighthouses or base stations does the user have?

Graphics Card. According to the Steam system report, what is the primary GPU of the user's PC?

Background

Music. Have you ever skillfully played a musical instrument?

Music. If you have ever skillfully played a musical instrument, list the instrument(s).

Dance. Have you ever skillfully practiced or exhibited a recognized form of dance?

Rhythm Games. Have you ever played a rhythm game other than Beat Saber?

Rhythm Games. If you have ever played a rhythm game other than Beat Saber, list the game(s).

Athletics. Have you ever competitively participated in an individual or team-based athletic sport?

Athletics. If you have ever competitively participated in an individual or team-based athletic sport, list the sport(s).

Health

Eyesight. Do you regularly wear prescription glasses or contact lenses?

Lenses. Do you usually wear prescription glasses or contact lenses while playing Beat Saber?

Color Blindness. Do you have any form of color blindness or color vision deficiency?

Physical Disabilities. Have you ever been diagnosed with a physical disability or other physical health condition?

Mental Disabilities. Have you ever been diagnosed with a mental disability or other mental health condition?

Illness. In the past year, have you experienced COVID-19?

Habits

Grip. Which of the following grips do you prefer to use on standalone VR devices (e.g., Oculus Quest 2, PICO Neo3)?

Preparation. Which of the following activities, if any, do you perform immediately before playing Beat Saber? Select all that apply.

Physical Fitness. How would you rate your current level of overall physical fitness?

Caffinated Items. Approximately how many caffeinated foods or beverages (e.g., Coffee, Black Tea, Energy Drinks, etc.) do you consume on a regular basis?

Caffeine Consumption. Do you usually consume caffeine in the 3 hours before starting to play Beat Saber?

Substance Use. How often do you play Beat Saber while under the influence of any intoxicating substance?

Environment

Venue. In which location do you most often play Beat Saber?

Room Size. What are the dimensions of the play area in which you most often play Beat Saber?

Location. What is the name of the country in which you most often play Beat Saber?

Location. What is the name of state or territory in which you most often play Beat Saber?

Location. What is the name of the city in which you most often play Beat Saber?

Anthropometrics

Height. What is your exact height in centimeters?

Left Arm. What is the exact length of your left arm in centimeters?

Right Arm. What is the exact length of your right arm in centimeters?

Wingspan. What is your exact wingspan in centimeters?

Handedness. Are you left or right handed?

Weight. What is your approximate weight in kilograms?

Interpupillary Distance. What is your exact interpupillary distance (IPD) in millimeters?

Foot Size. What is the exact length of your foot in centimeters?

Hand Length. What is the exact length of your hand in centimeters?

Reaction Time. What is your average reaction time in milliseconds?

Clothing

Lower Body. What clothing, if any, do you typically wear on your lower body when playing Beat Saber?

Upper Body. What clothing, if any, do you typically wear on your upper body when playing Beat Saber?

Footwear. What footwear, if any, do you typically wear when playing Beat Saber?

5.5 Evaluation

In this section, we describe our method for determining which of the survey responses are inferrable from VR telemetry data. Specifically, we describe a machine learning model architecture that attempts to infer user data attributes based on a sequential input containing their head and hand motion. Importantly, our goal in this section is not to describe an optimal architecture for inferring any particular attribute, such as age or gender, from motion data. Rather, we aim to describe a general-purpose method for producing binary classifications from VR motion data, and use this method to determine which attributes are present in the motion data with high statistical significance.

Binary Classifications

Our survey results contain a variety of attribute types, including categorical variables such as ethnicity or languages spoken, and numerical variables like height or age, all with different observed distributions. We began by choosing 50 attributes that we speculated had a reasonable chance of being inferred from motion patterns. To simplify our analysis, we then turned each of these attributes into a binary classification problem. For example, marital status was turned into a binary classification of “never married” versus all other responses (married, divorced, etc.). For continuous attributes, such as height, a wide rejection band was usually incorporated. The exact binary splits for all attributes are given in §5.6.

This approach allows us to use a single binary classification model architecture and statistical analysis technique for all attributes being considered. This simplified method is sufficient for our purposes of demonstrating whether the attribute can be inferred from VR motion data, though regression or multi-class classification models may be more suitable for use in a real-world deployment.[1]

Our choices of attributes and binary splits to include in this study were guided by a series of informal conversations with experts in the XR privacy domain. We chose to include a mix of attributes that experts believed were “obviously” inferable (e.g., height), “likely” inferable (e.g., age), and “possibly” inferable (e.g., footwear), with the goal of illustrating a spectrum of attribute inferability.

Model Architecture

We evaluated the efficacy of a variety of machine learning architectures, including Random Forest, CNN, LSTM, and Transformer models, for performing our binary classification task using the sequential motion data. We found the Transformer-based models to be most effective at inferring a majority of the chosen attributes.

[1] We stress that deployment in a context where the user has not knowingly agreed to this type of monitoring would raise ethical concerns, particularly if the data remained linked to the user’s identity.

The Transformer model [283] incorporates an attention mechanism to capture relationships within an input sequence. Unlike other deep learning models that process the elements sequentially, the Transformer simultaneously processes all elements in parallel, allowing it to weigh the importance of each element in the context of the whole sequence.

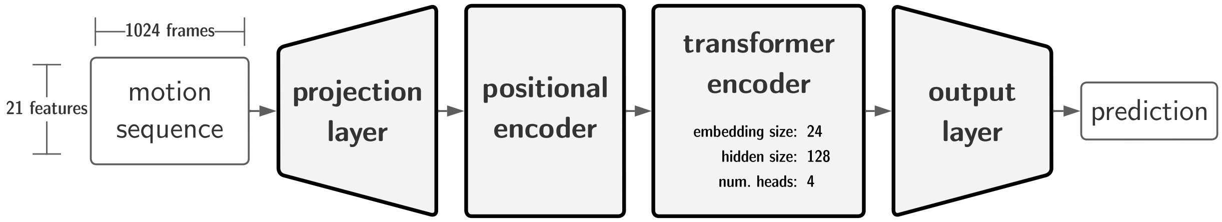


Figure 5.2: Transformer model architecture.

FIG. 5.2 illustrates our Transformer implementation. Input sequences first go through a projection layer that prepares the features by increasing the dimensionality to the embedding size of the Transformer. Following this, the data pass an encoding layer that applies a sinusoidal positional encoding, which adds information about the relative position of each element in the input sequence. This step is important, as Transformers do not have an inherent notion of order or position. Next, we use the Transformer encoder component to generate a contextualized representation of the input sequence. Finally, a dense output layer reduces the encoder output to a scalar value, which provides the binary classification.

Model Input

An advantage of transformer models is that they are intrinsically well-suited for handling time-series data. We thus chose to use a sequential featurization method to encode the motion of VR users in 3D space over a period of time. At a given time step ("frame"), we capture the position and orientation of three objects (the user's head and two hands) in 3D space. Each tracked object is captured using three positional coordinates and four orientation coordinates expressed as a quaternion. In total, 7 coordinates are taken for each object, resulting in a total of 21 values captured per frame. For each motion recording, we sample the first 1,024 frames to provide as input to our model. Thus, any given recording is represented by a (21×1024)-dimensional input; recordings with less than 1024 frames were zero-padded. The frames were sampled at their original frame rate without interpolation or normalization, as the model's normalization layer already allows it to rescale inputs internally.

As the computational complexity of transformer models increases quadratically with the sequence length, a sequence length of 1024 frames empirically provided the best balance of having sufficient information for accurate profiling while still being efficiently computable. While 1024 frames may seem to provide an overly brief and potentially irrelevant data capture, our final meta-classification is based on several such sequences per user.

Furthermore, we specifically refrained from manually restricting data capture to "key moments," such as waiting periods or instances of failure, as the attention mechanism of our transformer models is already well suited to automatically determine what the "key moments" are for a given user. Providing a sufficiently broad data sample and allowing the model to automatically determine the key events via its attention mechanism results in a more generalizable approach than selecting key moments based on external heuristics.

Data Split

Using the BeatLeader database, we downloaded the 100 most recent motion recordings from each of the users who responded to our survey. We then converted each recording into a (21×1024)-dimensional input using the featurization method of §5.5.

For each of the 50 attributes, we selected 20 users from each of the two classes to split between testing and validation sets, with the remaining users being used for training. We then resampled the training sequences to select 10,000 recordings for training each class. As such, all three sets were perfectly balanced between both classes of every binary attribute, with 10,000 recordings for training each class and 1,000 recordings for validating and testing each class. In addition to the sequence-level results, we produced a meta-classification for each user using a logarithmic sum of probabilities. An average of 100 recordings were included for each user, corresponding to between 10 and 30 minutes of real-time data capture per user. This process was repeated across 3 to 7 Monte Carlo cross-validations [309] for each attribute to assess statistical significance.

Training

We evaluated the machine learning approach by using PyTorch v2.0.1 to train and test one binary classification model for each of the 50 selected attributes. We utilized the Adam optimization algorithm [141] with a binary cross-entropy (BCE) loss function. Each model was trained across 100 epochs, with the best-performing epoch then being selected using a validation set. The evaluation was performed using a single machine with an AMD Ryzen 9 5950X 16-core CPU, 128 GB of DDR4 RAM, and an NVIDIA GeForce RTX 3090 GPU. With this setup, each model took an average of 37 minutes to train and test, with the entire evaluation taking approximately 31 hours.

Evaluation Metrics

Because our sampling technique always includes the same number of sequences and users in each class, the statistical significance of these results can be evaluated using a cumulative binomial test where n is the number of samples, K is the number of correct predictions, and $p_\emptyset$ is 0.5. We use this as our primary target metric in §5.7. The use of completely balanced training, testing, and validation sets substantially diminishes the need for more nuanced statistical tests, such as the F_1-score [240] or Cohen's kappa [44].

Our emphasis on null hypothesis significance testing (NHST) is in alignment with our overall goal of presenting a preliminary investigation exploring the potential for inferring various types of attributes from VR motion data. With 50 attributes under consideration, our evaluation cannot reasonably present the best-case inference results for each attribute, and instead represents a first step of identifying which attributes can be inferred with better than random accuracy. Doing so provides a reference that can be used to definitively argue the presence of data privacy risks from VR motion data, with no equally large-scale and comprehensive study existing to serve this purpose until now.

Hyperparameter Tuning

We performed a tuning sweep of the relevant hyperparameters (hidden size, learning rate, etc.) using just two attributes, `StandaloneGrip` and `Sex`. The hyperparameters that maximized the significance of these attributes per the metrics in §5.5 were then used throughout our evaluation and are provided below:

- Input Shape: (1024×21)
- Embedding Size: 24
- Hidden Size: 128
- Number of Heads: 4
- Number of Layers: 2
- Output Size: 1
- Learning Rate: 0.00002
- Epochs: 100
- Batch Size: 32

5.6 Response Distributions

Age	**585**
18-20	242 (41.4%)
21-24	147 (25.1%)
25-29	80 (13.7%)
30-39	71 (12.1%)
40-49	28 (4.8%)
≥ 50	17 (2.9%)

AnyAthletics	**1006**
No	548 (54.5%)
Yes	458 (45.5%)

AnyMentalDisabilities	**1006**
No	783 (77.8%)
Yes	223 (22.2%)

AnyMusic	**1006**
No	558 (55.5%)
Yes	448 (44.5%)

AnyPhysicalDisabilities	**1006**
No	850 (84.5%)
Yes	156 (15.5%)

AnyRhythmGames	**1006**
Yes	660 (65.6%)
No	346 (34.4%)

AnyVRRhythmGames	**1006**
Yes	27 (2.7%)
No	979 (97.3%)

Athletics	**858**
Swimming	114 (13.3%)
Soccer	101 (11.8%)
Basketball	83 (9.7%)
Tennis	49 (5.7%)
Baseball	47 (5.5%)
Track	36 (4.2%)
Football	34 (4.0%)
Cross Country	23 (2.7%)
Badminton	20 (2.3%)
Golf	19 (2.2%)
Volleyball	15 (1.7%)
Hockey	15 (1.7%)
Karate	14 (1.6%)
Judo	14 (1.6%)
Handball	10 (1.2%)
Table Tennis	10 (1.2%)
Rugby	9 (1.0%)
Gymnastics	8 (0.9%)
Ice Hockey	8 (0.9%)
Other	229 (25.6%)

CaffinatedBeverages	**974**
None (or Rarely)	393 (40.3%)
1-2 Items Weekly	211 (21.7%)
1-2 Items Daily	278 (28.5%)
3-4 Items Daily	61 (6.3%)
5+ Items Daily	31 (3.2%)

KEY: CLASS A – CLASS B

ColorBlindness	**971**
No	888 (91.5%)
Yes	52 (5.4%)
Maybe	31 (3.2%)

Controller	**1006**
Oculus Quest	168 (16.7%)
Valve Index	164 (16.3%)
Other	674 (67.0%)

Country	**926**
United States	376 (40.6%)
Canada	55 (5.9%)
United Kingdom	48 (5.2%)
Australia	36 (3.9%)
Germany	34 (3.7%)
France	33 (3.6%)
England	30 (3.2%)
Japan	22 (2.4%)
Netherlands	16 (1.7%)
Finland	15 (1.6%)
Poland	15 (1.6%)
New Zealand	12 (1.3%)
Denmark	11 (1.2%)
Austria	10 (1.1%)
China	9 (1.0%)
Other	204 (22.0%)

Dance	**966**
No	808 (83.6%)
Yes, recreationally	134 (13.9%)
Yes, professionally or competitively	24 (2.5%)

EducationalStatus	**955**
Less than high school	345 (36.1%)
High school graduate	246 (25.8%)
Some college	169 (17.7%)
4 year degree	108 (11.3%)
Professional degree	47 (4.9%)
2 year degree	34 (3.6%)
Doctorate	6 (0.6%)

EmploymentStatus	**970**
Student	521 (53.7%)
Employed full time	229 (23.6%)
Employed part time	96 (9.9%)
Unemployed looking for work	77 (7.9%)
Unemployed not looking for work	37 (3.8%)
Disabled	8 (0.8%)
Other	2 (0.2%)

Ethnicity	**976**
White	760 (77.9%)
Asian	109 (11.2%)
Black or African American	26 (2.7%)
American Indian or Alaska Native	8 (0.8%)
Native Hawaiian or Pacific Islander	3 (0.3%)
Other	70 (7.2%)

FootSize	**811**
< 24.0 cm	63 (7.8%)
24.0-24.9 cm	62 (7.6%)
25.0-25.9 cm	140 (17.3%)
26.0-26.9 cm	183 (22.6%)
27.0-27.9 cm	178 (21.9%)
28.0-28.9 cm	98 (12.1%)
29.0-29.9 cm	39 (4.8%)
≥ 30.0 cm	48 (5.9%)

Footwear	**883**
Typically Barefoot	350 (39.6%)
Typically Wear Socks	347 (39.3%)
Inconsistent/Varies	105 (11.9%)
Typically Wear Shoes	81 (9.2%)

HadCOVID	**1006**
No	655 (65.1%)
Yes	351 (34.9%)

HandLength	**760**
< 16.0 cm	41 (5.4%)
16.0-16.9 cm	57 (7.5%)
17.0-17.9 cm	132 (17.4%)
18.0-18.9 cm	151 (19.9%)
19.0-19.9 cm	185 (24.3%)
20.0-20.9 cm	99 (13.0%)
21.0-21.9 cm	32 (4.2%)
≥ 22.0 cm	63 (8.3%)

Handedness	**900**
Right Handed	737 (81.9%)
Left Handed	103 (11.4%)
Ambidextrous	60 (6.7%)

Headset	**1006**
Oculus Quest 2	499 (49.6%)
Valve Index	150 (14.9%)
Other	357 (35.5%)

Height	**838**
< 1.70 m	191 (22.8%)
1.70-1.79 m	321 (38.3%)
1.80-1.89 m	288 (34.4%)
≥ 1.90 m	38 (4.5%)

IPD	**737**
< 58.0 mm	28 (3.8%)
58.0-62.9 mm	153 (20.8%)
63.0-67.9 mm	373 (50.6%)
68.0-71.9 mm	143 (19.4%)
≥ 72.0 mm	40 (5.4%)

Income	**908**
Less than $10,000	583 (64.2%)
$10,000 to $19,999	94 (10.4%)
$20,000 to $29,999	39 (4.3%)
$30,000 to $39,999	45 (5.0%)
$40,000 to $49,999	39 (4.3%)
More than $50,000	108 (11.9%)

Languages	**1406**
English	845 (60.1%)
German	71 (5.0%)
French	58 (4.1%)
Spanish	49 (3.5%)
Japanese	37 (2.6%)
Dutch	33 (2.3%)
Polish	22 (1.6%)
Other	270 (19.2%)

LeftArm	**692**
< 0.60 m	62 (9.0%)
0.60-0.69 m	256 (37.0%)
0.70-0.79 m	306 (44.2%)
≥ 0.80 m	68 (9.8%)

Lenses	**972**
Never	721 (74.2%)
Often	231 (23.8%)
Sometimes	20 (2.1%)

LowerBody	**880**
Knee-Height Garment	350 (39.8%)
Ankle-Height Garment	248 (28.2%)
Inconsistent/Varies	194 (22.0%)
Undergarment Only	88 (10.0%)

MaritalStatus	**959**
Never married	896 (93.4%)
Married	39 (4.1%)
Divorced	16 (1.7%)
Separated	6 (0.6%)
Widowed	2 (0.2%)

Music	**714**
Piano	195 (27.3%)
Guitar	107 (15.0%)
Drums	63 (8.8%)
Trumpet	58 (8.1%)
Violin	41 (5.7%)
Saxophone	26 (3.6%)
Trombone	23 (3.2%)
Clarinet	21 (2.9%)
Flute	18 (2.5%)
Cello	16 (2.2%)
Bass	14 (2.0%)
Recorder	11 (1.5%)
Ukulele	8 (1.1%)
Viola	7 (1.0%)
Percussion	6 (0.8%)
French Horn	5 (0.7%)
Tuba	4 (0.6%)
Other	91 (12.7%)

PhysicalFitness	**972**
Far below average	51 (5.2%)
Below average	252 (25.9%)
Average	387 (39.8%)
Above average	242 (24.9%)
Far above average	40 (4.1%)

PoliticalOrientation	**934**
Independent or Neither	454 (48.6%)
Liberal or Left Wing	334 (35.8%)
Conservative or Right Wing	72 (7.7%)
Other	74 (7.9%)

Preparation	**1006**
None	247 (24.6%)
Warmup Only	244 (24.3%)
Warmup and Stretches	192 (19.1%)
Other	323 (32.1%)

ReactionTime	**852**
$<$ 125 ms	8 (0.9%)
125-174 ms	59 (6.9%)
175-199 ms	166 (19.5%)
200-224 ms	233 (27.3%)
225-274 ms	230 (27.0%)
275-324 ms	103 (12.1%)
$\geq$ 325 ms	53 (6.2%)

RhythmGames	**1850**
Osu	383 (20.7%)
Guitar Hero	107 (5.8%)
A Dance Of Fire And Ice	121 (6.5%)
Geometry Dash	100 (5.4%)
Quaver	82 (4.4%)
Muse Dash	65 (3.5%)
Dance Dance Revolution	42 (2.3%)
Synth Riders	38 (2.1%)
Arcaea	36 (1.9%)
Pistol Whip	27 (1.5%)
Rock Band	25 (1.4%)
Other	824 (44.7%)

RightArm	**692**
$<$ 0.60 m	57 (8.2%)
0.60-0.69 m	256 (37.0%)
0.70-0.79 m	305 (44.1%)
$\geq$ 0.80 m	74 (10.7%)

RoomArea	**589**
$<$ 2.0 m^2	82 (13.9%)
2.0-3.9 m^2	187 (31.7%)
4.0-5.9 m^2	161 (27.3%)
6.0-7.9 m^2	99 (16.8%)
$\geq$ 8.0 m^2	60 (10.2%)

Sex	**979**
Male	806 (82.3%)
Female	91 (9.3%)
Other	82 (8.4%)

StandaloneGrip	**475**
Default Grip	229 (48.2%)
Claw Grip	154 (32.4%)
Standard M-Grip	13 (2.7%)
Yoshi M-Grip	10 (2.1%)
Other	68 (14.5%)

SteamComputerFormFactor	**568**
Desktop	513 (90.3%)
Laptop	55 (9.7%)

SteamLighthouses	**194**
1	13 (6.7%)
2	133 (68.6%)
3	35 (18.0%)
4	13 (6.7%)

SteamOperatingSystemVersion	**574**
Windows 10 (64 bit)	385 (67.1%)
Windows 11 (64 bit)	177 (30.8%)
Other	12 (2.1%)

SteamProcessorCPUVendor	**572**
AMD	333 (58.2%)
Intel	239 (41.8%)

SteamProcessorLogicalCores	**569**
4	8 (1.4%)
6	22 (3.9%)
8	48 (8.4%)
12	199 (35.0%)
16	199 (35.0%)
20	30 (5.3%)
24	45 (7.9%)
32	18 (3.2%)

SubstanceUse	**970**
Never	778 (80.2%)
Rarely	130 (13.4%)
Somewhat Often	24 (2.5%)
Often	38 (3.9%)

TotalPlayTime	**1006**
$<$ 100 Hours	200 (19.9%)
100-499 Hours	358 (35.6%)
500-999 Hours	208 (20.7%)
1000-1999 Hours	182 (18.1%)
$\geq$ 2000 Hours	58 (5.8%)

UpperBody	**764**
Short Sleeve Garment	526 (68.8%)
Inconsistent/Varies	133 (17.4%)
Sleeveless Garment	52 (6.8%)
Undergarment Only	23 (3.0%)
Long Sleeve Garment	20 (2.6%)
Multiple Layers	10 (1.3%)

Weight	**834**
$<$ 40.0 kg	5 (0.6%)
40.0-49.9 kg	51 (6.1%)
50.0-59.9 kg	154 (18.5%)
60.0-69.9 kg	189 (22.7%)
70.0-79.9 kg	193 (23.1%)
80.0-89.9 kg	103 (12.4%)
90.0-99.9 kg	70 (8.4%)
$\geq$ 100.0 kg	69 (8.3%)

Wingspan	**710**
$<$ 1.60 m	133 (18.7%)
1.60-1.69 m	149 (21.0%)
1.70-1.79 m	205 (28.9%)
1.80-1.89 m	167 (23.5%)
$\geq$ 1.90 m	56 (7.9%)

	Per Sequence				Per User			
Attribute	**Total #**	**Test #**	**Accuracy**	**Significance**	**Total #**	**Test #**	**Accuracy**	**Significance**
StandaloneGrip	31,100	6,000	85.9%	p <0.001	311	60	91.7%	p <0.001
Height	19,100	6,000	76.5%	p <0.001	191	60	86.7%	p <0.001
Controller	33,200	6,000	81.2%	p <0.001	332	60	85.0%	p <0.001
Weight	9,800	6,000	73.6%	p <0.001	98	60	85.0%	p <0.001
FootSize	9,100	6,000	73.2%	p <0.001	91	60	85.0%	p <0.001
Country	33,300	6,000	60.3%	p <0.001	333	60	81.7%	p <0.001
RhythmGames	10,900	6,000	63.5%	p <0.001	109	60	80.0%	p <0.001
Age	62,300	6,000	64.9%	p <0.001	623	60	78.3%	p <0.001
TotalPlayTime	34,400	6,000	67.7%	p <0.001	344	60	78.3%	p <0.001
Headset	65,000	6,000	66.9%	p <0.001	650	60	76.7%	p <0.001
LeftArm	10,300	6,000	65.2%	p <0.001	103	60	76.7%	p <0.001
RightArm	10,200	6,000	64.9%	p <0.001	102	60	75.0%	p <0.001
Athletics	8,700	6,000	59.1%	p <0.001	87	60	75.0%	p <0.001
MaritalStatus	81,400	6,000	60.2%	p <0.001	814	60	73.3%	p <0.001
EmploymentStatus	64,200	6,000	65.1%	p <0.001	642	60	71.7%	p <0.001
AnyRhythmGames	83,000	6,000	54.8%	p <0.001	830	60	70.0%	p <0.001
Ethnicity	73,900	6,000	59.7%	p <0.001	739	60	70.0%	p <0.001
SteamComputerFormFactor	51,300	6,000	58.5%	p <0.001	513	60	70.0%	p <0.001
Footwear	36,700	6,000	60.5%	p <0.001	367	60	70.0%	p <0.001
AnyVRRhythmGames	83,000	8,000	56.8%	p <0.001	830	80	68.8%	p <0.001
Income	76,700	8,000	55.0%	p <0.001	767	80	68.8%	p <0.001
Wingspan	16,000	8,000	59.9%	p <0.001	160	80	68.8%	p <0.001
Handedness	71,600	10,000	55.2%	p <0.001	716	100	66.0%	p <0.001
HandLength	51,000	8,000	58.5%	p <0.001	510	80	66.3%	p = 0.002
SubstanceUse	69,200	10,000	55.9%	p <0.001	692	100	64.0%	p = 0.002
Preparation	39,400	8,000	58.2%	p <0.001	394	80	65.0%	p = 0.005
LowerBody	29,500	8,000	55.9%	p <0.001	295	80	65.0%	p = 0.005
Lenses	80,900	8,000	55.3%	p <0.001	809	80	65.0%	p = 0.005
Languages	80,700	8,000	56.5%	p <0.001	807	80	65.0%	p = 0.005
SteamOperatingSystemVersion	50,800	8,000	58.4%	p <0.001	508	80	65.0%	p = 0.005
Music	29,600	8,000	53.6%	p <0.001	296	80	65.0%	p = 0.005
AnyMentalDisabilities	83,000	10,000	52.6%	p <0.001	830	100	63.0%	p = 0.006
Sex	76,300	10,000	56.5%	p <0.001	763	100	63.0%	p = 0.006
AnyPhysicalDisabilities	83,000	10,000	54.5%	p <0.001	830	100	62.0%	p = 0.010
ReactionTime	9,800	14,000	53.1%	p <0.001	98	140	60.0%	p = 0.011
AnyMusic	83,000	8,000	55.7%	p <0.001	830	80	62.5%	p = 0.016
AnyAthletics	19,900	8,000	55.7%	p <0.001	199	80	61.3%	p = 0.016
EducationalStatus	62,200	8,000	57.1%	p <0.001	622	80	60.0%	p = 0.028
IPD	6,700	8,000	55.8%	p <0.001	67	80	60.0%	p = 0.028
Dance	82,000	10,000	52.3%	p <0.001	820	100	59.0%	p = 0.028
PoliticalOrientation	33,100	10,000	53.5%	p <0.001	331	100	58.0%	p = 0.044
UpperBody	47,200	10,000	52.0%	p <0.001	472	100	57.0%	p = 0.067
SteamProcessorLogicalCores	33,500	10,000	51.0%	p = 0.021	335	100	56.0%	p = 0.136
HadCOVID	83,000	10,000	54.4%	p <0.001	830	100	55.0%	p = 0.136
CaffinatedBeverages	40,800	10,000	52.9%	p <0.001	408	100	55.0%	p = 0.136
RoomArea	33,100	8,000	50.5%	p = 0.183	331	80	56.3%	p = 0.157
PhysicalFitness	7,800	12,000	54.2%	p <0.001	78	120	55.0%	p = 0.158
SteamProcessorCPUVendor	51,600	10,000	49.2%	p = 0.953	516	100	53.0%	p = 0.309
SteamLighthouses	5,500	8,000	48.6%	p = 0.993	55	80	52.5%	p = 0.369
ColorBlindness	79,800	10,000	50.4%	p = 0.227	798	100	52.0%	p = 0.382

Table 5.1: Accuracy of inferring 50 attributes from VR motion data, with statistical significance calculated via binomial tests.

5.7 Results

After training a model for each of the tested attributes, we first generated a classification for all of the 100 sequences per user for every user in the testing set. Next, we generated a meta-classification for each user as described in §5.5. Table 5.1 shows the accuracy of the results per sequence and per user, along with the p-values corresponding to the metrics described in §5.5. Overall, 33 of the 50 attributes were predicted with high statistical significance ($p < 0.01$), and another 8 of 50 with moderate statistical significance ($p < 0.05$) on a per-user basis. On a per-sequence basis, 45 out of 50 attributes were highly significant ($p < 0.01$), and one was moderately significant ($p < 0.05$). This difference is largely accounted for by sample size; in total, 100 times more recordings were present than users.

Macro Significance

Given that we evaluated 50 attributes in this chapter, only a portion of which were inferred with significant accuracy, it remains to be demonstrated that the overall evaluation was statistically significant. To assess the overall significance of our result, we performed a secondary evaluation in which the trained models from our main evaluation were tested with randomly-generated fictitious inputs. We then performed a Wilcoxon signed-rank test to compare the distribution of classification accuracy values across the 50 attributes on these fictitious inputs with the distribution of true results in Table 5.1. We found $p < 0.0001$ on both a per-sequence and per-user basis, indicating a high overall statistical significance.

5.8 Discussion

In Chapter 4, we showed that motion data from a seemingly harmless VR rhythm game can be used to uniquely identify over 50,000 users. In this chapter, we have shown that the same motion data can be used to infer a wide variety of user characteristics. "Beat Saber," the game used in both chapters, is not particularly conducive to data harvesting, with a simple ruleset and interaction model. For instance, there are no in-game characters to interact with, which could reveal even more information than we already observed.

In comparison with prior work, the setting evaluated in this chapter represents a realistic and challenging threat scenario. Our data comes from real VR users around the world with a wide variety of devices and environments. We limited our models to only use head and hand motion data and used the weakest adversary class for our evaluations. Despite these limitations, a large number of personal attributes were accurately and consistently inferable XR motion data alone. These attributes go beyond the obvious anthropometric measurements to include a surprising amount of information about the player's background, demographics, environment, habits, and even health. Many of these attributes, such as disability status, could be considered highly private information by end users.

There are also a number of avenues adversaries could pursue to further improve VR profiling capabilities. Since Chapter 4 has demonstrated that VR motion patterns constitute uniquely identifiable biometrics, adversaries are not limited to collecting data from a single application. Rather, because Chapter 4 and this chapter both target user adversaries, attackers could combine the attacks of both chapters by leveraging the identifiability of VR users to track them across applications and usage sessions, building a rich user profile over time. These risks are further exacerbated with the introduction of additional sensors, such as microphones, cameras, LIDAR arrays, and eye and body tracking, all of which may provide data beyond the head and hand motion considered herein.

Limitations

The motion recordings used in this study originate entirely from a single game. While "Beat Saber" is the most popular VR game to date [303], and is a representative example of a non-adversarial VR game, we cannot yet demonstrate that our findings will generalize to other types of VR applications. Furthermore, we chose to only survey existing Beat Saber players, and are unsure whether novice players, who would potentially demonstrate less consistent movement patterns, would be equally susceptible to these inferences.

We used the game recordings to infer a series of attributes that were self-reported via an online survey, and were thus subject to the biases typically associated with self-reported data. The participants in this survey were also not representative of the general population; for example, over 80% of respondents were male. However, the sample is fairly representative of the current VR user population [69]. The distribution of each attribute is given in Appendix 5.6. Unbalanced distributions did not inflate the reported results, as each binary class was rebalanced prior to training and testing.

Finally, a portion of the reported findings may be the result of hidden correlations rather than direct inference. For example, it is likely that some attributes like employment or marital status are not directly observable from motion, but are correlated to age, which can be inferred from motion data. These correlations could apply generally to human motion, but may also represent sampling or response biases. The following section (§5.9) shows the correlation between each pair of attributes. Due to the difficulty of explaining the internal function of deep learning models, we cannot easily determine the mechanisms of causality associated with each result. However, we consider the potential to infer this data from VR users to be noteworthy and concerning, regardless of the cause.

5.9 Response Correlations

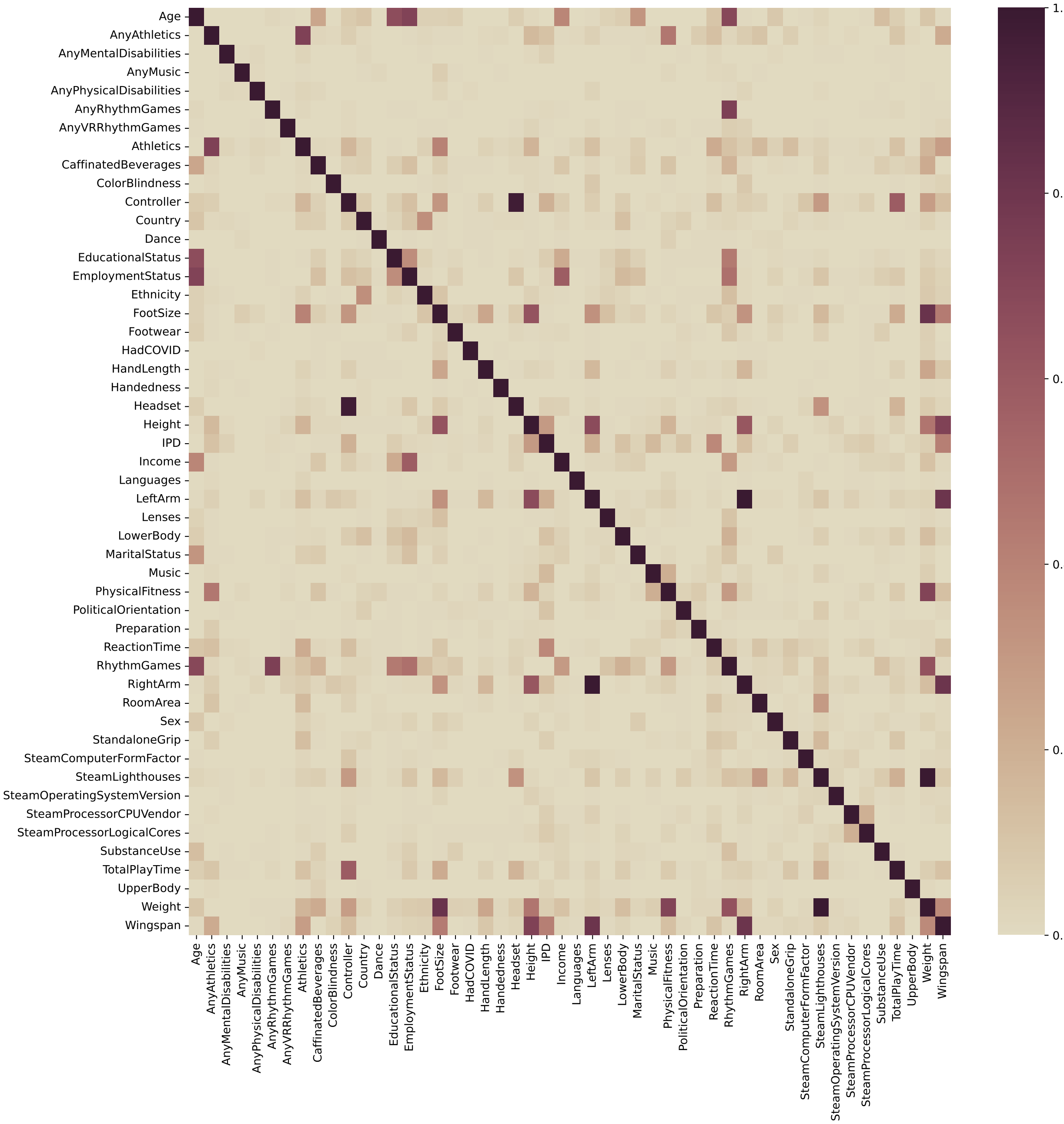


Figure 5.3: Correlation coefficient (R^2) between all pairs of attributes.

5.10 Conclusion

With major new products like the Meta Quest 3 and Apple Vision Pro on the horizon, XR technologies are on track to soon become a ubiquitous means of accessing the internet. For the foreseeable future, motion tracking "telemetry" data will remain at the core of nearly all extended reality and metaverse experiences. In the previous chapter, we showed that even in non-adversarial applications, motion data is a strong biometric signal that can be used to uniquely identify tens of thousands of users. In this chapter, we further demonstrated this data stream carries significant data privacy implications for XR users.

As of now, we have only explored the capabilities available to weak XR adversaries, such as game servers or other end users, through passive observation alone. However, these threats, despite their severity, are not the most pernicious risks applicable to XR. In the next chapter, we shift our focus to the application-level adversary, which is capable of conducting active attacks through adversarial VR game design. We show, through a series of examples, how such adversaries can conduct even more detailed and accurate data harvesting attacks than those demonstrated in this dissertation so far.

Chapter 6

Exploring the Privacy Risks of Adversarial Virtual Reality Game Design

6.1 Introduction

In the previous chapters of this dissertation, we have demonstrated that individuals exhibit distinct biomechanical motion patterns that can be used to identify them or infer their personal attributes. While our work thus far has largely focused on passive observation of VR users, the success of games specifically designed to harvest user data [121] on conventional social platforms motivates us to now investigate similar active attacks in VR.

This chapter aims to shed light on the significant privacy risks associated with adversarially designed VR games that appear innocuous to end users. We have identified over 25 examples of private data attributes that attackers can covertly harvest from VR users. We incorporate all of these attacks into "MetaData," an open-source adversarial VR game we created as a proof of concept of active adversarial attacks in VR. We then experimentally demonstrate the efficacy of this adversary through a 50-person user study. Many of the attributes recovered by our adversarial game would be difficult to observe passively but can be obtained with high fidelity by prompting users to unknowingly reveal more information about themselves via carefully designed interactive game elements.

While motion data is the primary focus of this dissertation, this chapter also investigates other data modalities, such as audio and network data, that are typically available to VR applications. Our findings highlight that these additional sensors, and even output devices like displays and haptics, can be combined with motion data to create enhanced privacy risks, allowing us to not lose sight of the bigger picture of the VR privacy landscape.

6.2 Method

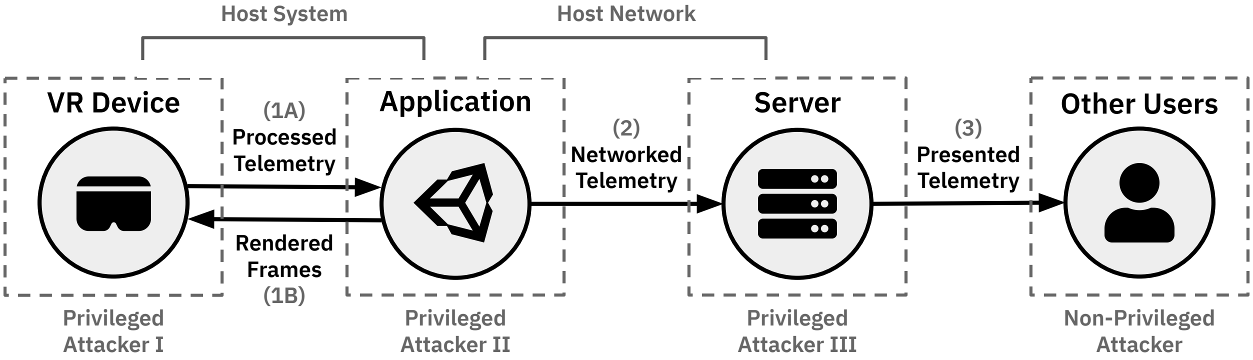


Figure 6.1: Virtual reality information flow and threat model (see Chapter 2).

VR Adversaries

Referring back to the VR information flow and threat model of Chapter 2, we now turn our focus to the capabilities of an adversarial VR application. As such, we now evaluate all threats from the perspective of the application adversary (labeled “Privileged Attacker II” above), while noting which other entities may be capable of performing similar attacks.

Gamified Data Harvesting

In 2018, the British political consulting firm Cambridge Analytica was revealed to be in possession of personal data from up to 87 million Facebook users. Subsequent analysis revealed that most of this data was collected through Facebook quizzes designed to seem like fun personality assessments while actually building a detailed profile of user data [121, 243, 21]. Gamified data collection mechanisms bypass the normal cognitive filters associated with data privacy by taking advantage of users’ innate desire to perform optimally when completing challenges. Presenting a data-revealing question as a puzzle or element serving a legitimate role in a broader game has proven effective at obscuring the hidden intent to collect personal information [51, 117]. In this chapter, we seek to combine the possibility of inferring private data attributes from VR motion data, demonstrated in Chapter 5, with the success of gamified data harvesting in conventional social platforms to explore data harvesting attacks made possible by adversarial game design in VR. Gaming is the predominant driver of VR adoption today [260], providing ample opportunity to disguise data collection mechanisms as VR game elements. Simultaneously, the social platforms exploited by Cambridge Analytica are now dominant players in the AR/VR space. It is therefore natural to assume that some may have incentives to use the same techniques that have proven successful on conventional social media platforms in the data-rich environment of VR.

6.3 Active VR Privacy Attacks

We begin by describing specific examples of adversarial game elements designed to extract VR user data, corresponding to the broad observable attribute classes detailed in Chapter 2. The goal of this section is not to be exhaustive with respect to the wide variety of interactive elements that can reveal user information, but rather to exemplify strategies for collecting various attributes using specific mechanisms that we later evaluate in our user study (§6.4).

Biometrics

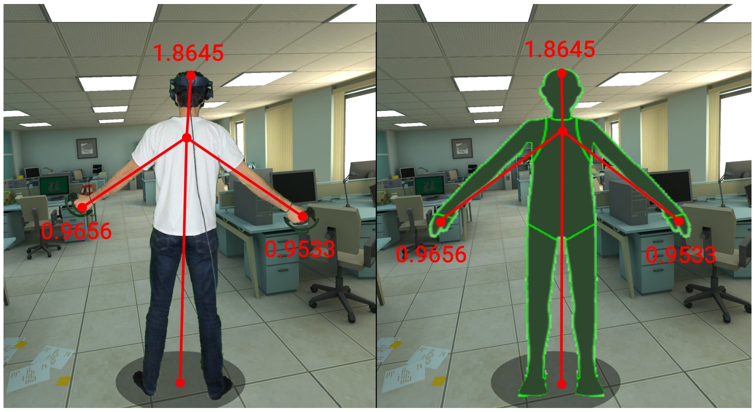


Figure 6.2: Measuring user anthropometrics from telemetry.

Continuous Anthropometrics. Basic anthropometrics provide a simple yet compelling example of the dangers of adversarial design. Fig. 6.2 illustrates how attackers can passively measure a user's height and wingspan from VR telemetry. However, users are unlikely to naturally stand in a position that readily facilitates the measurement of wingspan. Therefore, Fig. 6.3 depicts a pose-based game element designed to subtly induce a standing position more conducive to precise anthropometric measurement.

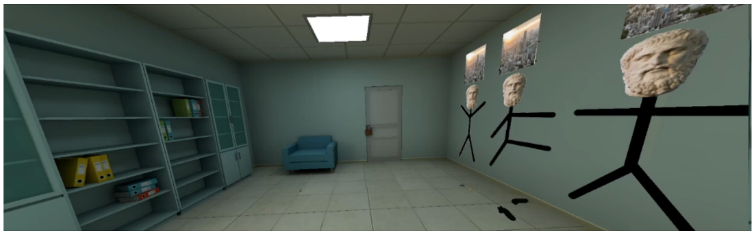


Figure 6.3: Adversarial measurement of wingspan.

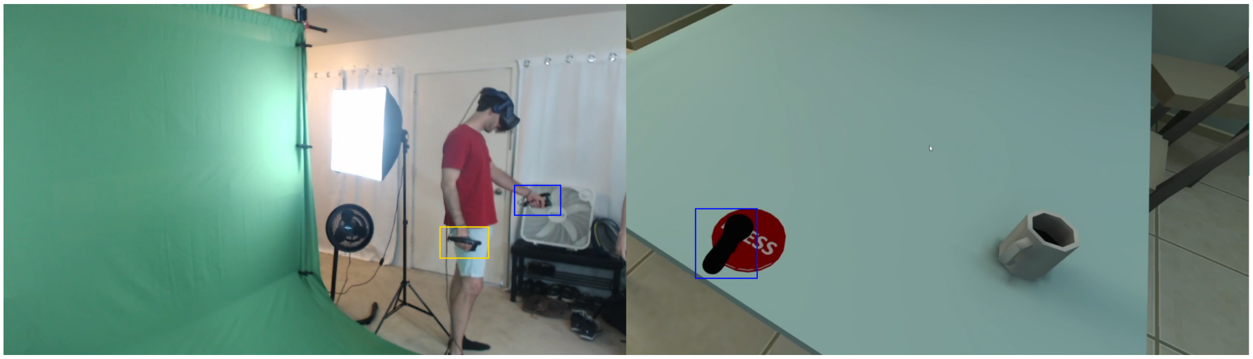


Figure 6.4: Estimating handedness from behavior.

Binary Anthropometrics. An attacker can collect binary anthropometrics, which include characteristics such as longer-arm and dominant handedness, both directly from telemetry (e.g., "which hand moves more?") and from behavior (e.g., "which hand is used to press a button?"). Fig. 6.4 illustrates an example process of determining a user's handedness by including a small button that requires precise manipulation, suggesting the use of one's dominant hand; catching or throwing a ball is an equivalent idea.

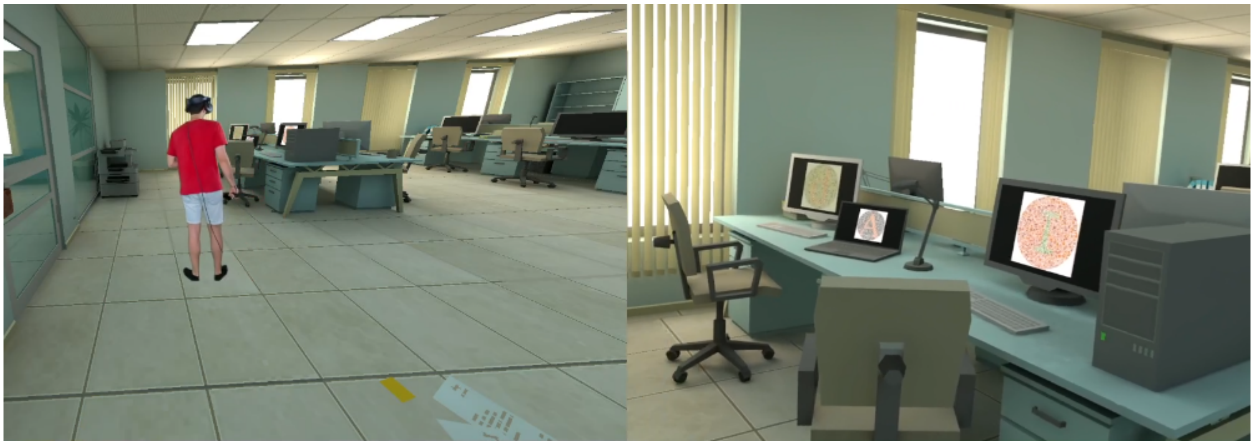


Figure 6.5: VR puzzle revealing deuteranopia.

Vision. VR attackers can carefully construct interactive elements that secretly reveal aspects of a player's visual acuity, such as nearsightedness, farsightedness, or color blindness. For example, Fig. 6.5 shows a puzzle element of a VR game that appears innocuous to most users but is not solvable by users with red-green color blindness (deuteranopia), thus revealing the presence of that condition.

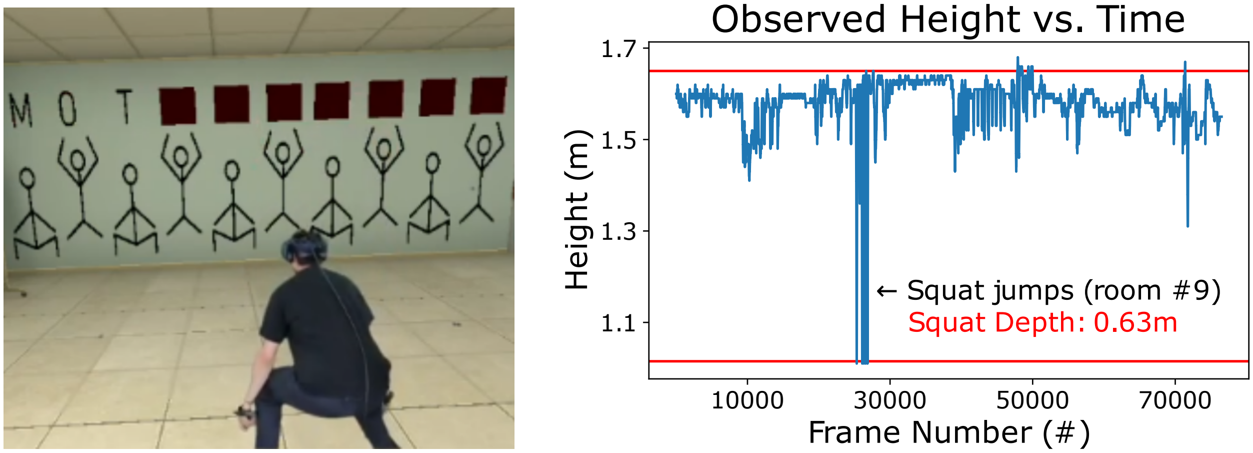


Figure 6.6: Measurement of physical fitness.

Fitness. Attackers can also use behavioral and telemetric measurements to asses a subject's degree of physical fitness. Fig. 6.6 illustrates a virtual room designed to elicit physical activity and shows the resulting metric of physical fitness measurable on a headset position (y-coordinate) vs. time graph. We observed that a squat depth of less than 25% of height corresponded to low physical fitness, though other metrics can also be used. An extreme lack of fitness may reveal a participant's age or the presence of physical disabilities.

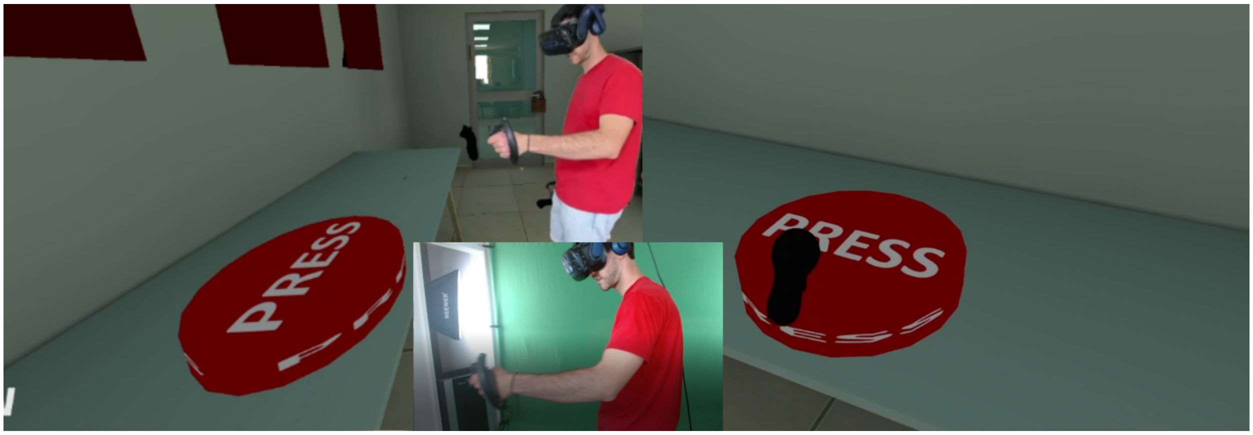


Figure 6.7: VR puzzle measuring reaction time.

Reaction Time. Fig. 6.7 shows a VR environment adversarially constructed to reveal the participant's reaction time by measuring the time interval between a visual stimulus and motor response. Reaction time is strongly correlated with age [305].

Environment

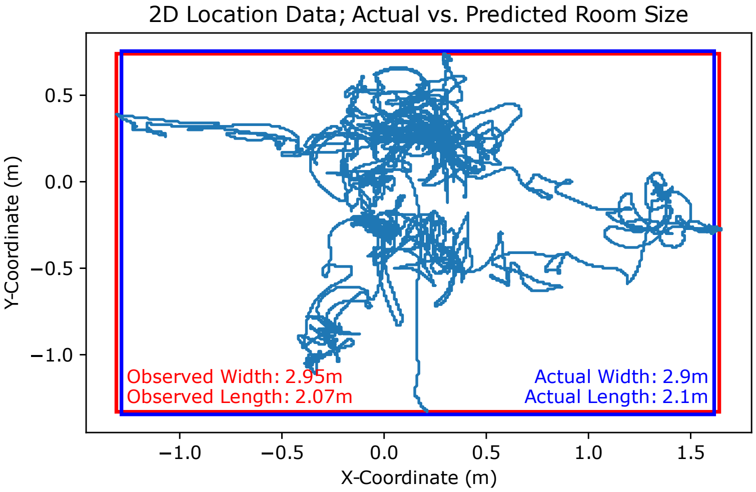


Figure 6.8: Estimating room size from spatial telemetry.

Room Size. Fig. 6.8 shows how an attacker could estimate the size of a user's physical environment by tracking their virtual movements. Virtual environments can be designed to contain interactive elements which specifically encourage the participant to explore the boundaries of their physical environment.

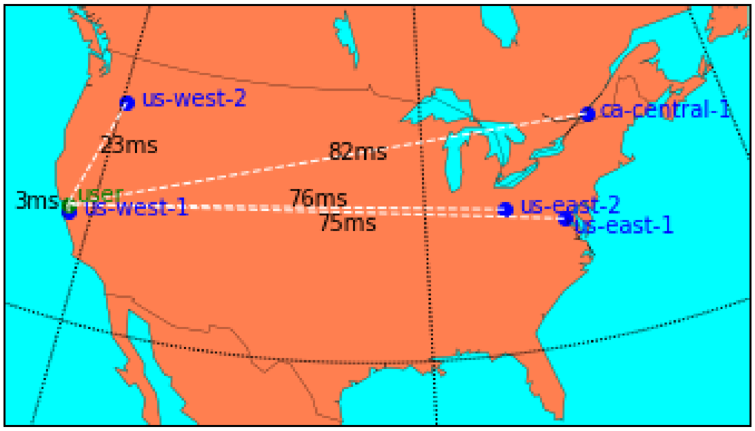
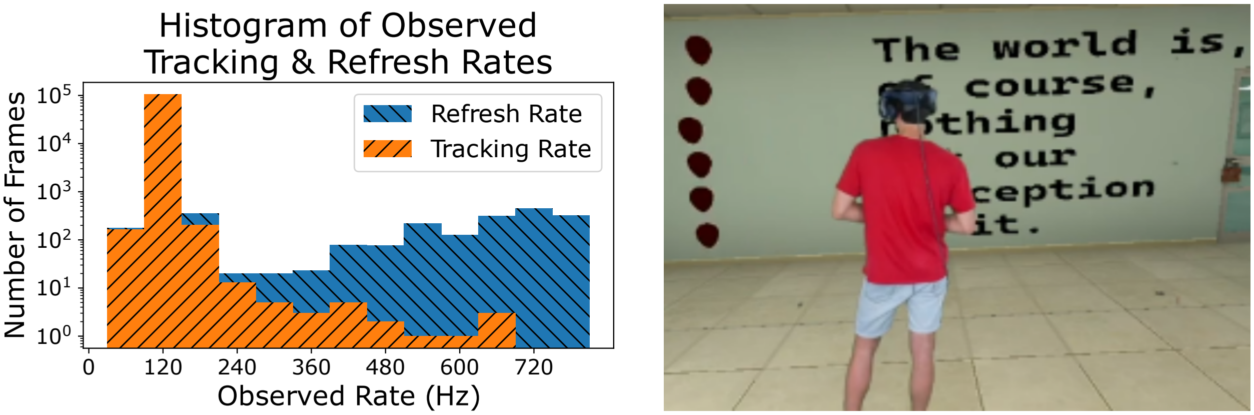


Figure 6.9: Estimating user location from network latency.

Geolocation. Fig. 6.9 shows how observing the round-trip delay between a client device and multiple game servers (proximity) can reveal an end user's location (locality) via multilateration. A non-privileged attacker could use the round trip delay of audio signals as an approximate measure of latency.

Device Specifications



(a) Measuring headset and tracking refresh rate from device API throughput (log scale).

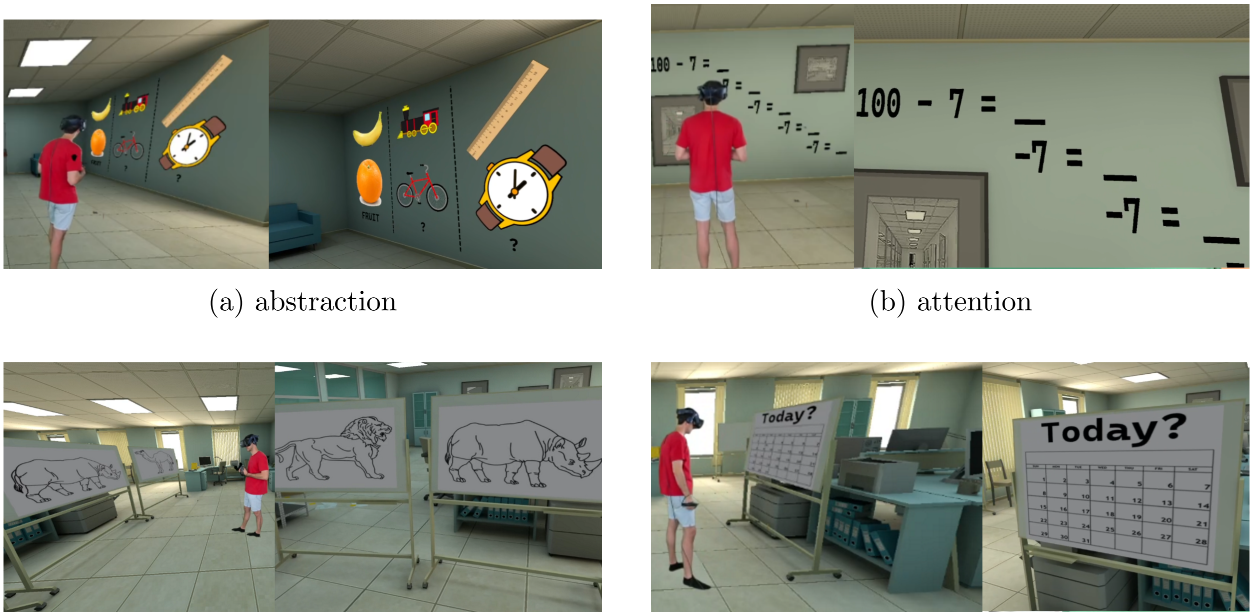
(b) Environment designed to reveal headset refresh rate via perceived motion differences.

Figure 6.10: Methods of attaining VR device metrics.

VR Device. We assume that privileged attackers I and II have intrinsic knowledge of the VR device specifications via direct API interaction. Fig. 6.10a shows how privileged attackers may use the observed update frequency of telemetry data to determine the polling rate of a target user's controller tracking. Further, Fig. 6.10b shows how even a non-privileged attacker can construct a virtual environment that replicates the "UFO test" [24], which users perceive differently depending on their devices' refresh rate (see puzzle 15 in Appendix 6.5). Currently, determining refresh rate, resolution, and field of view is sufficient to reveal the exact model of the VR device.

Host Device. Privileged attackers can also embed a variety of standardized benchmarks in their source code to assess the quality of the target user's host device (gaming computer). An attacker can use metrics such as CPU power, GPU power, and network bandwidth to reveal the age and price tier of the system and, thus, potentially correlate the spending power of the target user.

Acuity (MoCA)



(a) abstraction

(b) attention

(c) naming

(d) orientation

Figure 6.11: Methods of measuring cognitive acuity.

A number of standardized cognitive, diagnostic, and aptitude tests can be adapted for (and hidden within) VR environments. Fig. 6.11 illustrates VR environments designed to covertly asses four categories of the Montreal Cognitive Assessment (MoCA): abstraction (6.11a), attention (6.11b), naming (6.11c), and orientation (6.11d).

Demographics

Vocal Characteristics. Listening to the voice of a user may reveal key demographic attributes such as age, gender, and ethnicity [19, 92]. Shared VR environments with voice streaming provide a strong opportunity to exploit voice analysis, as attackers can cue target users to speak certain words or phrases that reveal more information, such as by requiring players to speak passwords aloud.

Language. There are a number of ways to ascertain a user's spoken language(s) in VR, including via speech recognition. Fig. 6.12 illustrates how a non-privileged attacker can observe a user's direction of gaze while solving a puzzle to reveal the languages they speak.

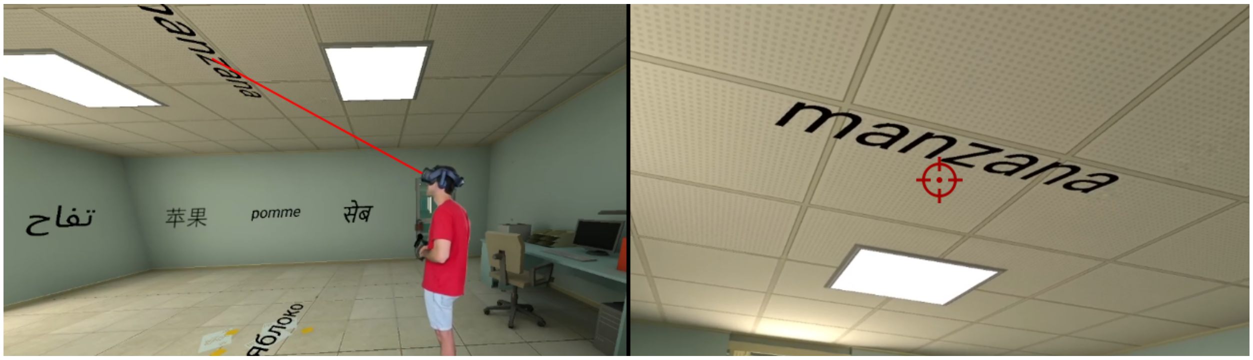


Figure 6.12: Determining language from user behavior.

Inferred Attributes. While most demographic attributes cannot be observed directly from VR data, attackers can often accurately infer them from primary data attributes. For example, height, wingspan, and IPD correlate strongly with gender, while eyesight, reaction time, and fitness correlate with age. While not possible to measure accurately in this study, we also suggest that in practice, information about room size, VR device type, and computing power could be used together to infer the income or wealth of a user.

Summary

Table 6.1 summarizes the VR privacy attacks presented in this section along with the VR device sensors or sources of information associated with each attack. The incredible volume of information exposed by a metaverse user, with at least 18 telemetry values collected 60 times per second or more, provides a vast amount of data from which adversarial inferences can be made. In all, we have identified dozens of unique data attributes, ranging from biometrics and demographics to behavioral and environmental measurements, that can be observed from users in VR via adversarial game design.

Of course, these attacks are by no means exhaustive, with many further attributes likely being observable that we have not discussed. Instead, our examples serve to illustrate the wide scope of observations available to VR adversaries and the ability to capture a comprehensive user attribute profile that would otherwise have involved aggregating data across several different devices.

Having discussed in great detail the theoretical information flow, adversaries, and plausible inferences of metaverse environments, the remainder of this chapter focuses on the experimental validation and quantification of these threats.

	Head			Left			Right			Device	Microphone	Behavior
	X	Y	Z	X	Y	Z	X	Y	Z			
Height		✓										
Left Arm	✓		✓	✓		✓						
Right Arm	✓		✓				✓		✓			
Longer Arm	✓		✓	✓		✓	✓		✓			
Handedness				✓	✓	✓	✓	✓	✓			
Wingspan				✓		✓	✓		✓			
Room Length	✓											
Room Width			✓									
Room Size	✓		✓									
IPD										✓		
Eyesight												✓
Color Blindness												✓
Locality											✓	
Device Refresh Rate												✓
Tracking Refresh Rate										✓		
Device Resolution										✓		
Device FOV										✓		
VR Device										✓		✓
Computing Power										✓		
Languages												✓
Physical Fitness		✓										
Reaction Time												✓
MOCA												✓
Gender		✓		✓		✓	✓		✓	✓	✓	
Age		✓										✓
Ethnicity		✓									✓	✓
Disability Status (Mental)												✓
Disability Status (Physical)		✓										
Identity	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 6.1: VR device sensors associated with each attack.

While a detailed description of the evaluated attacks is necessary for the completeness of this study, it is not our intention to focus on any particular attributes. Rather, our goal, as highlighted by the experimental design, is to generally demonstrate the extent to which adversarial game design can enable the collection of sensitive data attributes in VR.

6.4 Experimental Design

In this section, we describe "MetaData," a virtual reality "escape room" game designed as a case study for understanding how adversarial game design enhances attacker capabilities in virtual reality. The question we aim to answer is whether, and to what degree, an attacker can use data collected from consumer-grade VR devices to accurately extract and infer users' private information when aided by the capability to adversarially construct the virtual world and application rather than merely relying on passive observation.

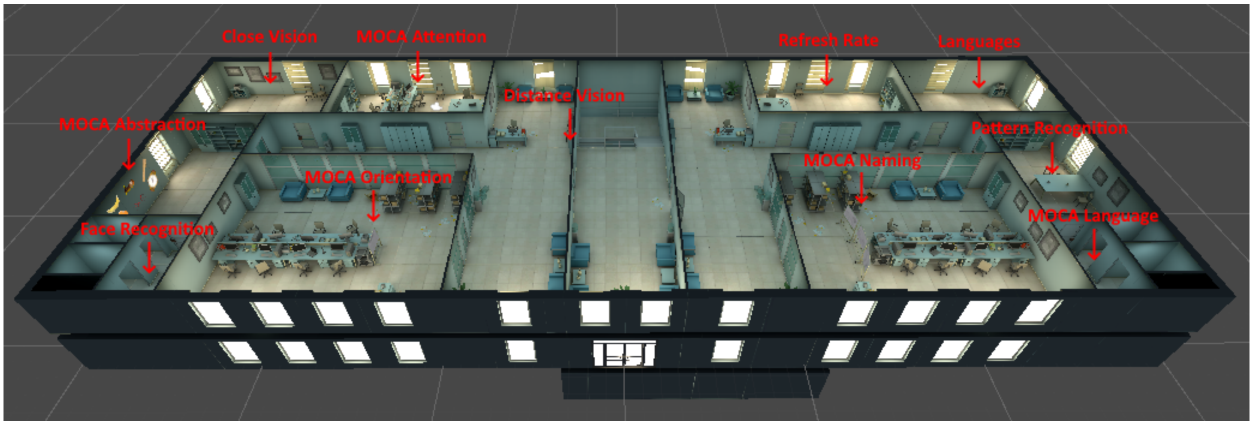


Figure 6.13: Virtual office building of the “MetaData” game, hosting the puzzle rooms.

This section details the experimental design, technical setup, and protocol used to answer this important question. After identifying the privacy-sensitive variables we believed to be accessible within VR (as detailed in §6.3), we implemented systematic methods to collect and analyze these variables from within VR applications.

To test the efficacy of these attacks, we designed an ”escape room”-style VR game themed as an office building (see Fig. 6.13). We then disguised the attacks as a set of puzzles within the game, which users were highly motivated to solve to the best of their ability in order to unlock a sequence of doors and win the game. We describe and illustrate the exact puzzles in detail in the following section (§6.5). We designed the experiment such that it did not bluntly reveal the ulterior goal, thereby illustrating how other VR applications could also accomplish the same goal covertly. To this end, we also added innocuous (i.e., “noisy”) rooms which did not necessarily collect meaningful personal information, but instead served to camouflage the data-harvesting puzzles.

Setup and Protocol

We recruited 50 individuals for the experiments (participant distribution given in §6.7). After completing a thorough informed consent and orientation process, we helped the participants don a VR headset (HTC Vive, Vive Pro 2, or Oculus Quest 2) and its hand-held controllers (Vive Controllers, Valve Index Controllers, or Oculus Quest Controllers, respectively), after which the participant proceeded to play the VR game (see the laboratory room layout in Fig. 6.14). Finally, the participants completed a post-game survey to collect the “ground truth” values for attributes of interest. The methods for collecting the true attribute values are summarized in §6.8.

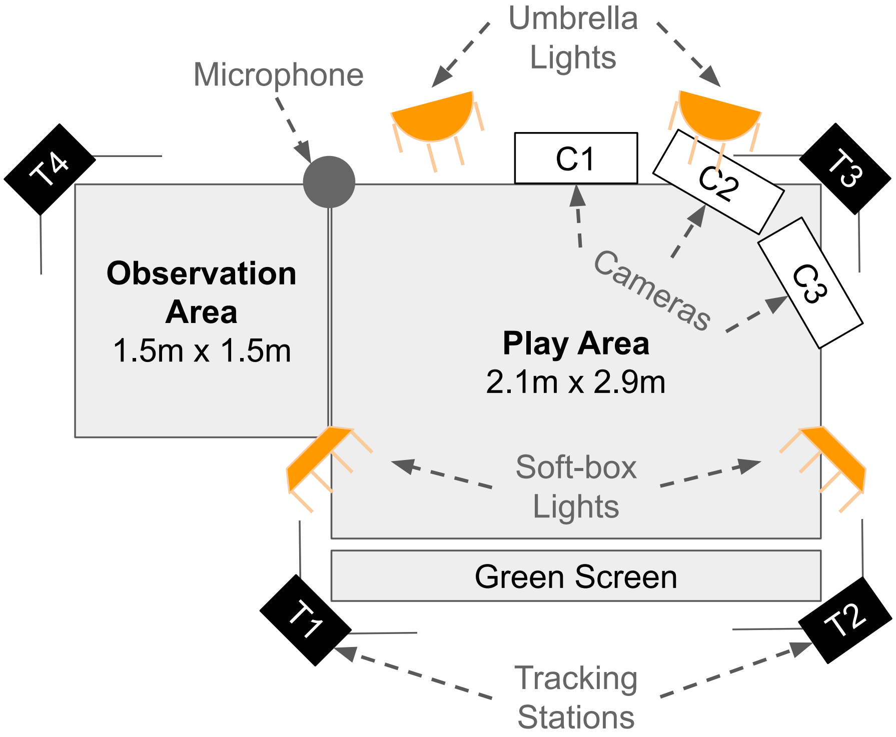


Figure 6.14: VR laboratory room layout.

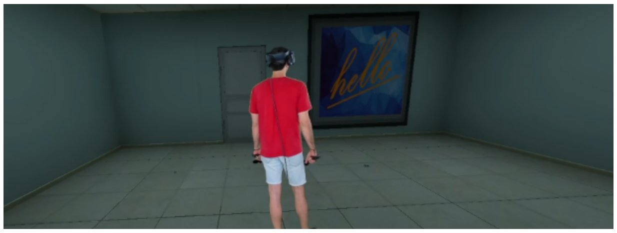
We tested three devices to determine if there were any noteworthy differences in the findings, which we did not observe other than in IPD (see §6.6), and to provide distinct classes for device identification. Each headset was paired with a gaming computer sufficiently powerful to run it at full fidelity; the main experimental setup had 64 GB of RAM, an AMD Ryzen 9 5950X CPU, and an Nvidia RTX 3090 GPU. To produce accurate results for room size and geolocation, we also conducted our experiment across four geographically distinct laboratories. Each experiment lasted approximately 10–20 minutes within VR, plus around 10 minutes for completing the survey. Throughout the experiments, we minimized the interactions with the participants and ensured their safety by intervening when they approached a wall in the room. The experiments remained the same for all participants; we did not alter the game play-through or logic. The game collected the targeted data points in CSV format during the play-through. Furthermore, the researchers manually annotated data points for data collection that required game development beyond what is reasonable for this study, e.g., automating voice recognition to register the escape room "passwords" (solutions) the participants articulated aloud. The researchers pressed keys on a keyboard to trigger animations in the virtual environment and teleport the player between rooms. These elements could be automated in a production-ready VR game.

Once the experiment ended, the participants filled out a form with their ground truth, which we used to validate the accuracy of the proposed privacy attacks. To collect the ground truth unknown to the participants themselves, we performed onsite measurements, e.g., we annotated the VR device and VR-room area, tested their reaction time with a desktop app, and measured their height and wingspan with a metric tape. Furthermore, knowing that researchers have studied the use of cognitive assessments in the diagnosis of attention disorders [231], autism [126], PTSD [159], and dementia [299], we chose the Montreal cognitive assessment (MoCA) [132] as a simple example of what advanced, immersive VR games could hide in their play-throughs. We randomized the order of the VR experiment and paper MoCA test (with half the participants taking the MoCA before and with the other half after the experiment) to neutralize potential biases in either direction. The exact method of collecting "ground truth" measurements for each attribute value is described in Appendix 6.8. Once we collected the ground truth, we ran our analysis scripts (privacy attacks) over the collected data to compile and infer data points, which we compared to the ground truth to assess the attacks' accuracy. The results of these experiments are described in §6.6.

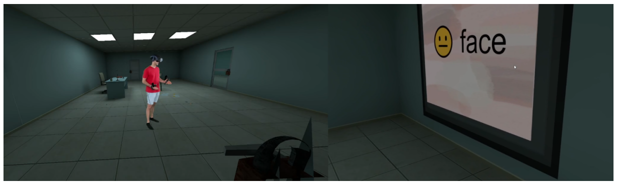
Ethics. We identified three primary ethical risks in our protocol: (i) the risk of discomfort using a VR device, (ii) the risk of a confidentiality breach of participant data, and (iii) the risk that participants might not have wished to disclose certain information about themselves during the course of the study. To address the first risk (i), we used high-fidelity VR devices and appropriately powerful gaming computers for all participants, together capable of consistently providing 120 frames per second, well above the minimum specifications recommended to mitigate the risk of VR sickness [245]. We designed our VR game to avoid distressing elements such as horror, claustrophobia, or flickering/strobing lights. Furthermore, a researcher was present to ensure participants did not collide with real-world objects during each play-through. To address the second risk (ii), we anonymized all collected data using random codes that we could not reasonably trace back to a participant's identity. Moreover, we avoided collecting any highly sensitive data that could potentially damage participants in a breach. Lastly, we normalized biometric measurements on a scale of 0 to 1 to avoid revealing exact measurements (e.g., in Fig. 6.15). The photos included in this chapter are not of actual participants. To address the third risk (iii), we made sure participants clearly understood the nature of the study. We emphasize that this is not a deception study. Our claims about the non-obviousness of the presented attacks should not be construed to imply that participants were unaware that their data was being collected during the study. Participants were informed that their data was being collected, including a description of the categories of data being observed. After completing the VR portion of the study, participants were made aware of the exact attributes being collected. They were explicitly given the opportunity to withdraw consent without penalty at any point in the process, including after having detailed knowledge of the data attributes involved, in which case their data would not have been included in the results. In light of these considerations, the study was deemed a minimal-risk behavioral intervention and was granted an IRB-exempt certification under 45 C.F.R. § 46.104(d)(3) by an OHRP-registered IRB.

6.5 Adversarial Puzzles

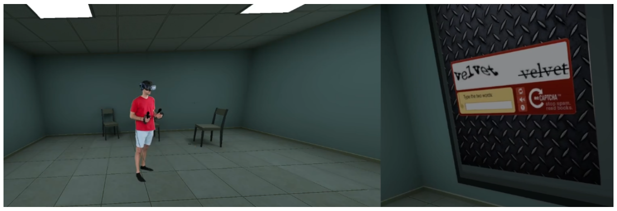
This section describes the experiment design in detail. Our experiment consists of puzzles located in VR rooms that the participants visit. The puzzles are artifacts that facilitate collecting privacy-sensitive variables that might not otherwise be evident. The rooms are themed as a virtual office. Before initiating the game, we explained to the participants that they would find the password by solving a puzzle, thereby "escaping" the room. As developing a full-fledged game with voice recognition or virtual password pads is out of scope, the participants spoke the passwords aloud so that the researchers could press a key and "teleport" them to the next room. We include five "noisy" rooms, i.e., rooms that do not serve the purpose of facilitating the measurement of sensitive information but help to mask the rooms that do. Nonetheless, noisy rooms habituate the player to the game mechanics, e.g., looking around the room or immersing the player further in the game. If the player gets stuck in one room, we press a key to teleport the participant to the next room. We request the users to remove their glasses or contact lenses for puzzles 23 and 24, measuring eyesight. While influencing players in such a way is not possible in a real scenario, these puzzles could at least identify the players who do not have good eyesight, i.e., they do not wear glasses/contacts when playing.



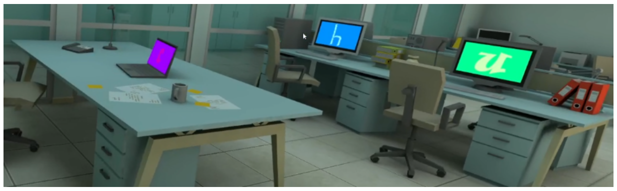
Puzzle 1: The first room introduces the player to the dynamics of the game, containing only a door and a poster with the word "*hello*", which is the password. Upon instinctively reading the word aloud, the player is teleported to the next room.



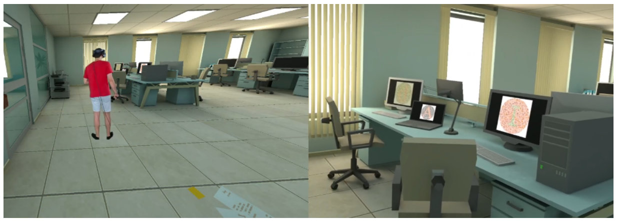
Puzzle 2: The second room contains a poster with the password "*face*". The player spawns facing the opposite wall of the poster; thus, we accustom the player to turn and explore the virtual environment to find the password and reinforce finding and speaking the password aloud.



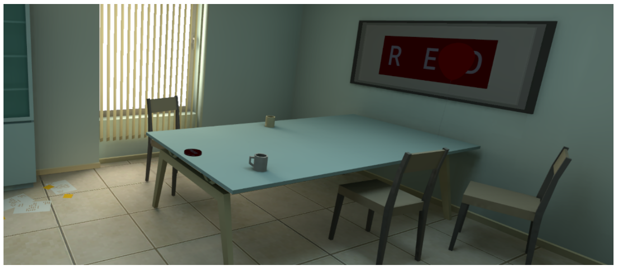
Puzzle 3: Similarly, a poster depicts a captcha with the word "*velvet.*"



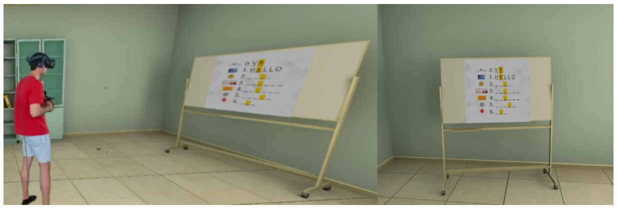
Puzzle 4: The room contains several tables with monitors, on whose screens are letters spelling "*church*" appropriately ordered from left to right.



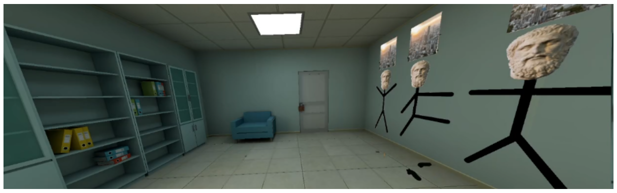
Puzzle 5: This room tests for color blindness. Similarly to puzzle 4, monitors display letters on Ishihara color test plates. Without colorblindness, the player would read "*daisy*"; with colorblindness, the player would read "*as*" instead. Each of these passwords unlocks the room.



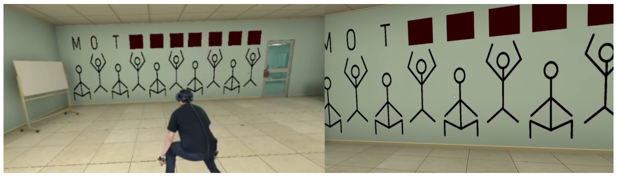
Puzzle 6: There is a button on a table; upon pressing it three times, the three balloons next to the opposite wall pop sequentially, revealing the password "*red*".



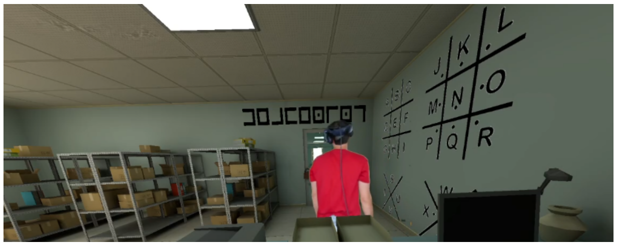
Puzzle 7: The puzzle tests the short-term memory of the participants (MoCA memory). A whiteboard displays seven rows arranged vertically, each with fill-in blanks. The first two rows contain the already filled-in words "*VR*" and "*hello*", respectively. The last five rows correspond to the previous passwords from puzzles 2 to 6. Connecting the highlighted letters sequentially from up to bottom, the participant reveals the password "*recluse*".



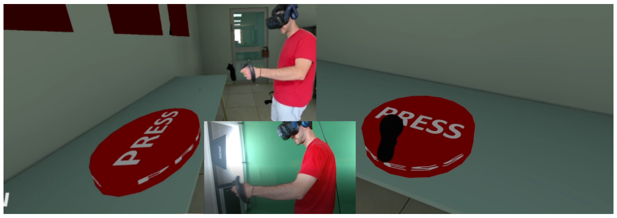
Puzzle 8: To measure wingspan, we depict on a wall four human stick figures with different poses. The participant must mimic the poses on the wall to uncover the four letters of the password "*cave*". One of the poses is a T-pose, which facilitates wingspan measurement.



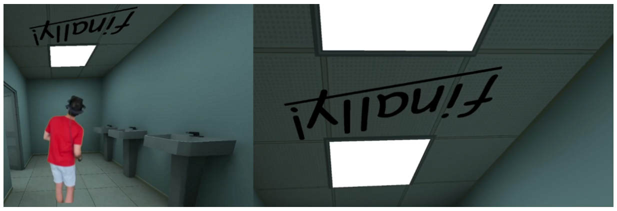
Puzzle 9: The participant must mimic the sequence of poses on the wall, a set of squats. For every squat, the participant uncovers two letters of the password "*motivation*". We correlate the distance traveled during the squats to fitness.



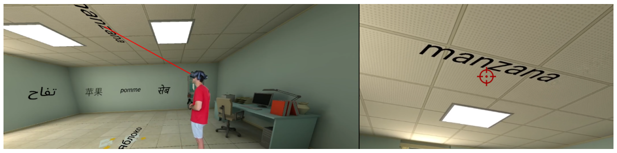
Puzzle 10: The (noisy) room depicts on a wall a pigpen cipher hiding the password "*deafening*".



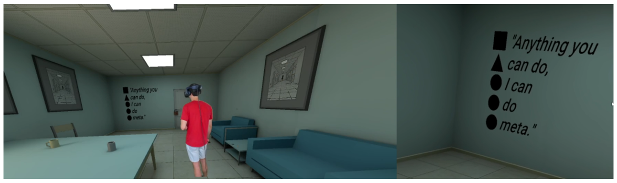
Puzzle 11: The player presses a button on a table in time with a visual input, revealing their reaction time.



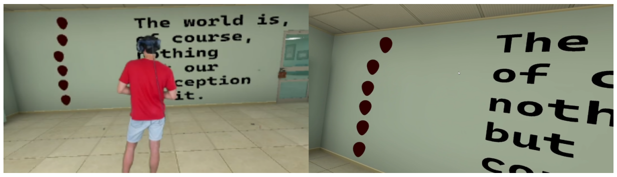
Puzzle 12: The room presents the password "*finally*" on the ceiling, habituating the user to also look upwards.



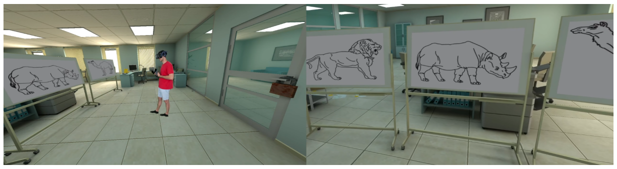
Puzzle 13: The room depicts the word "*apple*" in Hindi, Mandarin, French, Japanese, Russian, Spanish, Portuguese, and Arabic. The direction of gaze of the player when speaking the password reveals which language the participant recognizes.



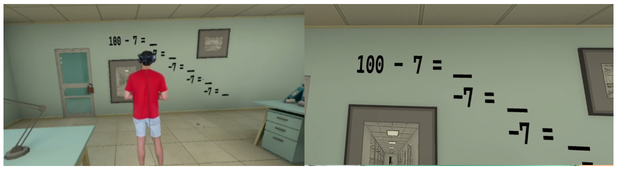
Puzzle 14: This room presents the sentence "*Everything you can do, I can do meta*" broken down vertically into five rows. To the left of each row, there is a shape. The last three shapes are the same (circles). To solve the puzzle, the participant must read aloud the words next to the first instance of the repeated shape "*I can.*"



Puzzle 15: Similarly to puzzle 14 and inspired by screen refresh rate tests [24], we present a number of balloons moving at different refresh rates. Depending on the refresh rate of the VR device, users cannot distinguish between some balloons.

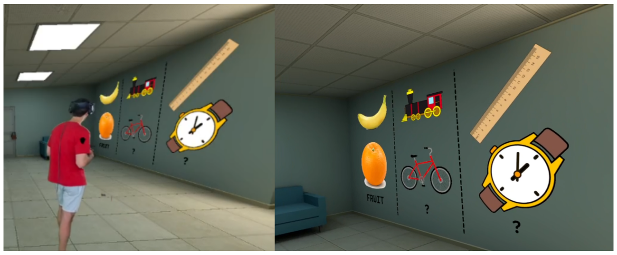


Puzzle 16: To deploy the "naming" MoCA task, the room presents three whiteboards depicting three animals.



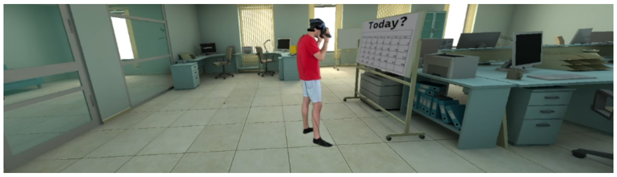
Puzzle 17: To measure an "attention" task from MoCA, we present a serial seven subtraction starting at 100, the password is the sequence of numbers that lead to the final answer: "65."

Puzzle 18: This room contains puzzle 7, thereby measuring delayed recall from the MoCA test.



Puzzle 19: This room tests MoCA abstraction.

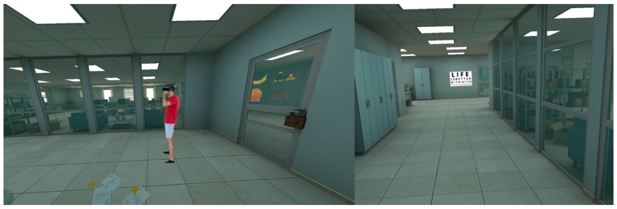
Puzzle 21: The (noisy) room depicts three pictures of a famous physicist—"*Albert Einstein*" is the password.



Puzzle 22: The room presents calendar days on a whiteboard with "*Today?*" as the header and without disclosing the year, month, weekday, or date, which prompts the participant to identify the date of the experiment, thereby measuring one variable of the orientation task in MoCA.

THE
CODE IS
EQUAL TO

Puzzle 23: We measure whether a participant can read the text at a close distance. We write the sentence "*The code is equal to three times four*" in four lines on the screen of a monitor, each line smaller than above.



Puzzle 24: Similarly, we measure whether a participant can read the sentence "*Life is better within the digital playground*" at a long distance.

Attribute	Type / Source	Precision	Accuracy	Statistics	Attackers
Height	Primary Telemetry	1 cm	76% within 5 cm 94% within 7 cm	$R^2 = 0.75$	Privileged I-III Non-Privileged*
Longer Arm	Primary Telemetry	boolean	58% for ≥ 2 cm difference 100% for ≥ 3 cm difference	$F_1 = 0.67$ $F_1 = 1.00$	Privileged I-III Non-Privileged*
Interpupillary Distance	Primary Telemetry	0.1 mm	96% within 0.5 mm (Vive Pro 2) 58% within 0.5 mm (All Devices)	$R^2 = 0.99$ $R^2 = 0.58$	Privileged I-II
Wingspan	Secondary Telemetry	1 cm	78% within 7 cm 98% within 12 cm	$R^2 = 0.68$	Privileged I-III Non-Privileged*
Room Size	Secondary Telemetry	1 m^2	70% within 2 m^2 96% within 3 m^2	$R^2 = 0.97$	Privileged I-III Non-Privileged*
Geolocation	Primary Network	100 km	50% within 400 km 94% within 500 km	N/A	Privileged II-III
HMD Refresh Rate	Primary Device	1 Hz	100% within 3 Hz (Privileged Attacker) 88% wtihin 60 Hz (Unprivileged Attacker)	$R^2 = 0.99$ $R^2 = 0.75$	Privileged I-II Privileged III* Non-Privileged*
Controller Tracking Rate	Primary Device	1 Hz	100% within 2.5 Hz	$R^2 = 0.99$	Privileged I-II Privileged III* Non-Privileged*
Device Resolution (MP)	Primary Device	0.1 MP	100% within 0.1 MP	$R^2 = 1.00$	Privileged I-II
Device FOV	Primary Device	10°	100% within 10°	$R^2 = 0.92$	Privileged I-II Privileged III* Non-Privileged*
Computational Power	Primary Device	0.1 GHz 10 Mh/s	CPU: 100% within 0.4 GHz GPU: 100% within 20 Mh/s	$R^2 = 0.92$ $R^2 = 0.81$	Privileged I-II
VR Device	Secondary Device	categorical	100%	$p = 0.00$	Privileged I-III Non-Privileged*
Handedness	Primary Behavior	boolean	96%	$F_1 = 0.98$	Privileged I-III Non-Privileged
Eyesight	Primary Behavior	boolean	72% (Hyperopia) 80% (Myopia)	$F_1 = 0.73$ $F_1 = 0.75$	Privileged I-III Non-Privileged
Color Blindness	Primary Behavior	boolean	100%	$F_1 = 1.00$	Privileged I-III Non-Privileged
Languages	Primary Behavior	boolean	90%	$p = 0.08$	Privileged I-III Non-Privileged
Physical Fitness	Primary Behavior	boolean	86%	$F_1 = 0.92$	Privileged I-III Non-Privileged
Reaction Time	Primary Behavior	categorical	88%	$F_1 = 0.90$	Privileged I-II Privileged III* Non-Privileged*
Acuity (MoCA)	Primary Behavior	1 point	94% within 2 points 100% diagnostic accuracy	$F_1 = 1.00$	Privileged I-III Non-Privileged
Gender	Inferred Classification	boolean	98%	$F_1 = 0.98$	Privileged I-III Non-Privileged
Age	Inferred Regression	1 yr	100% within 1.5 yr	$R^2 = 0.99$	Privileged I-III Non-Privileged
Ethnicity	Inferred Classification	categorical	98%	$p = 0.01$	Privileged I-III Non-Privileged
Disability Status	Inferred Classification	boolean	100%	$F_1 = 1.00$	Privileged I-III Non-Privileged
Identity	Inferred Classification	categorical	100%	$p = 0.00$	Privileged I-III Non-Privileged

* With degraded accuracy.

Table 6.2: Selected attributes collected and analyzed during the experiment, with accuracy and R^2, F_1, or p values from χ^2 tests.

6.6 Results

In this section, we present the empirical effectiveness of the privacy attacks introduced in §6.3, as summarized in Table 6.2.

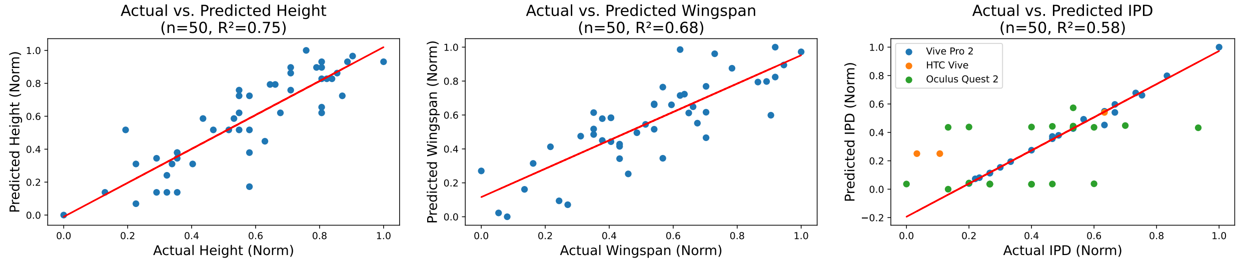


Figure 6.15: Actual and predicted user anthropometrics.

Biometrics

Continuous Anthropometrics. Fig. 6.15 shows (scaled) actual and predicted values for *height* ($R^2 = 0.75$), *wingspan* ($R^2 = 0.68$), and *interpupillary distance* (IPD) ($R^2 = 0.58$). IPD measurements were most accurate on the Vive Pro 2, with $R^2 = 0.99$ when excluding other devices. In general, we could accurately determine these three metrics for most users from just a few seconds of telemetry. We were not, however, able to accurately predict the individual lengths of the left and right arms ($R^2 = 0.02$ and $R^2 = 0.01$ respectively), due to the lack of a reliable center point from which to measure.

Binary Anthropometrics. Although absolute arm lengths were not discernible, relative lengths were accurate enough that we could usually identify which of the participant's arms was longer. We observed increasing accuracy for participants with greater differences in length, reaching 100% accuracy for the 12% of participants with a difference of at least 3 cm. Handedness can also be determined accurately from behavioral observations; we note, however, that 94% of our participants reported being right-handed.

Vision. Our vision tests achieved diagnostic accuracies for *hyperopia* (farsightedness), *myopia* (nearsightedness), and *deuteranopia* (red-green color blindness) of 72%, 80%, and 100% respectively. The overall accuracy of detecting a visual deficiency was 80%, in part because some users of contact lenses could not remove their contacts for the experiment.

Fitness. Using squat depth as a correlate of *physical fitness* identified "low" fitness with an accuracy of 86%; our method was not able to distinguish "moderate" and "high" fitness.

Reaction Time. We measured *reaction time* to a precision of one recorded frame (16.6 ms). We were able to detect whether a participant's reaction time was above or below 250 ms (the approximate median reaction time) with an accuracy of 88%.

Environment

Room Size. The *length* and *width* of each of three testing rooms was determined to within 1.0 m with accuracies of nearly 90%. This allowed true room area to be found within 3 m^2 in 96% of trials. Taking the average estimated area for each tested room vs. the true accessible room area yields $R^2 = 0.97$.

Geolocation. Using the server latency multilateration (hyperbolic positioning) technique for *geolocation* yielded a mean longitudinal and latitudinal error of around 2.5° across four tested locations. This was sufficient to locate the test subject to within 500 km in 94% of cases, and within the correct state in 100% of cases.

Device Specifications

Tracking Rate. We found that privileged attackers could determine various VR device specifications (namely, *display refresh rate*, *display resolution*, *field of view*, and *tracking rate*) with high accuracy. Tracking rate (the number of unique telemetry measurements taken per second) is a particularly interesting metric, as the top four VR headsets, together accounting for over 75% market share [82], all have different default HMD refresh rates (72/144/80/90 Hz).

VR Device. Using the above device specifications and the highly heterogeneous nature of VR device specifications, privileged attackers can determine the type of VR device with 100% accuracy. We also found that non-privileged attackers could determine the refresh rate to within 30 Hz with an accuracy of 30% and to within 60 Hz with an accuracy of 88%; however, this was not sufficient to accurately determine the type of device.

Host Device. We found that an attacker benchmarking host device specifications can determine *GPU power* with 100% accuracy to within 20 Mh/s (daggerhashimoto) and *CPU clock speed* to within 0.4 GHz, allowing them to estimate the price tier of the host device.

Acuity (MoCA)

Table 6.3 summarizes the numerical (continuous, i.e., the score of each category) and diagnostic (binary, i.e., passing or failing a category) accuracy of the *Montreal Cognitive Assessment* (MoCA) we conducted in the VR experiments. We achieved a diagnostic accuracy of 90% or greater for 5 of the 7 scored MoCA categories (excluding visuospatial/executive and delayed recall), with an overall diagnostic accuracy of 100%.

MoCA Category	**Accuracy (Numerical)**	**Accuracy (Diagnostic)**
Executive	N/A	N/A
Naming	100%	100%
Memory	78%	84%
Serial 7	90%	100%
Attention	88%	100%
Repetition	74%	96%
Language	74%	96%
Abstraction	100%	100%
Recall	60%	90%
Orientation	100%	100%
Overall	80% within 1 point 94% within 2 points	100%

Table 6.3: Accuracy of each MoCA category.

Demographics

Language. The visual focus method of *language* determination correctly identified a spoken language (other than English) in 90% of multilingual participants.

Vocal Characteristics. We used existing machine learning models to determine the gender [19] and ethnicity [92] of participants from their voice with an accuracy of 98% and 66% respectively; these accuracy values improved to 100% when combined with other attributes such as height and wingspan as described in "Inferred Attributes" below.

Inferred Attributes. We used Azure Automated Machine Learning [174] to determine the optimal preprocessor, model architecture, and input metrics for inferring several demographic attributes. Table 6.4 summarizes the results of this meta-analysis. For identity, we used the best-performing technique of Miller et al. [178]. Using the identified optimized models and parameters, we determined the participant's gender, ethnicity, disability status, age (within 1.5 years), and identity with nearly 100% accuracy across several Monte Carlo cross-validations; importantly, users were never simultaneously present in the training and testing datasets, other than for inferring identity, and it is not possible that the demographic inferences were a result of simply identifying users.

With respect to disability, there was one reported physical disability and three reported mental disabilities amongst our 50 participants; we were able to identify these disabilities individually with 100% accuracy ($F_1 = 1.00$). In each case, the model far outperformed any individual attribute; for example, ethnicity was 98% accurate despite its most significant input (voice) being only 66% accurate on its own.

Attribute (Prediction)	**Inputs**	**Preprocessing / Model**
Gender (Classification)	Voice, Height, Wingspan, Interpupillary Distance (IPD)	TruncatedSVDWrapper SVM
Age (Regression)	Close Vision, Reaction Time, Height, Test Duration, Acuity	MaxAbsScaler ExtremeRandomTrees
Ethnicity (Classification)	Voice, Language, Height	StandardScalerWrapper LightGBM
Disabilities (Classification)	Vision, Fitness, Acuity	MaxAbsScaler NaiveBayes
Identity (Classification)	Height, Wingspan, Acuity, IPD, Vision, Reaction Time	StandardScalerWrapper RandomForest

Table 6.4: Inputs and methodology of inferred attributes.

By following the Azure Automated ML Workflow [174], we avoided the biases of manually selecting features for demographic inference.

The AutoML workflow begins by clustering users into training, validation, and testing sets. Using the validation set, a variety of preprocessing techniques are first evaluated. Next, a variety of classical machine learning models are trained using the preprocessed features, using the validation set to evaluate the accuracy of each algorithm. For the most accurate architecture, a hyperparameter sweep is performed to optimize the model for the feature set.

Using the best-performing model, an explainability analysis is conducted to select the most important input features for inferring each attribute. For example, rather than instructing the model to infer gender from voice, height, wingspan, etc., we initially provided the system with all available primary attributes and allowed it to determine on its own which are relevant to a given inference task. Finally, the testing set is used to evaluate the selected approach, consisting of automatically determined features, preprocessing, architecture, and hyperparameters. The process is repeated across several Monte Carlo cross-validations

6.7 Participant Distribution

Demographics

Gender	**50**
Male	26 (52.0%)
Female	24 (48.0%)

Age	**50**
18–23	24 (48.0%)
24–27	20 (40.0%)
28–64	6 (12.0%)

Nationality	**50**
American	23 (46.0%)
Chinese	8 (16.0%)
Indian	6 (12.0%)
German	3 (6.0%)
Canadian	2 (4.0%)
Brazilian	1 (2.0%)
British	1 (2.0%)
Portuguese	1 (2.0%)
Spanish	1 (2.0%)
Swiss	1 (2.0%)
Undisclosed	3 (6.0%)

Ethnicity	**50**
Asian	30 (60.0%)
White	14 (30.0%)
Black	3 (6.0%)
Hispanic	3 (6.0%)

Income	**50**
≤ $25k	20 (40.0%)
$25k–$50k	15 (30.0%)
$50k–$100k	7 (14.0%)
≥ $100k	3 (6.0%)
Undisclosed	5 (10.0%)

Disability Status	**50**
None	46 (92.0%)
Mental	3 (6.0%)
Physical	1 (2.0%)

Languages	**50**
Chinese	20 (40.0%)
Spanish	14 (28.0%)
French	13 (26.0%)
Hindi	7 (14.0%)
None	6 (12.0%)
Portuguese	2 (4.0%)
Arabic	1 (2.0%)

Biometrics

Height	**50**
150 cm – 165 cm	18 (36.0%)
166 cm – 175 cm	16 (32.0%)
176 cm – 189 cm	16 (32.0%)

Wingspan	**50**
100 cm – 169 cm	21 (42.0%)
170 cm – 179 cm	18 (36.0%)
180 cm – 191 cm	11 (22.0%)

Longer Arm	**50**
Left	26 (52.0%)
Right	18 (36.0%)
Same	6 (12.0%)

Reaction Time	**50**
> 250 ms	27 (54.0%)
< 250 ms	23 (46.0%)

IPD	**50**
< 63 mm	26 (52.0%)
63 mm – 66 mm	21 (41.0%)
> 66 mm	3 (6.0%)

Fitness	**50**
Moderate	32 (64.0%)
High	10 (20.0%)
Low	8 (16.0%)

Colorblindness	**50**
None	48 (96.0%)
Deuteranopia	2 (4.0%)

Hyperopia	**50**
None	28 (56.0%)
Severe	13 (26.0%)
Mild	9 (18.0%)

Myopia	**50**
Severe	32 (66.0%)
None	14 (28.0%)
Mild	4 (8.0%)

MoCA	**50**
Pass (> 26)	43 (86.0%)
Fail (≤ 26)	7 (14.0%)

Handedness	**50**
Right	47 (94.0%)
Left	3 (6.0%)

Environment

Location	**50**
Location A	26 (52.0%)
Location B	20 (40.0%)
Location C	4 (8.0%)

Room Size	**50**
$5\ m^2$–$7\ m^2$	24 (48.0%)
$> 8\ m^2$	20 (40.0%)
$< 5\ m^2$	4 (8.0%)

Duration	**50**
≤ 15 min	22 (44.0%)
16 min–20 min	18 (36.0%)
20 min–30 min	10 (20.0%)

Device	**50**
Vive Pro 2	26 (52.0%)
Oculus Quest 2	21 (42.0%)
HTC Vive	3 (6.0%)

6.8 Sources of Ground Truth

Gender	Self-Reported
Age	Self-Reported
Nationality	Self-Reported
Ethnicity	Self-Reported
Income	Self-Reported
Disability Status	Self-Reported
Languages	Self-Reported
Height	Stadiometer
Wingspan	Measuring Tape
Longer Arm	Measuring Tape
Reaction Time	Application
IPD	Pupilometer
Fitness	Self-Reported
Colorblindness	Self-Reported
Hyperopia	Self-Reported
Myopia	Self-Reported
MoCA	Administered
Handedness	Self-Reported
Location	GPS
Room Size	Measuring Tape
Duration	Chronometer
Device	Observed

6.9 Discussion

In Chapter 5, we demonstrated that a passive attacker can use data collected from consumer-grade VR devices to accurately infer users' private information. In this study, we have shown active attackers can do so with even greater fidelity, with moderate to high accuracy values for most of the aggregated and inferred data points presented in Table 6.2. While we were required to condense our attack into a concise 20-minute experiment for logistical reasons, real-world adversaries could gain increased accuracy and covertness by integrating our identification method of Chapter 4 to aggregate data collected over longer time periods.

Participant Awareness

In sections 6.3 and 6.4, we argued that a developer could design VR environments and games to facilitate the covert collection of targeted data points disguised as normal game elements. For ethical reasons, our participants were informed that they were participating in an adversarial experiment where their personal information would be collected. However, they were not told exactly which attributes were being collected until the end of the experiment.

Our experiment reflects a realistic scenario in which users may be generally aware of privacy threats but would not necessarily know the attributes collected by a given application. Upon debriefing, we asked the participants whether they could identify which of the attributes we were attempting to measure. All 50 participants reported not knowing exactly which attributes were being collected and inferred during the game, with no participant able to correctly identify more than 3 of the attributes.

When we revealed the list of attributes, many participants expressed surprise at the breadth of information that could be collected within VR, but none expressed particular shock at the existence of some degree of data harvesting (perhaps having already grown accustomed to these practices on the web). Even if participants were aware of the attributes being measured, we believe it would have been difficult to counteract many of the attacks, due to the deeply subconscious nature of many of the observed behaviors.

Adversarial Capabilities

In §6.3, we provided wingspan as an initial example of an attribute that is far easier to observe in an adversarial application; while users rarely adopt a posture in which their arms are completely outstretched, an adversarial game can easily drive users to adopt such a posture through the use of an interactive element.

Reflecting now on the entire set of attributes inferred in this study, we observe that nearly all of them are aided by the introduction of adversarial puzzles in our escape room. For example, most behavioral observations, including the entire MoCA acuity assessment, relied on observing the user's responses to specific adversarially-constructed puzzles.

Our major conclusion from this finding is that the privacy risk of VR devices stems not only from their sensors, such as accelerometers and gyroscopes, but also from the immersive nature of their displays, which can be used to totally control a user's virtual environment to influence the information they reveal. Thus, while this dissertation is primarily focused on the capability to infer various attributes from motion-tracking data, it is equally important to consider the threat posed by adversarial VR application design.

Societal Implications

While for ethical reasons we limited our attacks to relatively benign data points, an attacker could potentially track and infer additional information about other more critically sensitive personality traits, like sexual, religious, or political orientation, educational level, and illnesses, among others, to enhance practices such as surveillance advertisement [46] or pushing political agendas [206]. Given how immersive and emotionally engaging VR environments can be [199, 139, 296, 155], such practices could become more pernicious and effective than with current mobile and desktop applications. By combining profiling techniques with identification methods, an attacker, or even a group of colluding attackers, could attempt to aggregate user profiles from data across many VR sessions.

Limitations

The results of this study should be understood in light of a few limitations. Unfortunately, our sample of participants was not perfectly representative of the general population; for example, college students were overrepresented. For logistical reasons, we could not modify VR device firmware and thus could not consider hardware-level attacks in this dissertation. Therefore, privileged attackers I and II, while different in theory, had identical capabilities within the scope of our experiments. While ground truth values for user attributes were measured by the researchers when possible, many were also self-reported (see §6.8) and thus potentially biased. Lastly, the researchers were forced to interact with participants outside of VR on some occasions, such as to warn of nearby obstacles. While we did attempt to minimize such occurrences, these interactions could nevertheless have biased certain results.

6.10 Conclusion

In this chapter, we shed more light on the serious privacy risks of the metaverse by showing how VR can be turned against its users. Specifically, we demonstrated the practicality and accuracy of active adversarial attacks by designing and conducting experiments with 50 participants using consumer-grade VR devices. The participants played our "escape room" VR game, which was secretly designed to collect personal information, like biometrics, demographics, and VR device and network details, among numerous other data points. The results demonstrate high information leakage with moderate to high accuracy values over most identified vulnerable attributes, with just a handful of these attributes being sufficient to uniquely identify a user [232, 264, 197, 91, 144, 75, 265, 13].

The alarming accuracy and covertness of these attacks and the push of data-hungry companies towards metaverse technologies indicate that data collection and inference practices in VR environments may soon become more pervasive in our daily lives. Furthermore, the breadth of possible VR applications, increasing quality of VR devices, and relative simplicity of our demonstration, all suggest that more sophisticated attacks with a higher success rate are possible and perhaps on the horizon. Therefore, the remainder of this dissertation investigates privacy-preserving technologies for VR, and in particular, proposes countermeasures for new and existing privacy attacks in the metaverse.

Chapter 7

Going Incognito in the Metaverse and Achieving Theoretically Optimal Privacy-Usability Tradeoffs in Virtual Reality

7.1 Introduction

In the first several chapters of this dissertation, we have painted a striking picture of the security and privacy consequences of VR motion data. Specifically, we have shown that seemingly-anonymous VR users can easily and accurately be deanonymized (Chapter 4) and profiled (Chapter 5) from just a few minutes of tracking data, and that these threats become even more dangerous when adversarial VR applications are involved (Chapter 6). Despite these risks, users are currently less broadly aware of security and privacy risks in VR than they are of similar risks in traditional platforms like social media [45, 87].

Of course, data privacy challenges are not unique to VR. For example, on the web, browser cookies pose a widely understood risk to privacy by attaching identifiers and tracking users across websites [32]. However, the maturation of web technologies has also brought an enhanced understanding of, and countermeasures to, such attacks, with technologies private browsing (or "incognito") mode in browsers providing users with vital defensive tools for reclaiming control of their data. By contrast, equivalent comprehensive privacy defenses have yet to be developed for the metaverse. We thus find ourselves now in the dangerous situation of facing unprecedented privacy threats in VR while lacking the defensive resources we have become accustomed to on the web.

In this chapter, we aim to begin addressing this disparity by designing and implementing the first "incognito mode" for VR. Our method leverages local ε-differential privacy to provide quantifiable resilience against known VR privacy attacks according to a user-adjustable privacy parameter ε. In doing so, it allows for inherent privacy and usability trade-offs to be dynamically rebalanced, along a theoretically optimal continuum, according to the risks and requirements of each VR application, with a focus on the targeted addition of noise to those parameters that are most vulnerable. We provide an open-source implementation of our solution as a Unity plugin, which we then use to replicate three existing VR privacy attack studies. Our results show a significant degradation of attacker capabilities when using our extension. Finally, we provide statistical bounds for the perceived error that users may experience when using our technique. These bounds are well within the range that VR users can naturally adapt to according to past research on homuncular flexibility in VR [304].

7.2 Method

VR Adversaries

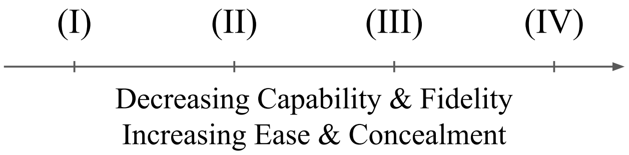

(I)
(II)
(III)
(IV)
Decreasing Capability & Fidelity
Increasing Ease & Concealment

Figure 7.1: VR privacy adversary model (see Chapter 2).

In this chapter, we present algorithmic defenses for vulnerable attributes that can be implemented at either the device firmware or client software level. Tab. 7.1 shows the attackers covered by each implementation possibility. In practice, lacking any special access to VR device firmware, our evaluated systems were all implemented at the software level.

	Attackers			
	I	**II**	**III**	**IV**
Software Incognito			✓	✓
Firmware Incognito		✓	✓	✓

Table 7.1: Coverage of proposed defenses.

Overall, the "VR incognito mode" defenses proposed in this chapter are unable to address the threat of hardware and firmware attackers. We argue that this is a necessary concession of a software-based defense, and that unlike the client, server, and user attackers we cover, hardware and firmware attacks can be discovered via reverse engineering. Still, in an ideal world, VR devices would contain hardware-based mechanisms for ensuring user privacy.

Private Web Browsing

We now detour briefly to the more mature field of private web browsing to seek inspiration from the web privacy solutions which have stood the test of time.

The research community has surveyed the field of web privacy [184, 278], and identified observable attributes ranging from tracking cookies [32] and HTTP headers [150] to browsing histories [153] and motion sensor data [310]. As in VR, these attributes can be combined to achieve profiling [78, 98], fingerprinting [150] and deanonymization [311]. Further, the attack model used by web privacy researchers resembles the metaverse threat model presented in Chapter 2, with most defenses focusing on web servers and other users, some on client-side applications, and relatively few on the underlying hardware.

In response to these threats, proposed solutions have included proxies, VPNs [133], Tor [63, 135], and, of course, private browsing or "incognito" mode in browsers, as well as dedicated private browsers and search engines, e.g., Brave [28] and DuckDuckGo [68]. Of these solutions, "incognito mode" stands out due to its ease of use: a wide range of defensive modifications to protocols, APIs, cookies, and browsing history can all be deployed with a single click [5]. Due perhaps to this outward simplicity, surveys of web privacy protections used in practice have found private browsing mode to be by far the most popular [108].

In summary, web privacy is highly analogous to metaverse privacy; although the data attributes being protected are vastly different, the threat of combining attributes to profile and deanonymize users is a constant, as is the threat model used to characterize both fields. On the other hand, the size and scope of data collection in VR potentially exceed that of the web [193], while users are simultaneously less aware of the threat in VR [164], and the equivalent privacy tools are not generally available. We are motivated by the popularity of incognito mode on the web to seek an equivalent for VR, with the same fundamental goal as in browsers: allowing users, at the flick of a switch, to become harder to link across sessions.

Differential Privacy

Having established our motivation for pursuing a metaverse equivalent to "incognito mode," we now lay out the tools necessary to enable its realization. Chief among these is differential privacy [74], which provides a context-agnostic mathematical definition of privacy that statistically bounds the information gained by a hypothetical adversary from the output of a given function $\mathcal{M}(\cdot)$:

Definition 1. *(ε-Differential Privacy [73]). A randomized function $\mathcal{M}(\cdot)$ is ε-differentially private if for all input datasets D and D' differing on at most one element, and for all possible outputs $\mathcal{S} \subseteq$ Range$(\mathcal{M})$:* $\Pr[\mathcal{M}(D) \in \mathcal{S}] \leq e^{\varepsilon} \times \Pr[\mathcal{M}(D') \in \mathcal{S}]$.

A function $\mathcal{M}(\cdot)$ fulfills differential privacy if its outputs with and without the presence of an individual input element are indistinguishable with respect to the privacy parameter $\varepsilon \geq 0$. In practice, a randomized function $\mathcal{M}(\cdot)$ typically ensures differential privacy by adding calibrated random noise to the output of a deterministic function, $\mathcal{M}(x) = f(x) + \text{Noise}$. Lower ε values correspond to higher noise, making it harder to distinguish outputs and strengthening the privacy protection. In addition to ε, the required noise is affected by the *sensitivity* (Δ) of the deterministic function, which quantifies the maximum difference between a function's outputs between D and D'.

Another aspect worth highlighting is *sequential composition* [73]: if $\mathcal{M}(\cdot)$ is computed n times over D with ε_i, the total *privacy budget* consumed is $\sum \varepsilon_i$. Thus, users' attributes become less protected with every query execution. Differentially private outputs are also *immune to post-processing* [73]; an adversary can compute any function on the output (e.g., rounding) without reducing privacy.

In practice, differential privacy can be used *centrally*, whereby a server adds noise to an aggregation function computed over data from multiple clients, or *locally*, whereby clients add noise to data points before sharing them with a server. While local differential privacy is noisier than the central variant, it also requires less trust of the server. Since servers are considered potential adversaries in our threat model (§7.2), we use local differential privacy to protect VR users in this chapter. Specifically, we implement local differential privacy using the Bounded Laplace Mechanism [115, 73] for continuous attributes and randomized response [295] for Boolean attributes.

Bounded Laplace Mechanism. The Laplace mechanism [73], also known as the "workhorse of differential privacy," [115] is a popular method of implementing local differential privacy for continuous attributes. Laplacian noise satisfies a stronger notion of ε-differential privacy than Gaussian noise, which only satisfies a weaker (ε, δ)-differential privacy [313]. However, its unbounded noise can yield semantically absurd edge cases (e.g., a negative value for the height attribute). Thus, in this chapter, we use the Bounded Laplace mechanism [115], which transforms the noise distribution according to the privacy parameters and deterministic value, then samples outputs until a value falls within pre-determined bounds without compromising differential privacy. Inputs that fall outside the bounds are automatically clamped to the nearest bound. Additionally, we employ the modified sampling technique of Holohan et. al [114] to avoid a known vulnerability associated with the use of finite floating-point in other differential privacy implementations [183].

Randomized Response. To achieve local differential privacy for Boolean attributes, we can apply the randomized response method from Warner [295]: (i) the client flips a coin, (ii) if heads, the client sends a truthful response, (iii) else, the client flips a coin again and sends "true" if heads and "false" if tails. This method has been shown to be ($\varepsilon = \ln 3$)-differentially private with a fair coin [73], though one can vary ε by changing the bias of the first coin flip.

Homuncular Flexibility

While differential privacy can be used to quantifiably address the problem of data leakage from VR telemetry, it does so by introducing noise to the VR data, thus potentially degrading the user experience. However, past research on "homuncular flexibility" has shown that users can learn to control bodies that are different from their own, particularly in virtual reality [304, 1]. Thus, the remainder of this work focuses on deploying differential privacy in VR in a way that users can rapidly learn to ignore. By transforming the virtual object hierarchy according to known usable non-linear interaction techniques (e.g., the Go-Go technique [218]), the corresponding attributes (e.g., wingspan) can be obscured while allowing users to flexibly adapt to their new environment.

7.3 VR Privacy Defenses

In this section, we provide a differentially-private framework for user data attribute protection in VR. We define each attribute defense in terms of abstract coordinate transformations, without regard to any specific method of implementation. Later, in §7.4, we describe a concrete system for implementing these defenses within VR applications via a Unity plugin.

Our "incognito mode" defenses aim to prevent adversaries from tracking VR users across sessions in the metaverse. In practice, this means limiting the number of data attributes adversaries can reliably harvest from users and use to infer their identity. Local differential privacy (LDP) is the primary tool that allows us to achieve this with a mathematically quantifiable degree of privacy. LDP has the effect of significantly widening the range of attribute values observed by an adversary given a particular ground truth attribute value of a user. In doing so, it ensures that the observable attribute profile of a user always significantly overlaps with that of at least several other users, thus making a precise determination of identity infeasible. The noise added by LDP may have some negative impacts on user experience, as is the case with incognito mode in browsers. However, users can tune the privacy parameter (ε) to reduce the impact of noise on user experience as required.

Upon initiating a new metaverse session (i.e., connecting to a VR server), the defenses generate a random set of "offset" values, which are then used throughout the session to obfuscate attributes within the VR telemetry data stream through a set of deterministic coordinate transformations. The re-randomization of offset values at the start of each session ensures that all usage sessions of a user are statistically unlinkable.[1] On the other hand, these offsets remain consistent within a session to ensure adversaries never receive more than one view of sensitive attribute values.

[1] Methods for tracking users that are not unique to VR (such as via their IP addresses) are not considered to be within the scope of this dissertation; corresponding defenses like VPNs are already widespread.

What follows are the specific differentially-private coordinate transformations that protect user data attributes (and thus allow them to "go incognito") in VR. While for simplicity this section considers the protections for each attribute in isolation, in practice, our implementation uses a relative transformation hierarchy to allow any set of enabled defenses to seamlessly combine with each other (see §7.4). The coordinates used throughout this chapter refer to the left-handed, Y-up Unity coordinate system, pictured in Fig. 7.2.

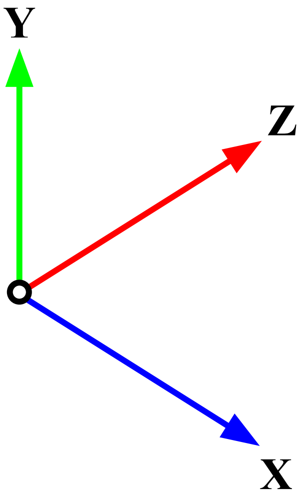


Figure 7.2: Left-handed, Y-up Unity 3D coordinate system.

Preliminaries

In our setting, local differential privacy protects against adversaries with knowledge of observed attributes across all user sessions except for the current session of a target user (D'). Sequential composition allows us to provide an upper bound for a user's privacy budget as the sum of each ε value used per attribute.

We identified the Bounded Laplace mechanism [115] as our tool of choice for protecting continuous attributes like *height*, *wingspan*, and *room size* in VR because it produces random noise centered around the sensitive value (e.g., *height*) while preserving the semantic consistency of the attribute (e.g., *height* > 0). The Laplacian noise distribution is preferable over, e.g., simply imbuing uniformly distributed random noise, because it has the property of minimizing the mean-squared error of any attribute at a given privacy level (ε) [145], thereby minimizing its impact the user experience by this metric.

Where Boolean attributes are concerned, we use randomized response [295] with a weighted coin to provide ε-differential privacy for chosen values of ε. The use of randomized response over simpler mechanisms (e.g., a single coin flip) aligns Boolean attributes with the same ε-differential privacy framework as continuous attributes, and thus allows the ε values of multiple attributes to be combined into a single "privacy budget" if desired.

Throughout this chapter, we use the following standard variable notation in our algorithms:

- v: sensitive deterministic value ("ground truth")
- (l_v, u_v): population bounds of v
- $\varepsilon \geq 0$: differential privacy parameter
- p: randomized response coin bias
- (x_h, y_h, z_h): headset coordinates
- (x_r, y_r, z_r): right controller coordinates
- (x_l, y_l, z_l): left controller coordinates

For a given attribute a (e.g., *height*), we use a' (e.g., *height'*) to denote the LDP-protected value an adversary observes. Our use of local differential privacy requires Δ to cover the entire range of the bounded interval $[l, u]$ ($\Delta = |u - l|$). Alg. 1 contains helper functions for the mechanisms discussed here that will be used throughout §7.3.

Algorithm 1: Preliminaries for privacy defenses.

```
Function LDPNoisyOffset($v$, $\varepsilon$, $l_v$, $u_v$):
    return BoundedLaplacianNoise($v$, $|u_v - l_v|$, $\varepsilon$, $l_v$, $u_v$)
Function RandomizedResponse($v$, $p$):
    if $Random(0,1) \leq p$ then
        return $v$
    else
        return $Random(0,1) \leq 0.5$
Function PolarTransform($x_r, z_r, x_l, z_l$):
    $\vec{d_r} = \langle x_r, z_r \rangle - \langle \frac{x_r + x_l}{2}, \frac{z_r + z_l}{2} \rangle$
    $\vec{d_l} = \langle x_l, z_l \rangle - \langle \frac{x_r + x_l}{2}, \frac{z_r + z_l}{2} \rangle$
    $d_r, d_l = |\vec{d_r}|, |\vec{d_l}|$
    $\alpha_r, \alpha_l = \text{ArcTan}(\vec{d}_{r_x}, \vec{d}_{r_z}),\ \text{ArcTan}(\vec{d}_{l_x}, \vec{d}_{l_z})$
    return $d_r, d_l, \alpha_r, \alpha_l$
```

Continuous Attributes

Using the preliminaries established above, and in particular the Bounded Laplace mechanism, we now describe coordinate transformations for protecting continuous attributes in VR. Each defense begins by calculating an *offset* using the LDPNoisyOffset helper function before diverging into two distinct categories: *additive offset* defenses, which protect attributes such as interpupillary distance (IPD) that are not expected to change over the course of a session, and *multiplicative offset* defenses, which protect attributes like observed height that might be updated each frame.

Additive Offset

Some continuous attributes (e.g., *interpupillary distance*) can be protected by simply adding a fixed *offset* value to the ground truth as a one-time transformation. The use of an additive offset is sufficient to protect these attributes without impacting usability due to the relatively static nature of such attributes throughout any given usage session. The resulting static defenses are shown in Alg. 2.

IPD. We start with IPD as it is amongst the easiest attributes to defend due to the fact that it should not reasonably be expected to change during a session. Our suggested countermeasure to attacks on IPD defends the player by scaling their avatar such that when an adversary measures the gap between their left and right eyes, the distance will correspond to a differentially private value.

Algorithm 2: Local differential privacy for continuous numerical attributes with additive offsets.

```
Function IPD($IPD, \varepsilon, l_i, u_i$):
    offset = LDPNoisyOffset($IPD, \varepsilon, l_i, u_i$)
    $IPD' = IPD +$ offset
    return $IPD'$
Function Pitch($pitch, \varepsilon, l_p, u_p$):
    offset = LDPNoisyOffset($pitch, \varepsilon, l_p, u_p$)
    $pitch' = pitch +$ offset
    return $pitch'$
```

Multiplicative Offset

We now turn our attention to the bulk of attributes for which a multiplicative offset is required. Consider, for example, the case of *wingspan*, where the perceived distance between a user's hands should appear to be 0 when their hands are touching, but should reflect $wingspan + offset$ when their hands are fully extended. Simply adding *offset* to the distance in all cases, as per the *additive offset* approach, is insufficient to achieve this property. Instead, we scale the entire range of values by v'/v as shown in Fig. 7.3. As a result, observable attributes attain a differentially-private value at their extremes, while their zero-point is maintained. We present in this section *multiplicative offset* defenses for a variety of attributes, as summarized in Alg. 3.

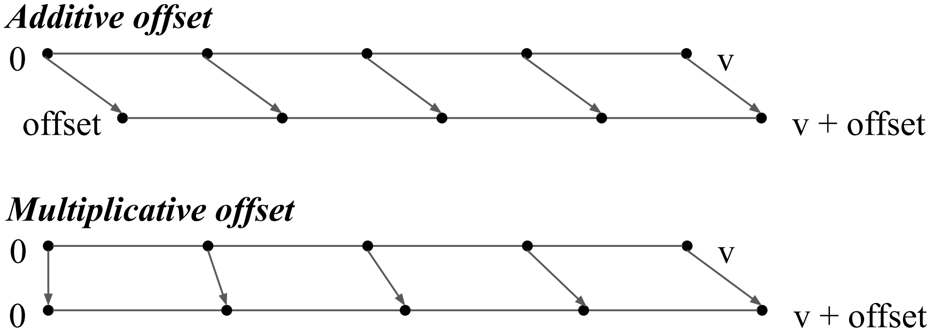


Figure 7.3: Additive vs. multiplicative offset transformations.

Height. A typical method for inferring the height of a VR user is to record the y-coordinate of the VR headset (y_h) over the course of a session, and then use the highest observed coordinate (or, e.g., the 99th percentile) as a direct linear correlate of height. This attack is effective because $y_h = height$ when a user is standing upright, which they generally are for a large portion of their session.

While one may be tempted to simply adjust y_h by *offset* at all times, doing so could cause the relative error of a fixed offset can grow to become disproportionate in applications where users are required to get close to the ground. In fact, in an extreme scenario where a user decides to lie flat on the ground, an adversary may observe $y_h' = 0 + offset$, which could defeat the privacy of this method by revealing *offset*.

Therefore, our suggested countermeasure is to use a *multiplicative offset*, whereby $y_h' = y_h * (height'/height)$. When $y_h = height$, the adversary now observes the differentially-private value $y_h' = height + offset$, while $y_h' = 0$ when $y_h = 0$ as shown in Fig. 7.4. We also suggest adjusting y_r and y_l such that the relative distance between the user's head and hands appears to remain unchanged.

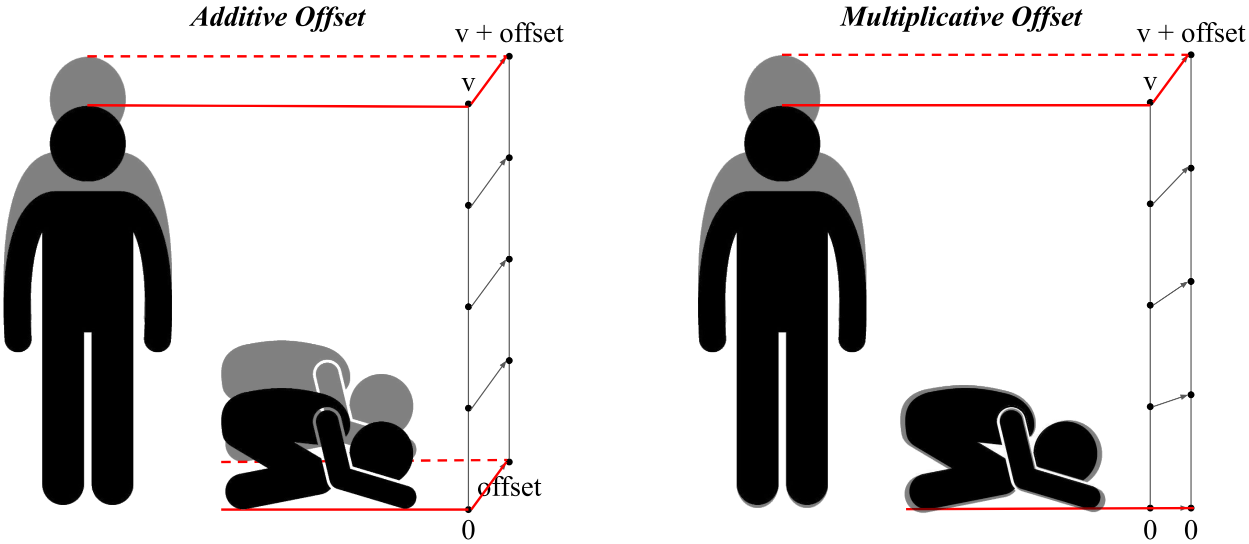


Figure 7.4: Use of additive vs. multiplicative offset for height.

Squat Depth. In Chapter 6, we have shown that an adversary can assess a proxy of a user's physical fitness by covertly prompting the users to squat and measuring their *squat depth*, i.e., $depth = height - y_h$, where y_h is the lowest headset coordinate recorded during the squat. The aim of this defense is to ensure that an adversary can only observe a differentially private *depth* value. While this could be achieved by setting a strict lower bound on y_h, doing so has the potential to be disorienting and could potentially have a negative impact on the VR user experience perspective. Instead, our suggested defense offsets y_h using the following transformation (independent of any defenses to *height*):

$$y'_h = height - (height - y_h) * (depth'/depth)$$

Consequently, y'_h smoothly transitions from $height$ to $height - depth + noise$ as y_h goes from $height$ to $height - depth$, obscuring the user's actual squat depth.

Wingspan. The wingspan attribute is harvested in a similar way to height, with an adversary monitoring the distance d between the left and right controllers over the course of a usage session and using the maximum observed value of d as a strong correlate of the user's wingspan. A VR application could require a user to fully extend their arms for seemingly legitimate gaming purposes, thus revealing their wingspan to potential attackers. The defense must therefore modify the observed distance d when the user's arms are extended. However, as discussed at the start of this section, simply adding a fixed offset to d does not allow $d = 0$ when the user's hands are touching, which is desirable for UX.

Algorithm 3: Local DP for continuous attributes with multiplicative offsets.

```
Function Height($y_h, y_r, y_l, height, \varepsilon, l_h, u_h$):
    $height' = height +$ LDPNoisyOffset($height, \varepsilon, l_h, u_h$)
    offset $= y_h * (height'/height) - y_h$
    return $y'_h, y'_r, y'_l = y_h+$ offset $, y_r+$ offset $, y_l+$ offset
Function Depth($y_h, y_r, y_l, height, depth, \varepsilon, l_d, u_d$):
    $depth' = depth+$ LDPNoisyOffset($depth, \varepsilon, l_d, u_d$)
    offset $= (height - ((height - y_h)/depth) * depth') - y_h$
    return $y'_h, y'_r, y'_l = y_h+$ offset, $y_r+$ offset, $y_l+$ offset
Function Wingspan($x_r, z_r, x_l, z_l, arm_R, arm_L, \varepsilon, l_w, u_w$):
    $span = arm_R + arm_L$
    $span' = span+$ LDPNoisyOffset($span, \varepsilon, l_w, u_w$)
    $d_r, d_l, \alpha_r, \alpha_l =$ PolarTransform($x_r, z_r, x_l, z_l$)
    $\text{offset}_\text{r} = (d_r/arm_R) * (span'/2) - d_r$
    $\text{offset}_\text{l} = (d_l/arm_L) * (span'/2) - d_l$
    $\text{offset}_{\text{r}_\text{x}}, \text{offset}_{\text{r}_\text{z}} = \text{offset}_\text{r} * cos(\alpha_r), \text{offset}_\text{r} * sin(\alpha_r)$
    $\text{offset}_{\text{l}_\text{x}}, \text{offset}_{\text{l}_\text{z}} = \text{offset}_\text{l} * cos(\alpha_l), \text{offset}_\text{l} * sin(\alpha_l)$
    $x'_r, z'_r = x_r + \text{offset}_{\text{r}_\text{x}}, z_r + \text{offset}_{\text{r}_\text{z}}$
    $x'_l, z'_l = x_l + \text{offset}_{\text{l}_\text{x}}, z_l + \text{offset}_{\text{l}_\text{z}}$
    return $x'_r, z'_r, x'_l, z'_l$
Function Arms($x_r, z_r, x_l, z_l, arm_R, arm_L, \varepsilon, l_{rat}, u_{rat}$):
    $span = arm_R + arm_L$
    $ratio = arm_R/span$
    $ratio' = ratio+$ LDPNoisyOffset($ratio, \varepsilon, l_{rat}, u_{rat}$)
    $d_r, d_l, \alpha_r, \alpha_l =$ PolarTransform($x_r, z_r, x_l, z_l$)
    $\text{offset}_\text{r} = (d_r/arm_R) * span * ratio' - d_r$
    $\text{offset}_\text{l} = (d_l/arm_L) * span * (1/ratio') - d_l$
    $\text{offset}_{\text{r}_\text{x}}, \text{offset}_{\text{r}_\text{z}} = \text{offset}_\text{r} * cos(\alpha_r), \text{offset}_\text{r} * sin(\alpha_r)$
    $\text{offset}_{\text{l}_\text{x}}, \text{offset}_{\text{l}_\text{z}} = \text{offset}_\text{l} * cos(\alpha_l), \text{offset}_\text{l} * sin(\alpha_l)$
    $x'_r, z'_r = x_r + \text{offset}_{\text{r}_\text{x}}, z_r + \text{offset}_{\text{r}_\text{z}}$
    $x'_l, z'_l = x_l + \text{offset}_{\text{l}_\text{x}}, z_l + \text{offset}_{\text{l}_\text{z}}$
    return $x'_r, z'_r, x'_l, z'_l$
Function Room($x_h, z_h, x_r, z_r, x_l, z_l, L, W, \varepsilon, l, u$):
    $L' = L+$ LDPNoisyOffset($L, \varepsilon, l, u$)
    $W' = W+$LDPNoisyOffset($W, \varepsilon, l, u$)
    $\text{offset}_\text{x}, \text{offset}_\text{z} = (x_h/W) * W' - x_h, (z_h/L) * L' - z_h$
    $x'_h, x'_r, x'_l = x_h + \text{offset}_x, x_r + \text{offset}_x, x_l + \text{offset}_x$
    $z'_h, z'_r, z'_l = x_h + \text{offset}_z, z_r + \text{offset}_z, z_l + \text{offset}_z$
    return $x'_h, x'_r, x'_l, z'_h, z'_r, z'_l$
```

In function Wingspan of Alg. 3, we formally introduce our recommended defense, where arm_R and arm_L are the arm length measurements in VR. As with our protection of squat depth, we ensure that the noise scales smoothly to preserve the user experience. As a result, when the user's hands are at the same coordinates, the observed distance is 0; thus, when the user touches their physical hands, the virtual hands also touch. On the other hand, when the arms are extended completely, the real-time distances between the controllers and their midpoint become $d_r = arm_R$ and $d_l = arm_L$, where $d_r + d_l = span$. In such a position, the observed wingspan becomes differentially private:

$$\text{offset} = \frac{d_r}{arm_R} * \frac{span'}{2} - d_r + \frac{d_l}{arm_L} * \frac{span'}{2} - d_l$$
$$\therefore \tfrac{span'}{2} - d_r + \tfrac{span'}{2} - d_l = span' - (d_r + d_l) = \text{offset}$$

The defense adds half the total offset to each arm. Consequently, the adversary will only observe a differentially private wingspan value when using the controllers' coordinates ((x_r, z_r) and (x_l, z_l)) to calculate the distance:

$$|\langle x_r, z_r\rangle - \langle x_l, z_l\rangle| = \tfrac{span'}{2} + \tfrac{span'}{2} = span'$$

In VR research, this is known as the "go-go technique" [218]; here, we use a small scale factor to obscure the user's wingspan (rather than to extend reach). As with the other multiplicative offset defenses, post-processing immunity protects the sensitive values when multiplied by $\frac{w}{v} \in [0, 1]$, and the adversary can only learn $span'$ from the observed distances in the range $[0, span']$.

Arm Length Ratio. If an adversary manages to measure the wingspan of a user, determining the arm length ratio is possible by using the headset as an approximate midpoint. As function Arms of Algorithm 3 shows, the corresponding defense is almost equivalent to that of the user's *wingspan*, but while the wingspan protection adds noise symmetrically to both arms, in this case, we add noise asymmetrically to obfuscate the ratio of arm lengths. This reflects a unique deployment of the go-go technique with different scale factors used for each arm to obscure length asymmetries.

Room Size. Lastly, in Chapter 6, we demonstrated that an adversary can determine the dimensions of a user's play area by observing the range of their movement. Once again, an additive offset would fail to defend against this attack by simply shifting the user's position rather than affecting their movement range. We therefore employ a similar technique as with the other multiplicative offset transformations in that the dynamic noise at the center of the room is 0, which increases as the user approaches the edges of their play area.

When the user is at the center of the room, $(x_h, z_h) = (0, 0)$, the offsets are 0. When the user is at a corner of the room, e.g., at $(x_h, z_h) = (\frac{width}{2}, \frac{length}{2})$, the offsets become half the noise added to each room dimension $(\frac{\text{Noise}_x}{2}, \frac{\text{Noise}_z}{2})$. Consequently, the adversary can only collect the noisy room dimensions, e.g., for width: $x'_h = x_h + \text{offset}_\text{x} = \frac{width/2}{width} * width' = \frac{width'}{2}$. Thus, the adversary would only learn a differentially private room dimension from observing x'_h in the range $[0, \frac{width'}{2}]$, with the same being true of *length*. Note that offsets added to x_h and z_h are intentionally chosen independently so that the adversary cannot even learn the proportions of the room.

Security Arguments. We conclude by arguing why the multiplicative offset approach maintains differential privacy, emphasizing that applying a fixed *offset* multiplicatively is very different from re-sampling the random *offset* value.

Proposition 1. *Given an single individual's ground truth value $v \in [l, u]$ collected locally once, where l and u are the lower and upper bounds of possible values of v, and an offset* N *sampled once from a differentially private distribution, broadcasting any $v' = \frac{w}{v}(v + \text{N})$ to a server protects v with differential privacy, where $w \in [0, v]$ is a real-time value continuously generated locally.*

Proof: Firstly, an adversary cannot learn the sensitive value from the ratio $\frac{w}{v} \in [0, 1]$ without knowing w. Thus, an adversary can only learn $v + \text{N}$ from the possible stream of broadcasted values $v' = \{0, ..., v + \text{N}\}$ sent to the server. Given that N is sampled from a differentially private distribution s.t. $v + N$ is centered around v, $v + \text{N}$ is immune to post-processing and is thus differentially private [73]. □

To provide a concrete example, consider again the attribute of *height*: $v = height, v' = height + offset, w = y_h$. Given that $height'$ is differentially private, an adversary who does not know the user's current y_h value (between 0 and *height*) will only be able to observe the current y'_h value (between 0 and $height'$), which cannot be used to find *height*.

Binary Attributes

We now switch our focus to attributes like *handedness* which can be represented as Boolean variables. For such attributes, we deploy the RandomizedResponse function of Alg. 1. If randomized response suggests an untruthful response, the user's virtual avatar is mirrored for other users, as is their view of the virtual world. While the user can still interact with the world and other avatars normally, we found that this approach comes at the cost of all text appearing to be backwards absent any special corrective measures.

Handedness. An adversary may observe a user's behavior, e.g., which hand they use to interact with virtual objects, to determine their handedness over time. Mirroring the user's avatar randomly on each VR session obfuscates handedness.

Arm Length Asymmetry. Using a mirrored avatar also provides plausible protection against adversaries observing which arm is longer; however, there is a large degree of overlap between this defense and that of *arm length ratio*.

Summary

While our aim in this section was to be as thorough as possible with regard to covering known VR privacy attacks, we by no means claim to have comprehensively addressed every possible VR privacy threat vector. Instead, we hope to have accomplished two simple goals. Firstly, we believe the combined defenses of this section are sufficient to significantly hinder attempts to deanonymize users in the metaverse. Within a large enough group of users, adversaries may have to combine dozens of unique attributes to reliably identify individuals; the absence of the low-hanging attributes discussed herein should obstruct their ability to do so. Secondly, we hope that the attributes covered in this section were diverse enough, and the corresponding defenses flexible enough, to be extended to future VR privacy threats.

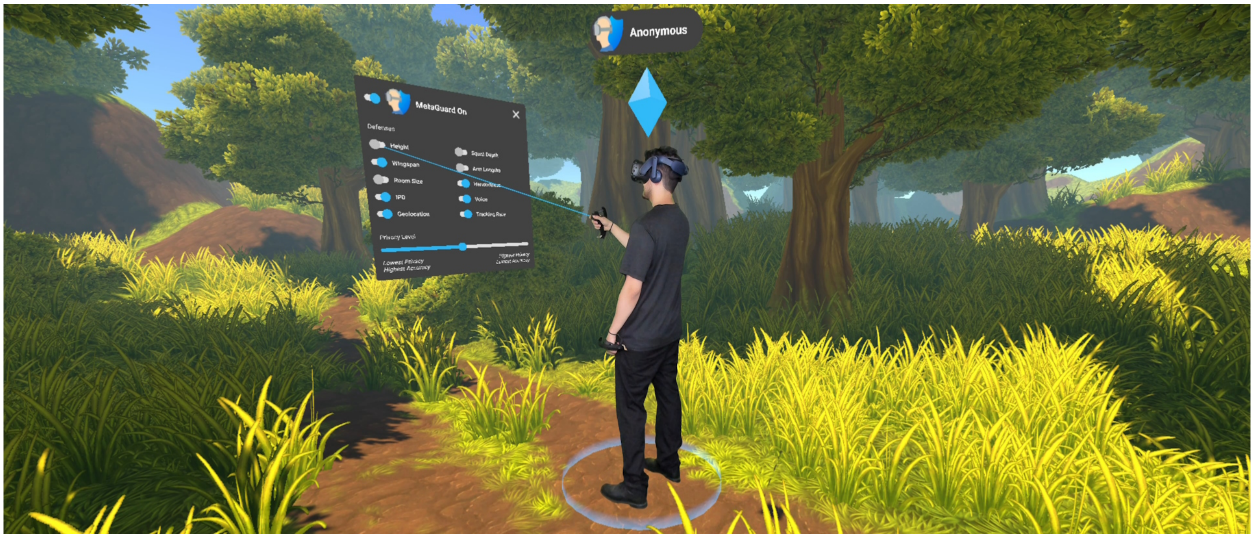


Figure 7.5: Mixed reality photo of "MetaGuard," our incognito mode for VR.

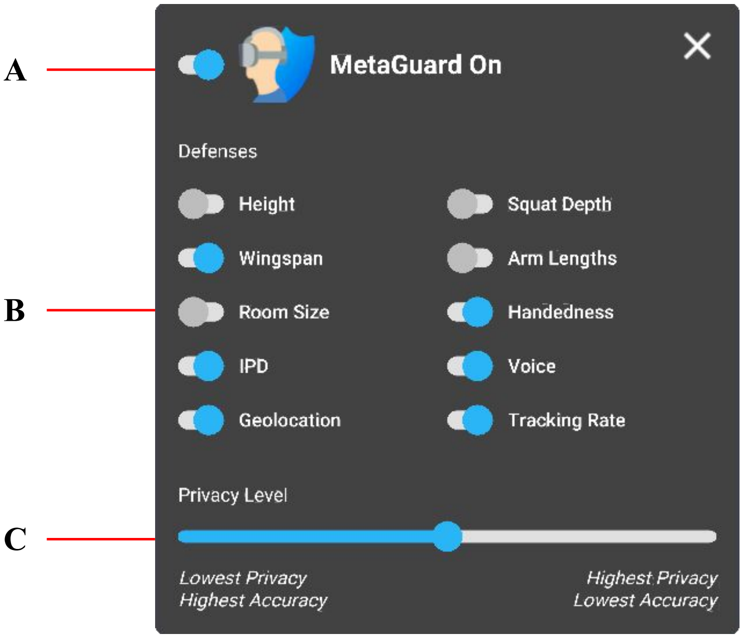
7.4 VR Incognito Mode

In this section, we introduce "MetaGuard,"[2] our practical implementation of the defenses presented in §7.3 and the first known "incognito mode" for the metaverse. We built MetaGuard as an open-source Unity (C#) plugin that can easily be patched into virtually any VR application using MelonLoader [168].[3] Fig. 7.5 shows a mixed reality photo of a player using the MetaGuard VR plugin within a VR game.

[2] Short for "Metaverse Guard."

[3] Unlike mobile apps, desktop VR apps can be modified by end users.

We begin by describing the options and interface made available to MetaGuard users. We then discuss our choice of DP parameters (ε, bounds, etc.) and outline how MetaGuard calibrates noise to each user. Finally, we describe the concrete game object transformations applied to the virtual world to implement the defenses of §7.3.



☑ MetaGuard On

Defenses

☐ Height	☐ Squat Depth
☑ Wingspan	☐ Arm Lengths
☐ Room Size	☑ Handedness
☑ IPD	☑ Voice
☑ Geolocation	☑ Tracking Rate

Privacy Level

Lowest Privacy
Highest Accuracy

Highest Privacy
Lowest Accuracy

Figure 7.6: VR user interface of MetaGuard plugin.

Settings & User Interface

The main objective of MetaGuard is to protect VR user privacy while minimizing usability impact. The flexible interface of MetaGuard (shown in Fig. 7.6) reflects this goal, allowing users to tune the defense profile according to their preferences and to the needs of the particular VR application in use. Specifically, we expose the following options:

(A) Master Toggle. The prominent master switch allows users to "go incognito" at the press of a button, with safe defaults that invite (but don't require) further customization.

(B) Feature Toggles. The feature switches allow users to toggle individual defenses according to their needs; e.g., in a game like Beat Saber [88], users may wish to disable defenses that interfere with gameplay (i.e., wingspan and arm lengths), while keeping the other defenses enabled.

(C) Privacy Slider. Lastly, we present users with a "privacy level" slider that adjusts the privacy parameter (ε) for each defense, allowing users to dynamically adjust the inherent trade-off between privacy and accuracy when using the defenses of §7.3. Users can choose from the following options, which we generally refer to simply as the "low," "medium," and "high" privacy settings:

- **High Privacy**, intended for virtual telepresence applications such as VRChat [119] and others [173, 169].
- **Balanced**, intended for casual gaming applications, such as virtual board games requiring some dexterity [89].
- **High Accuracy**, intended for noise-sensitive competitive gaming applications [248] such as Beat Saber [88].

Selecting Epsilon Values & Attribute Bounds

As discussed in §7.2, the level of privacy provided by the defenses of §7.3 depends on the appropriate selection of DP parameters, namely ε, Δ, and attribute bounds. Although our approach in MetaGuard is to allow users to adjust the privacy parameter (ε) according to their preferences, we must nevertheless translate the semantic settings of "low,“ "medium,“ and "high“ privacy into concrete ε-values, noting that a given privacy level may translate to a different ε-value for each attribute depending on its sensitivity to noise. Furthermore, the specific lower bound (l) and upper bound (u) of each attribute (and thus $\Delta = |u - l|$) must be determined in order to use the Bounded Laplace mechanism. This section outlines our method of selecting these values, with the results shown in Tab. 7.2.

Selecting ε-Values & Clamps

Continuous Anthropometrics. We conducted a small empirical analysis to select appropriate ε-values for each of the continuous anthropometric attributes at each privacy level. We began by selecting three VR applications (VRChat [119], Tabletop Simulator [89], and Beat Saber [88]) that represent the most popular examples of the intended use cases for the high, medium, and low privacy modes respectively. We then tested a wide range of ε-values for each attribute in each application while monitoring their effect on usability. For example, in Beat Saber, we had both a novice and expert-level player complete the same challenges at different ε-values to evaluate the impact of noise on in-game performance. By contrast, in VRChat, we were simply interested in the impact of noise on the ability to hold a conversation (e.g., to maintain virtual "eye contact").

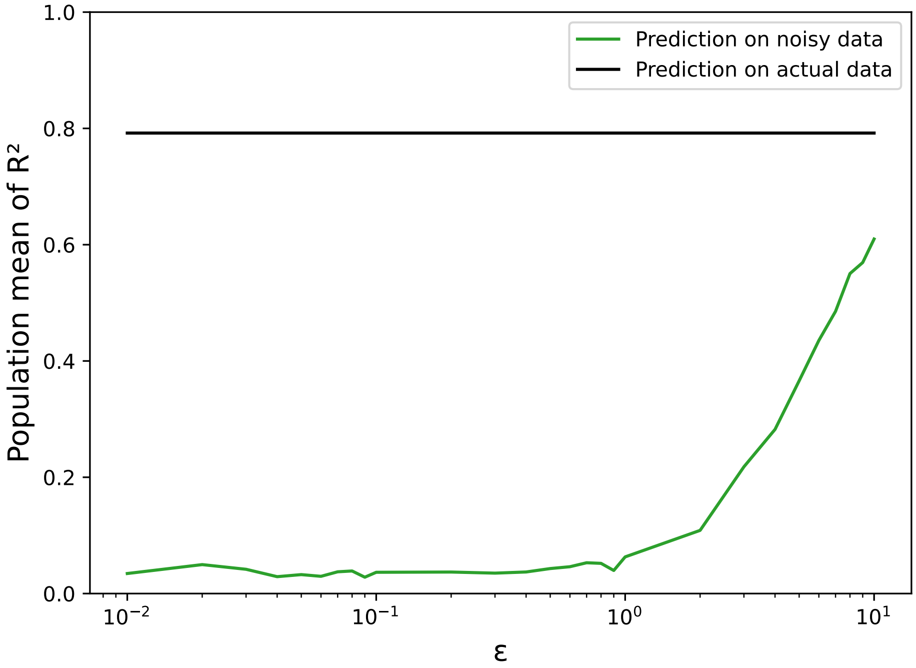


Figure 7.7: Coefficients of determination of height from predictions as ε increases.

Next, we analyzed the concrete privacy impact of candidate ε choices by simulating attackers at a variety of ε-values. Fig. 7.7 illustrates that for the height attribute, the vast majority of privacy benefit is already realized at $\varepsilon = 1$. We combined these results with the findings of our usability analysis to produce the final ε-values shown in Tab. 7.2 according to the appropriate balance of privacy and usability for the intended use of each level.

Binary Anthropometrics. For attributes where the defenses of §7.3 suggest the use of randomized response, we selected ε-values such that the corresponding prediction accuracy was degraded by 15%, 50% and 85% at the low, medium, and high privacy levels.

Clamps. Finally, for attributes where the corresponding defense of §7.3 suggests clamping, we chose clamp values which have the effect of anonymizing users within progressively larger groups. For example, for refresh/tracking rate, we selected clamps which hide users within the set of high (90Hz [90]), medium (72Hz [172]), and low (60Hz [256]) fidelity VR devices. For the latency-related attributes, we selected values below the perceptible 100ms threshold [33, 182, 188] that significantly decreased prediction accuracy.

Selecting Attribute Bounds

Finally, beyond ε, the Bounded Laplace mechanism also requires attribute bounds to constrain the outputs to semantically consistent values. We used public datasets to obtain the 95th percentile bounds for anthropometric measurements [35, 64, 228, 239]; our use of local DP causes Δ to reflect the full range of possible values. For room size, we extracted the bounds from official VR setup specifications [287]. We list the bounds and corresponding references in Tab. 7.2 below.

Data Point	**Bounds**		**Privacy Levels**		
	Lower	**Upper**	**Low**	**Medium**	**High**
Height [35]	1.496m	1.826m	ϵ=5	ϵ=3	ϵ=1
IPD [64]	55.696mm	71.024mm	ϵ=5	ϵ=3	ϵ=1
Voice Pitch [228]	85 Hz	255 Hz	ϵ=6	ϵ=1	ϵ=0.1
Squat Depth [193]	0m	0.913m	ϵ=5	ϵ=3	ϵ=1
Wingspan [239]	1.556m	1.899m	ϵ=3	ϵ=1	ϵ=0.5
Arm Ratio [193]	0.95	1.05	ϵ=3	ϵ=1	ϵ=0.5
Room Size [287]	0m	5m	ϵ=3	ϵ=1	ϵ=0.1
Handedness	0	1	ϵ=1.28	ϵ=0.88	ϵ=0.73
Latency (Geolocation)	Clamped		25ms	30ms	50ms
Reaction Time	Clamped		10ms	20ms	100ms
Refresh/Tracking Rate	Clamped		90 Hz	72 Hz	60 Hz

Table 7.2: Selected ε, clamps, and attribute bound values.

We emphasize that the sole purpose of our informal experimentation in this section is to set a reasonable range of ε-values that cover a variety of VR use cases. Given the lack of consensus on a formal method for selecting DP parameters [72], our choices simply serve to establish a plausible spectrum of ε-values corresponding to our perceived boundaries of the privacy-usability trade-off. The power to select exactly which point on this spectrum is best suited for a particular application remains with the end user.

Rerandomization & Linkability

By default, we suggest randomly resampling offset values according to the algorithms of §7.3 at the start of each session. Assuming that MetaGuard users cannot be linked across sessions, adversaries will be unable to aggregate measurements across multiple sessions to obtain user data. Alternatively, one-time randomization can be used, allowing cross-session linkability but guaranteeing that no attribute leakage occurs.

Calibration & Noise Centering

One final parameter is required to successfully implement the continuous attribute defenses of §7.3: the ground truth attribute values of the end user. Centering the Laplacian noise distribution around the ground truth attribute values of the current user has the effect of minimizing noise for as many users as possible, particularly those who are outliers, thus achieving theoretically optimal usability.

To achieve this, the MetaGuard extension calculates instantaneous ground truth estimates upon instantiation using the method shown in Fig. 7.8. Specifically, the OpenVR API [281] provides MetaGuard with one-time snapshot locations of the user's head, left and right eyes, left and right hands, and a plane representing the play area. Estimates for the ground truth values of height, wingspan, IPD, room size, and left and right arm lengths can then be derived from these measurements. We note that the privacy of MetaGuard is not dependent on the accuracy of the ground truth estimates, which exist only to ensure that the added noise is not more than the level necessary to protect a given user.

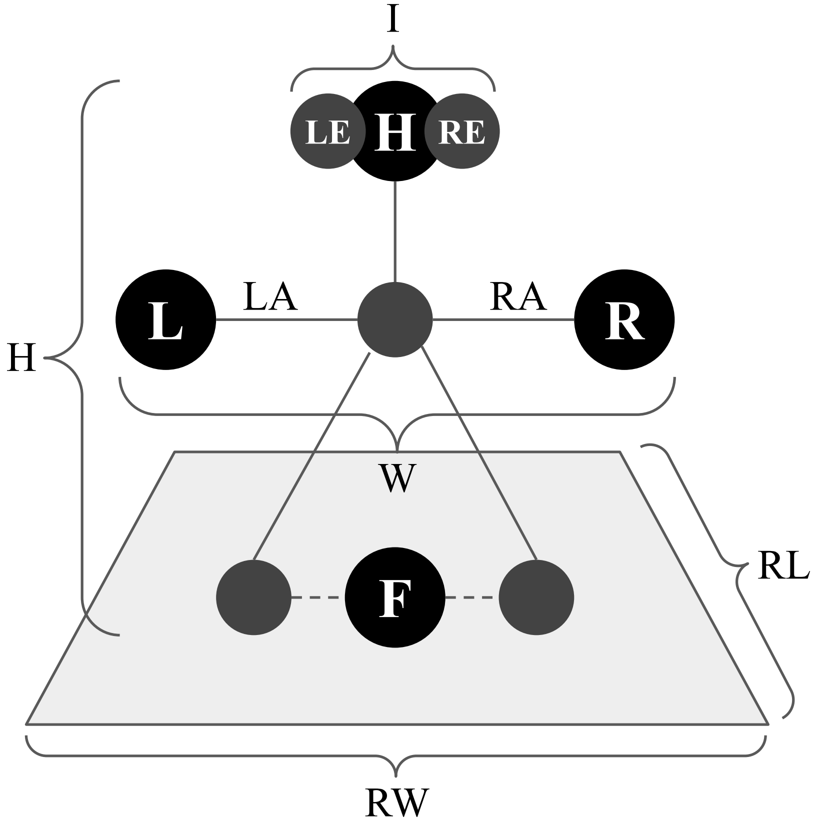


Figure 7.8: Instantaneous calibration of ground truth for height (H), left arm (LA), right arm (RA), wingspan (W), IPD (I), room width (RW), and room length (RL), using head (H), floor (F), left/right controllers (L/R); figure not to scale.

Defense Implementation

We now finally provide a complete description of our "VR Incognito Mode" system for implementing the defenses of §7.3 in light of the interface, ε-values, bounds, and calibration procedures described above. Our implementation follows two phases: a *setup phase*, which executes exactly once on the frame when a defense is enabled, and an *update phase*, which executes every frame thereafter.

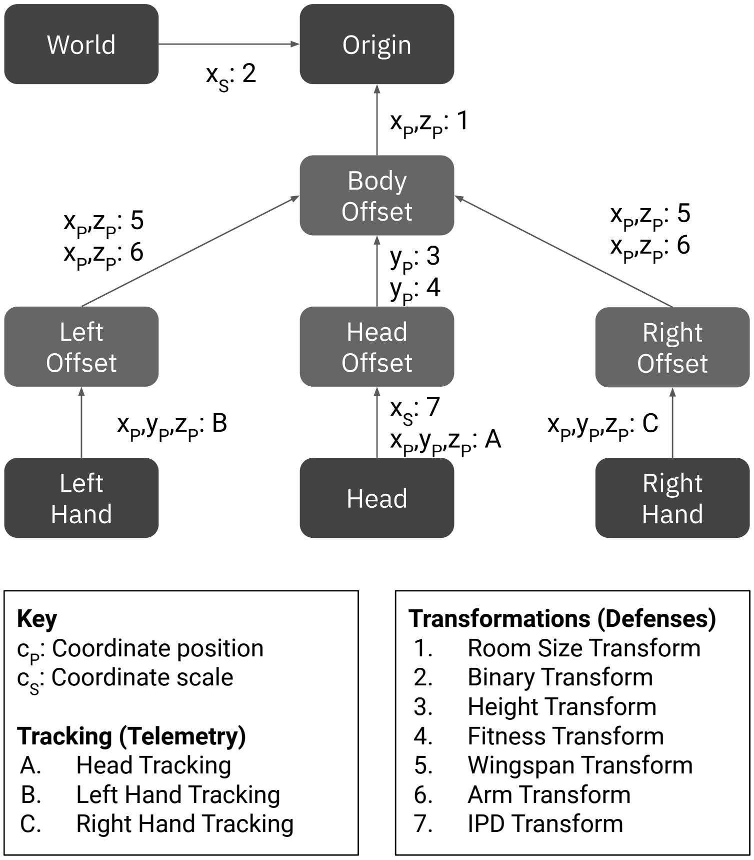


Figure 7.9: Game object hierarchy with existing (dark grey) and inserted (light grey) game objects, and coordinate transformations used to implement VR Incognito Mode defenses.

Setup Phase. When a defense is first enabled, MetaGuard uses the calibration procedures of §7.4 to estimate the ground truth attribute values of the user. These values are then used in combination with the ε-values and bounds of §7.4 to calculate noisy offsets corresponding to each privacy level using the methods outlined in §7.3, and are then immediately discarded from program memory (with only offsets retained) so as to minimize the chance of unintentional data leakage. By default, the Unity game engine uses telemetry data from OpenVR [282] to position game objects within a virtual environment, which are then manipulated by a VR application. During the setup phase, the system modifies the game object hierarchy by inserting intermediate "offset" objects as shown in Fig. 7.9.

Update Phase. During the update phase, the system first checks which defenses the user has enabled in the interface (see §7.4). For all disabled attributes, the corresponding offset transformations in the game object hierarchy (as shown in Fig. 7.9) are set to the identity matrix. For each enabled feature, the system implements the corresponding defense of §7.3 by fetching the noisy attribute value calculated during the setup phase for the currently-selected privacy level and enabling the relevant coordinate transformation on the inserted offset objects such that the observable attribute value matches the noisy attribute value. Specifically, Fig. 7.9 illustrates how the position of each game object is defined with respect to another object in the hierarchy, and how the defenses modify the relative position or scale of each object with respect to its parent.

7.5 System Evaluation

In this section, we demonstrate the effectiveness of the defenses introduced in §7.3 by evaluating their impact on the accuracy of a theoretical attacker. To do so, we replicated the attacks of the TTI [178], MetaData (Chapter 6), and 50k (Chapter 4) studies to measure their accuracy both with no defenses and with the MetaGuard extension at the low, medium, and high privacy levels. The results of this evaluation are summarized in Tab. 7.3 of §7.6. The presented accuracy values represent what a server attacker could achieve, and also provide an upper bound for the capabilities of user attackers.

Evaluation Method

We obtained from the original authors anonymized frame-by-frame telemetry data recordings of the 511 users from the TTI [178] study. We also used our own data from the MetaData study (Chapter 6) and 50k study (Chapter 4). Using this data, we could virtually "replay" the original sessions exactly as they occurred, and were able to reproduce the identification and inference attacks described in the original studies with nearly identical results. Next, we repeated this process for each session with MetaGuard enabled at the low, medium, and high privacy levels. The resulting decrease in attack accuracy for each attribute at each privacy level is shown in §7.6.

To emulate a realistic metaverse threat environment, we streamed telemetry data from the client to a remote game server via a WebSocket. The MetaGuard extension was allowed to clamp the bandwidth and latency of this data stream as discussed in §7.3. The network-related attacks were then run on the server side.

Beyond the attacks which deterministically harvest sensitive data attributes, all three studies use machine learning to identify users or profile their demographics. We used sklearn to replicate the published methods as closely as possible, using the same model types and parameters as in the original papers. Once again, we replicated the original results with similar accuracy, with the decrease in identification corresponding to the use of the low, medium, and high privacy levels of MetaGuard being shown in Tab. 7.3C of §7.6.

Ethical Considerations

Other than the ε-calibration effort described in §7.4, which was performed by the authors, this chapter does not involve any new research with human subjects. Instead, our results rely on the replication of prior studies using anonymous data obtained either from public online repositories or directly from the authors of those studies. All original studies from which we obtained data were non-deceptive and were each subject to individual ethics review processes by OHRP-registered institutional review boards. Furthermore, the informed consent documents of the original studies explicitly included permission to re-use collected data for follow-up studies, and we strictly followed the data handling requirements of the original consent documentation, such as the promise to only publish statistical aggregates rather than individual data points.

Primary & Secondary Attributes

Continuous Anthropometrics. Tab. 7.3A shows that our defenses effectively reduce the coefficients of determination to values below 0.5 for the targeted continuous attributes. We found that physical fitness (squat depth) is the most challenging attribute to protect while preserving user experience, as it shows the smallest drops in prediction accuracy. The remaining attributes show significant decreases in attack accuracy even at the low privacy level: IPD (−67.53%), room size (−55.89% within 2m^2), wingspan (−33.07% within 7 cm) and height (−16.93% within 5 cm).

Binary Anthropometrics. An advantage of the randomized response technique is precise control over attacker accuracy levels by choosing the values of ε. Unsurprisingly, the prediction accuracy of handedness (92.5%, 75%, and 57.5% for the low, medium, and high privacy levels) corresponded to the chosen ε-values.

Inferred Attributes

The machine learning models of the MetaData study primarily use the attributes discussed above as model inputs to infer demographics. Clearly, the reduction in accuracy of these primary attributes will have a negative impact on the accuracy of inferences based on them; nonetheless, we ran the models on the noisy attributes to quantify this impact. The results show significant accuracy drops in predicting gender (−23.5%), age (−58.25%), ethnicity (−48.75%), and income (−73.85%), even at the lowest privacy setting. Most importantly, the three identification models simulating an attacker identifying a user amongst a group all had a significant drop in accuracy (see Tab. 7.3C); thus, MetaGuard empirically succeeds at its primary goal of preventing users from being deanonymized.

7.6 Results

Table 7.3A: Primary and Secondary Attributes (MetaData [193] Study)

Attribute	**Metric**	**No Privacy**	**Low Privacy**	**Medium Privacy**	**High Privacy**
Height	Within 5cm	70%	53.07% ±2.41%	45.00% ±2.35%	32.63% ±2.3%
	Within 7cm	100%	68.6% ±2.18%	58.17% ±2.09%	44.47% ±2.43%
	R^2	0.79	0.37 ±0.040	0.22 ±0.035	0.06 ±0.020
Physical Fitness	Categorical	90%	86.11% ±2.65%	79.11% ±2.60%	61.56% ±4.15%
IPD (Vive Pro 2)	Within 0.5mm	96%	18.53% ±1.76%	13.40% ±1.33%	11.10% ±1.24%
	R^2	0.991	0.399 ±0.041	0.165 ±0.031	0.068 ±0.019
IPD (All Devices)	Within 0.5mm	87%	19.47% ±1.81%	14.17% ±1.35%	12.17% ±1.26%
	R^2	0.857	0.318 ±0.038	0.134 ±0.027	0.068 ±0.017
Wingspan	Within 7cm	87%	53.93% ±3.61%	42.13% ±3.32%	40.80% ±2.80%
	Within 12cm	100%	78.80% ±2.76%	66.00% ±3.31%	65.46% ±3.14%
	R^2	0.669	0.134 ±0.042	0.047 ±0.019	0.036 ±0.021
Room Size	Within $2m^2$	78%	22.11% ±2.85%	16.33% ±2.74%	12.66% ±2.98%
	Within $3m^2$	97%	33.52% ±3.80%	23.44% ±3.08%	19.53% ±2.92%
	R^2	0.974	0.406 ±0.153	0.495 ±0.171	0.360 ±0.136
Longer Arm	$\geq$ 1cm Difference	63%	58.63% ±5.79%	52.35% ±6.83%	54.90% ±5.12%
	$\geq$ 3cm Difference	100%	77.78% ±13.46%	62.22% ±15.09%	53.33% ±15.64%
Handedness	Categorical	97%	92.5%	75%	57.5%
Reaction Time	Categorical	87.50%	79.20%	62.50%	54.20%
HMD Refresh Rate	Within 3 Hz	100%	0%	0%	0%
Tracking Refresh Rate	Within 2.5 Hz	100%	0%	0%	0%
VR Device	Categorical	100%	10%	0%	0%

Table 7.3B: Inferred Attributes (MetaData [193] Study)

Attribute	**Metric**	**No Privacy**	**Low Privacy**	**Medium Privacy**	**High Privacy**
Gender	Categorical	100%	76.5% ±1.29%	70.47% ±1.85%	57.19% ±2.20%
Age	Within 1yr	100%	41.75% ±1.65%	36.09% ±1.87%	24.28% ±1.87%
Ethnicity	Categorical	100%	51.25% ±2.70%	40.75% ±2.36%	31.37% ±2.40%
Income	Within $10k	100%	26.15% ±1.41%	28.00% ±1.87%	26.06% ±2.11%

Table 7.3C: Identity (TTI [178], MetaData [193], and 50k [196] Studies)

Attribute	**Dataset**	**No Privacy**	**Low Privacy**	**Medium Privacy**	**High Privacy**
Identity	TTI (Miller et al.)	95%	81.10% ±5.78%	45.29% ±5.48%	26.51% ±1.37%
Identity	MetaData (Chapter 6)	100%	5.44% ±0.68%	4.59% ±0.76%	4.0% ±0.67%
Identity	50k (Chapter 4)	94.33%	15.59% ±4.50%	6.10% ±1.76%	2.19% ±1.17%

Table 7.3: Main Results (accuracy and R^2 values with **99**% confidence intervals)

7.7 Discussion

In this study, we set out to design, implement, and evaluate a comprehensive suite of VR privacy defenses to protect VR users against a wide range of known attacks. In the absence of any defenses, these attacks demonstrated the ability to not only infer specific sensitive attributes, but also to combine these attributes to infer demographics and even deanonymize users entirely. Through our evaluation of MetaGuard, our practical implementation of a "VR incognito mode" plugin, we have demonstrated that ε-differential privacy can pose an effective countermeasure to such attacks. Our results show a considerable accuracy reduction in the identification and profiling of users using real VR user data from 56,082 participants across three popular VR privacy studies. By evaluating our system using telemetry data from these existing studies, we were able to independently measure the performance of each defense at each supported privacy level, a feat that would otherwise have required an infeasible number of separate laboratory trials.

MetaGuard allows users to "go incognito" by randomizing their fictitious measurements, such as height and wingspan, at the start of each new session, thus thwarting cross-session likability. Alternatively, if users do not mind being linked across sessions, they do not need to re-randomize their fictitious measurements between sessions, allowing adversaries to track them across sessions without revealing their true attribute values in the process.

Our use of bounded Laplacian noise allows us to achieve a theoretically optimal balance between privacy and usability, minimizing the mean squared tracking error a user is expected to experience for a given privacy level (ε) [115, 73]. This, in turn, allows us to leverage homuncular flexibility to implement the defenses in a way that users can rapidly learn to ignore [304, 1]. For example, the average wingspan offset at the medium privacy level is 4.5 cm, which is well within the range that VR users can flexibly adapt to [218]. Even those transformations which do not directly affect the player model can be thought of as equivalent to body modifications. For example, room size is not necessarily implemented as a body manipulation, but changing the room-to-avatar ratio can be thought of as equivalent to changing the size of the entire avatar and thereby hiding the relative size of the room. As such, we expect homuncular flexibility to be applicable to such transformations as well.

Overall, MetaGuard constitutes the first attempt at producing a privacy-preserving "incognito mode" solution for VR. Grounded in theoretical privacy, and demonstrated using thorough empirical evaluation, we aim to provide a solid foundation for future work in this area. The importance of privacy-enhancing software like MetaGuard will become more pronounced as current market trends make virtual reality increasingly ubiquitous and shape the next generation of the social internet, the so-called "metaverse" [189, 233, 258]. As it stands, VR device manufacturers have been observed selling VR hardware at losses of up to \$10 billion per year [219], presumably with the goal of recouping this investment through software-based after-sale revenue, such as via targeted advertisement [46, 4].

Despite using the terms “attacker,” and “adversary” throughout our writing, it’s possible that VR data harvesting could be entirely above board, with users agreeing (knowingly or otherwise) to have their data collected. It is more important than ever to give users the ability to protect their data through technological means, independent of any warranted data privacy regulations, in a way that is as easy to use as privacy tools for the web.

Limitations. Our decision to base our evaluation on data from prior studies means that we inherit the biases of the original studies. In particular, the test subjects of the studies from which our data is derived were not perfectly representative of the general population of VR users. While our evaluation method replicates the telemetry stream that would have been generated by the original participants were they using the MetaGuard extension, it does so under the assumption that their use of MetaGuard would not have changed their behavior. The accuracy of MetaGuard could be somewhat diminished if it turns out that users modify their behavior to compensate for the added noise. Further, our study considers a limited set of data attributes, which may not be comprehensive with respect to the attributes inferable in VR. MetaGuard may not be effective at protecting attributes beyond those that we directly considered. Finally, the mean-squared-error definition of “usability” by which our system is theoretically optimal may in some cases fail to align with the true user experience in VR.

7.8 Conclusion

In this chapter, we have presented the first comprehensive “incognito mode for VR.” Specifically, we designed a suite of defenses that quantifiably obfuscate a variety of sensitive user data attributes with ε-differential privacy. We then implemented these defenses as a universal Unity VR plugin that we call “MetaGuard.” Our implementation, which is compatible with a wide range of popular VR applications, gives users the power to “go incognito” in the metaverse with a single click, with the flexibility of adjusting the privacy level and set of enabled defenses for each application as they see fit.

Upon replicating VR privacy attacks using real user data from prior studies, including the attacks of Chapters 4 and 6, we demonstrated a significant decrease in attacker capabilities across a wide range of metrics. In particular, the ability of an attacker to deanonymize a VR user was degraded by as much as 96.0% while using the MetaGuard extension.

Over the course of decades of research in web privacy, private browsing mode has remained amongst the most ubiquitous privacy tools in popular use today. We were inspired by the success of “incognito mode” on the web to produce a metaverse equivalent that is just as user-friendly, while serving the same fundamental purpose of helping users remain untraceable across multiple sessions. We hope our open-source MetaGuard plugin and promising results serve as a foundation for other privacy practitioners to continue exploring usable privacy solutions in this important field.

Chapter 8

Deep Motion Masking for Secure, Usable, and Scalable Real-Time Anonymization of Virtual Reality Motion Data

8.1 Introduction

As demonstrated in this dissertation, the head and hand motion data captured by a VR device can be used to uniquely identify its user across a variety of applications [178, 241, 274], over long periods of time [177, 221], and at a rate of over 1 in 50,000 (see Chapter 4), comparable to that of a fingerprint scan [301]. Moreover, a variety of potentially sensitive user data attributes can be inferred directly from VR telemetry streams (see Chapter 5). Such results raise serious questions about whether XR devices can be used without involuntarily revealing a plethora of personal information to the device, application, and other XR users. Researchers have proposed a number of methods for anonymizing VR motion data, as summarized in Chapter 2. Most recently, in Chapter 7, we proposed a differential privacy approach for anonymizing VR motion data. However, all anonymization methods discussed thus far underestimate the identifiability of motion data when using sophisticated models trained on large datasets. In this chapter, we present a best-in-class VR identification model that achieves over 90% cross-session identification accuracy with 500 users, even when using existing countermeasures. We then propose "deep motion masking," a technique that uses deep learning to effectively anonymize VR motion data.

Deep motion masking represents a multi-axis improvement over prior VR anonymization methods. Through a comprehensive evaluation, we demonstrate a 2.7× improvement in the indistinguishability of anonymized motion data, and an over 20× improvement in cross-session unlinkability. Our proposed system is capable of low-latency real-time anonymization of VR telemetry streams, making it practical for deployment in new and existing VR systems.

8.2 Method

VR Adversaries

For a final time, we revisit the information flow and threat model of Chapter 2. As in most of this dissertation, our emphasis in this chapter is on protecting the motion data visible to external adversaries, namely VR game servers and other VR users. These adversaries are considered "weaker" in the threat model, meaning that attacks available to them are typically available to all other adversaries. Moreover, attacks performed by these adversaries are generally the hardest to detect due to their remote and decentralized nature.

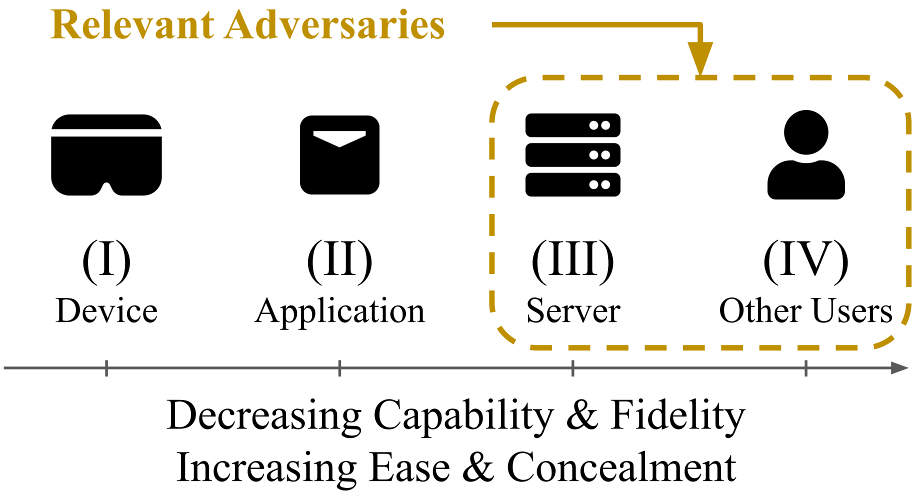


Figure 8.1: VR privacy threat model and relevant adversaries.

In summary, the focus of this chapter is on the threat posed by broadcasting head and hand motion data to servers and external users in multi-user VR applications. These threats are amongst the most realistic, universal, and pernicious VR privacy challenges.

Dataset

As before, this work is based on the BOXRR-23 dataset, described in Chapter 3. In particular, only the BeatLeader portion of the data is used in this chapter. Our motivation for selecting this dataset is threefold. First, BOXRR-23 is multiple orders of magnitude larger than the next largest VR motion dataset, making it an obvious choice for training deep learning models. Additionally, the authors explicitly endorse using the dataset for security and privacy research, and state that the dataset underwent stringent ethical and legal review for those purposes prior to its release. Finally, using an already-public dataset will improve the transparency, reproducibility, and extensibility of this work.

8.3 Motivation

We now present a series of introductory experiments on motion-based identification in VR using the dataset of Chapter 3. We describe the basic principles behind existing VR identification models and then show that with a sufficiently large volume of data, models can be trained that are far more robust and capable than those discussed in prior work. The aim of this section is not to serve as the main contribution of this chapter but rather to motivate our new defensive approach by demonstrating the insufficiency of existing countermeasures.

Prevailing Architectures

At present, most existing papers on VR user identification utilize classical machine learning models, such as those based on the Random Forest [29] and LightGBM [137] architectures. The motivation for using these models over theoretically more powerful deep learning approaches is that deep learning typically requires a significantly larger volume of data to successfully train and converge, whereas tree-based architectures can produce generalizable classifiers with fewer samples per user.

On the other hand, the sequential time-series format of VR motion data streams is not a natural fit for tree-based models, which usually require a one-dimensional tabular data format. As such, prior works suggest deliberate feature engineering to convert motion data streams into tabular samples by using summary statistics to eliminate the time dimension.

Specifically, Pfeuffer et al. [214] suggest dividing motion data into one-second chunks, and then converting each chunk into a flat feature vector by taking four statistics (min, max, mean, and standard deviation) across each tracked dimension. Miller et al. [179] use a very similar approach, but also include the median of each axis. Moore et al. [186] use identical features to Miller, while our own approach in Chapter 4 uses similar features but adds contextual data specific to the VR application. At a high level, many prior works have found the basic idea of summarizing one-second chunks of motion to be highly effective.

Surprisingly, the method of using one-second summary statistics has in some instances outperformed sequential deep learning models even when sufficiently large datasets are present. For example, our identification study in Chapter 4 found that LightGBM with tabular summary statistics outperformed MLP, GRU, and LSTM models despite using a fairly large amount of data.

For reasons yet unknown, the basic notion of summarizing one-second subsequences of larger motion recordings seems uniquely well-suited for identifying VR users. Thus, we are motivated to replicate this approach using deep learning architectures in order to achieve better identification performance.

LSTM Funnel Architecture

In this section, we propose a new deep learning architecture that aims to internally replicate the idea of summarizing one-second motion subsequences by using a combination of Long Short-Term Memory (LSTM) [112] and Multi-Layer Perceptron (MLP) [99] layers. Figure 8.2 illustrates how the proposed architecture may be used to identify VR motion sequences. The model receives as input a 30-second motion sequence normalized to 30 frames per second, thus containing 900 frames in total. Using an LSTM layer, each frame is converted into a 256-dimensional feature vector. Then, an average pooling layer combines each one-second (30-frame) subsequence into a 256-dimensional summary. Next, another LSTM layer combines the sequence of 30 256-dimensional summaries into a flat 256-dimensional embedding. Finally, a fully connected MLP layer with softmax activation produces a classification output, with optional additional dense layers in between.

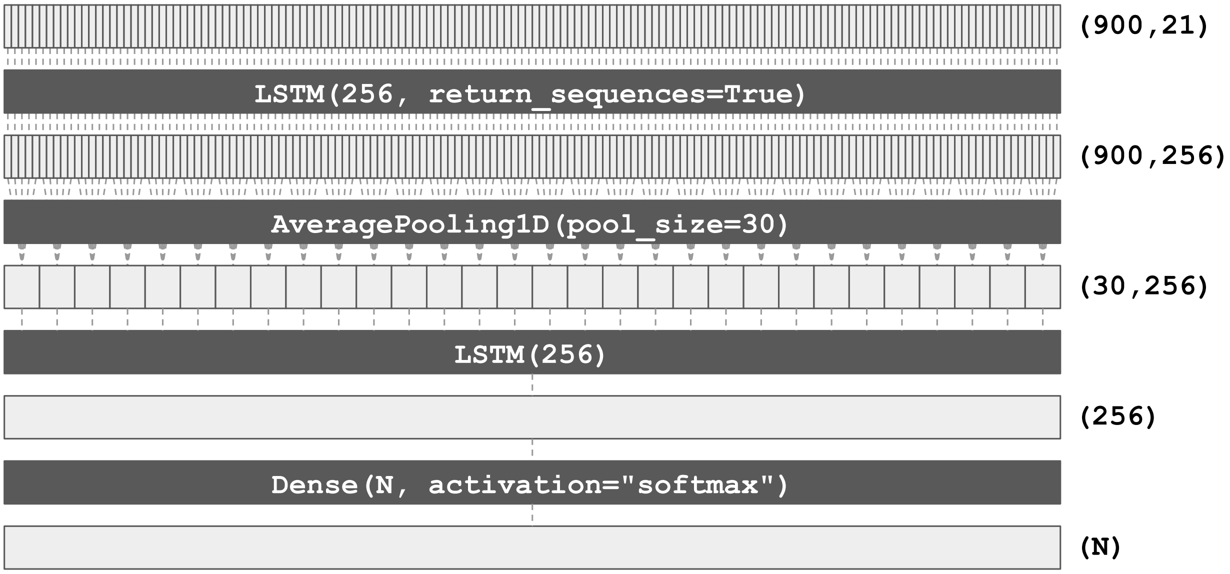

(900,21)
LSTM(256, return_sequences=True)
(900,256)
AveragePooling1D(pool_size=30)
(30,256)
LSTM(256)
(256)
Dense(N, activation="softmax")
(N)

Figure 8.2: “LSTM funnel” identification architecture.

In essence, the architecture described above continues to represent VR motion sequences using summary statistics taken across one-second chunks, yet is able to outperform prior approaches for a few major reasons. First, instead of manually specifying summary statistics to be taken, such as mean, standard deviation, etc., the model is allowed to learn its own relevant statistics via the first LSTM layer. Second, instead of manually specifying how to summarize the classification of each subsequence, such as via a logarithmic sum of probabilities, the model is allowed to learn its own meta-classification method via the second LSTM and subsequent MLP layers. Moreover, the “featurization” and “classification” parts of the model are trained together in an end-to-end fashion, allowing the model to learn how to create complex statistics that result in optimal classification results.

We call this approach the "LSTM funnel" architecture due to the dimensionality reduction performed by the average pooling layer. While the method seems fairly simple overall, to the best of our knowledge, this architecture has either not yet been disclosed in general, or at least has not been used for similar purposes.

Worst-Case Identifiability

We now demonstrate how the LSTM funnel architecture can be used to drive significant improvements in motion-based identification accuracy, provided a large amount of training data per user is available. Using the dataset of Chapter 3, we first found the 500 users for which the greatest number of individual recordings were available. For these top 500 users, an average of 821 recordings were available per user, with each recording averaging about three minutes in length. We used the 500 most recent recordings of each user for our evaluation, with 400 of these recordings being used for training, 50 for validation, and the remaining 50 being used for testing. To conform to the architecture of §8.3, only the first 30 seconds of each recording were utilized, and recordings were normalized to a constant 30 frames per second by using a numerical linear interpolation for positional coordinates and a spherical linear interpolation for orientation quaternions.

To evaluate the performance of the LSTM funnel architecture on this particular dataset, we implemented the architecture of Figure 8.2 in Keras v2.10.1 [138] and trained it for 500 epochs on the described dataset using the Adam optimizer [141] with a learning rate of 0.001. The validation dataset was used for early stopping after 25 epochs of no improvement. For the sake of comparison, we also trained and tested several previously proposed identification model architectures using the same dataset, the results of which were as follows:

- Our new LSTM funnel architecture achieves a per-sample accuracy of 98.12% and a per-user accuracy of 100.00%.
- The LightGBM-based architecture proposed in Chapter 4 achieves a per-sample accuracy of 71.66% and a per-user accuracy of 100.00%.
- The Miller et al. [177] architecture achieves a per-sample accuracy of 56.59% and a per-user accuracy of 97.60%.

As evidenced by the above results, our architecture substantially exceeds the identification performance of the most notable prior models when using identical datasets. This, on its own, is not entirely surprising, given that we used over three hours of training data per user to perform this demonstration, which also exceeds all prior works; the previously proposed models and featurization approaches were not designed to take full advantage of this volume of data. However, the robustness of our new architecture to reductions in input dimensionality is, to our knowledge, unprecedented:

- The original representation with the full 21 features ($\{head, left\_hand, right\_hand\} \times \{x, y, z, i, j, k, w\}$) gives a sample accuracy of 98.12% and a user accuracy of 100.00%.

- Removing the head, the remaining 14 features ($\{left\_hand, right\_hand\} \times \{x, y, z, i, j, k, w\}$) reduce sample accuracy to 94.76% (and still 100% user accuracy).

- Using only hand rotations, the remaining 8 features ($\{left\_hand, right\_hand\} \times \{i, j, k, w\}$) give a sample accuracy of 93.42% and a user accuracy of 100.00%.

- Using only left hand rotations, the remaining 4 features ($\{left\_hand\} \times \{i, j, k, w\}$) still result in a sample accuracy of 92.77% and a user accuracy of 100.00%.

- Using only left hand rotational magnitude, the single feature ($\{left\_hand\} \times \{w\}$) still results in a sample accuracy of 84.23% and a user accuracy of 100.00%.

In other words, by observing just the absolute magnitude of the rotation of one hand of a user for just 30 seconds, the model can still correctly identify the user out of 500 options with nearly 85% accuracy, provided it was first trained on over 3 hours of data per user.

Today, obtaining 200 minutes of motion capture data for a user may seem like an absolute worst-case scenario from a privacy perspective, with the 500 individuals used in our demonstration perhaps being amongst the only individuals in the world for which this amount of data is readily accessible. However, if extended reality truly replaces existing mobile devices as a default method of human-computer interaction for millions of users in the near future, having multiple hours of cumulative time spent using XR devices may soon come to represent an average or even below-average usage pattern.

Prevailing Defenses

In light of the new findings discussed above, we now briefly revisit and reevaluate the existing proposals for countermeasures against motion-based identification in VR:

- Miller et al. [178] have suggested transmitting only certain rotational dimensions rather than positional data. However, as demonstrated by the results of §8.3, hand rotation values alone are now sufficient to accurately deanonymize users.
- Moore et al. [186] suggest transmitting velocity data rather than positions. However, one can recover rotational magnitude by integrating angular velocities, which we have shown is sufficient for identification. Others have found that joint velocities are actually more identifiable than positions [241].
- MetaGuard (see Chapter 7) suggests using differential privacy to randomize particular anthropometric measurements like height and wingspan. This method has no impact on rotation values, which we have shown are sufficient to deanonymize users.

Each of the existing countermeasures was not designed with the understanding that any individual axis of motion data could be sufficient on its own to deanonymize users if a large enough amount of training data is utilized. With this in mind, a truly effective solution must comprehensively anonymize every individual axis present in the motion telemetry stream, as well as all of the identifiable relationships between those dimensions. Manually engineering an adequate solution for each dimension is already on the edge of feasibility with the 21 dimensions tracked by current systems, and becomes completely impractical when given the hundreds of dimensions measured by next-generation full-body tracking systems. Therefore, we are motivated to investigate the use of deep learning to comprehensively anonymize VR telemetry data and construct a more scalable motion anonymization system.

Problem Statement

Having motivated our reasons for wanting to improve VR anonymization techniques beyond the current state of the art, we present in this chapter a new "deep motion masking" approach to VR motion anonymization, which we use to create an improved motion anonymization system. The goals of our new system and approach are as follows:

- **Anonymity**: The primary goal of the system is to prevent users from being identified based on their motion data. Specifically, we invoke the same notion of anonymity as used in MetaGuard (Chapter 7), *cross-session unlinkability*; given motion data with known user identities in a first session, the adversaries relevant to this chapter (see §8.2) should not be able to identify the same set of users using their anonymized motion data from a second session. As in MetaGuard, we assume that adversaries have no other means of linking participant identities across sessions, such as IP addresses.
- **Usability**: The system must not significantly degrade the user experience by anonymizing user motion data. Specifically, we target the strong notion of *indistinguishability* of anonymized motion data from unmodified VR motion data.

We contend that these properties are both necessary and sufficient for a practical VR motion privacy system. Clearly, anonymity is a necessary property of a motion privacy system in order to protect the identity of VR users. In particular, the cross-session unlinkability definition we use prevents adversaries from tracking users from one usage context to another and aggregating an increasingly detailed profile of the user over time. Of course, as discussed in Chapter 2, known VR attacks go beyond the identification of users, and include the ability to profile various personal attributes. However, if anonymized, such attributes will no longer be linkable to the identity of a particular user. Further, a system that is effective at anonymizing users must, in practice, also effectively obscure any set of personal attributes that can be correlated to their identity.

Similarly, the usability of the resulting system is sufficiently ensured by the indistinguishability of anonymized motion data, as anonymized motion data that is indistinguishable from unmodified natural human motion data cannot negatively impact the user experience. If the anonymized motion data diminished the usability of the VR system in any way, it would be distinguishable from unmodified human motion data by virtue of causing said diminution.

In addition to the main properties described above, we note two further "soft" requirements that influenced our design choices. While these properties are technically already encapsulated in the above goals, they serve to further constrain the design of our system and to distinguish its capabilities from those of previous defensive systems like MetaGuard:

- **Scalability**: The anonymization system should comprehensively anonymize every axis of motion data without manually engineering a solution for each feature.
- **Interactivity**: The system should minimize the perceived impact of the anonymization process on the interaction of the user with objects in the virtual world.

With these properties in mind, we now describe our new proposed deep learning architectures for building a "deep motion masking" system.

8.4 Architecture

At a high level, our proposed method involves decomposing the plausible variance of human motion sequences into action-related variance and user-related variance. For this purpose, we train an "action encoder" model, which learns an embedding for the action a user is taking while ignoring the user's identity, and a "user encoder" model, which learns an embedding for the user's identity while ignoring the action they are taking. We then train an "anonymizer" model that anonymizes motion sequences by changing their user embedding without changing their action embedding. Finally, we train a "normalizer" model to remove unwanted noise added by the anonymizer. Each of the models we describe was implemented in Keras [138] and trained using the Adam optimizer [141] with a diminishing learning rate scheduler and early stopping based on an independent validation set. For each training step, and throughout the entirety of this chapter, we provide benchmarking results in §8.7.

Action Similarity

First, we describe our method for measuring the similarity of the "action" performed in two separate VR motion sequences. To achieve this, we train an "action similarity" model using the architecture shown in Figure 8.3. The model is trained as a binary classifier that receives two 30-second telemetry sequences (900 × 21) as input. Each of the sequences is passed through an encoder using the LSTM funnel architecture described in §8.3 to generate a 256-dimensional embedding. The Euclidean distance between these embeddings is then used to output a 1 if the two sequences correspond to the same action, and a 0 otherwise.

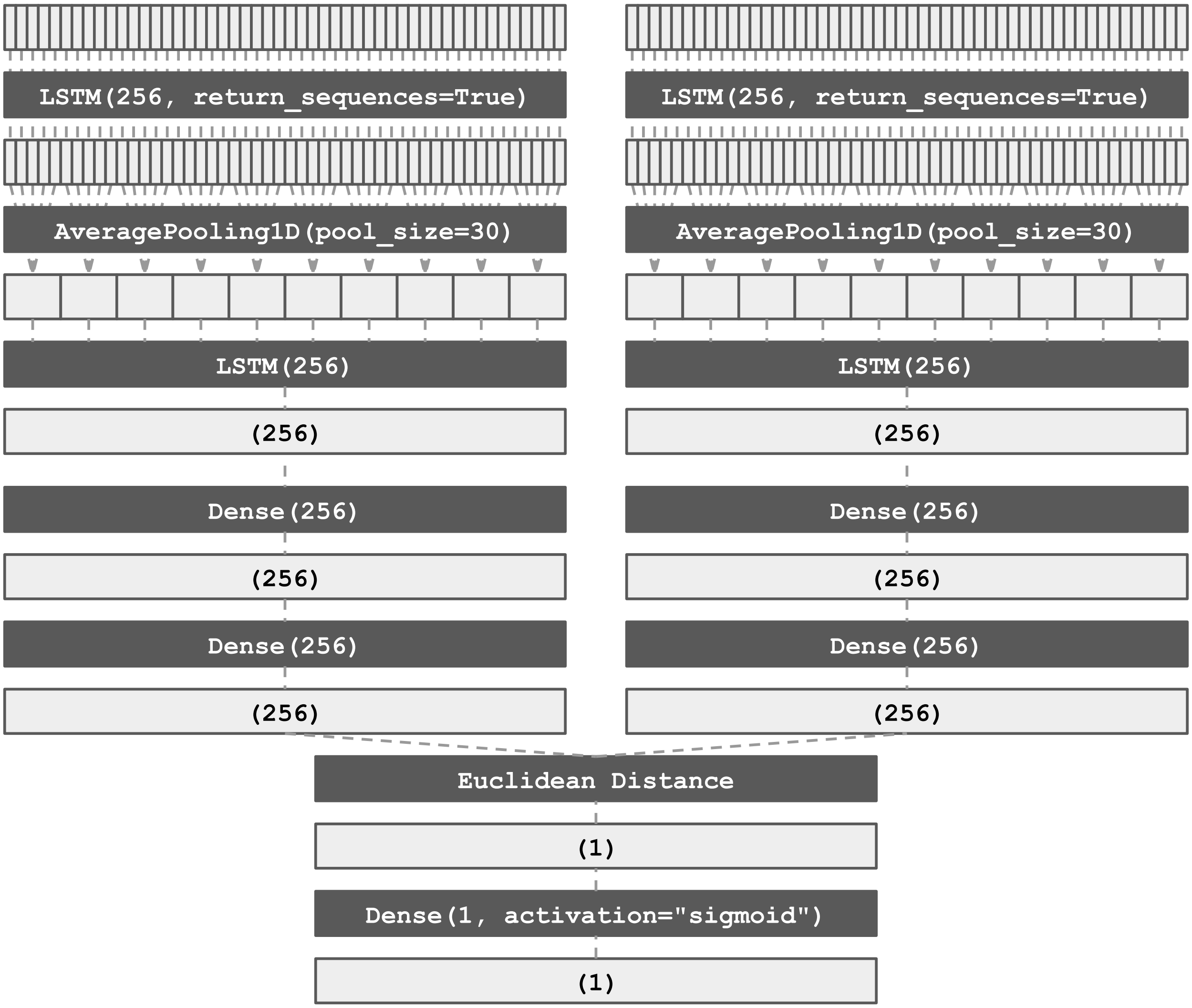

LSTM(256, return_sequences=True)
LSTM(256, return_sequences=True)
AveragePooling1D(pool_size=30)
AveragePooling1D(pool_size=30)
LSTM(256)
LSTM(256)
(256)
(256)
Dense(256)
Dense(256)
(256)
(256)
Dense(256)
Dense(256)
(256)
(256)
Euclidean Distance
(1)
Dense(1, activation="sigmoid")
(1)

Figure 8.3: Siamese architecture for similarity models.

The approach illustrated in Figure 8.3 is sometimes known as a “Siamese neural network” [38]. Siamese architectures have previously been used in VR identification models [181], albeit with CNN layers rather than our LSTM funnel architecture. An advantage of this approach is that while it is trained as a binary classifier for “action similarity,” a limb of the model can later be used on its own as an “action encoder,” such that the Euclidean distance between two output embeddings reveals the similarity of actions in the inputs.

To train the action similarity model, we randomly sampled 50,000 distinct pairs of “similar” motion sequences from the dataset of Chapter 3, and another 50,000 distinct pairs of “dissimilar” motion sequences. An additional 5,000 similar and 5,000 dissimilar pairs were sampled for validation, with a further 5,000 similar and 5,000 dissimilar pairs for testing. For the purpose of defining similarity, we use the “software.activity.id” attribute of the recordings provided in BOXRR-23 [192]. In this case, the attribute corresponds to the exact map the user is playing (see Chapter 3). In every instance, the two motion sequences constituting a pair of inputs originate from different users. The model is thus tasked to classify whether two different users are playing identical or different in-game levels.

When training the action similarity model on the 200,000 motion sequences (50,000 pairs × 2 classes), early stopping occurred after the 156th epoch. The model achieved 100.00% training accuracy, 99.53% validation accuracy, and 99.40% testing accuracy. Therefore, we now have (1) a binary classifier that can determine with 99.4% accuracy whether two motion sequences correspond to the same map, and (2) an action encoder that has learned an approximate metric for measuring the similarity of two motion sequences.

User Similarity

Next, we train a “user similarity” model, which is essentially the inverse of the action similarity model described above. Using the same architecture as before (Figure 8.3), we now randomly sampled 50,000 pairs of motion sequences from the same user, and another 50,000 distinct pairs of motion sequences from different users. Again, an additional 5,000 similar and 5,000 dissimilar pairs were sampled for validation, and 5,000 similar and 5,000 dissimilar pairs for testing. In every instance, the two motion sequences constituting a pair of inputs originate from different in-game maps. The model is thus now tasked to ignore the action and classify whether two motion samples originate from the same or different users.

When training the user similarity model on the 200,000 motion sequences (50,000 pairs × 2 classes) discussed above, early stopping occurred after the 27th epoch. The model achieved 97.94% training accuracy, 92.60% validation accuracy, and 92.81% testing accuracy. Therefore, in addition to the (1) action similarity and (2) action encoder models, we also have (3) a user similarity classifier that can determine with 92.8% accuracy whether two motion sequences correspond to the same user, and (4) a user encoder that has learned a metric for characterizing the user from a motion sequence.

Anonymizer

Using the trained action similarity and user similarity models described above, we can now train the “anonymizer” model that performs the core deep motion masking functionality. The anonymizer model receives as input a 30-second motion telemetry sequence (900×21), and a 32-dimensional noise vector containing random Gaussian noise. It uses these values to output a corresponding 30-second motion sequence (900×21) that is an anonymized version of the input. Our anonymizer model architecture is illustrated in Figure 8.4.

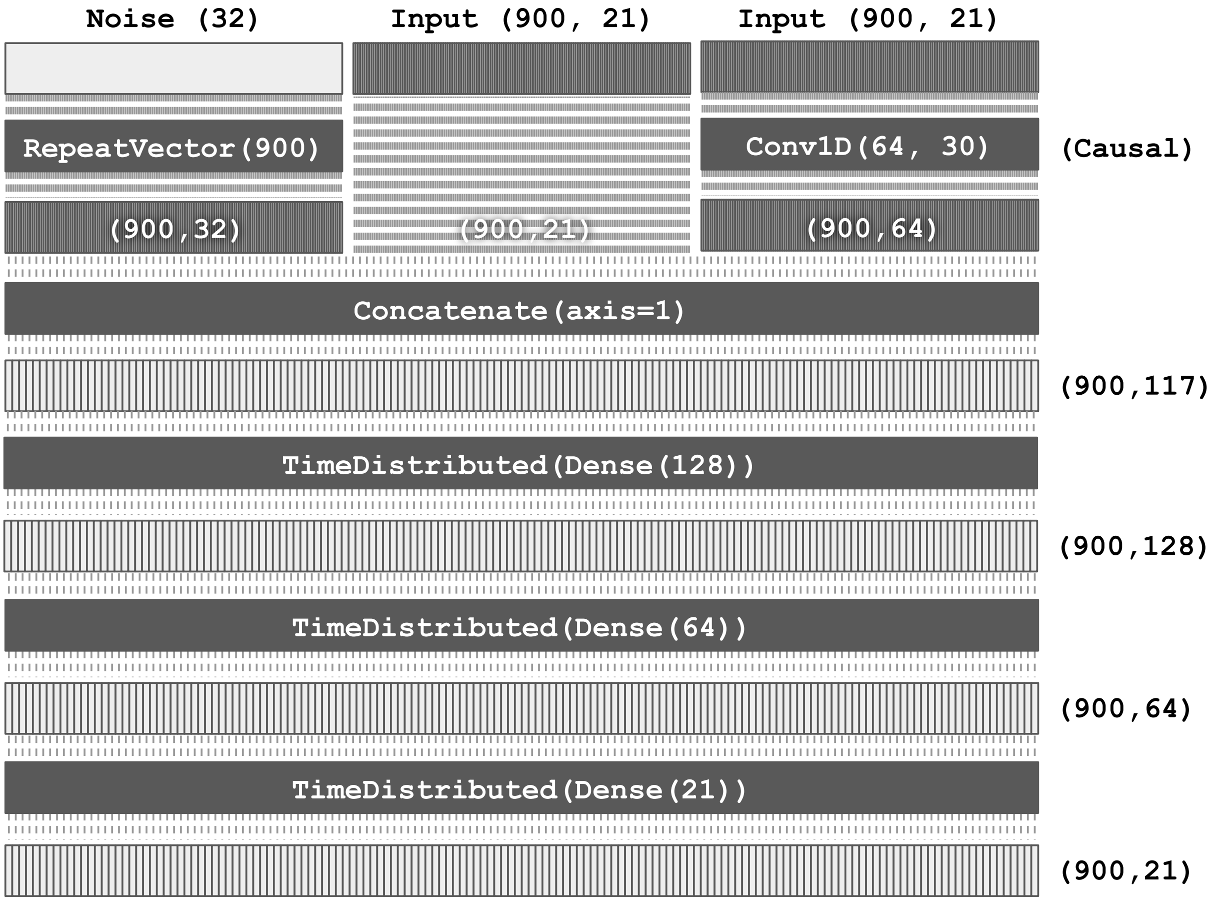


Figure 8.4: Architecture used for anonymizer model.

In addition to the motion input (900×21) and noise (32), which is repeated to produce a (900×32) sequence, a learned 1D convolution (900×64) of the motion input is produced. These three sequences are then vertically concatenated to produce a (900×117) hybrid sequence. Multiple time-distributed dense layers are then used to reduce this sequence back to a (900×21) output motion sequence.

The intuition behind this architecture is that the dense layers effectively combine the noise and motion data to anonymize the motion data in a way that is consistent across each frame, creating a smooth and continuous motion output. This allows the motion to be anonymized in 3D space, but not across the time domain. Therefore, the 1D convolution is added to allow limited manipulation of time-series relationships in the data within a sliding one-second window. Importantly, every component of this architecture respects causality; the model does not have the capability to "look into the future" when producing any output frame. For example, the 1D convolution uses causal padding such that only frames $N - 30$ through N are used in the output of frame N. After training, this allows the resulting anonymizer model to be deployed in real-time on a frame-by-frame basis.

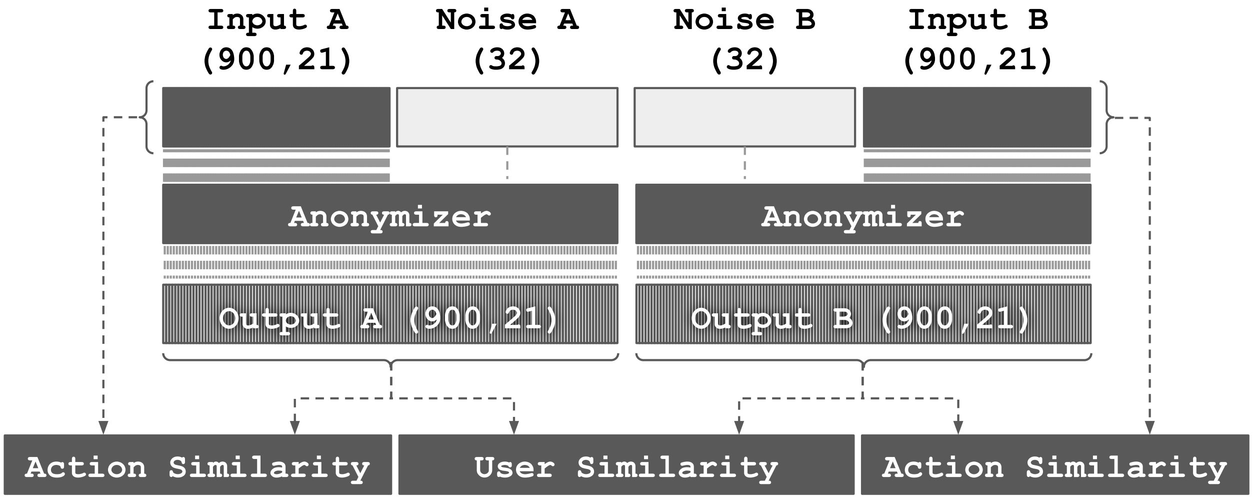


Figure 8.5: Siamese architecture for training anonymizer model.

Figure 8.5 shows how the action similarity and user similarity models are used to train the anonymizer model. First, the anonymizer is pre-trained for 20 epochs as an autoencoder with MSE loss, such that the output frames are initially nearly identical to the inputs, regardless of which noise values are provided. Then, a Siamese architecture is once again used. Leveraging the trained action and user similarity models (the weights of which are now frozen), the anonymizer is trained with the following loss function components:

1. The action embeddings of input_A and output_A should always be as close as possible, irrespective of noise_A; i.e., input_A and output_A are the same action.
2. Similarly, input_B and output_B should always have as close of an action embedding as possible; i.e., input_B and output_B are the same action.
3. If $\mathsf{user}_A = \mathsf{user}_B$ and $\mathsf{noise}_A = \mathsf{noise}_B$, the user embedding for output_A and output_B should be as close as possible; i.e., ouput_A and output_B represent the same faux user.
4. If $\mathsf{user}_A = \mathsf{user}_B$ and $\mathsf{noise}_A \neq \mathsf{noise}_B$, the user embedding for output_A and output_B should be far apart; i.e., ouput_A and output_B represent different faux users.

In other words, the action represented by an anonymized motion sequence should remain unchanged from the original motion sequence, helping to achieve the indistinguishability goal of our model. Furthermore, the intended use of the noise value is to be randomly sampled at the start of each new session, and then to remain consistent within that session. Thus, a user should assume a consistent faux identity within a session, but should assume distinct apparent identities across sessions, achieving cross-session unlinkability. Importantly, by using the adversarial training method in Figure 8.5, the anonymizer receives precise differentiable feedback from the action and user similarity models on how to achieve both of these goals.

An additional advantage of this training method is that it provides a tunable security parameter that can be used to adjust the balance of anonymity and usability while training the model. If additional usability is needed, more weight can be placed on loss components (1) and (2), causing the output motion to appear more similar to the input motion. On the other hand, if more anonymity is required, further weight can be put on loss components (3) and (4), emphasizing cross-session unlinkability of outputs. In our evaluation, we use equal weights for both components, meaning that indistinguishability and cross-session unlinkability are equally important goals.

To train the anonymizer, we randomly sampled 50,000 pairs of motion sequences, with both samples in any given pair coming from the same user. We then randomly sampled 50,000 pairs of random Gaussian noise vectors. For half of the pairs, the noise inputs are identical ($\mathsf{noise}_A = \mathsf{noise}_B$), while for the other half, they are different ($\mathsf{noise}_A \neq \mathsf{noise}_B$), per the loss function described above. An additional 5,000 pairs were sampled for testing. No validation set was used; the model was trained for a full 500 epochs without early stopping.

The model achieved user similarity accuracy of 95.54% on the training data and 94.71% on the testing data. In other words, 94.71% of the time, the model correctly predicted that $\mathsf{user}_A = \mathsf{user}_B$ when $\mathsf{noise}_A = \mathsf{noise}_B$ and that $\mathsf{user}_A \neq \mathsf{user}_B$ when $\mathsf{noise}_A \neq \mathsf{noise}_B$. These numbers should be interpreted in light of the user similarity model's baseline accuracy of 92.81%. Importantly, on both datasets, the model achieved an action similarity accuracy of 100.00%; in every training and testing sample, the action similarity model correctly described the input and output motion as containing the same action.

Normalizer

While the anonymizer is effective at obscuring the identity of a VR user while keeping their actions looking the same, it introduces some undesirable noise to the telemetry signal (at the frame level) due to the lack of an incentive against doing so. One idea for combating this would be to use an adversarial architecture (e.g., GAN [100]) with a discriminator network that provides feedback to the anonymizer by attempting to distinguish anonymized motion from unmodified motion sequences. Unfortunately, we found this idea difficult to apply for our use case as discussed further in §8.8. Instead, we use a normalizer model that aims to reverse the effects of the anonymizer using the architecture in Figure 8.6.

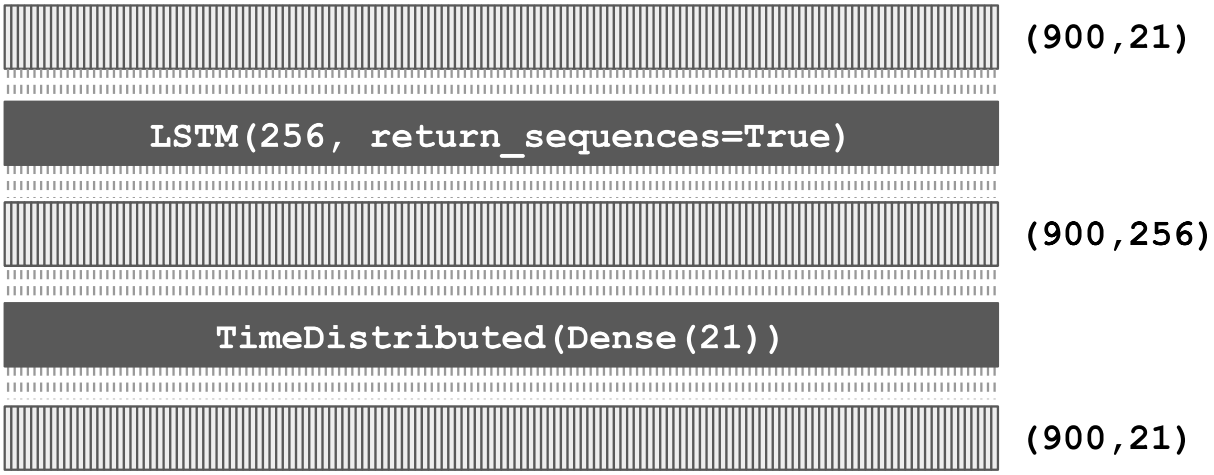


Figure 8.6: Normalizer model architecture.

The normalizer receives as input an anonymized motion sequence (900 × 21) and outputs a smoother-looking normalized anonymous motion sequence (900×21). The relatively simple architecture consists of an LSTM layer that returns a 256-dimensional state for each frame and a time-distributed dense layer that converts each state back to a 21-dimensional output. As with the anonymizer, the architecture obeys causality (e.g., no bidirectional layers) and can therefore be deployed in a real-time setting.

To train the normalizer, we randomly sampled 50,000 motion sequences from random users and maps and anonymized each of them using random noise vectors. We then trained the normalizer using a subset of the anonymized motion sequences as inputs and the corresponding original motion sequences as the target outputs, with a mean squared error loss function. Using a portion of the sequences reserved for testing, we found that the mean squared error between input and output samples after z-score normalizing every dimension was reduced by about one order of magnitude.

Importantly, the normalizer model is not provided with the noise values used to anonymize the original motion sequences, and, during inference, does not have access to the original motion data. Therefore, it will never be able to fully recover the original motion sequences, and cannot reduce the anonymity of the motion sequences, as any deterministic algorithm that could undo the anonymization without access to the original motion or noise values could also be deployed by an adversary to defeat anonymized motion sequences. Instead, the normalizer network can only remove any component of the noise added by the anonymizer that is consistent or predictable across all anonymized motion sequences, which does not affect the actions or anonymity of any particular user.

The entire deep motion masking system architecture, with about 2.2 million parameters, is shown in §8.5. Of these, 290k parameters are in the normalizer, with the action and user similarity models containing nearly one million parameters each. The anonymizer contains only about 65k trainable parameters, allowing it to run extremely quickly on its own.

Deployment

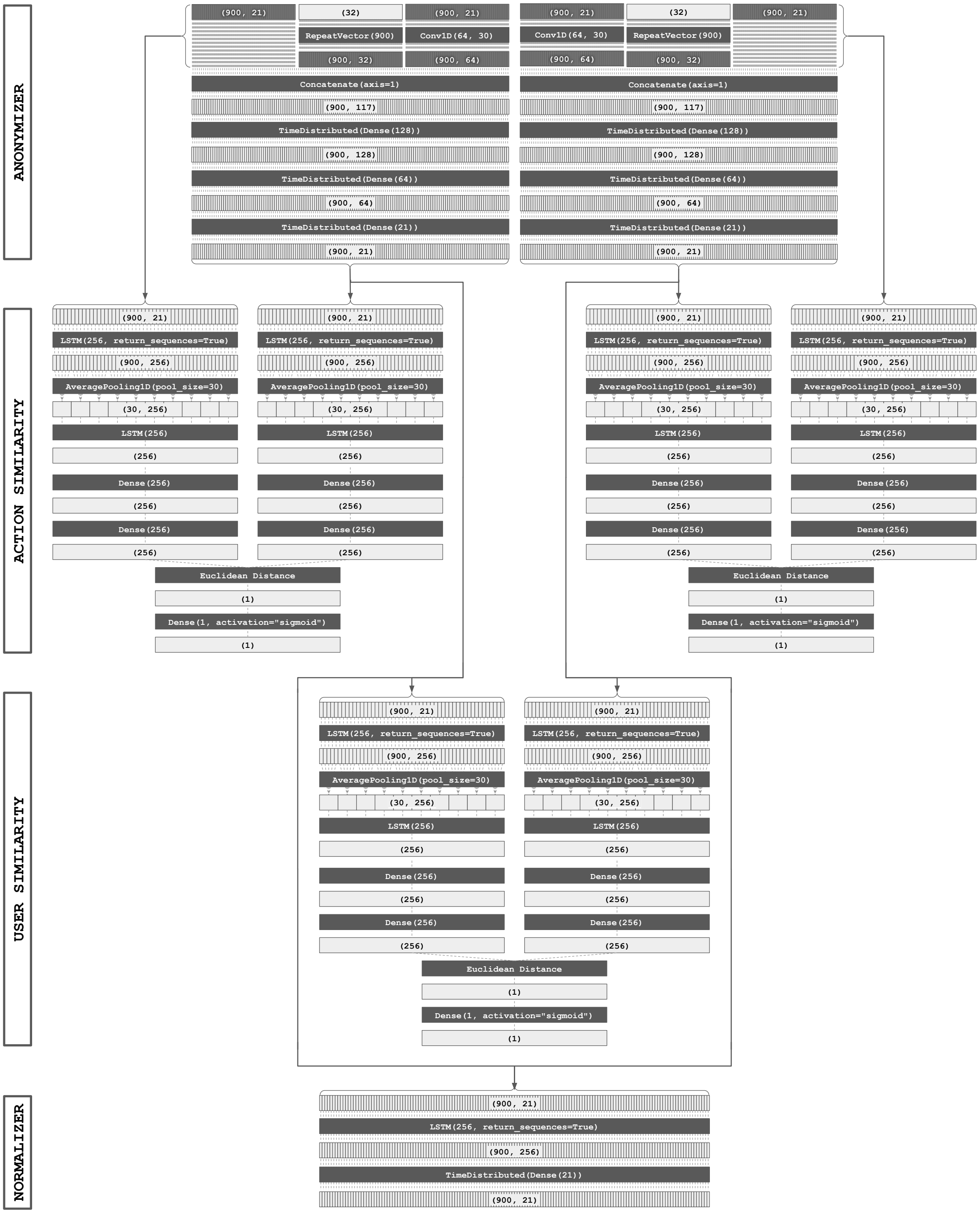
Deploying the trained models for post-hoc anonymization of motion recordings is now as simple as randomly sampling 32 Gaussian noise values, invoking the anonymizer model on the input motion sequence and randomly sampled noise values, and then running the normalizer model on the output of the anonymizer model.

Based on our observations, we suggest a few simple optimizations to the above process. First, we observe that it is better for indistinguishability if the population mean and standard deviation of each motion dimension in anonymized recordings match the population mean and standard deviation of each motion dimension in unmodified motion. This population-level shift does not impact the anonymity of any individual user. Second, we recommend duplicating the first frame of motion 30 times before including the subsequent motion input. This ensures the 1D convolution buffer of the anonymizer model is always filled with real data, reducing apparent noise and instability in the first second of the anonymized output. Finally, the quaternions representing rotational dimensions of the output should be normalized to unit magnitude to maintain validity.

The deep motion masking system can also be used in a real-time (streaming) setting. To do so, a buffer of the last 30 frames should be maintained and initially filled with 30 copies of the first frame. For each new frame, a corresponding anonymized frame can be produced by running the anonymizer's learned 1D convolution on the frame buffer, then concatenating its 64-dimensional output to the 21-dimensional input and 32-dimensional noise vector to produce a 117-dimensional hybrid vector. That hybrid vector can then be converted into a 21-dimensional anonymized output frame using the dense layers of the anonymizer.

Next, the optional optimization of shifting the population mean and standard deviation of each motion dimension back to that of the general population can be applied. Finally, the resulting frame can be fed into the LSTM layer of the normalizer, and the 256-dimensional LSTM state can be used by the dense layer of the normalizer to recover a final 21-dimensional anonymized and normalized output frame. Again, the quaternions should be normalized to unit magnitude. Overall, the real-time deployment of deep motion masking adds no delay other than the computational delay of invoking the anonymizer and normalizer models, which we found to be less than 1 ms per frame. Due to our causal design, the anonymous output in the streaming setting is identical to the result of the post-hoc anonymization process.

8.5 Full Architecture



8.6 Evaluation

Having fully described our proposed deep motion masking approach, we now present a detailed evaluation of the privacy and usability of the resulting system. Our evaluation directly compares the cross-session unlinkability and indistinguishability of our system to that of the MetaGuard system described in the previous chapter.

Anonymity

First, we analyze the impact of our deep motion masking system on cross-session linkability. If the system is effective at anonymizing VR motion data, it should be able to trick our LSTM funnel classification model (§8.3) into wrongly classifying anonymized users in most instances. However, to ensure that our anonymizer didn't overfit by only fooling our own classification model, we also include the Random Forest identification model of Miller et al. [178] and LightGBM-based identification model of Chapter 4.

Furthermore, we train each model both as an oblivious adversary, which is trained on unmodified motion sequences from each user and tested on anonymized motion sequences, and as an adaptive adversary, which is trained on anonymized motion sequences from within a session and tested from anonymized motion sequences in another session. Per our definition of cross-session unlinkability in §8.3, none of the models are trained on multiple independent sessions of anonymized motion, as we operate under the assumption that no external identifiers can be used to link sessions together.

To perform the evaluation, we randomly selected 1,000 users from the dataset of Chapter 3. In order to be representative of average VR users, we only include users for which between 30 and 100 recordings were present; about 20,000 such users exist in the dataset. For each user, we selected 10 recordings to constitute the first session (for training) and another 10 recordings to constitute the second session (for testing). We then anonymized either one or both sessions (depending on the type of adversary), using either MetaGuard or the full post-hoc anonymization pipeline detailed in §8.4. The results of training and testing each of the considered identification models on each set of data are summarized in Table 8.1 below.

	Miller et al. [179]		**50k (Chapter 4)**		**LSTM Funnel (§8.3)**	
	Oblivious	*Adaptive*	*Oblivious*	*Adaptive*	*Oblivious*	*Adaptive*
Unmodified	90.3%	90.3%	91.0%	91.0%	96.5%	96.5%
MetaGuard (§7.3)	57.4%	79.5%	67.0%	84.3%	81.3%	96.3%
DMM (§8.4)	1.5%	1.2%	3.1%	3.5%	3.7%	0.1%

Table 8.1: Identification accuracy for oblivious and adaptive adversaries with three model architectures, with and without anonymization.

As demonstrated by the results of Table 8.1, deep motion masking is significantly better than MetaGuard at anonymizing users across sessions. While MetaGuard users remain up to 96% identifiable, deep motion masking reduces identification accuracy to less than 4%, representing a 20× to over 100× improvement in anonymity depending on the model.

As expected, adaptive adversaries are usually better at identifying anonymized users across sessions, as information about what the user looks like when using the anonymity tool of choice (albeit with different noise values) can be incorporated into the identification model. In the case of MetaGuard, this allows the LSTM funnel architecture to perform at nearly full accuracy, as the model learns to ignore anonymized dimensions and identify users by the unmodified dimensions. It is worth noting that the performance of MetaGuard appears worse now than in Chapter 7, as just 1,000 users are present rather than over 50,000.

Interestingly, however, the LSTM funnel model actually performs significantly worse with the deep motion masking samples when trained adaptively. This is likely because component (3) of the loss function used to train the anonymizer model (§8.4) is measured by a user encoder based on the LSTM funnel architecture. The anonymizer model therefore is particularly good at tricking the LSTM funnel architecture into learning fictitious user attributes and consequently becoming worse at identifying users.

Usability

Next, to evaluate the indistinguishability of motion data anonymized with deep motion masking, we conducted a large-user study (N=182). The study consisted of an online survey in which users were asked to watch VR motion recordings from the game Beat Saber in the Beat Saber web replay viewer tool [224] after reading and agreeing to an informed consent document. Four types of treatments were tested:

1. As a negative control group, we included unmodified VR motion recordings from the dataset of Chapter 3 that will certainly be indistinguishable from natural human motion.
2. As a positive control group, we included completely AI-generated motion recordings created by CyberRamen [227], a machine learning model trained to play Beat Saber. As it stands, these recordings are easily distinguishable from natural motion, serving as a good test of response quality.
3. As a baseline treatment group, we included recordings anonymized using the method of Chapter 7 with the "height," "wingspan," and "room size" defenses enabled at the "medium" privacy settings.
4. As our new treatment group, we included recordings anonymized with deep motion masking using the same models and processes as the anonymity evaluation (§8.6).

As shown in Figure 8.7, users were given one set of recordings at a time, consisting of four recordings of different users playing the same map in Beat Saber (see Chapter 3). To remove confounding variables, all recordings in all sets were first normalized to 30 FPS and trimmed to the first 30 seconds. One of the four recordings in each set was additionally treated (i.e., "anonymized") using one of the four treatments listed above.

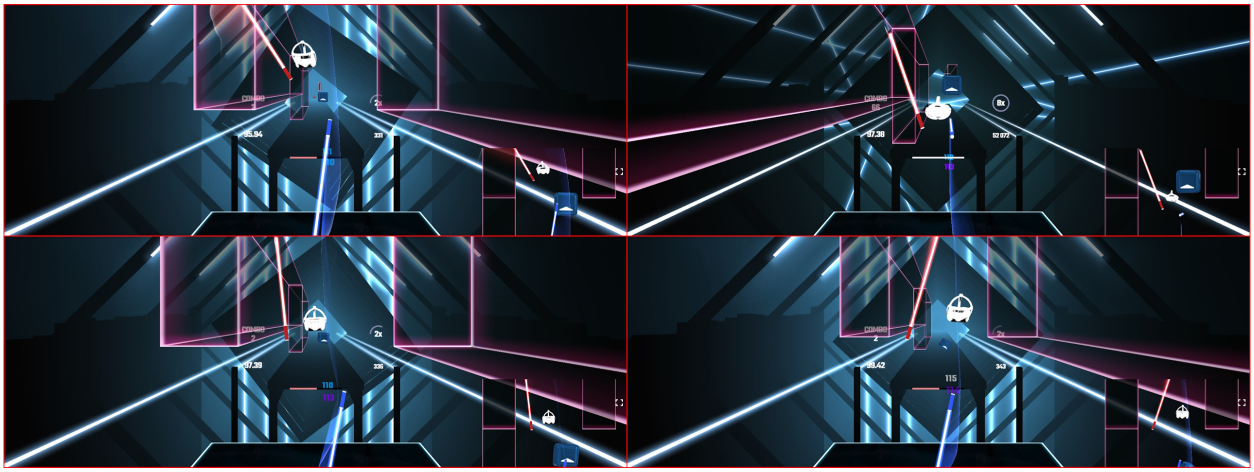


Figure 8.7: A set of Beat Saber replays shown to participants.

Each user was shown 12 such sets of recordings in a randomized order, corresponding to a slow, medium, and fast song for each of the four treatment groups described above. For each set, their task was to decide which (if any) of the four recordings was modified. To aid their decision, users could view each replay in slow motion, zoom in on particular areas, and turn to view the motion from a variety of perspectives.

When recruiting participants for our study, we focused primarily on finding VR users with significant Beat Saber experience, as such users are more familiar with what natural VR motion data should look like, and thus are likely to be more challenging and discerning critics of our system. With that in mind, we primarily recruited participants through social media pages related to VR, and through VR interest groups like CVRE [49]. However, we also wanted to ensure that some number of novice users participated in the study, and recruited a small number of participants from a broader general population for that purpose.

The study ran for two weeks, from September 20th, 2023 through October 3rd, 2023, and received 241 responses in that time. We removed the 59 responses that were either blank or answered all six of the control questions incorrectly, leaving 182 valid responses. Of those, 149 were from expert Beat Saber players (with 100 or more hours of in-game experience), and the remaining 33 participants were novices (with 0 to 100 hours of experience). Figure 8.8 shows the observed distinguishability for each of the evaluated treatments.

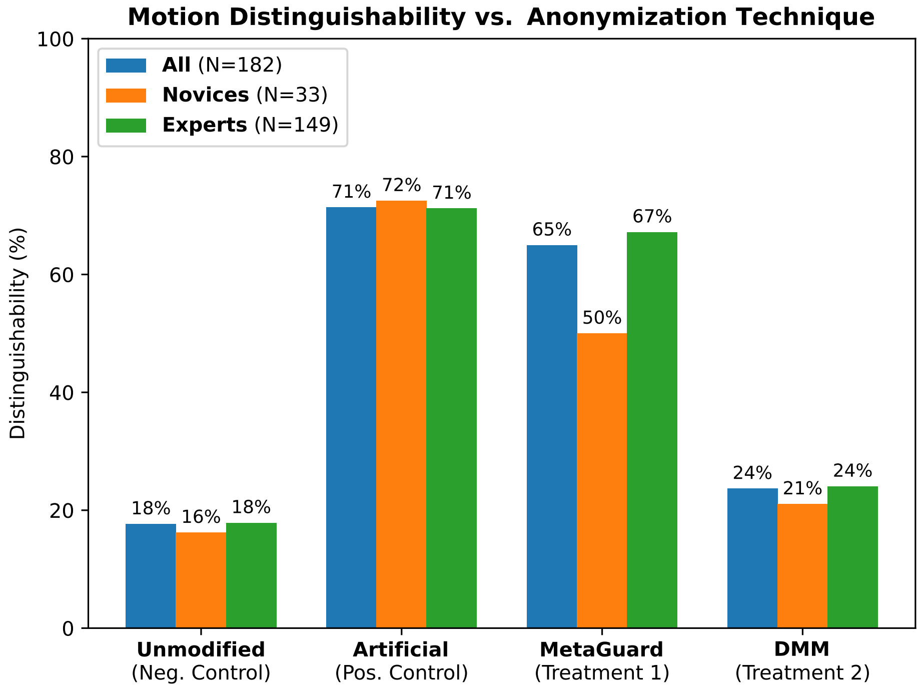


Figure 8.8: Results of indistinguishability user study.

The negative control group has a surprisingly high rate of distinguishability in our results (18%). This indicates that when unsure about which replay was modified, users in our study were prone to randomly guessing one of the four replays rather than indicating that all four replays were unmodified. With that in mind, unfortunately, the replays anonymized with deep motion masking were still not perfectly indistinguishable from natural motion, but were only marginally more distinguishable than the negative control group. Moreover, deep motion masking represents a significant improvement over the MetaGuard system, with nearly a 3× reduction in the rate of distinguishability, particularly for expert users. Using both a standard χ^2 test and Fisher's exact test [84], the difference between MetaGuard and deep motion masking is highly statistically significant with $p < 0.01$.

Interactivity

The indistinguishability study of §8.6 already demonstrates that the deep motion masking anonymizer has minimal impact on observed interactions between users and virtual objects, as participants in that study could view users interacting with blocks in Beat Saber when determining whether a motion sample was modified. However, to enhance the explainability of the user study results and further demonstrate that our deep motion masking system satisfies the stated goal of interactivity, we conducted additional in vitro experiments in which we simulated the effects of deep motion masking on interactions with virtual objects.

SimSaber [47] is a Python library that simulates Beat Saber gameplay by faithfully replicating the physics and collision detection algorithms used by Beat Saber and the Unity game engine [280], as shown in Figure 8.9. We randomly sampled 1,000 Beat Saber replays from the dataset of Chapter 3, and anonymized them with deep motion masking. We then ran the original and modified replays through the simulator to evaluate what impact the anonymization process had on user interactions with the virtual blocks in Beat Saber.

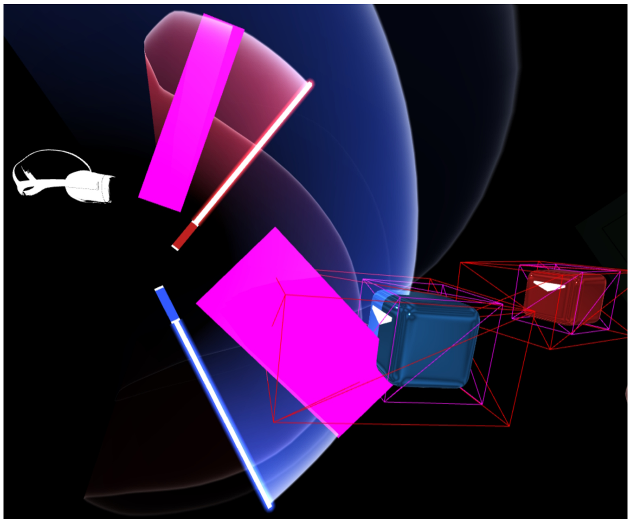


Figure 8.9: Collision modeling for Beat Saber objects.

In Beat Saber, the cutting of a block with a saber is typically characterized by the player's pre-swing angle, post-swing angle, and accuracy (closeness to the center of the block). The combination of these three factors is used to calculate the player's score. At a minimum, a usable anonymization tool should not significantly impact these three measurements in order to avoid substantially affecting the user's performance.

In our evaluation of 1,000 replays, we found that anonymized players had a mean absolute difference in pre-swing angle of about 5°, and an average relative difference of 4.5%. The mean absolute difference in post-swing angle was about 4°, and an average relative difference of 6.7%. The closeness to the center of the block was modified by a mean absolute difference of about 6.5 cm, resulting in an average relative accuracy difference of 14%. Overall, the mean absolute difference in the player's score after anonymization was only 0.7%, a difference that should be unnoticeable for all but the most experienced players. Still, our system may not be suitable for situations requiring extreme precision (see §8.8).

These findings complement our indistinguishability results of §8.6 by demonstrating that the deep motion masking system is able to maintain approximate apparent interactions with in-game objects, despite having no direct information about virtual object positions and geometries. By incorporating an action similarity metric (§8.4), the model simply learns to avoid making changes that are likely to change the semantic meaning of the motion. As a result, viewers struggle to distinguish the anonymized motion from that of a real user.

Ethics

The primary source of data for this study is the BOXRR-23 dataset of Chapter 3, a publicly available dataset intended for use in VR research, including security and privacy research. This dataset has already been used in published research papers in the VR security and privacy domain [196]. It contains built-in privacy measures, such as pseudonymization of participants, and was reviewed by the legal and ethics boards of its authors prior to release. We specifically only use the BeatLeader part of the dataset in our research; these users explicitly consent to the use of their data for "research topics such as VR security, privacy, and usability" in the BeatLeader privacy policy [223].

Other than the BOXRR-23 dataset, the only additional data used in this chapter is from our usability study in §8.6. All participants in the survey were adults over the age of 18, and no vulnerable populations were specifically targeted in this study. Participants consented to their inclusion in academic research by reading and agreeing to an informed consent document before proceeding in the survey. Users optionally provided their Beat Saber username, but no further identifiable information was collected. Information collected consisted exclusively of the users' selections of which recordings they believed were modified. Therefore, the likelihood of any harm to participants, either through participating or through a later breach of confidentiality, is exceedingly low.

All aspects of this study, including our use of the public BeatLeader data and our collection of survey responses in §8.6, were also independently reviewed and approved by an OHRP-certified IRB under protocol number 2023-06-16467.

8.7 Benchmarking

For all experiments described in this chapter, we used a desktop computer running Windows 10 v22H2 with 128 GB of 2133 MHz DDR4 RAM, an AMD Ryzen 9 5950X CPU (16 cores, 3.40 GHz), and an NVIDIA GeForce RTX 3090 GPU (10496 CUDA cores, 24 GB VRAM). The time required to run the experiments in each section was as follows:

Motivation (§8.3)

- Preprocessing the BOXRR-23 dataset to sample and normalize 500 replays each for 500 users took **37h 14m**.
- Training and testing the LSTM funnel models took **3h 50m**.
- Featurization for the Miller et al. [178] Random Forest model took **2h 54m**, and training and testing the model took **4m 8s**.
- Featurization for the LightGBM model from Chapter 4 took **16h 32m**, and training and testing the model took **13m 47s**.

Method (§8.4)

- Preprocessing the BOXRR-23 dataset to sample the action similarity features took **55h 40m**. Training and testing the action similarity model took **3h 26m**.
- Preprocessing the BOXRR-23 dataset to sample the user similarity features took **57h 12m**. Training and testing the user similarity model took **1h 18m**.
- Training and testing the anonymizer model took **3h 15m**.
- Training and testing the normalizer model took **56m 51s**.

Anonymity (§8.6)

- Preprocessing the BOXRR-23 dataset to sample and normalize 20 replays each for 1000 users took **3h 24m**.
- Featurization for the Miller et al. [178] Random Forest model took **40m 51s**, and training and testing the model took **5m 9s**.
- Featurization for the LightGBM model from Chapter 4 took **3h 42m**, and training and testing the model took **1h 12m**.
- Training and testing the LSTM funnel model took **8m 52s**.

Interactivity (§8.6)

- Preprocessing the dataset to sample and anonymize 1,000 replays took **37m 10s**.
- Using SimSaber to simulate the 1,000 replays before and after masking took **4m 32s**.

Overall, the total compute time required was about **192h 52m**.

8.8 Discussion

Anonymizing VR motion data inherently involves diverging from the original motion data to some extent. The approach detailed in §8.4 ensures that such deviations correspond mostly to apparent differences not in the actions being taken but rather in the user taking the actions. This results in the system being highly suitable for motion data intended for consumption by human observers, as demonstrated in §8.6.

On the other hand, there will always be VR applications in which very high precision is required, such as telemedicine, competitive e-sports, or remote operation of equipment. In such situations, the average discrepancies measured in §8.6 of 6.5 cm (position) and 5° (rotation) may be intolerable. If anonymity is still desired in such an application, an alternative solution, such as MPC or TEEs, may be more suitable.

Overall, we recommend a two-channel approach for VR motion data, with one system handling real-time anonymization of low-fidelity motion for human eyes, and another handling precise motion data for asynchronous computational use. Deep motion masking presents a secure, usable, and scalable solution for the former scenario, while the latter merits further investigation by researchers in future work.

Finally, one may wonder why a GAN architecture was not used in this work. While GANs theoretically could be a great way to ensure anonymized motion data remains indistinguishable, we found them to not work well in practice for this dataset, because the goal of computational indistinguishability is too strong to be practical. While data anonymized with our deep motion masking system is almost perfectly indistinguishable to the human eye, it can still be distinguished by a machine learning classifier with almost 100% accuracy. Thus, regardless of which combinations of architectures and learning rates we tried, a GAN always resulted in the generator ceasing to make progress as the discriminator reached 100% accuracy. However, we leave open the possibility that a GAN could work in this application, and perhaps produce even better results, if used in a way that we did not consider.

Limitations

One major limitation of our system is that it has only been trained on data from a single VR application, Beat Saber. This is because there are currently about four orders of magnitude more motion data available from Beat Saber than any other VR application, with deep learning models benefiting from large amounts of training data. Unlike prior work using this dataset, we don't allow our model to see anything specific to Beat Saber, such as block positions and timings. Therefore, it should be possible to train a deep motion masking model, using the present architecture, on motion data from any VR application, if enough motion data were available. However, without such data, we cannot confidently claim that the evaluation results will generalize to other applications.

Another major limitation of deep motion masking is that it loses the provable security properties of MetaGuard, as highlighted in Chapter 7. One of the most significant features of MetaGuard is that it obeys ε-differential privacy, and thus provides provable security and privacy properties. However, that provability only extends to the specific anthropometric measurements that we consider in Chapter 7. As demonstrated in §8.3, this creates a weakness, as rotational dimensions are excluded entirely. Thus, while proving the security of our deep learning approach is significantly harder, the method empirically provides better cross-session unlinkability than MetaGuard as demonstrated in §8.6.

8.9 Conclusion

Deep learning is increasingly emerging as a powerful method for the usable real-time anonymization of sequential data (e.g., voice anonymization [57]). In this chapter, we've shown that deep learning can also be an effective tool for anonymizing VR telemetry data by developing a technique we call deep motion masking, which is analogous to a real-time voice changer for movement patterns. By decomposing the space of motion variability into action-related variation and user-related variation, our model is effective at hiding user identity while maintaining action similarity, leading to empirically better indistinguishability and cross-session unlinkability the differential privacy method we presented in Chapter 7.

Chapter 9

Conclusion

9.1 Summary of Contributions

In this dissertation, we have shown that head and hand motion, the data most fundamental to nearly all XR applications, carries an unprecedented set of security and privacy risks. We presented a comprehensive information flow and threat model for XR privacy research, illustrating the various entities that all have the potential to misuse XR telemetry data.

Next, we presented BOXRR-23, a new XR motion dataset with orders of magnitude more users than any comparable research dataset. We then used this dataset to drive improvements in VR identification, achieving a 94.33% identification accuracy of 55,541 VR users from head and hand motion data, as well as profiling, inferring over 40 private personal user data attributes from over 1,000 users. We further demonstrated that an active adversary can perform even more invasive attacks by manipulating the immersive virtual environment.

Finally, we presented two distinct approaches for anonymizing VR motion data, one based on local ε-differential privacy, and one based on deep learning. The former offers provable privacy guarantees that adhere to a theoretical optimality, while the latter offers better empirical performance and the potential to scale to full-body motion data.

Through the course of this research, we have not only gained a better understanding of the current XR security and privacy posture but also armed the community with tools to combat existing privacy risks. We are optimistic that at this early stage in XR development, it is not too late to build privacy protections into the core of the XR technology stack.

9.2 Availability

As discussed in Chapter 2, the lack of reproducibility is a significant problem in the VR privacy domain. This dissertation thus prioritizes offering transparent, reproducible results. Every study in this dissertation is associated with free, open-source code, offered under permissive licenses, and can be replicated using the open-access BOXRR-23 dataset.

BOXRR-23 Dataset (Chapter 3): Researchers may access the BOXRR-23 dataset, subject to a license agreement and data use agreement, through our website:

`https://rdi.berkeley.edu/metaverse/boxrr-23`

The permanent DOI is `https://doi.org/10.25350/B5NP4V`. The source code to parse the XROR files in the dataset is available here: `https://github.com/metaguard/xror`

Identification Study (Chapter 4): The source code for all parts of our 50,000 user identification study is available here: `https://github.com/MetaGuard/Identification`

This artifact was reviewed by the USENIX Security '23 Artifact Evaluation Committee (AEC), and received all three artifact badges.

ARTIFACT EVALUATED usenix ASSOCIATION AVAILABLE

ARTIFACT EVALUATED usenix ASSOCIATION FUNCTIONAL

ARTIFACT EVALUATED usenix ASSOCIATION REPRODUCED

Profiling Study (Chapter 5): The source code for the transformer models in our attribute inference study is available here: `https://github.com/MetaGuard/Profiling`

MetaData Study (Chapter 6): The Unity project files for our adversarial escape room game are available here: `https://github.com/metaguard/metadata`

This artifact was reviewed and approved by the PETS '23 Artifact Review Committee.

Artifact

MetaGuard Study (Chapter 7): The source code for the MetaGuard study is available here: `https://github.com/metaguard/metaguard`

Deep Motion Masking Study (Chapter 8): The source code for all aspects of the Deep Motion Masking study is available here: `https://github.com/metaguard/metaguardplus`

9.3 Future Work

VR Privacy Attacks

A major focus of this dissertation was on identifying XR users from their head and hand motion data. While this capability was presented as a potential privacy risk, it may also be productively deployed for use cases such as passive authentication. There are also several interesting applications of our techniques to Beat Saber specifically, as well as VR gaming in general. These include advanced cheating detection, score prediction, skill-based matchmaking, and map recommendation engines.

Despite our best efforts, the actual functionality of the identification and inference models presented in this dissertation remains somewhat opaque. While deep learning models are notoriously difficult to explain, we hope to see future work that uses advanced model explainability techniques to better understand the mechanisms underlying our results.

Beyond identification and inference, we presented examples of adversarial XR application design at a proof of concept stage. Future work could demonstrate how developers can design XR games or applications that make privacy attacks even more stealthy, including by integrating these attacks into daily tasks in future XR environments, or by integrating additional data modalities that we did not consider, such as eye tracking and full-body tracking. Most concerningly, future work could demonstrate that an active attacker can not only predict but actually change users' opinions on sensitive subjects. On the other hand, researchers should also study analysis techniques for revealing hidden data collection mechanisms (where possible) to make these attacks harder to achieve.

Beyond head and hand motion data, there are many XR threat vectors that were not explored in this dissertation. These of course include full-body tracking and eye tracking, which have been explored by other researchers, but also more obscure attacks, such as fingerprinting users based on the background audio in their environment. Comprehensive XR privacy solutions will need to understand and manage the privacy implications of the entire multi-modal data stream generated by XR devices.

While we demonstrated the feasibility of adversarial VR game design for inferring rich user attributes, our defenses were primarily focused on passive observers. In future work, researchers should develop concrete countermeasures against malicious XR content design while achieving an appropriate balance between flexibility and consumer protection. Once again, such countermeasures must understand the risks of each output modality, including audio, stereoscopic vision, haptic feedback, etc.

Lacking access to VR device firmware, we implemented the defenses described in this dissertation at the client software layer, providing an effective defense against server and user attackers. In future work, we believe the same defenses could be applied at the firmware level, allowing data to also be protected from client attackers. However, protecting data from hardware or firmware-level adversaries will likely require entirely different methods to the ones presented in this dissertation.

VR Privacy Defenses

An important aspect of the MetaGuard system is the ability for users to toggle individual VR defenses according to the requirements of the application being used. While this process is manual in our implementation, in the future, the "incognito mode" system could be configured to automatically profile VR applications and determine which defenses are appropriate for a given scenario. Furthermore, the application could incorporate the differential privacy concept of a "privacy budget," adding more noise to enabled attributes to compensate for the privacy loss of disabled attributes and maintain the same level of overall anonymity.

One important area of future work in this field is extending motion anonymization systems support to full-body tracking data. Deep motion masking is particularly suitable for this purpose, as it doesn't involve manually engineering features between pairs of tracked objects, and may in fact be immediately applicable to full-body telemetry streams. At present, we lack a large full-body motion capture dataset to use for training. However, as next-generation VR devices adopt full-body tracking, such data may become more available, and the importance of full-body motion anonymization will simultaneously increase.

On the subject of data, future work may focus on procuring large-scale VR motion datasets from applications other than Beat Saber. Demonstrating the generalizability of deep motion masking to a wide variety of VR games and applications is an important step toward the potential adoption of such a system. Other machine learning architectures, such as diffusion or transformer models, could also be useful, although inference latency may become a concern. We hope to see future work that explores various other architectures and techniques for masking VR motion data.

Finally, another defense worth exploring is the use of trusted execution environments (TEEs) to provide auditability for metaverse servers that utilize telemetry data. TEEs like Intel's SGX could provide a hardware-based attestation mechanism that allows users to check that servers only use their motion data for legitimate purposes. At the server or client level, TEEs might enhance privacy without modifying XR motion data in ways that the defenses proposed in this dissertation cannot.

9.4 Moving Forward

XR technology is currently on track to become a ubiquitous means of accessing the internet, with AR devices having the potential to replace most of the existing portable electronic devices a consumer would typically carry today. With the forthcoming introduction of Apple into the XR device market, plus tens of billions of dollars in annual research and development expenditure from existing players like Meta, Microsoft, Google, Valve, and HTC, some of the largest and most influential technology companies on earth are clearly betting big on XR playing a significant role in the future of human-computer interaction.

Given that several of the major players in the metaverse space have their roots in advertising, the temptation will surely exist to leverage existing sales channels to monetize metaverse user data. Thus, we are currently at a crossroads. If nothing is done to improve the metaverse's present security and privacy posture, it is poised to inherit an exaggerated version of the privacy issues that are prevalent on the web. However, if we take the opportunity to learn from the history of browser-based attacks and defenses, security and privacy practitioners can prioritize research in this field and build privacy-preserving mechanisms into the fabric of the metaverse before the theoretical threats actually become widespread.

The strong incentives against XR privacy today make technical solutions unlikely to be sufficient on their own. On the other hand, XR policy should be constructed carefully to avoid crippling this burgeoning industry. Technologists and policymakers must therefore work hand in hand to develop and implement user-centric solutions that lay the groundwork for a ubiquitous metaverse that is secure and private while remaining usable and personal.

Bibliography

[1] Parastoo Abtahi et al. “Beyond Being Real: A Sensorimotor Control Perspective on Interactions in Virtual Reality”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI ’22. New Orleans, LA, USA: Association for Computing Machinery, 2022. ISBN: 9781450391573. DOI: 10.1145/3491102.3517706. URL: https://doi.org/10.1145/3491102.3517706.

[2] *ACCAD MoCap System and Data*. URL: https://accad.osu.edu/research/motion-lab/mocap-system-and-data.

[3] ACM. *ACM Digital Library*. https://dl.acm.org/. Online; accessed 10 Aug 2022.

[4] Devon Adams et al. “Ethics Emerging: the Story of Privacy and Security Perceptions in Virtual Reality”. In: *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. Baltimore, MD: USENIX Association, 2018, pp. 427–442. ISBN: 978-1-939133-10-6. URL: https://www.usenix.org/conference/soups2018/presentation/adams.

[5] Gaurav Aggarwal et al. “An Analysis of Private Browsing Modes in Modern Browsers”. In: *Proceedings of the 19th USENIX Conference on Security*. USENIX Security’10. Washington, DC: USENIX Association, 2010, p. 6.

[6] Ijaz Akhter and Michael J. Black. “Pose-conditioned joint angle limits for 3D human pose reconstruction”. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 1446–1455. DOI: 10.1109/CVPR.2015.7298751.

[7] Rachel Albert et al. “Latency Requirements for Foveated Rendering in Virtual Reality”. In: *ACM Trans. Appl. Percept.* 14.4 (2017). ISSN: 1544-3558. DOI: 10.1145/3127589. URL: https://doi.org/10.1145/3127589.

[8] Khaled Albishre, Mubarak Albathan, and Yuefeng Li. “Effective 20 newsgroups dataset cleaning”. In: *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. Vol. 3. IEEE. 2015, pp. 98–101. DOI: 10.1109/WI-IAT.2015.90.

[9] Nadisha-Marie Aliman and Leon Kester. “Malicious Design in AIVR, Falsehood and Cybersecurity-oriented Immersive Defenses”. In: *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 2020, pp. 130–137. DOI: 10.1109/AIVR50618.2020.00031.

[10] Amelia. *Virtual Reality Solution for Mental Health Professionals*. Online; accessed 22 Sep 2022. URL: https://ameliavirtualcare.com/virtual-reality-solution-psychology/.

[11] Leonardo Angelini et al. "Towards an Emotionally Augmented Metaverse: A Framework for Recording and Analysing Physiological Data and User Behaviour". In: *13th Augmented Human International Conference*. AH2022. Winnipeg, MB, Canada: Association for Computing Machinery, 2022. ISBN: 9781450396592. DOI: 10.1145/3532530.3532546. URL: https://doi.org/10.1145/3532530.3532546.

[12] Abdullah Al Arafat, Zhishan Guo, and Amro Awad. "VR-Spy: A Side-Channel Attack on Virtual Key-Logging in VR Headsets". In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 2021, pp. 564–572. DOI: 10.1109/VR50410.2021.00081.

[13] Maryam Archie et al. "Who′s Watching? De-anonymization of Netflix Reviews using Amazon Reviews". In: 2018. URL: https://courses.csail.mit.edu/6.857/2018/project/Archie-Gershon-Katchoff-Zeng-Netflix.pdf.

[14] R. Azuma et al. "Recent advances in augmented reality". In: *IEEE Computer Graphics and Applications* 21.6 (2001), pp. 34–47. DOI: 10.1109/38.963459.

[15] Ronald T. Azuma. "A Survey of Augmented Reality". In: *Presence: Teleoper. Virtual Environ.* 6.4 (1997), pp. 355–385. ISSN: 1054-7460. DOI: 10.1162/pres.1997.6.4.355. URL: https://doi.org/10.1162/pres.1997.6.4.355.

[16] Stefano Baldassi et al. *Challenges and New Directions in Augmented Reality, Computer Security, and Neuroscience – Part 1: Risks to Sensation and Perception*. 2018. DOI: 10.48550/ARXIV.1806.10557. URL: https://arxiv.org/abs/1806.10557.

[17] Matthew Ball. *The Metaverse: And How It Will Revolutionize Everything*. Minneapolis: Norton & Company, 2022.

[18] *BeatSaver*. en. URL: https://beatsaver.com/ (visited on 01/30/2023).

[19] Kory Becker. *primaryobjects/voice-gender*. original-date: 2016-06-09T14:30:44Z. 2022. URL: https://github.com/primaryobjects/voice-gender (visited on 05/25/2022).

[20] Steve Benford et al. "Shared Spaces: Transportation, Artificiality, and Spatiality". In: *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work*. CSCW '96. Boston, Massachusetts, USA: Association for Computing Machinery, 1996, pp. 77–86. ISBN: 0897917650. DOI: 10.1145/240080.240196. URL: https://doi.org/10.1145/240080.240196.

[21] Hal Berghel. "Malice domestic: The Cambridge analytica dystopia". In: *Computer* 51.05 (2018), pp. 84–89.

[22] Guillermo Bernal et al. "Galea: A physiological sensing system for behavioral research in Virtual Environments". In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2022, pp. 66–76. DOI: 10.1109/VR51125.2022.00024.

[23] bHaptics. *TactSuit X40*. Online; accessed 22 Sep 2022. URL: `https://www.bhaptics.com/tactsuit/tactsuit-x40`.

[24] Blur Busters. *UFO Motion Tests*. `https://www.testufo.com/`. Online; accessed 30 April 2022.

[25] Courtney Bowman et al. *The Architecture of Privacy. On Engineering Technologies that can deliver trustworthy safeguards*. O'Reilly, 2015.

[26] Efe Bozkir et al. "Differential privacy for eye tracking with temporal correlations". In: *PLOS ONE* 16.8 (Aug. 17, 2021). Ed. by Luca Citi, e0255979. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0255979`. URL: `https://dx.plos.org/10.1371/journal.pone.0255979` (visited on 09/20/2022).

[27] Bracket Foundation. *Gaming and the Metaverse: The Alarming Rise of Online Sexual Exploitation and Abuse of Children Within the NEw Digital Frontier*. Online; accessed 10 Oct 2022. URL: `https://www.weprotect.org/wp-content/uploads/Gaming_and_the_Metaverse_Report_final.pdf`.

[28] Brave Software, Inc. *Brave*. `https://brave.com/`. Online; accessed 21 July 2022. 2023.

[29] Leo Breiman. "Random forests". In: *Machine learning* 45 (2001), pp. 5–32. DOI: `10.1023/A:1010933404324`.

[30] Karla Brkic et al. "I Know That Person: Generative Full Body and Face De-identification of People in Images". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2017, pp. 1319–1328. DOI: `10.1109/CVPRW.2017.173`.

[31] US Census Bureau. *2020 Census Results*. Section: Government. 2020. URL: `https://www.census.gov/2020results` (visited on 06/08/2023).

[32] Aaron Cahn et al. "An Empirical Study of Web Cookies". In: *Proceedings of the 25th International Conference on World Wide Web*. WWW '16. Montréal, Québec, Canada: International World Wide Web Conferences Steering Committee, 2016, pp. 891–901. ISBN: 9781450341431. DOI: `10.1145/2872427.2882991`. URL: `https://doi.org/10.1145/2872427.2882991`.

[33] Stuart K. Card, George G. Robertson, and Jock D. Mackinlay. "The Information Visualizer, an Information Workspace". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '91. New Orleans, Louisiana, USA: Association for Computing Machinery, 1991, pp. 181–186. ISBN: 0897913833. DOI: `10.1145/108844.108874`. URL: `https://doi.org/10.1145/108844.108874`.

[34] Peter Casey, Ibrahim Baggili, and Ananya Yarramreddy. "Immersive Virtual Reality Attacks and the Human Joystick". In: *IEEE Transactions on Dependable and Secure Computing* 18.2 (2021), pp. 550–562. DOI: `10.1109/TDSC.2019.2907942`.

[35] Center of Disease Control and Prevention. *Percentile Data Files with LMS Values.* `https://www.cdc.gov/growthcharts/percentile_data_files.htm`. Online; accessed 17 July 2022. 2022.

[36] Aayush Kumar Chaudhary and Jeff B Pelz. "Privacy-Preserving Eye Videos Using Rubber Sheet Model". In: *ACM Symposium on Eye Tracking Research and Applications.* ETRA '20 Short Papers. Stuttgart, Germany: Association for Computing Machinery, 2020. ISBN: 9781450371346. DOI: `10.1145/3379156.3391375`. URL: `https://doi-org.eaccess.ub.tum.de/10.1145/3379156.3391375`.

[37] Tianqi Chen et al. "Xgboost: extreme gradient boosting". In: *R package version 0.4-2* 1.4 (2015), pp. 1–4. URL: `https://cran.microsoft.com/snapshot/2017-12-11/web/packages/xgboost/vignettes/xgboost.pdf`.

[38] Davide Chicco. "Siamese Neural Networks: An Overview". In: *Artificial Neural Networks.* New York, NY: Springer US, 2021, pp. 73–94. ISBN: 978-1-0716-0826-5. DOI: `10.1007/978-1-0716-0826-5_3`. URL: `https://doi.org/10.1007/978-1-0716-0826-5_3`.

[39] Hang Chu et al. "Expressive Telepresence via Modular Codec Avatars". In: *Computer Vision – ECCV 2020.* Ed. by Andrea Vedaldi et al. Cham: Springer International Publishing, 2020, pp. 330–345. ISBN: 978-3-030-58610-2.

[40] Pietro Cipresso et al. "The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature". In: *Frontiers in Psychology* 9 (Nov. 6, 2018), p. 2086. ISSN: 1664-1078. DOI: `10.3389/fpsyg.2018.02086`. URL: `https://www.frontiersin.org/article/10.3389/fpsyg.2018.02086/full` (visited on 09/21/2022).

[41] Clarivate. *ISI Web of Science Digital Library.* `https://www.webofknowledge.com`. Online; accessed 10 Aug 2022.

[42] Linda A Clark and Daryl Pregibon. "Tree-based models". In: *Statistical models in S.* Routledge, 2017, pp. 377–419. DOI: `10.1201/9780203738535`.

[43] *CMU Graphics Lab Motion Capture Database.* URL: `http://mocap.cs.cmu.edu/`.

[44] Jacob Cohen. "A coefficient of agreement for nominal scales". In: *Educational and psychological measurement* 20.1 (1960), pp. 37–46.

[45] Morning Consult. *National Tracking Poll 2203015.* en. 2022.

[46] Matthew Crain. *Profit Over Privacy. How Surveillance Advertising Conquered the Internet.* Minneapolis: University of Minnesota Press, 2021.

[47] Atticus Cull and Vivek Nair. *SimSaber: Python-based Beat Saber replay simulator and scoring validator.* en. URL: `https://github.com/MetaGuard/SimSaber` (visited on 10/07/2023).

[48] James E. Cutting and Lynn T. Kozlowski. “Recognizing friends by their walk: Gait perception without familiarity cues”. en. In: *Bulletin of the Psychonomic Society* 9.5 (May 1977), pp. 353–356. ISSN: 0090-5054. DOI: `10.3758/BF03337021`. (Visited on 02/05/2023).

[49] *Collegiate VR Esports League (CVRE)*. en. URL: `https://cvreleague.com/` (visited on 01/30/2023).

[50] Loris D’Antoni et al. “Operating System Support for Augmented Reality Applications”. In: *14th Workshop on Hot Topics in Operating Systems (HotOS XIV)*. Santa Ana Pueblo, NM: USENIX Association, 2013. URL: `https://www.usenix.org/conference/hotos13/session/d%7B%5Ctextquoteright%7Dantoni`.

[51] *Data Collection Through Gamification*. en. URL: `https://www.othot.com/blog/data-collection-through-gamification` (visited on 03/01/2023).

[52] B. David-John et al. “A privacy-preserving approach to streaming eye-tracking data”. In: *IEEE Transactions on Visualization & amp; Computer Graphics* 27.05 (2021), pp. 2555–2565. ISSN: 1941-0506. DOI: `10.1109/TVCG.2021.3067787`.

[53] Brendan David-John, Kevin Butler, and Eakta Jain. “For Your Eyes Only: Privacy-Preserving Eye-Tracking Datasets”. In: *2022 Symposium on Eye Tracking Research and Applications*. ETRA ’22. Seattle, WA, USA: Association for Computing Machinery, 2022. ISBN: 9781450392525. DOI: `10.1145/3517031.3529618`. URL: `https://doi.org/10.1145/3517031.3529618`.

[54] Jaybie A. De Guzman, Kanchana Thilakarathna, and Aruna Seneviratne. “Security and Privacy Approaches in Mixed Reality: A Literature Survey”. In: *ACM Comput. Surv.* 52.6 (2019). ISSN: 0360-0300. DOI: `10.1145/3359626`. URL: `https://doi-org.eaccess.ub.tum.de/10.1145/3359626`.

[55] Alexandre Défossez et al. *Decoding speech from non-invasive brain recordings*. 2022. DOI: `10.48550/ARXIV.2208.12266`. URL: `https://arxiv.org/abs/2208.12266`.

[56] Jia Deng et al. “Imagenet: A large-scale hierarchical image database”. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255. DOI: `10.1109/CVPR.2009.5206848`.

[57] Jiangyi Deng et al. “V-Cloak: Intelligibility-, Naturalness- & Timbre-Preserving Real-Time Voice Anonymization”. In: *32nd USENIX Security Symposium (USENIX Security 23)*. Anaheim, CA: USENIX Association, Aug. 2023, pp. 5181–5198. ISBN: 978-1-939133-37-3. URL: `https://www.usenix.org/conference/usenixsecurity23/presentation/deng-jiangyi-v-cloak`.

[58] Li Deng. “The mnist database of handwritten digit images for machine learning research [best of the web]”. In: *IEEE signal processing magazine* 29.6 (2012), pp. 141–142. DOI: `10.1109/MSP.2012.2211477`.

[59] Mina Deng et al. “A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements”. In: *Requirements Engineering* 16.1 (Mar. 1, 2011), pp. 3–32. ISSN: 1432-010X. DOI: 10.1007/s00766-010-0115-7. URL: https://doi.org/10.1007/s00766-010-0115-7.

[60] Shaquitta Dent et al. “The effect of music on body sway when standing in a moving virtual environment”. In: *PLOS ONE* 16 (Sept. 2021), e0258000. DOI: 10.1371/journal.pone.0258000.

[61] Roberto Di Pietro and Stefano Cresci. “Metaverse: Security and Privacy Issues”. In: *2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*. 2021, pp. 281–288. DOI: 10.1109/TPSISA52974.2021.00032.

[62] Ellysse Dick. “Balancing User Privacy and Innovation in Augmented and Virtual Reality”. In: *INFORMATION TECHNOLOGY* (2021), p. 28. URL: https://itif.org/publications/2021/03/04/balancing-user-privacy-and-innovation-augmented-and-virtual-reality/.

[63] Roger Dingledine, Nick Mathewson, and Paul Syverson. “Tor: The Second-Generation Onion Router”. In: *13th USENIX Security Symposium (USENIX Security 04)*. San Diego, CA: USENIX Association, Aug. 2004. URL: https://www.usenix.org/conference/13th-usenix-security-symposium/tor-second-generation-onion-router.

[64] Neil A Dodgson. *Variation and extrema of human interpupillary distance.* SPIE, 2004.

[65] S.A. Douglas and A.K. Mithal. *The Ergonomics of Computer Pointing Devices.* Applied Computing. Springer London, 2012. ISBN: 9781447109174. URL: https://books.google.com/books?id=clLlBwAAQBAJ.

[66] Liang Du et al. “GARP-face: Balancing privacy protection and utility preservation in face de-identification”. In: *IEEE International Joint Conference on Biometrics*. 2014, pp. 1–8. DOI: 10.1109/BTAS.2014.6996249.

[67] Yuming Du et al. *Avatars Grow Legs: Generating Smooth Human Motion from Sparse Tracking Inputs with Diffusion Model.* 2023. arXiv: 2304.08577 [cs.CV].

[68] Duck Duck Go, Inc. *Duck Duck Go.* https://duckduckgo.com/. Online; accessed 21 July 2022. 2023.

[69] Joe Durbin. *Report: Vive Users Are 95 Percent Male And Spend 5 Hours Per Week in VR.* en. Feb. 2017. URL: https://www.uploadvr.com/vive-users-94-9-percent-male-spend-5-hours-week-vr-average/ (visited on 05/24/2023).

[70] Reyhan Düzgün et al. “SoK: A Systematic Literature Review of Knowledge-Based Authentication on Augmented Reality Head-Mounted Displays”. In: *Proceedings of the 17th International Conference on Availability, Reliability and Security*. ARES ’22. Vienna, Austria: Association for Computing Machinery, 2022. ISBN: 9781450396707. DOI: 10.1145/3538969.3539011. URL: https://doi.org/10.1145/3538969.3539011.

[71] Yogesh K. Dwivedi et al. “Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy”. In: *International Journal of Information Management* 66 (2022), p. 102542. ISSN: 0268-4012. DOI: https://doi.org/10.1016/j.ijinfomgt.2022.102542. URL: https://www.sciencedirect.com/science/article/pii/S0268401222000767.

[72] Cynthia Dwork, Nitin Kohli, and Deirdre Mulligan. “Differential Privacy in Practice: Expose your Epsilons!” In: *Journal of Privacy and Confidentiality* 9.2 (Oct. 2019). DOI: 10.29012/jpc.689. URL: https://journalprivacyconfidentiality.org/index.php/jpc/article/view/689.

[73] Cynthia Dwork and Aaron Roth. “The Algorithmic Foundations of Differential Privacy”. en. In: *Foundations and Trends in Theoretical Computer Science* 9.3-4 (2013), pp. 211–407. ISSN: 1551-305X, 1551-3068. DOI: 10.1561/0400000042. URL: http://www.nowpublishers.com/articles/foundations-and-trends-in-theoretical-computer-science/TCS-042 (visited on 01/30/2022).

[74] Cynthia Dwork et al. “Calibrating Noise to Sensitivity in Private Data Analysis”. In: *Theory of Cryptography*. Ed. by Shai Halevi and Tal Rabin. Online; accessed 30 December 2021. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 265–284. ISBN: 978-3-540-32732-5. URL: https://link.springer.com/chapter/10.1007/11681878_14.

[75] Cynthia Dwork et al. “Exposed! A Survey of Attacks on Private Data”. In: *Annual Review of Statistics and Its Application* 4.1 (Mar. 7, 2017). Publisher: Annual Reviews, pp. 61–84. ISSN: 2326-8298. DOI: 10.1146/annurev-statistics-060116-054123. URL: https://doi.org/10.1146/annurev-statistics-060116-054123 (visited on 07/08/2021).

[76] Peter Eckersley. “How Unique Is Your Web Browser?” en. In: *Privacy Enhancing Technologies*. Ed. by Mikhail J. Atallah and Nicholas J. Hopper. Vol. 6205. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 1–18. ISBN: 978-3-642-14526-1. DOI: 10.1007/978-3-642-14527-8_1. URL: http://link.springer.com/10.1007/978-3-642-14527-8_1 (visited on 05/14/2023).

[77] Elservier. *SCOPUS Digital Library*. https://www.scopus.com/. Online; accessed 10 Aug 2022.

[78] Steven Englehardt et al. *Cookies That Give You Away: The Surveillance Implications of Web Tracking.* Florence Italy, May 18, 2015. DOI: 10.1145/2736277.2741679. URL: https://dl.acm.org/doi/10.1145/2736277.2741679 (visited on 07/11/2022).

[79] Morgane Evin et al. "Personality trait prediction by machine learning using physiological data and driving behavior". In: *Machine Learning with Applications* 9 (2022), p. 100353. ISSN: 2666-8270. DOI: https://doi.org/10.1016/j.mlwa.2022.100353. URL: https://www.sciencedirect.com/science/article/pii/S2666827022000548.

[80] Ben Falchuk, Shoshana Loeb, and Ralph Neff. "The Social Metaverse: Battle for Privacy". In: *IEEE Technology and Society Magazine* 37.2 (2018), pp. 52–61. DOI: 10.1109/MTS.2018.2826060.

[81] Linxi Fan et al. "MineDojo: Building Open-Ended Embodied Agents with Internet-Scale Knowledge". In: *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track.* 2022. URL: https://openreview.net/forum?id=rc8o_j8I8PX.

[82] Jamie Feltham. *Valve Index Is Now The Second Most Used Headset On Steam.* en-US. Section: VR Hardware. Oct. 2021. URL: https://uploadvr.com/valve-index-second-most-used-headset-on-steam/ (visited on 12/05/2022).

[83] Lucas Silva Figueiredo et al. "Prepose: Privacy, Security, and Reliability for Gesture-Based Programming". In: *2016 IEEE Symposium on Security and Privacy (SP).* 2016, pp. 122–137. DOI: 10.1109/SP.2016.16.

[84] R. A. Fisher. "On the Interpretation of X2 from Contingency Tables, and the Calculation of P". In: *Journal of the Royal Statistical Society* 85.1 (1922), pp. 87–94. ISSN: 09528385. URL: http://www.jstor.org/stable/2340521 (visited on 10/08/2023).

[85] Batya Friedman and Peter H. Kahn. "New Directions: A Value-Sensitive Design Approach to Augmented Reality". In: *Proceedings of DARE 2000 on Designing Augmented Reality Environments.* DARE '00. Elsinore, Denmark: Association for Computing Machinery, 2000, pp. 163–164. ISBN: 9781450373265. DOI: 10.1145/354666.354694. URL: https://doi.org/10.1145/354666.354694.

[86] Christoph Fromm and Edward V Evarts. "Relation of motor cortex neurons to precisely controlled and ballistic movements". In: *Neuroscience letters* 5.5 (1977), pp. 259–265.

[87] Andrea Gallardo et al. "Speculative Privacy Concerns About AR Glasses Data Collection". In: *Proceedings on Privacy Enhancing Technologies* 4 (2023), pp. 416–435.

[88] Beat Games. *Beat Saber.* en. https://beatsaber.com/. URL: https://beatsaber.com/ (visited on 01/31/2023).

[89] Berserk Games. *Tabletop Simulator.* https://www.tabletopsimulator.com. Online. 2022.

[90] Gamespot. *Valve and HTC Reveal Vive VR Headset*. https://www.gamespot.com/articles/valve-and-htc-reveal-vive-vr-headset/1100-6425606/. Online; accessed 17 July 2022. 2015.

[91] Xianyi Gao et al. "Elastic pathing: your speed is enough to track you". en. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct*. Seattle, Washington: ACM Press, 2014, pp. 975–986. ISBN: 978-1-4503-2968-2. DOI: 10.1145/2632048.2632077. URL: http://dl.acm.org/citation.cfm?doid=2632048.2632077 (visited on 11/28/2020).

[92] Yatharth Garg. *Speech-Accent-Recognition*. original-date: 2018-06-21T07:55:52Z. 2022. URL: https://github.com/yatharthgarg/Speech-Accent-Recognition (visited on 05/25/2022).

[93] Rod Garratt and Maarten R.C. van Oordt. "Privacy as a Public Good: A Case for Electronic Cash". In: *Journal of Political Economy* (2018). DOI: 10.1086/714133.

[94] Gonzalo Munilla Garrido, Vivek Nair, and Dawn Song. *SoK: Data Privacy in Virtual Reality*. arXiv:2301.05940 [cs]. Jan. 2023. URL: http://arxiv.org/abs/2301.05940 (visited on 01/30/2023).

[95] Gonzalo Munilla Garrido et al. "Revealing the landscape of privacy-enhancing technologies in the context of data markets for the IoT: A systematic literature review". In: *Journal of Network and Computer Applications* 207 (2022), p. 103465. ISSN: 1084-8045. DOI: https://doi.org/10.1016/j.jnca.2022.103465. URL: https://www.sciencedirect.com/science/article/pii/S1084804522001126.

[96] Saeed Ghorbani et al. "MoVi: A large multi-purpose human motion and video dataset". In: *PLOS ONE* 16.6 (2021). Ed. by Peter Andreas Federolf, e0253157. DOI: 10.1371/journal.pone.0253157. URL: https://doi.org/10.1371%5C%2Fjournal.pone.0253157.

[97] Alberto Giaretta. *Security and Privacy in Virtual Reality – A Literature Survey*. 2022. DOI: 10.48550/ARXIV.2205.00208. URL: https://arxiv.org/abs/2205.00208.

[98] Roberto Gonzalez et al. "User Profiling by Network Observers". In: *Proceedings of the 17th International Conference on Emerging Networking EXperiments and Technologies*. CoNEXT '21. Virtual Event, Germany: Association for Computing Machinery, 2021, pp. 212–222. ISBN: 9781450390989. DOI: 10.1145/3485983.3494859. URL: https://doi-org.eaccess.ub.tum.de/10.1145/3485983.3494859.

[99] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016. URL: https://www.deeplearningbook.org/.

[100] Ian Goodfellow et al. "Generative adversarial networks". In: *Communications of the ACM* 63.11 (2020), pp. 139–144.

[101] P. Grother et al. *IREX III: Performance of Iris Identification Algorithms*. 2012. URL: https://www.nist.gov/publications/irex-iii-performance-iris-identification-algorithms.

[102] Aniket Gulhane et al. "Security, Privacy and Safety Risk Assessment for Virtual Reality Learning Environment Applications". In: *2019 16th IEEE Annual Consumer Communications Networking Conference (CCNC)*. 2019, pp. 1–9. DOI: 10.1109/CCNC.2019.8651847.

[103] Rajesh Gupta et al. "Machine Learning Models for Secure Data Analytics: A taxonomy and threat model". In: *Computer Communications* 153 (2020), pp. 406–440. ISSN: 0140-3664. DOI: https://doi.org/10.1016/j.comcom.2020.02.008. URL: https://www.sciencedirect.com/science/article/pii/S0140366419318493.

[104] Jaybie A. de Guzman, Kanchana Thilakarathna, and Aruna Seneviratne. "A First Look into Privacy Leakage in 3D Mixed Reality Data". In: *Computer Security – ESORICS 2019*. Ed. by Kazue Sako, Steve Schneider, and Peter Y. A. Ryan. Cham: Springer International Publishing, 2019, pp. 149–169. ISBN: 978-3-030-29959-0.

[105] Jaybie A. de Guzman, Kanchana Thilakarathna, and Aruna Seneviratne. *Conservative Plane Releasing for Spatial Privacy Protection in Mixed Reality.* 2020. DOI: 10.48550/ARXIV.2004.08029. URL: https://arxiv.org/abs/2004.08029.

[106] Jaybie Agullo de Guzman, Kanchana Thilakarathna, and Aruna Seneviratne. "SafeMR: Privacy-aware Visual Information Protection for Mobile Mixed Reality". In: *2019 IEEE 44th Conference on Local Computer Networks (LCN)*. 2019, pp. 254–257. DOI: 10.1109/LCN44214.2019.8990850.

[107] Jaybie Agullo de Guzman, Aruna Seneviratne, and Kanchana Thilakarathna. "Unravelling Spatial Privacy Risks of Mobile Mixed Reality Data". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5.1 (2021). DOI: 10.1145/3448103. URL: https://doi.org/10.1145/3448103.

[108] Hana Habib et al. "Away From Prying Eyes: Analyzing Usage and Understanding of Private Browsing". In: *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. Baltimore, MD: USENIX Association, Aug. 2018, pp. 159–175. ISBN: 978-1-939133-10-6. URL: https://www.usenix.org/conference/soups2018/presentation/habib-prying.

[109] haptx. *Haptx Gloves DK2.* Online; accessed 22 Sep 2022. URL: https://haptx.com/.

[110] Ragib Hasan et al. "Toward a Threat Model for Storage Systems". In: *Proceedings of the 2005 ACM Workshop on Storage Security and Survivability*. StorageSS '05. Fairfax, VA, USA: Association for Computing Machinery, 2005, pp. 94–102. ISBN: 159593233X. DOI: 10.1145/1103780.1103795. URL: https://doi.org/10.1145/1103780.1103795.

[111] Olli I. Heimo et al. "Augmented reality - Towards an ethical fantasy?" In: *2014 IEEE International Symposium on Ethics in Science, Technology and Engineering*. 2014, pp. 1–7. DOI: 10.1109/ETHICS.2014.6893423.

[112] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[113] Sarah Holland et al. *The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards*. 2018. arXiv: 1805.03677 [cs.DB].

[114] Naoise Holohan and Stefano Braghin. "Secure Random Sampling in Differential Privacy". en. In: *Computer Security – ESORICS 2021*. Ed. by Elisa Bertino, Haya Shulman, and Michael Waidner. Vol. 12973. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 523–542. ISBN: 978-3-030-88427-7. DOI: 10.1007/978-3-030-88428-4_26. URL: https://link.springer.com/10.1007/978-3-030-88428-4_26 (visited on 01/24/2022).

[115] Naoise Holohan et al. "The Bounded Laplace Mechanism in Differential Privacy". In: *Journal of Privacy and Confidentiality* 10.1 (Dec. 2019). ISSN: 2575-8527. DOI: 10.29012/jpc.715. (Visited on 01/31/2021).

[116] Ludovic Hoyet et al. "Sleight of Hand: Perception of Finger Motion from Reduced Marker Sets". In: *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. I3D '12. Costa Mesa, California: Association for Computing Machinery, 2012, pp. 79–86. ISBN: 9781450311946. DOI: 10.1145/2159616.2159630. URL: https://doi.org/10.1145/2159616.2159630.

[117] *HR Magazine - Why Cambridge Analytica's techniques could kill gamification*. en. Apr. 2018. URL: https://www.hrmagazine.co.uk/content/features/why-cambridge-analytica-s-techniques-could-kill-gamification/ (visited on 03/01/2023).

[118] IEEE. *IEEE Xplore*. https://ieeexplore.ieee.org. Online; accessed 10 Aug 2022.

[119] VRChat Inc. *VRChat*. https://hello.vrchat.com/. Online; accessed 17 May 2022. 2022.

[120] Catalin Ionescu et al. "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.7 (2014), pp. 1325–1339. DOI: 10.1109/TPAMI.2013.248.

[121] Jim Isaak and Mina J. Hanna. "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection". In: *Computer* 51.8 (2018), pp. 56–59. DOI: 10.1109/MC.2018.3191268.

[122] Syem Ishaque et al. "Physiological Signal Analysis and Classification of Stress from Virtual Reality Video Game". In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 2020, pp. 867–870. DOI: 10.1109/EMBC44109.2020.9176110.

[123] Eakta Jain et al. "Is the Motion of a Child Perceivably Different from the Motion of an Adult?" en. In: *ACM Transactions on Applied Perception* 13.4 (July 2016), pp. 1–17. ISSN: 1544-3558, 1544-3965. DOI: 10.1145/2947616. (Visited on 02/05/2023).

[124] Suman Jana, Arvind Narayanan, and Vitaly Shmatikov. "A Scanner Darkly: Protecting User Privacy from Perceptual Applications". In: *2013 IEEE Symposium on Security and Privacy*. 2013, pp. 349–363. DOI: `10.1109/SP.2013.31`.

[125] Suman Jana et al. "Enabling Fine-Grained Permissions for Augmented Reality Applications with Recognizers". In: *22nd USENIX Security Symposium (USENIX Security 13)*. Washington, D.C.: USENIX Association, 2013, pp. 415–430. ISBN: 978-1-931971-03-4. URL: `https://www.usenix.org/conference/usenixsecurity13/technical-sessions/presentation/jana`.

[126] W. Jarrold et al. "Social attention in a virtual public speaking task in higher functioning children with autism." In: *Autism Res.* (2013). DOI: `10.1002/aur.1302`.

[127] Carter Jernigan and Behram F.T. Mistree. "Gaydar: Facebook friendships expose sexual orientation". In: *First Monday* 14.10 (2009). DOI: `10.5210/fm.v14i10.2611`. URL: `https://journals.uic.edu/ojs/index.php/fm/article/view/2611`.

[128] Jiaxi Jiang et al. *AvatarPoser: Articulated Full-Body Pose Tracking from Sparse Motion Sensing*. 2022. arXiv: `2207.13784 [cs.CV]`.

[129] Brendan John, Sanjeev Koppal, and Eakta Jain. "EyeVEIL: Degrading Iris Authentication in Eye Tracking Headsets". In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & amp; Applications*. Denver, Colorado, 2019. ISBN: 9781450367097. DOI: `10.1145/3314111.3319816`. URL: `https://doi.org/10.1145/3314111.3319816`.

[130] Brendan John et al. "The Security-Utility Trade-off for Iris Authentication and Eye Animation for Social Virtual Avatars". In: *IEEE Transactions on Visualization and Computer Graphics* 26.5 (2020), pp. 1880–1890. DOI: `10.1109/TVCG.2020.2973052`.

[131] Roberto Jordaney et al. "Transcend: Detecting Concept Drift in Malware Classification Models". en. In: 2017, pp. 625–642. ISBN: 978-1-931971-40-9. URL: `https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/jordaney` (visited on 05/14/2023).

[132] Parunyou Julayanont and Ziad S. Nasreddine. "Montreal Cognitive Assessment (MoCA): Concept and Clinical Review". In: *Cognitive Screening Instruments*. Ed. by A. J. Larner. Cham: Springer International Publishing, 2017, pp. 139–195. ISBN: 978-3-319-44774-2. DOI: `10.1007/978-3-319-44775-9_7`. URL: `http://link.springer.com/10.1007/978-3-319-44775-9_7` (visited on 05/19/2022).

[133] Nesrine Kaaniche, Maryline Laurent, and Sana Belguith. *Privacy enhancing technologies for solving the privacy-personalization paradox: Taxonomy and survey*. 2020.

[134] Christos Kalloniatis, Evangelia Kavakli, and Stefanos Gritzalis. "Addressing privacy requirements in system design: the PriS method". In: *Requirements Engineering* 13.3 (Sept. 1, 2008), pp. 241–255. ISSN: 1432-010X. DOI: `10.1007/s00766-008-0067-3`. URL: `https://doi.org/10.1007/s00766-008-0067-3`.

[135] Ishan Karunanayake et al. “De-Anonymisation Attacks on Tor: A Survey”. In: *IEEE Communications Surveys & Tutorials* 23.4 (2021), pp. 2324–2350. DOI: 10.1109/COMST.2021.3093615.

[136] Christina Katsini et al. “The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–21. ISBN: 9781450367080. DOI: 10.1145/3313831.3376840. URL: https://doi.org/10.1145/3313831.3376840.

[137] Guolin Ke et al. “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

[138] *Keras: Deep Learning for humans*. en. URL: https://keras.io/ (visited on 10/04/2023).

[139] Orin S Kerr. “Criminal Law in Virtual Worlds”. In: (2008), p. 17.

[140] Yoonsang Kim et al. “Erebus: Access Control for Augmented Reality Systems”. In: *32nd USENIX Security Symposium (USENIX Security 23)*. Anaheim, CA: USENIX Association, Aug. 2023, pp. 929–946. ISBN: 978-1-939133-37-3. URL: https://www.usenix.org/conference/usenixsecurity23/presentation/kim-yoonsang.

[141] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[142] Adam G. Kirk, James F. O’Brien, and David A. Forsyth. “Skeletal Parameter Estimation from Optical Motion Capture Data”. In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) 2005*. June 2005, pp. 782–788. URL: http://graphics.cs.berkeley.edu/papers/Kirk-SPE-2005-06/.

[143] Bernard Koch et al. *Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research*. 2021. arXiv: 2112.01716 [cs.LG].

[144] Daniel Kondor et al. “Towards Matching User Mobility Traces in Large-Scale Datasets”. en. In: *IEEE Transactions on Big Data* 6.4 (2020), pp. 714–726. ISSN: 2332-7790, 2372-2096. DOI: 10.1109/TBDATA.2018.2871693. URL: https://ieeexplore.ieee.org/document/8470173/ (visited on 11/27/2020).

[145] Fragkiskos Koufogiannis, Shuo Han, and George J. Pappas. *Optimality of the Laplace Mechanism in Differential Privacy*. 2015. DOI: 10.48550/ARXIV.1504.00065. URL: https://arxiv.org/abs/1504.00065.

[146] Lynn T. Kozlowski and James E. Cutting. “Recognizing the sex of a walker from a dynamic point-light display”. en. In: *Perception & Psychophysics* 21.6 (Nov. 1977), pp. 575–580. ISSN: 1532-5962. DOI: 10.3758/BF03198740. (Visited on 02/05/2023).

[147] Alex Krizhevsky and Geoff Hinton. "Convolutional deep belief networks on cifar-10". In: *Unpublished manuscript* 40.7 (2010), pp. 1–9. URL: `https://www.cs.toronto.edu/~kriz/conv-cifar10-aug2010.pdf`.

[148] Jacob Leon Kröger, Otto Hans-Martin Lutz, and Florian Müller. "What Does Your Gaze Reveal About You? On the Privacy Implications of Eye Tracking". In: *Privacy and Identity Management. Data for Better Living: AI and Privacy: 14th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Windisch, Switzerland, August 19–23, 2019, Revised Selected Papers*. Cham: Springer International Publishing, 2020, pp. 226–241. ISBN: 978-3-030-42504-3. DOI: `10.1007/978-3-030-42504-3_15`. URL: `https://doi.org/10.1007/978-3-030-42504-3_15`.

[149] Jesse Lake. "Hey, You Stole My Avatar!: Virtual Reality and Its Risks to Identity Protection". In: *EMORY LAW JOURNAL* 69 (2020), p. 48.

[150] Pierre Laperdrix et al. "Browser Fingerprinting: A Survey". In: *ACM Trans. Web* 14.2 (2020). ISSN: 1559-1131. DOI: `10.1145/3386040`. URL: `https://doi-org.eaccess.ub.tum.de/10.1145/3386040`.

[151] Sarah M. Lehman et al. "Hidden in Plain Sight: Exploring Privacy Risks of Mobile Augmented Reality Applications". In: *ACM Trans. Priv. Secur.* 25.4 (2022). ISSN: 2471-2566. DOI: `10.1145/3524020`. URL: `https://doi.org/10.1145/3524020`.

[152] Jingjie Li et al. "Kaleido: Real-Time Privacy Control for Eye-Tracking Systems". In: *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 2021, pp. 1793–1810. ISBN: 978-1-939133-24-3. URL: `https://www.usenix.org/conference/usenixsecurity21/presentation/li-jingjie`.

[153] Bin Liang et al. *Scriptless Timing Attacks on Web Browser Privacy*. 2014. DOI: `10.1109/DSN.2014.93`.

[154] Jonathan Liebers et al. "Understanding User Identification in Virtual Reality Through Behavioral Biometrics and the Effect of Body Normalization". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021. ISBN: 9781450380966. DOI: `10.1145/3411764.3445528`. URL: `https://doi.org/10.1145/3411764.3445528`.

[155] Junsu Lim et al. "Mine Yourself!: A Role-Playing Privacy Tutorial in Virtual Reality Environment". In: *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. CHI EA '22. New Orleans, LA, USA: Association for Computing Machinery, 2022. ISBN: 9781450391566. DOI: `10.1145/3491101.3519773`. URL: `https://doi-org.eaccess.ub.tum.de/10.1145/3491101.3519773`.

[156] Zhen Ling et al. "I Know What You Enter on Gear VR". In: *2019 IEEE Conference on Communications and Network Security (CNS)*. 2019, pp. 241–249. DOI: `10.1109/CNS.2019.8802674`.

[157] Ao Liu et al. "Differential Privacy for Eye-Tracking Data". In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & amp; Applications*. ETRA '19. Denver, Colorado: Association for Computing Machinery, 2019. ISBN: 9781450367097. DOI: 10.1145/3314111.3319823. URL: https://doi-org.eaccess.ub.tum.de/10.1145/3314111.3319823.

[158] Matthew Loper, Naureen Mahmood, and Michael J. Black. "MoSh: Motion and Shape Capture from Sparse Markers". In: *ACM Trans. Graph.* 33.6 (2014). ISSN: 0730-0301. DOI: 10.1145/2661229.2661273. URL: https://doi.org/10.1145/2661229.2661273.

[159] L. Loucks et al. "You can do that?!: Feasibility of virtual reality exposure therapy in the treatment of PTSD due to military sexual trauma." In: *Anxiety Disord.* (2019). DOI: 10.1016/j.janxdis.2018.06.004..

[160] Shiqing Luo et al. "OcuLock: Exploring Human Visual System for Authentication in Virtual Reality Head-mounted Display". In: *NDSS*. 2020.

[161] A. Martin M. Przybocki. *Speaker Recognition Evaluation Chronicles*. 2004. URL: https://nist.gov/publications/nist-speaker-recognition-evaluation-chronicles.

[162] Magic Leap, Inc. *Magic Leap 2*. Online; accessed 4 Oct 2022. URL: https://www.magicleap.com/magic-leap-2.

[163] Naureen Mahmood et al. "AMASS: Archive of Motion Capture as Surface Shapes". In: *International Conference on Computer Vision*. 2019, pp. 5442–5451.

[164] Divine Maloney, Samaneh Zamanifard, and Guo Freeman. "Anonymity vs. Familiarity: Self-Disclosure and Privacy in Social Virtual Reality". In: *26th ACM Symposium on Virtual Reality Software and Technology*. VRST '20. Virtual Event, Canada: Association for Computing Machinery, 2020. ISBN: 9781450376198. DOI: 10.1145/3385956.3418967. URL: https://doi.org/10.1145/3385956.3418967.

[165] Christian Mandery et al. "The KIT whole-body human motion database". In: *2015 International Conference on Advanced Robotics (ICAR)*. 2015, pp. 329–336. DOI: 10.1109/ICAR.2015.7251476.

[166] Ivan Martinovic et al. "On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces". In: *21st USENIX Security Symposium (USENIX Security 12)*. Bellevue, WA: USENIX Association, 2012, pp. 143–158. URL: https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/martinovic.

[167] Ifigeneia Mavridou et al. "Towards an Effective Arousal Detection System for Virtual Reality". In: *Proceedings of the Workshop on Human-Habitat for Health (H3): Human-Habitat Multimodal Interaction for Promoting Health and Well-Being in the Internet of Things Era*. H3 '18. Boulder, Colorado: Association for Computing Machinery, 2018. ISBN: 9781450360753. DOI: 10.1145/3279963.3279969. URL: https://doi.org/10.1145/3279963.3279969.

[168] MelonLoader community. *Melon Loader*. `https://melonwiki.xyz/`. Online; accessed 22 July 2022. 2022.

[169] Meta. *Horizon Worlds*. `https://www.oculus.com/horizon-worlds/`. Online; accessed 17 May 2022. 2022.

[170] Meta. *Meta Quest 2*. Online; accessed 22 Sep 2022. URL: `https://www.meta.com/es/en/quest/`.

[171] Meta. *Meta Quest Pro*. Online; accessed 21 October 2022. URL: `https://www.oculus.com/blog/meta-quest-pro-price-release-date/?intern_source=blog&intern_content=/meta-quest-pro-privacy`.

[172] Meta. *Oculus Go Features*. `https://www.oculus.com/go/features/`. Online; accessed 17 July 2022. 2022.

[173] Microsoft. *AltspaceVR*. `https://altvr.com`. Online; accessed 17 May 2022. 2022.

[174] Microsoft. *Azure Automated Machine Learning - AutoML — Microsoft Azure*. en. URL: `https://azure.microsoft.com/en-us/services/machine-learning/automatedml/` (visited on 05/25/2022).

[175] Stuart E. Middleton, David C. De Roure, and Nigel R. Shadbolt. "Capturing Knowledge of User Preferences: Ontologies in Recommender Systems". In: *Proceedings of the 1st International Conference on Knowledge Capture*. K-CAP '01. Victoria, British Columbia, Canada: Association for Computing Machinery, 2001, pp. 100–107. ISBN: 1581133804. DOI: `10.1145/500737.500755`. URL: `https://doi.org/10.1145/500737.500755`.

[176] Paul Milgram and Fumio Kishino. "A Taxonomy of Mixed Reality Visual Displays". In: *IEICE Transactions on Information and Systems* 77.12 (1994), pp. 1321–1329.

[177] Mark Roman Miller et al. *A Large-Scale Study of Personal Identifiability of Virtual Reality Motion Over Time*. 2023. arXiv: `2303.01430 [cs.CR]`.

[178] Mark Roman Miller et al. "Personal identifiability of user tracking data during observation of 360-degree VR video". en. In: *Scientific Reports* 10.1 (Oct. 2020). Number: 1 Publisher: Nature Publishing Group, p. 17404. ISSN: 2045-2322. DOI: `10.1038/s41598-020-74486-y`. URL: `https://www.nature.com/articles/s41598-020-74486-y` (visited on 01/31/2023).

[179] Robert Miller, Natasha Banerjee, and Sean Banerjee. "Within-System and Cross-System Behavior-Based Biometric Authentication in Virtual Reality". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. Mar. 2020, pp. 311–316. DOI: `10.1109/VRW50115.2020.00070`.

[180] Robert Miller, Natasha Kholgade Banerjee, and Sean Banerjee. "Combining Real-World Constraints on User Behavior with Deep Neural Networks for Virtual Reality (VR) Biometrics". In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2022, pp. 409–418. DOI: `10.1109/VR51125.2022.00060`.

[181] Robert Miller, Natasha Kholgade Banerjee, and Sean Banerjee. "Using Siamese Neural Networks to Perform Cross-System Behavioral Authentication in Virtual Reality". In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 2021, pp. 140–149. DOI: 10.1109/VR50410.2021.00035.

[182] Robert B. Miller. "Response Time in Man-Computer Conversational Transactions". In: *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*. AFIPS '68 (Fall, part I). San Francisco, California: Association for Computing Machinery, 1968, pp. 267–277. ISBN: 9781450378994. DOI: 10.1145/1476589.1476628. URL: https://doi.org/10.1145/1476589.1476628.

[183] Ilya Mironov. "On significance of the least significant bits for differential privacy". en. In: *Proceedings of the 2012 ACM conference on Computer and communications security - CCS '12*. Raleigh, North Carolina, USA: ACM Press, 2012, p. 650. ISBN: 978-1-4503-1651-4. DOI: 10.1145/2382196.2382264. URL: http://dl.acm.org/citation.cfm?doid=2382196.2382264 (visited on 12/25/2020).

[184] Kelley Misata, Raymond A. Hansen, and Baijian Yang. "A Taxonomy of Privacy-Protecting Tools to Browse the World Wide Web". In: *Proceedings of the 3rd Annual Conference on Research in Information Technology*. RIIT '14. Atlanta, Georgia, USA: Association for Computing Machinery, 2014, pp. 63–68. ISBN: 9781450327114. DOI: 10.1145/2656434.2656446. URL: https://doi-org.eaccess.ub.tum.de/10.1145/2656434.2656446.

[185] *mocapdata.com*. URL: http://mocapdata.com/.

[186] Alec G. Moore et al. *Personal Identifiability and Obfuscation of User Tracking Data From VR Training Sessions*. 2021. DOI: 10.1109/ISMAR52148.2021.00037.

[187] Meinard Müller et al. "Documentation Mocap database HDM05". In: June 2007.

[188] Brad A. Myers. "The Importance of Percent-Done Progress Indicators for Computer-Human Interfaces". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '85. San Francisco, California, USA: Association for Computing Machinery, 1985, pp. 11–17. ISBN: 0897911490. DOI: 10.1145/317456.317459. URL: https://doi.org/10.1145/317456.317459.

[189] Stylianos Mystakidis. *Metaverse*. Feb. 10, 2022. DOI: 10.3390/encyclopedia2010031. URL: https://www.mdpi.com/2673-8392/2/1/31 (visited on 04/30/2022).

[190] Vivek Nair, Gonzalo Munilla Garrido, and Dawn Song. "Going Incognito in the Metaverse: Achieving Theoretically Optimal Privacy-Usability Tradeoffs in VR". In: *36th ACM Symposium on User Interface Software and Technology (UIST 23)*. 2023.

[191] Vivek Nair, Viktor Radulov, and James F. O'Brien. *Results of the 2023 Census of Beat Saber Users: Virtual Reality Gaming Population Insights and Factors Affecting Virtual Reality E-Sports Performance*. 2023. arXiv: 2305.14320 [cs.HC].

[192] Vivek Nair et al. *Berkeley Open Extended Reality Recordings 2023 (BOXRR-23): 4.7 Million Motion Capture Recordings from 105,852 Extended Reality Device Users.* 2023. arXiv: `2310.00430 [cs.HC]`.

[193] Vivek Nair et al. "Exploring the Privacy Risks of Adversarial VR Game Design". In: *23rd Privacy Enhancing Technologies Symposium (PETS 23)*. 2023. DOI: `10.56553/popets-2023-0108`. (Visited on 10/05/2023).

[194] Vivek Nair et al. *Inferring Private Personal Attributes of Virtual Reality Users from Head and Hand Motion Data.* 2023. arXiv: `2305.19198 [cs.HC]`.

[195] Vivek Nair et al. *Truth in Motion: The Unprecedented Risks and Opportunities of Extended Reality Motion Data.* 2023. arXiv: `2306.06459 [cs.HC]`.

[196] Vivek Nair et al. "Unique Identification of 50,000+ Virtual Reality Users from Head & Hand Motion Data". In: *32nd USENIX Security Symposium (USENIX Security 23)*. Anaheim, CA: USENIX Association, Aug. 2023, pp. 895–910. ISBN: 978-1-939133-37-3. URL: `https://www.usenix.org/conference/usenixsecurity23/presentation/nair-identification`.

[197] Arvind Narayanan and Vitaly Shmatikov. "Robust De-anonymization of Large Sparse Datasets". en. In: *2008 IEEE Symposium on Security and Privacy (sp 2008)*. ISSN: 1081-6011. Oakland, CA, USA: IEEE, 2008, pp. 111–125. ISBN: 978-0-7695-3168-7. DOI: `10.1109/SP.2008.33`. URL: `http://ieeexplore.ieee.org/document/4531148/` (visited on 12/29/2020).

[198] Andreas Nautsch et al. "Preserving privacy in speaker and speech characterisation". In: *Computer Speech & Language* 58 (2019), pp. 441–480. ISSN: 0885-2308. DOI: `https://doi.org/10.1016/j.csl.2019.06.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0885230818303875`.

[199] John William Nelson. "A Virtual Property Solution: How Privacy Law Can Protect the Citizens of Virtual Worlds". In: (2010), p. 24.

[200] Neurospec. *DSI-VR300.* Online; accessed 22 Sep 2022. URL: `https://wearablesensing.com/dsi-vr300/`.

[201] Naheem Noah, Sommer Shearer, and Sanchari Das. "Security and Privacy Evaluation of Popular Augmented and Virtual Reality Technologies". In: *In Proceedings of the 2022 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence, and Neural Engineering (IEEE MetroXRAINE 2022).* Association for Computing Machinery, 2020. DOI: `http://dx.doi.org/10.2139/ssrn.4173372`. URL: `https://ssrn.com/abstract=4173372`.

[202] James F. O'Brien et al. "Automatic Joint Parameter Estimation from Magnetic Motion Capture Data". In: *Proceedings of Graphics Interface 2000.* May 2000, pp. 53–60. URL: `http://graphics.cs.berkeley.edu/papers/Obrien-AJP-2000-05/`.

[203] Fiachra O'Brolcháin et al. "The Convergence of Virtual Reality and Social Networks: Threats to Privacy and Autonomy". In: *Science and Engineering Ethics* 22.1 (2016), pp. 1–29. DOI: 10.1007/s11948-014-9621-1. URL: https://doi.org/10.1007/s11948-014-9621-1.

[204] Blessing Odeleye et al. "Detecting framerate-oriented cyber attacks on user experience in virtual reality". In: *USENIX Symposium on Usable Privacy and Security (SOUPS)* (2021), p. 5.

[205] Blessing Odeleye et al. "Virtually Secure: A taxonomic assessment of cybersecurity challenges in virtual reality environments". In: *Computers & Security* (2022), p. 102951. ISSN: 0167-4048. DOI: https://doi.org/10.1016/j.cose.2022.102951. URL: https://www.sciencedirect.com/science/article/pii/S0167404822003431.

[206] UK's Information Commissioner's Office. *Audits of data protection compliance by UK political parties.* https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2020/11/uk-political-parties-must-improve-data-protection-practices/. Online; accessed 17 May 2022.

[207] Ilesanmi Olade, Charles Fleming, and Hai-Ning Liang. "BioMove: Biometric User Identification from Human Kinesiological Movements for Virtual Reality Systems". In: *Sensors* 20.10 (2020). ISSN: 1424-8220. DOI: 10.3390/s20102944. URL: https://www.mdpi.com/1424-8220/20/10/2944.

[208] Seiya Otsuka, Kanami Kurosaki, and Mitsuhiro Ogawa. "Physiological measurements on a gaming virtual reality headset using photoplethysmography: A preliminary attempt at incorporating physiological measurement with gaming". In: *TENCON 2017 - 2017 IEEE Region 10 Conference*. 2017, pp. 1251–1256. DOI: 10.1109/TENCON.2017.8228049.

[209] OVR Technology. *Scent Technology for Virtual Reality.* Online; accessed 22 Sep 2022. URL: https://ovrtechnology.com/.

[210] Tiago Palma Pagano et al. *Bias and unfairness in machine learning models: a systematic literature review.* 2022. arXiv: 2202.08176 [cs.LG].

[211] Marcus Pendleton et al. "A Survey on Systems Security Metrics". In: *ACM Comput. Surv.* 49.4 (2016). ISSN: 0360-0300. DOI: 10.1145/3005714. URL: https://doi.org/10.1145/3005714.

[212] Tao Peng, Christopher Leckie, and Kotagiri Ramamohanarao. "Survey of Network-Based Defense Mechanisms Countering the DoS and DDoS Problems". In: *ACM Comput. Surv.* 39.1 (2007), 3–es. ISSN: 0360-0300. DOI: 10.1145/1216370.1216373. URL: https://doi.org/10.1145/1216370.1216373.

[213] Marco Pennacchiotti and Ana-Maria Popescu. "A Machine Learning Approach to Twitter User Classification". In: *Proceedings of the International AAAI Conference on Web and Social Media* 5.1 (2021), pp. 281–288. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14139.

[214] Ken Pfeuffer et al. "Behavioural Biometrics in VR: Identifying People from Body Motion and Relations in Virtual Reality". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–12. ISBN: 9781450359702. DOI: `10.1145/3290605.3300340`. URL: `https://doi.org/10.1145/3290605.3300340`.

[215] Frank E. Pollick et al. "Gender recognition from point-light walkers". In: *Journal of Experimental Psychology: Human Perception and Performance* 31 (2005). Place: US Publisher: American Psychological Association, pp. 1247–1265. ISSN: 1939-1277. DOI: `10.1037/0096-1523.31.6.1247`.

[216] *Polygone Art*. en. URL: `https://polygone.art/` (visited on 01/30/2023).

[217] Jose Luis Ponton et al. "Combining Motion Matching and Orientation Prediction to Animate Avatars for Consumer-Grade VR Devices". In: *Computer Graphics Forum* 41.8 (2022), pp. 107–118. ISSN: 1467-8659. DOI: `10.1111/cgf.14628`.

[218] Ivan Poupyrev et al. "The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR". In: *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*. UIST '96. Seattle, Washington, USA: Association for Computing Machinery, 1996, pp. 79–80. ISBN: 0897917987. DOI: `10.1145/237091.237102`. URL: `https://doi.org/10.1145/237091.237102`.

[219] Michael L. Hicks published. *Despite Quest 2 sales success, Meta lost $10.2 billion on VR/AR last year*. en. Feb. 2022. URL: `https://www.androidcentral.com/despite-quest-2-sales-success-meta-lost-102-billion-vrar-last-year` (visited on 05/29/2022).

[220] Ihsan Rabbi and Sehat Ullah. "A Survey on Augmented Reality Challenges and Tracking". In: *Acta Graphica* 24 (2016), pp. 29–46.

[221] Christian Rack et al. *Extensible Motion-based Identification of XR Users using Non-Specific Motion Data*. 2023. arXiv: `2302.07517 [cs.HC]`.

[222] Viktor Radulov. *BeatLeader*. en. URL: `https://www.beatleader.xyz/` (visited on 01/30/2023).

[223] Viktor Radulov. *BeatLeader Privacy Policy*. en. URL: `https://www.beatleader.xyz/privacy` (visited on 01/30/2023).

[224] Viktor Radulov et al. *Beat Saber Web Replays*. en. URL: `https://github.com/BeatLeader/BeatSaber-Web-Replays/graphs/contributors` (visited on 10/07/2023).

[225] Muhammad Usman Rafique and Sen-ching S. Cheung. "Tracking Attacks on Virtual Reality Systems". In: *IEEE Consumer Electronics Magazine* 9.2 (2020), pp. 41–46. DOI: `10.1109/MCE.2019.2953741`.

[226] Md Mustafizur Rahman et al. "An Information Retrieval Approach to Building Datasets for Hate Speech Detection". In: *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*. 2021. URL: https://openreview.net/forum?id=jI_BbL-qjJN.

[227] Dziugas Ramonas. *CyberRamen*. en. URL: https://www.beatleader.xyz/u/165749 (visited on 01/30/2023).

[228] Daniel E. Re et al. *Preferences for Very Low and Very High Voice Pitch in Humans*. Ed. by David Reby. Mar. 5, 2012. DOI: 10.1371/journal.pone.0032719. URL: https://dx.plos.org/10.1371/journal.pone.0032719 (visited on 07/17/2022).

[229] Derek Reilly et al. "SecSpace: Prototyping Usable Privacy and Security for Mixed Reality Collaborative Environments". In: *Proceedings of the 2014 ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. EICS '14. Rome, Italy: Association for Computing Machinery, 2014, pp. 273–282. ISBN: 9781450327251. DOI: 10.1145/2607023.2607039. URL: https://doi.org/10.1145/2607023.2607039.

[230] *Report: Vive Users Are 95 Percent Male And Spend 5 Hours Per Week in VR*. en. Feb. 2017. URL: https://www.uploadvr.com/vive-users-94-9-percent-male-spend-5-hours-week-vr-average/ (visited on 05/24/2023).

[231] A.A. Rizzo et al. "Diagnosing attention disorders in a virtual classroom". In: *Computer* 37.6 (2004), pp. 87–89. DOI: 10.1109/MC.2004.23.

[232] Luc Rocher, Julien M. Hendrickx, and Yves-Alexandre de Montjoye. "Estimating the success of re-identifications in incomplete datasets using generative models". In: *Nature Communications* 10.1 (Dec. 2019), p. 3069. ISSN: 2041-1723. DOI: 10.1038/s41467-019-10933-3. URL: http://www.nature.com/articles/s41467-019-10933-3 (visited on 05/18/2022).

[233] Black Rock. *The metaverse: Investing in the future now*. https://www.blackrock.com/us/individual/insights/metaverse-investing-in-the-future. Online; accessed 27 October 2022.

[234] Rafael A. Rodríguez-Gómez, Gabriel Maciá-Fernández, and Pedro García-Teodoro. "Survey and Taxonomy of Botnet Research through Life-Cycle". In: *ACM Comput. Surv.* 45.4 (2013). ISSN: 0360-0300. DOI: 10.1145/2501654.2501659. URL: https://doi.org/10.1145/2501654.2501659.

[235] Franziska Roesner, Tadayoshi Kohno, and David Molnar. "Security and Privacy for Augmented Reality Systems". In: *Commun. ACM* 57.4 (2014), pp. 88–96. ISSN: 0001-0782. DOI: 10.1145/2580723.2580730. URL: https://doi.org/10.1145/2580723.2580730.

[236] Louis Rosenberg. "Marketing in the Metaverse and the Need for Consumer Protections". In: *2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. 2022, pp. 0035–0039. DOI: 10.1109/UEMCON54665.2022.9965661.

[237] Louis B. Rosenberg. "Regulating the Metaverse, a Blueprint for the Future". In: *Extended Reality*. Ed. by Lucio Tommaso De Paolis, Pasquale Arpaia, and Marco Sacco. Cham: Springer International Publishing, 2022, pp. 263–272. ISBN: 978-3-031-15546-8.

[238] Justas Šalkevicius et al. "Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals". In: *Electronics* 8.9 (2019). ISSN: 2079-9292. DOI: 10.3390/electronics8091039. URL: https://www.mdpi.com/2079-9292/8/9/1039.

[239] Amitav Sarma et al. *Correlation between the arm-span and the standing height among males and females of the Khasi tribal population of Meghalaya state of North-Eastern India*. 2020. DOI: 10.4103/jfmpc.jfmpc_1350_20. URL: https://journals.lww.com/jfmpc/Fulltext/2020/09120/Correlation_between_the_arm_span_and_the_standing.50.aspx (visited on 07/17/2022).

[240] Yutaka Sasaki et al. "The truth of the f-measure. 2007". In: *URL: https://www. cs. odu. edu/mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07. pdf [accessed 2021-05-26]* 49 (2007).

[241] Christian Schell, Andreas Hotho, and Marc Erich Latoschik. "Comparison of Data Encodings and Machine Learning Architectures for User Identification on Arbitrary Motion Sequences". In: *2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. ISSN: 2771-7453. Dec. 2022, pp. 11–19. DOI: 10.1109/AIVR56993.2022.00010.

[242] Christian Schell et al. *cschell/who-is-alyx: v2.0*. Version v2.0. 2023. DOI: 10.5281/zenodo.7663984. URL: https://doi.org/10.5281/zenodo.7663984.

[243] Christophe Olivier Schneble, Bernice Simone Elger, and David Shaw. "The Cambridge Analytica affair and Internet-mediated research". In: *EMBO reports* 19.8 (2018), e46579.

[244] Daniel Schneider et al. "ReconViguRation: Reconfiguring Physical Keyboards in Virtual Reality". In: *IEEE Transactions on Visualization and Computer Graphics* 25.11 (2019), pp. 3190–3201. DOI: 10.1109/TVCG.2019.2932239.

[245] NATO Science and Technology Organization. *Guidelines for Mitigating Cybersickness in Virtual Reality Systems*. URL: https://www.sto.nato.int/publications/STO%5C%20Technical%5C%20Reports/STO-TR-HFM-MSG-323/%5C$%5C$TR-HFM-MSG-323-ALL.pdf.

[246] ScienceDirect. *ScienceDirect Digital Library*. https://www.sciencedirect.com/. Online; accessed 10 Aug 2022.

[247] *ScoreSaber*. en. URL: https://www.scoresaber.com/ (visited on 01/30/2023).

[248] KW Studios Sector3 Studios. *RaceRoom*. https://www.raceroom.com/en/. Online; accessed 17 May 2022. 2022.

[249] *SFU Motion Capture Database*. URL: https://mocap.cs.sfu.ca/.

[250] Yiran Shen et al. "GaitLock: Protect Virtual and Augmented Reality Headsets Using Gait". In: *IEEE Transactions on Dependable and Secure Computing* 16.3 (2019), pp. 484–497. DOI: 10.1109/TDSC.2018.2800048.

[251] Cong Shi et al. "Face-Mic: Inferring Live Speech and Speaker Identity via Subtle Facial Dynamics Captured by AR/VR Motion Sensors". In: *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. MobiCom '21. New Orleans, Louisiana: Association for Computing Machinery, 2021, pp. 478–490. ISBN: 9781450383424. DOI: 10.1145/3447993.3483272. URL: https://doi-org.eaccess.ub.tum.de/10.1145/3447993.3483272.

[252] Mingyi Shi et al. *MotioNet: 3D Human Motion Reconstruction from Monocular Video with Skeleton Consistency*. arXiv:2006.12075 [cs]. June 2020. URL: http://arxiv.org/abs/2006.12075 (visited on 05/14/2023).

[253] Shiftall. *HaritoraX 1.1*. Online; accessed 22 Sep 2022. URL: https://en.shiftall.net/products/haritorax.

[254] Prakash Shrestha and Nitesh Saxena. "An Offensive and Defensive Exposition of Wearable Computing". In: *ACM Comput. Surv.* 50.6 (2017). ISSN: 0360-0300. DOI: 10.1145/3133837. URL: https://doi.org/10.1145/3133837.

[255] Leonid Sigal, Alexandru Balan, and Michael Black. "HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion". In: *International Journal of Computer Vision* 87 (Mar. 2010), pp. 4–27. DOI: 10.1007/s11263-009-0273-6.

[256] Sizescreens. *Samsung Gear VR 2017 detailed specifications*. https://www.sizescreens.com/samsung-gear-vr-2017-specifications/. Online; accessed 17 July 2022. 2017.

[257] Springer Link. *Springer Link Digital Library*. https://link.springer.com/. Online; accessed 10 Aug 2022.

[258] Morgan Stanley. *Metaverse: more evolutionary than revolutionary*. https://www.morganstanley.com/ideas/metaverse-investing. Online; accessed 17 May 2022. 2022.

[259] *Steam*. en. URL: https://store.steampowered.com/ (visited on 01/30/2023).

[260] SteamDB. *Most played VR Games Steam Charts*. en. URL: https://steamdb.info/charts/?tagid=21978 (visited on 02/05/2023).

[261] Julian Steil et al. "PrivacEye: Privacy-Preserving Head-Mounted Eye Tracking Using Egocentric Scene Image and Eye Movement Features". In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & amp; Applications*. ETRA '19. Denver, Colorado: Association for Computing Machinery, 2019. ISBN: 9781450367097. DOI: 10.1145/3314111.3319913. URL: https://doi.org/10.1145/3314111.3319913.

[262] Julian Steil et al. “Privacy-Aware Eye Tracking Using Differential Privacy”. In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & amp; Applications*. ETRA ’19. Denver, Colorado: Association for Computing Machinery, 2019. ISBN: 9781450367097. DOI: `10.1145/3314111.3319915`. URL: `https://doi.org/10.1145/3314111.3319915`.

[263] Sophie Stephenson et al. “SoK: Authentication in Augmented and Virtual Reality”. In: *2022 IEEE Symposium on Security and Privacy (SP)*. 2022, pp. 267–284. DOI: `10.1109/SP46214.2022.9833742`.

[264] Latanya Sweeney. “Simple Demographics Often Identify People Uniquely”. In: *Pittsburgh* (2000), p. 34.

[265] Latanya Sweeney, Akua Abu, and Julia Winn. “Identifying Participants in the Personal Genome Project by Name”. en. In: *SSRN Electronic Journal* (2013). ISSN: 1556-5068. DOI: `10.2139/ssrn.2257732`. URL: `http://www.ssrn.com/abstract=2257732` (visited on 11/28/2020).

[266] Philipp Sykownik et al. “Something Personal from the Metaverse: Goals, Topics, and Contextual Factors of Self-Disclosure in Commercial Social VR”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI ’22. New Orleans, LA, USA: Association for Computing Machinery, 2022. ISBN: 9781450391573. DOI: `10.1145/3491102.3502008`. URL: `https://doi.org/10.1145/3491102.3502008`.

[267] Piotr Szczuko. “Augmented Reality for Privacy-Sensitive Visual Monitoring”. In: *Multimedia Communications, Services and Security*. Ed. by Andrzej Dziech and Andrzej Czyżewski. Cham: Springer International Publishing, 2014, pp. 229–241. ISBN: 978-3-319-07569-3.

[268] Luma Tabbaa et al. “VREED: Virtual Reality Emotion Recognition Dataset Using Eye Tracking &amp; Physiological Measures”. In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5.4 (2022). DOI: `10.1145/3495002`. URL: `https://doi.org/10.1145/3495002`.

[269] Jerry Tang et al. “Semantic reconstruction of continuous language from non-invasive brain recordings”. In: *bioRxiv* (2022). DOI: `10.1101/2022.09.29.509744`. eprint: `https://www.biorxiv.org/content/early/2022/09/29/2022.09.29.509744.full.pdf`. URL: `https://www.biorxiv.org/content/early/2022/09/29/2022.09.29.509744`.

[270] Robert Templeman et al. “PlaceAvoider: Steering First-Person Cameras away from Sensitive Spaces”. In: *Proceedings 2014 Network and Distributed System Security Symposium*. Network and Distributed System Security Symposium. San Diego, CA: Internet Society, 2014. ISBN: 978-1-891562-35-8. DOI: `10.14722/ndss.2014.23014`. URL: `https://www.ndss-symposium.org/ndss2014/programme/placeavoider-steering-first-person-cameras-away-sensitive-spaces/` (visited on 09/21/2022).

[271] *Tilt Brush by Google*. en. URL: https://www.tiltbrush.com/ (visited on 01/30/2023).

[272] Fernando De la Torre et al. “Guide to the Carnegie Mellon University Multimodal Activity (CMU-MMAC) Database”. In: 2008.

[273] Andrew Trask et al. *Beyond Privacy Trade-offs with Structured Transparency*. 2020. arXiv: 2012.08347 [cs.CR]. URL: https://www.researchgate.net/publication/347300876_Beyond_Privacy_Trade-offs_with_Structured_Transparency.

[274] Pier Paolo Tricomi et al. “You Can’t Hide Behind Your Headset: User Profiling in Augmented and Virtual Reality”. In: *IEEE Access* 11 (2023), pp. 9859–9875. DOI: 10.1109/ACCESS.2023.3240071.

[275] Rahmadi Trimananda et al. “OVRseen: Auditing Network Traffic and Privacy Policies in Oculus VR”. In: *31st USENIX Security Symposium (USENIX Security 22)*. Boston, MA: USENIX Association, 2022, pp. 3789–3806. ISBN: 978-1-939133-31-1. URL: https://www.usenix.org/conference/usenixsecurity22/presentation/trimananda.

[276] Nikolaus F. Troje. “Decomposing biological motion: a framework for analysis and synthesis of human gait patterns.” In: *Journal of vision* 2 5 (2002), pp. 371–87.

[277] Matt Trumble et al. “Total Capture: 3D Human Pose Estimation Fusing Video and Inertial Sensors”. In: *2017 British Machine Vision Conference (BMVC)*. 2017.

[278] Nikolaos Tsalis et al. *Exploring the protection of private browsing in desktop browsers*. June 2017. DOI: 10.1016/j.cose.2017.03.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167404817300597 (visited on 07/11/2022).

[279] Wen-Jie Tseng et al. “The Dark Side of Perceptual Manipulations in Virtual Reality”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI ’22. New Orleans, LA, USA: Association for Computing Machinery, 2022. ISBN: 9781450391573. DOI: 10.1145/3491102.3517728. URL: https://doi.org/10.1145/3491102.3517728.

[280] *Unity Real-Time Development Platform: 3D, 2D, VR, and AR Engine*. en. URL: https://unity.com (visited on 01/30/2023).

[281] Unity. *Unity documentation*. https://docs.unity3d.com/Manual/VROverview.html. Online; accessed 17 July 2022. 2022.

[282] Valve. *OpenVR*. https://github.com/ValveSoftware/openvr. Online. 2022.

[283] Ashish Vaswani et al. “Attention is All You Need”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Long Beach, California, USA: Curran Associates Inc., 2017, pp. 6000–6010. ISBN: 9781510860964.

[284] Shiv Naga Prasad Vitaladevuni. “Human Movement Analysis: Ballistic Dynamics, and Edge Continuity for Pose Estimation”. In: (2007). URL: http://hdl.handle.net/1903/7610 (visited on 06/03/2023).

[285] VIVE. *Introducing VIVE Tracker*. Online; accessed 21 Sep 2022. URL: `https://www.vive.com/us/accessory/tracker3/`.

[286] Vive. *Facial Tracker*. Online; accessed 4 Oct 2022. URL: `https://www.vive.com/us/accessory/facial-tracker/`.

[287] Vive. *SteamVR Base Station 2.0*. `https://www.vive.com/us/accessory/base-station2/`. Online; accessed 17 July 2022. 2022.

[288] Vive. *Vive Cosmos*. Online; accessed 22 Sep 2022. URL: `https://www.vive.com/us/product/vive-cosmos/overview/`.

[289] Vive. *Vive Pro*. Online; accessed 22 Sep 2022. URL: `https://www.vive.com/us/product/vive-pro-full-kit/`.

[290] Vive. *Vive Pro Eye*. Online; accessed 22 Sep 2022. URL: `https://www.vive.com/us/product/vive-pro-eye/overview/`.

[291] M. Vondráček et al. "Rise of the Metaverse's Immersive Virtual Reality Malware and the Man-in-the-Room Attack & Defenses". In: *Computers & Security* (2022), p. 102923. ISSN: 0167-4048. DOI: `https://doi.org/10.1016/j.cose.2022.102923`. URL: `https://www.sciencedirect.com/science/article/pii/S0167404822003157`.

[292] VRChat. *Network Specs and Tips*. Online; accessed 4 Oct 2022. URL: `https://docs.vrchat.com/docs/network-details`.

[293] Junjue Wang et al. "A Scalable and Privacy-Aware IoT Service for Live Video Analytics". In: *Proceedings of the 8th ACM on Multimedia Systems Conference*. MMSys'17. Taipei, Taiwan: Association for Computing Machinery, 2017, pp. 38–49. ISBN: 9781450350020. DOI: `10.1145/3083187.3083192`. URL: `https://doi.org/10.1145/3083187.3083192`.

[294] Yuntao Wang et al. "A Survey on Metaverse: Fundamentals, Security, and Privacy". In: *IEEE Communications Surveys & Tutorials* (2022), pp. 1–1. DOI: `10.1109/COMST.2022.3202047`.

[295] Stanley L Warner. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias". In: *Journal of the American Statistical Association* 60.309 (1965), pp. 63–69. DOI: `10.1080/01621459.1965.10480775`.

[296] Ian Warren and Darren Palmer. "Crime risks of three-dimensional virtual environments". In: *Trends and issues in crime and criminal justice* 388 (2010), pp. 1–6.

[297] Yu-Chih Wei et al. "pISRA: privacy considered information security risk assessment model". In: *The Journal of Supercomputing* 76.3 (2020), pp. 1468–1481. DOI: `10.1007/s11227-018-2371-0`. URL: `https://doi.org/10.1007/s11227-018-2371-0`.

[298] Xing Wei and Chenyang Yang. "FoV Privacy-aware VR Streaming". In: *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. 2022, pp. 1515–1520. DOI: `10.1109/WCNC51071.2022.9771832`.

[299] P. Werner et al. “Use of the virtual action planning supermarket for the diagnosis of mild cognitive impairment: a preliminary study.” In: *Dement. Geriatr. Cogn. Disord.* (2009). DOI: 10.1159/000204915..

[300] Wiley. *Wiley InterScience Digital Library*. https://onlinelibrary.wiley.com/. Online; accessed 10 Aug 2022.

[301] C. L. Wilson. *Biometric Accuracy Standards*. 2003. URL: https://csrc.nist.gov/CSRC/media/Events/ISPAB-MARCH-2003-MEETING/documents/March2003-Biometric-Accuracy-Standards.pdf.

[302] Alexander Winkler, Jungdam Won, and Yuting Ye. *QuestSim: Human Motion Tracking from Sparse Sensors with Simulated Avatars*. 2022. DOI: https://doi.org/10.48550/arXiv.2209.09391. URL: https://arxiv.org/abs/2209.09391.

[303] Jan Wöbbeking. *Beat Saber generated more revenue in 2021 than the next five biggest apps combined*. en-US. Aug. 2022. URL: https://mixed-news.com/en/beat-saber-generated-more-revenue-in-2021-than-the-next-five-biggest-apps-combined/ (visited on 01/31/2023).

[304] Andrea Stevenson Won et al. “Homuncular flexibility in virtual reality”. In: *Journal of Computer-Mediated Communication* 20.3 (2015), pp. 241–259.

[305] David L. Woods et al. “Age-related slowing of response selection and production in a visual choice reaction time task”. In: *Frontiers in Human Neuroscience* 9 (2015). ISSN: 1662-5161. URL: https://www.frontiersin.org/article/10.3389/fnhum.2015.00193 (visited on 05/27/2022).

[306] Dongrui Wu et al. “Optimal Arousal Identification and Classification for Affective Computing Using Physiological Signals: Virtual Reality Stroop Task”. In: *IEEE Transactions on Affective Computing* 1.2 (2010), pp. 109–118. DOI: 10.1109/T-AFFC.2010.12.

[307] Felix T Wu. “Defining Privacy and Utility in Data Sets”. In: *84 University of Colorado Law Review 1117 (2013); 2012 TRPC* (2012), pp. 1117–1177. DOI: 10.2139/ssrn.2031808.

[308] Yifan Wu et al. “Privacy-Protective-GAN for Privacy Preserving Face De-Identification”. In: *Journal of Computer Science and Technology* 34.1 (Jan. 1, 2019), pp. 47–60. ISSN: 1860-4749. DOI: 10.1007/s11390-019-1898-8. URL: https://doi.org/10.1007/s11390-019-1898-8.

[309] Qing-Song Xu and Yi-Zeng Liang. “Monte Carlo cross validation”. In: *Chemometrics and Intelligent Laboratory Systems* 56.1 (2001), pp. 1–11.

[310] Chuan Yue. *Sensor-Based Mobile Web Fingerprinting and Cross-Site Input Inference Attacks*. 2016. DOI: 10.1109/SPW.2016.17.

[311] Mojtaba Zaheri, Yossi Oren, and Reza Curtmola. "Targeted Deanonymization via the Cache Side Channel: Attacks and Defenses". In: *31st USENIX Security Symposium (USENIX Security 22)*. Boston, MA: USENIX Association, Aug. 2022. URL: `https://www.usenix.org/conference/usenixsecurity22/presentation/zaheri`.

[312] Eisa Zarepour et al. "A context-based privacy preserving framework for wearable visual lifeloggers". In: *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. 2016, pp. 1–4. DOI: `10.1109/PERCOMW.2016.7457057`.

[313] Jun Zhao et al. *Reviewing and Improving the Gaussian Mechanism for Differential Privacy*. 2019. arXiv: `1911.12060 [cs.CR]`.

[314] Yingying Zhao et al. "Do Smart Glasses Dream of Sentimental Visions? Deep Emotionship Analysis for Eyewear Devices". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6.1 (2022). DOI: `10.1145/3517250`. URL: `https://doi.org/10.1145/3517250`.