Security and Privacy Approaches in Mixed Reality: A Literature Survey

JAYBIE A. DE GUZMAN, University of New South Wales, Australia and Data 61, CSIRO, Australia
KANCHANA THILAKARATHNA, University of Sydney and Data 61, CSIRO
ARUNA SENEVIRATNE, University of New South Wales and Data 61, CSIRO

Mixed reality (MR) technology development is now gaining momentum due to advances in computer vision, sensor fusion, and realistic display technologies. With most of the research and development focused on delivering the promise of MR, the privacy and security implications of this technology are yet to be thoroughly investigated. This survey article aims to put in to light these risks and to look into the latest security and privacy work on MR. Specifically, we list and review the different protection approaches that have been proposed to ensure user and data security and privacy in MR. We extend the scope to include work on related technologies such as augmented reality, virtual reality, and human-computer interaction as crucial components, if not the origins, of MR, as well as numerous related work from the larger area of mobile devices, wearables, and Internet-of-Things. We highlight the lack of investigation, implementation, and evaluation of data protection approaches in MR. Further challenges and directions on MR security and privacy are also discussed.

CCS Concepts: • Human-centered computing → Mixed / augmented reality; • Security and privacy → Privacy protections; Usability in security and privacy; • General and reference → Surveys and overviews;

Additional Key Words and Phrases: Mixed reality, augmented reality, privacy, security

ACM Reference format:
Jaybie A. de Guzman, Kanchana Thilakarathna, and Aruna Seneviratne. 2019. Security and Privacy Approaches in Mixed Reality: A Literature Survey. ACM Comput. Surv. 52, 6, Article 110 (October 2019), 37 pages.
https://doi.org/10.1145/3359626

1 INTRODUCTION

Mixed reality (MR) used to pertain to the various devices—specifically, displays—that encompass the reality-virtuality continuum as seen in Figure 1 (Milgram et al. 1994). This means that augmented reality (AR) systems and virtual reality (VR) systems are MR systems but, if categorized, will lie on different points along the continuum. Presently, mixed reality has a hybrid definition that combines aspects of AR and VR to deliver rich services and immersive experiences (Curtin 2017), and allow interaction of real objects with synthetic virtual objects and vice versa. By

J. A. de Guzman is pursuing his PhD with a scholarship from the Philippine government through its Engineering Research and Development for Technology (ERDT) program.

Authors’ addresses: J. A. de Guzman, Data 61, 13 Garden Street Eveleigh, NSW, 2015, Australia; email: j.deguzman@student.unsw.edu.au; K. Thilakarathna, School of Computer Science, University of Sydney, Australia; email: kanchana.thilakarathna@sydney.edu.au; A. Seneviratne, School of Electrical Engineering and Telecommunications, University of New South Wales; email: a.seneviratne@unsw.edu.au.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

0360-0300/2019/10-ART110 $15.00
https://doi.org/10.1145/3359626

ACM Computing Surveys, Vol. 52, No. 6, Article 110. Publication date: October 2019.
Fig. 1. Milgram and Kishino’s Reality and Virtuality Continuum (Milgram and Kishino 1994), and relative positions of where AR, VR, and AV are along the continuum. AR sits near the real environment as its primary intention is to augment synthetic objects to the physical world while VR sits near the virtual environment with varying degree of real-world information being feed in to the virtual experience. A third intermediary type, AV, sits in the middle where actual real-world objects are integrated into the virtual environment and thus intersecting with both AR and VR.

combining the synthetic presence offered by VR and the extension of the real world by AR, MR enables a virtually endless suite of applications that is not offered by current AR and VR platforms, devices, and applications.

Recent developments in computer vision (particularly in object sensing, tracking, and gesture identification), sensor fusion, and artificial intelligence has furthered the human-computer interaction as well as the machine understanding of the real-world. Likewise, technological advances in 3D rendering, optics (such as projections, and holograms), and display technologies have made possible the delivery of realistic virtual experiences. Despite varying environment sensing approaches and visual processing algorithms, most MR platforms, i.e., Google’s ARCore, Apple’s ARKit, and Windows’ Mixed Reality API, have interoperable spatial mapping data in the form of oriented point clouds. These spatial maps or point clouds can be utilized in conjunction with image, video, audio, and other sensor data to create a fully immersive MR experience. As a result, MR can now allow us to interact with machines and each other in a totally different manner: for example, using gestures in the air instead of swiping in screens or tapping on keys. The output of our interactions, also, will no longer be confined within a screen. Instead, outputs will now be mixed with our real-world experience, and soon we may not be able to tell what is real and what is synthetic. Recently released MR devices such as Microsoft’s Hololens and the Magic Leap demonstrates what these MR devices can do. They allow users to interact with holographic augmentations in a more seamless and direct manner.

Most of the work on MR for the past two decades has focused on delivering the necessary technology to make MR a possibility. As the necessary technology matures, MR devices, like any other technology, will become more available and affordable. Consequently, the proliferation of these devices may entail security and privacy implications that may yet not be known. For example, it has been demonstrated how facial images captured by a web camera can be cross-matched with publicly available online social network (OSN) profile photos to match the names with the faces and determine social security numbers (Acquisti 2011). With most MR devices coming out in a wearable, i.e., head-mounted, form-factor and having at least one camera to capture the environment, it will be easy to perform such facial matching tasks in the wild without the subjects knowing it. Security and privacy in these systems, more often than not, comes as an afterthought, as shown in this survey article.

To systematically capture the trend, we used the search tool of Scopus to initially gather AR and MR literature. We further identified works with security and privacy from this gathered list.

---

1 A point cloud is a collection of points in the three-dimensional Euclidean space accompanied by a triangle mesh or graph information to represent surfaces. The points may also be accompanied by a normal vector to determine their orientation, or their normal vectors may be estimated from the triangle mesh.
Although Scopus does not index literature from all resources, particularly in security and privacy research, the search tool can include secondary documents that have been cited by Scopus-indexed documents, which now effectively includes most security and privacy literature such as those from USENIX Security Symposium, the Internet Society’s Networks and Distributed System Security (NDSS) Symposium, and so on. Figure 2 shows the yearly percentage of these papers. Despite the increasing percentage, most only mention security and privacy, and much fewer actually discuss the impact, or present security and privacy. Likewise, among all security and privacy work, those that mentions AR/VR/MR are a lot less. Nonetheless, we supplement the gathered works from Scopus by separately searching for AR/VR/MR works with security and privacy from Google Scholar, IEEE Xplore, ACM Digital Library, and other specific venues covering computer vision, human-machine interfaces, and other related technologies.

1.1 Previous Surveys and Meta-analyses

Early surveys on AR and MR, have focused on categorizing the existing technologies. In 1994, a taxonomy for classifying mixed reality displays based on the user-interface—from monitor-based video displays to completely immersive environments—were presented and these devices were plotted along a reality-virtuality continuum (Milgram and Kishino 1994). However, in contrast to this one-dimensional continuum, two different classifications for mixed reality were also presented: (1) a two-dimensional categorization of shared space or collaborative mixed reality technologies according to concepts of transportation and artificiality, and (2) a one-dimensional classification based on spatiality (Benford et al. 1996).

Later works have focused on collecting all relevant technologies necessary for AR and VR applications. The various early challenges—such as matching the real and virtual displays, aligning the virtual objects with the real world, and the various errors that needs to be addressed such as optical distortion, misalignment, and tracking—have been discussed in (Azuma 1997). It was complemented with a following survey that focuses on the enabling technologies, interfacing, and visualization (Azuma et al. 2001). A much more recent survey updated the challenges in mR systems to be performance, alignment, interaction, mobility, and visualization (Rabbi and Ullah 2013). A review of the various head-mounted display (HMD) technologies for consumer electronics was also presented in (Kress and Starner 2013). Another study looked into a specific type of AR, mobile AR, and the different technologies that enable mobility with AR (Chatzopoulos et al. 2017). While all of these different challenges and technologies are important to enable AR, none of these studies have focused onto the fundamental issues of security and privacy in AR or MR.

Transportation is the extent to which the users are transported from their physical locality to a virtual or remote space; artificiality is the extent to which the user space has been synthesized from a real physical space; spatiality is the extent to which properties of natural space and movement is supported by the shared space.
A few others have pointed out the non-technical issues such as ethical considerations (Heimo et al. 2014) and value-sensitive design approaches (Friedman and Kahn Jr 2000) that highlights the need to consider data ownership, privacy, secrecy, and integrity. Another recent study has focused on the potential perceptual and sensory threats that can arise from MR outputs such as photosensitive epilepsy and motion-induced blindness (Baldassi et al. 2018). An earlier work emphasized the three aspects for protection in AR—input, data access, and output—over varying system complexity (from single to multiple applications, and, eventually, to multiple systems) (Roesner et al. 2014b). We expand from these three aspects, and include interaction and device protection as equally important aspects and dedicate separate discussions for both. A very recent survey collected offensive and defensive approaches on wearables (Shrestha and Saxena 2017) and it included some strategies for wearable eye-wears and HMDs. In this work, we expand the coverage to include all MR platforms and setups (including non-wearables) and present a more specific and, yet, wider exposition of MR security and privacy approaches.

1.2 Contributions
To the best of our knowledge, this is the first survey on the relevant security and privacy challenges and approaches on mixed reality. To this end, this work makes the following contributions:

1. We provide a data-centric categorization of the various works to five major aspects, and present generic system block diagrams to capture the different mechanisms of protection.
2. We include a collection of other relevant work not necessarily directed to mixed reality but is expected to be related to or be part of the security and privacy considerations for MR.
3. Last, we identify the provided security and privacy properties of these approaches as well as which data flow elements (i.e., data flow, process, storage, and/or data entity) are targeted, and present them in a summary table to show the distribution of strategies of protection among the properties.

Before proceeding to the review of the various security and privacy work, we clarify that we do not focus on network security and related topics. We will rely on the effectiveness of existing security and privacy measures that protects the communication networks, and data transmission, in general.

The rest of the survey proceeds as follows. Section 2 lists the different security and privacy properties, and explains the categorization used to sort the different approaches. The details of these approaches and which properties they address are discussed in Section 3. Last, in Section 4, the open challenges and future directions are discussed before finally concluding this survey in Section 5.

2 DEFINING AND CATEGORIZING SECURITY AND PRIVACY IN MIXED REALITY
In this section, we first present the important security and privacy requirements that need to be considered in designing or using MR. Then, we present an an overall categorization that focuses on the flow of data within an MR system.

2.1 General Security and Privacy Requirements
We derive security and privacy properties from three models and combine them to have an overarching model from which we can refer or qualify the different approaches (both defensive and offensive strategies) that will be discussed in this survey article. Table 1 lists the 13 combined security and privacy properties from Microsoft’s Security Development Lifecycle (or SDL) (Howard and Lipner 2006), PriS (Kalloniatis et al. 2008), and LINDDUN (Deng et al. 2011) and their corresponding threats. The SDL has been popularly used by industry to elicit security threat scenarios and
Table 1. Combined General Security and Privacy Properties and Their Corresponding Threats

| Property                  | Threat                      | Model                          |
|---------------------------|-----------------------------|--------------------------------|
|                           |                             | Microsoft’s SDL 2006 | PriS 2008 | LINDDUN 2011 |
| Security-oriented         |                             | ✓                              | ✓         | ✓            |
| Integrity                 | Tampering                   | ✓                              | ✓         | ✓            |
| Non-repudiation           | Repudiation                 | ✓                              | ✓         | ✓            |
| Availability              | Denial of Service           | ✓                              | ✓         | ✓            |
| Authorization             | Elevation of Privilege      | ✓                              | ✓         |             |
| Authentication            | Spoofing                    | ✓                              | ✓         |             |
| Identification            | Anonymity                   | ✓                              | ✓         |             |
| Confidentiality           | Disclosure of Information   | ✓                              | ✓         | ✓            |
| Anonymity & Pseudonymity  | Identifiability             | ✓                              | ✓         | ✓            |
| Unlinkability             | Linkability                 | ✓                              | ✓         |             |
| Unobservability & Undetectability | Detectability        | ✓                              | ✓         |             |
| Plausible Deniability     | Non-repudiation             | ✓                              | ✓         |             |
| Content Awareness         | Unawareness                 | ✓                              | ✓         |             |
| Policy & Consent Compliance | Non-compliance             | ✓                              | ✓         |             |

It is interesting to note that some security properties are conversely considered as privacy threats, such as non-repudiation and plausible deniability. This highlights the differences in priority that an organization, user, or stakeholder can put into these properties or requirements. However, these properties are not necessarily mutually exclusive and can be desired at the same time. Moreover, the target element to be secured or made private adds another dimension to this list. Specifically, these properties can be applied to the data flow elements data entities, data flow, process, and data storage as shown in Figure 3. We highlight these co-existence as well as example target elements as we go through each of these properties one-by-one.

1. **Integrity**—The data storage, flow, or process in MR is not and cannot be tampered or modified. This is to ensure that, for example, the visual targets are detected correctly and the appropriate augmentations are displayed accordingly. No unauthorized parties should be able to modify any of these elements in an MR system.

2. **Non-repudiation**—Any modification, or generation of data, flow, or process cannot be denied especially if the entity is essential or an adversary was able to perform such modifications or actions. When necessary, the modifier or generator of the action should be identified and cannot deny that it was their action. In privacy, however, the converse is desired.

3. **Availability**—All necessary data, flow, or process for an MR system should be available to satisfy and accomplish the targeted or intended service. An adversary should not be able to impede the availability of these entities or resources.
(4) Authorization and Access Control—All actions or processes should be originated from authorized and verifiable parties. The same actions are also actuated according to their appropriate access privileges. Or only the augmentations from applications that have been authorized to deliver augmentations should be rendered.

For example, only the authorized application should be able to access the cup and its data in our sample MR environment in Figure 3. The integrity of the cup and its data should be ensured and, if modifications were done, the modifying agent cannot refuse (non-repudiate) their action. The resulting modification should also not lead to non-availability of the service.

(5) Identification—All actions should be identified to the corresponding entity, i.e., user or party. In a security context, this is interrelated with authorization and authentication properties. Verified identities are used for authorizing access control. Unidentified parties can be treated as adversaries to prevent unidentifiable and untraceable attacks. In sensitive situations, e.g., an attack has occurred, anonymity is not desired.

(6) Authentication—Only the legitimate users of the device or service should be allowed to access the MR device or service. Their authenticity should be verified through an authentication method and, then, identification or authorization can follow after a successful authentication.

(7) Confidentiality—All actions involving sensitive or personal data, flow, or process should follow the necessary authorization, and access control policies. Parties that are not authorized should not have access to these confidential elements. All elements can be assumed as confidential especially storage or processing of personal and re-identifiable data.

(8) Anonymity and Pseudonymity—Entities should be able to remove their association or relationship to the data stored, flow, or process. Likewise, a pseudonym can be used to link the entities but should not be able to be linked back to specific identities. Moreover, an adversary should not be able to identify the user from combinations of information from
these elements. However, in the security context, anonymity is not desired especially when adversaries need to be identified.

9) **Unlinkability**—Any link or relationship of the entity, i.e., user or party, to the data stored, flow, or process as well as with other entities (e.g., data to data, data to flow, and so on) cannot be identified or distinguished by an adversary. For example, co-located MR users, who want to keep their co-location (linking) information private, should not be identifiable by an adversary.

10) **Undetectability and Unobservability**—Any entities’ existence cannot be ensured or distinguished by an attacker; or an entity can be deemed unobservable or undetectable by an adversary; or the entity cannot be distinguished from randomly generated entities. For example, an MR game like Pokemon Go needs access to the camera view of the user device and determine the ground plane to place the Pokemon on the ground as viewed by the user, but the game does not need to know what other objects are within view.

Unlinkability, undetectability, and unobservability properties can be extended to include *latent privacy* protection. It is the protection of entities that are not necessitated by an application or service but can be in the same domain as their target entities. This includes *bystander privacy*.

11) **Plausible Deniability**—An entity should be able to deny that they are the originator of a process, data flow or data storage. This is a converse of the non-repudiation security property, which is actually the corresponding threat for plausible deniability (or repudiation). However, plausible deniability is essential when the relationship of an entity (as originator) to stored or processed data should not be divulged; however, non-repudiation is essential if, for example, when an action needs to be identified.

In the case of the cup and its data in the previous example, if the authorized application connects with a health insurance provider and transmits aggregated *anonymized* data of the cup, the resulting aggregated data should not be *linkable* to the users from which they originated. Hence, the users are provided with plausible deniability for user-specific sensitive information that may have been included in the aggregated transmission, or, with the help of machine learning, inferred by, say, the health insurance provider.

12) **Content Awareness**—The user as an entity should be aware of all data flows or processes divulged especially those that are sensitive. And that they should be aware that they have released the information that was necessary for providing the required functionality. In the MR context, for example, the user should be made aware that the MR application not only captured the spatial information around them but also the visual information.

13) **Policy and consent Compliance**—An MR system should follow the policies that aims to protect the user’s privacy or security. There should be a guarantee that third-party applications or services follow these policies.

For every approach that will be discussed in the next section (Section 3), we will identify which properties they are trying to address. Moreover, there are other “soft” properties (i.e., reliability and safety) that we will liberally use in the discussions. We categorize these approaches in a *data- and data flow* -centric fashion as explained in the next subsection.

### 2.2 Categorizing the Threats

Figure 3 presents an example of an MR environment and shows how data flows from the *observable* environment, and *within* the MR device to process inputs and deliver augmentations. The left-half of the diagram shows the “view” of the mixed reality device, which, in this example, is a see-through MR head-mounted device (HMD) or, simply, an MR eye-wear. Within the view are the
physical objects that are “seen” by the MR device as indicated by the solid arrows. The synthetic augmentations are shown in the diagram which are represented by the broken arrows. A cloud- or web-based supporting service is also shown through which multiple MR users can collaborate or share MR experiences, say, through a social network which supports MR such as Snapchat and Pokemon Go.

The right-half of the figure shows the data flow diagram which follows the MR processing pipeline of detection → transformation → rendering. Physical entities (such as the desk, cup, or keyboard) from the environment are captured or detected. After detection, the resulting entities will be transformed or processed to deliver services accordingly. Depending on the service or application, different transformations are used. Finally, the results of the transformation are delivered to the user by rendering them (such as the virtual pet bird or the cup-contents indicator) through the device’s output interfaces.

(1) **Input Protection**—This first category focuses on the challenges in ensuring security and privacy of data that is gathered and inputted to the MR platform. These data can contain sensitive information. For example, in Figure 3, the MR eye-wear can capture the sensitive information on the user’s desktop screen (labelled 1) such as e-mails, chat logs, and so on. These are user-sensitive information that needs to be protected. Similarly, the same device can also capture information that may not be sensitive to the user but may be sensitive to other entities such as bystanders. This is called bystander privacy. Aside from readily sensitive objects, the device may capture other objects in the environment that are seemingly benign (or subtle) and were not intended to be shared but can be used by adversaries to infer knowledge about the users/bystanders. All the necessary protections can be mapped to privacy properties of confidentiality, unobservability and undetectability, and content awareness.

Another emerging class of threats in the multimedia space are adversarial input attacks which aim to deceive detection algorithms, especially machine learning-based vision algorithms (Goodfellow et al. 2014). This kind of attack undermines the integrity and availability of the intended MR service. Vision algorithms are presented with tampered inputs, whether through hardly perceptible digital (Kurakin et al. 2016), or physical (Sharif et al. 2016) perturbations, that result in different outputs or prevents a legitimate input from being recognized such as prevent a vision-based navigation MR application from recognizing a stop sign. Similar adversarial attacks have also been demonstrated on voice commands which allows an attacker to produce an unintelligible command to humans but are recognized by voice-enabled devices (Carlini et al. 2016).

The various input protection and defense approaches are discussed in Section 3.1.

(2) **Data Protection**—Data from multiple sensors or sources are aggregated, and, then, stored in a database or other forms of data storage. Applications, then, need to access the data to deliver output in the form of user-consumable information or services. However, almost all widely used computing platforms allows applications to collect and store data individually (as shown in the access of supporting data services labelled 2 in Figure 3) and the users may have no control over their data once it has been collected and stored by these applications. Majority of security and privacy risks have been raised concerning the access and use of user data by third party agents, particularly, on user data gathered from wearable (Felt et al. 2012), mobile (Lee et al. 2015), and on-line activity (Ren et al. 2016). Thus, MR systems faces even greater risks as richer information can be gathered using a wide variety of sensitive sensors—i.e., visual data from which spatial mapping information can be extracted to determine spatial features such as surfaces or physical objects on which augmentations are rendered over. For data protection, there are a lengthy list
of properties that needs to be maintained such as integrity, availability, confidentiality, unlinkability, anonymity and pseudonymity, and plausible deniability among others. Section 3.2 will present a discussion of the data protection approaches.

(3) Output Protection—After processing the data, applications send outputs to the mixed reality device to be displayed or rendered. However, in MR, applications may inadvertently have access to outputs of other applications. If an untrusted application has access to outputs other than those needed for its functionality, then they can potentially modify those outputs making them unreliable. For example, in the smart information (labelled 3) hovering over the cup in Figure 3, malicious applications can modify the sugar level information. Other adversarial output attacks include clickjacking, which deceives users to “clicking” on sensitive elements through transparent or misleading interfaces (Roesner et al. 2014b), and physiological attacks such as inducing epileptic seizures through a visual trigger (Baldassi et al. 2018). Furthermore, when one application’s output is another application’s input, this necessitates multiple application access to an output object. For output protection, the integrity, non-repudiation, availability, and policy compliance as well as reliability properties has to be maintained. The approaches are discussed in Section 3.3.

(4) User Interaction Protection—MR mixes or includes the utilization of other sensing and display interfaces to allow immersive interactions. Examples of these are room-scale interfaces, which allow multiple users to interact in the same MR space (combining virtual and physical spaces), or the virtual “transporation” of users in MR video-conferencing to allow them to seemingly co-exist in the same virtual space while being fully aware of their physical space. Thus, we expand the coverage of protection to ensure protected sharing and collaborations (labelled 4 in Figure 3) in MR.

In contrast to current widely adapted technologies like computers and smart phones, MR can enable entirely new and different ways of interacting with the world, with machines, and with other users. One of the key expectations is how users can share MR experiences with assurances of security and privacy of information. Similar to data protection, there is a number of properties that is necessary in interaction protection, namely, non-repudiation, authorization, authentication, identifiability, and policy and consent compliance. Details of the approaches in protecting user interactions are discussed in Section 3.4.

(5) Device Protection—This last category focuses on the actual physical MR device, and the input and output interfaces. This implicitly protects data that is used in the above four aspects by ensuring device-level protection. Authentication, authorization, and identifiability are among the most important properties for device protection. In Section 3.5, the different novel approaches in device access and physical display protection are discussed.

Figure 4 shows these categories and their subcategories. The first three categories are directly mapped to the associated risks with the main steps of the processing pipeline—protecting how applications, during the transformation stage, access real-world input data gathered during detection, which may be sensitive, and generate reliable outputs during rendering. Likewise, some input and output protection approaches are also applied in the transformation stage, while some data access protection approaches, e.g., data aggregation, are usually applied in the detection and rendering stages. Furthermore, the interaction protection and device protection approaches cannot be mapped along the pipeline unlike the other three as the intended target elements of these two categories may transcend this pipeline.

3Theoriz studio designed and developed a room-scale MR demonstration (http://www.theoriz.com/portfolio/mixed-reality-project/); Microsoft’s Holoportation demonstrates virtual teleportation in real-time (https://www.microsoft.com/en-us/research/project/holoportation-3/).
The presented categorization does not exclusively delineate the five aspects, and it is significant to note that most of the approaches that will be discussed can fall under more than one category or subcategory. Notwithstanding, this survey article complements the earlier surveys by presenting an up-to-date collection of security and privacy research and development for the past two decades on MR and related technologies and categorizing these various works according to the presented data-centric categorization. The next section proceeds in discussing these various approaches that have been done to address each of the five aspects discussed above.

3 SECURITY AND PRIVACY APPROACHES

In the following subsections, we present the various security and privacy work that has been done on MR and related technologies, especially on AR. We have organized these approaches according to the five major categories and, where applicable, to their further subcategories. Most approaches, especially input and output approaches, rely on inserting an intermediary protection layer, as shown in Figure 5, that enables in- and out-flow control of data from trusted elements to untrusted elements. There may be instances the presented solutions may address several aspects and may fall on more than one category. For these cases, we focus on the primary objective or challenges focused by the solution. We first discuss the threats specific to that category and, then, follow it with the security and privacy approaches, and strategies both in literature and those used in existing systems.
Fig. 6. Example strategies for input protection: (1) information reduction or partial sanitization, e.g., from RGB facial information to facial outline only; (2) complete sanitization or blocking; or (3) skeletal information instead of raw hand video capture. (The broken arrow indicates less privileged information flow.)

3.1 Input Protection

Perhaps the main threat to input protection, as well as to the other four categories, is the unauthorized and/or unintended disclosure of information—may it be of actual data entities or of their flow. These vulnerable inputs can be categorized in two based on the user intention. Targeted physical objects and other non-user-intended inputs are both captured from the environment usually for visual augmentation anchoring. We can collectively call these inputs as passive, while those that are intentionally provided by users, such as gestures, can be considered as active inputs. Approaches in protecting sensitive information from captured data are primarily aimed at access, i.e., authorization, control, and privacy. However, protection against adversarial inputs are aimed toward securing integrity and availability of the intended MR services. The protection is aimed at securing the integrity of the process (e.g., object recognition for navigation assistance) rather than the data itself.

3.1.1 Threats to Passive Inputs: Targeted and Non-intended Latent Data. Aside from threats to confidentiality (i.e., information disclosure), the two other main threats to sensitive, personally identifiable information during data capture are detectability and user content unawareness. Both stem from the fact that these MR systems (just like any other service that employs a significant number of sensors) collect a large amount of information, and among these are necessary and sensitive information alike. As more of these services becomes personalized, the sensitivity of these information increases. These threats are very evident with visual data. MR, as well as AR and VR, requires the detection of targets, i.e., objects or contexts, in the real environment, but other non-necessary and sensitive information are captured as well. These objects become detectable and users may not want for these latent information or contexts to be detected. Likewise, as they continuously use their MR device, users may not be made aware that the applications running on their device are also able to collect information about these objects and contexts.

Protection Approaches. The most common input protection approaches usually involves the removal of latent and sensitive information from the input data stream. These approaches are generally called input sanitization techniques (see samples labelled 1 and 2 in Figure 6). These are usually implemented as an intermediary layer between the sensor interfaces and the applications as shown in Figure 5. In general, this protection layer acts as an input access control mechanism aside from sanitization. These techniques can further be categorized according to the policy enforcement—whether intrinsic or extrinsic policies for protection are used. With intrinsic enforcement, the user, device, or system itself imposes the protection policies that dictates the input sanitization that is applied. However, extrinsic input protection arises from the need for sensitive objects external to the user that are not considered by the intrinsic policies. In reference to the data flow diagram in
Figure 3, intrinsic policies are only aimed at protecting the primary entities while extrinsic policies also includes external entities, e.g., bystanders. In the following subsections, the sanitization techniques are presented as either intrinsic or extrinsic.

(1) **Intrinsic input sanitization** policies are usually user-defined. For example, the DARKLY system (Jana et al. 2013b) for perceptual applications uses OpenCV in its intermediary input protection layer to implement a multi-level feature sanitation. The basis for the level or degree of sanitization are the user-defined policies. The users can impose different degrees of sensitivity permissions, which affects the amount of detail or features that can be provided to the applications, i.e., stricter policies mean less features are provided. For example, facial information can vary from showing facial feature contours (of eyes, nose, brows, mouth, and so on) to just the head contour depending on the user’s preferences. The user can actively control the level of information that is provided to the applications. Thus, aside from providing undetectability and unobservability and content awareness to users, DARKLY also provides a form of authorization through information access control, specifically a least privilege access control.

**Context-based Sanitization.** A context-based intrinsic sanitization framework (Zarepour et al. 2016) improves on the non-contextual policies of DARKLY. It determines if there are sensitive objects in the captured images, like faces or car registration plates, and automatically implements sanitization. Sensitive features are sanitized by blurring them out, while images of sensitive locations (e.g., bathrooms) are deleted entirely. Similarly, PLACEAVOIDER (Templeman et al. 2014) also categorizes images as sensitive or not, depending on the features extracted from the image, but deletion is not automatic and still depends on the user. Despite the context-based nature of the sanitization, the policy that governs how to interpret the extracted contexts are still user-defined, thus, we consider both sanitization techniques as intrinsic. However, intrinsic policy enforcement can be considered as self-policing, which can potentially have a myopic view of privacy preferences of other users and objects. Furthermore, intrinsic policies can only protect the inputs that are explicitly identified in the policies.

**Video Sanitization.** The previously discussed sanitization techniques were for generic capturing devices and were mostly sanitizing images and performs the sanitization after the image is stored. For MR platforms that require real-time video feeds, there is a need for live and on-the-fly sanitization of data. A privacy-sensitive visual monitoring (Szczykuko 2014) system was implemented by removing persons from a video surveillance feed and rendering 3D animated humanoids in place of the detected and visually removed persons. Another privacy-aware live video analytic system called OPENFACE-RTFACE (Wang et al. 2017) focused on performing fast video sanitization by combining it with face recognition. OPENFACE-RTFACE system lies near the edge of the network, or on cloudlets. Similar approaches to edge or cloud-assisted information sanitization can potentially be utilized for MR.

(2) **Extrinsic input sanitization** receives input policies, e.g., privacy preferences, from the environment. An early implementation (Truong et al. 2005) involved outright capture interference to prevent sensitive objects from being captured by unauthorized visual capturing devices. A camera-projector set up is used. The camera detects unauthorized visual capture devices, and the projector beams a directed light source to “blind” the unauthorized device. This technique can be generalized as a form of a physical access control, or, specifically, a deterrent to physical or visual access. However, this implementation requires a
dedicated set up for every sensitive space or object, and the light beams can be disruptive to regular operation.

Other approaches involve the use of existing communication channels or infrastructure for endorsing or communicating policies to capture devices, and to ensure that enforcement is less disruptive. The goal was to implement a fine-grained permission layer to “automatically” grant or deny access to continuous sensing or capture of any real-world object. A simple implementation on a privacy-aware see-through system (Hayashi et al. 2010) allowed other users detected to be blurred out or sanitized and shown as human icons only if the viewer is not their friend. However, this requires that users have access to the shared database and explicitly identify friends. Furthermore, enabling virtually anyone or, in this case, anything to specify policies opens new risks such as tampering, and malicious policies.

To address authenticity issues in this so called world-driven access control, policies can be transmitted as digital certificates (Roesner et al. 2014c) using a public key infrastructure (PKI). PKI provides cryptographic protection to media access and sanitization policy transmission. However, the use of a shared database requires that all possible users’ or sensitive objects’ privacy preferences have to be pushed to this shared database. Furthermore, it excludes or, unintentionally, leaves out users or objects that are not part of the database, which defeats the purpose of a world-driven protection.

I-pic (Aditya et al. 2016) removes the involvement of shared databases. Instead users endorse privacy choices via a peer-to-peer approach using Bluetooth Low Energy (BLE) devices. However, I-pic is only a capture-or-no system. PrivacyCamera (Li et al. 2016b) is another peer-to-peer approach but is not limited to BLE. Also, it performs face blurring, instead of just capture-or-no, using endorsed GPS information to determine if sensitive users are within camera view. However, CARDEA (Shu et al. 2016) allows users to use hand gestures to endorse privacy choices. In CARDEA, users can show their palms to signal protection while a peace-sign to signal no need for protection. These three approaches are targeted at bystander privacy protection, i.e., facial information sanitization.

MARKIT (Raval et al. 2014) can provide protection to any user or object through the use of privacy markers and gestures (similar to CARDEA) to endorse privacy preferences to cameras. It was integrated to Android’s camera subsystem to prevent applications from leaking private information (Raval et al. 2016) by sanitizing sensitive media. This is a step closer to automatic extrinsic input sanitization, but it requires visual markers for detecting sensitive objects. Furthermore, all these extrinsic approaches have only been targeted for visual capture applications and not with AR- or MR-specific ones.

3.1.2 Threats to Gestures and other Active User Inputs. Another essential input that needs to be protected is gesture input. We put a separate emphasis on this as gesture inputs entails a “direct” command to the system, while the previous latent and user inputs do not necessarily invoke commands. Currently, the most widely adopted user input interfaces are tactile such as the keyboard, mouse, and touch interfaces. However, these current tactile inputs are limited by the dimension\(^4\) of space that they are interacting with and some MR devices now don’t have such interfaces. Also, these input interface types are prone to a more physical threat such as \textit{external inference} or \textit{shoulder-surfing} attacks. From which, threats such as \textit{spoofing}, \textit{denial of service}, or \textit{tampering} may arise.

\(^4\)Keyboards and other input pads can be considered as one-dimensional interfaces, while the mouse and the touch interfaces provides two-dimensional space interactions with limited third dimension using scroll, pan, and zoom capabilities.
Furthermore, there is a necessity for new user input interfaces to allow three-dimensional inputs. Early approaches used gloves (Dorfmuller-Ulhaas and Schmalstieg 2001; Thomas and Piekarski 2002) that can determine hand movements, but advances in computer vision have led to tether- and glove-free 3D interactions. Gesture inference from smart watch movement have also been explored particularly on finger-writing inference (Xu et al. 2015). Vision-based natural user interfaces (NUI), such as the Leap Motion (Zhao and Seah 2016) and Microsoft Xbox Kinect, have long been integrated with MR systems to allow users to interact with virtual objects beyond two dimensions. This allows the use of body movement or gestures as input channels and move away from keypad and keyboards. However, the use of visual capture to detect user gestures or using smart watch movement to detect keyboard strokes means that applications that require gesture inputs can inadvertently capture other sensitive inputs (Maiti et al. 2016). Similar latent privacy risks such as detectability and content unawareness arise. Thus, as new ways of interacting in MR are being explored, security and privacy should also be maintained.

Protection Through Abstraction. PREPOSE (Figueiredo et al. 2016) provides secure gesture detection and recognition as an intermediary layer (as in Figure 5). The PREPOSE core only sends gesture events to the applications, which effectively removes the necessity for untrusted applications to have access to the raw input feed. Similar to DARKLY, it provides least privilege access control to applications, that is, only the necessary gesture event information is transmitted to the third party applications and not the raw gesture feed.

Some work prior to PREPOSE implemented the similar idea of inserting a hierarchical recognizer (Jana et al. 2013a) as an intermediary input protection layer. They inserted Recognizers to the Xbox Kinect to address input sanitization as well as to provide input access control. The Recognizer policy is user-defined, thus, an intrinsic approach. Similarly, the goal is to implement a least privilege approach to application access to inputs—applications are only given the least amount of information necessary to run. For example, a dance game in Xbox, e.g., Dance Central or Just Dance, only needs body skeletal (similar to sample labelled 3 in Figure 6) movement information, and it does not need facial information, thus, the dance games are only provided with the moving skeletal information and not the raw video feed of the user while playing. To handle multiple levels of input policies, the Recognizer implements a hierarchy of privileges in a tree structure, with the root having highest privilege, i.e., access to RGB and depth information, and the leaves having lesser privileges, i.e., access to skeletal information.

SEMA DROID (Xu and Zhu 2015), however, is a device level protection approach. It is a privacy-aware sensor management framework that extends the current sensor management framework of Android and allows users to specify and control fine-grained permissions to applications accessing sensors. Just like the other abstraction strategies, it is implemented as an intermediary protection layer that provides users application access control or authorization to sensors and sensor data. What differs is its application of auditing and reporting of potential leakage and applying them to a privacy bargain. This allows users to ‘trade’ their data or privacy in exchange for services from the applications. There are a significant number of work on privacy bargain and the larger area of privacy economics, and we refer the readers to Acquisti’s work (Acquisti et al. 2016).

Remaining Challenges in Protected Data Capture. Most of the approaches discussed so far are founded on the idea of least privilege. However, it requires that the intermediary layer, e.g., the Recognizers, must know what type of inputs or objects the different applications will require. PREPOSE addresses this for future gestures but not for future objects. For example, an MR painting application may require the detection of different types of brushes but the current recognizer does not know how to “see” or detect the brushes. Extrinsic approaches like MARKIT try to address this by using markers to tell which objects can and cannot be seen. What seemingly arises now
is the need to have a dynamic abstraction and/or sanitization of both pre-determined and future sensitive objects. In Section 3.5.2, we will focus on device-level protection approaches to protect user activity involving input interfaces.

3.1.3 Defending against Adversarial Inputs. The previous protection approaches are primarily focused on ensuring privacy of sensitive and personally identifiable information during data capture. However, adversarial inputs primarily pose security threats to the MR processes, or the entire MR system, rather than the sensitive information itself. While adversarial machine learning is applied in a broader space, most of the attack and defense work has been done over images, i.e., MNIST (LeCun et al. 2010) and ImageNet (Deng et al. 2009), which demonstrates how such threats can be easily realized in MR.

In the subsequent discussion, we present the major fronts of defense against adversarial inputs that have been presented in the literature; however, as the space of adversarial machine learning is fast developing, newer and much powerful attacks are presented against the previously presented defenses. We point the user to recent survey work for wider exposition on adversarial attacks and defenses (Akhtar and Mian 2018; Yuan et al. 2019). Nonetheless, the broad approaches of (1) adversarial detection, (2) architectural or model modification, and (3) adversarial training still remain as the major directions for adversarial defense.

1) Adversarial detection usually involves the use of a separate classifier to distinguish adversarial inputs from legitimate ones. This approach has been demonstrated to be effective on adversarial voice commands (Carlini et al. 2016). It can also be used as an initial defensive step as demonstrated in MagNet followed by a reformation step, using an autoencoder, to recover the legitimate input information from the detected adversarial input (Meng and Chen 2017).

2) Architectural modifications to the machine learning, i.e., neural network (NN or DNN for deep neural network), model has also incorporated the use of autoencoders to have input recovery within the model itself (Gu and Rigazio 2014). Other model modifications have proposed the use of visual causal features (Chalupka et al. 2014) for learning which are robust to adversarial examples. A RELU-based modification (Zantedeschi et al. 2017) took advantage of the linearity observed on adversarial examples (Goodfellow et al. 2014). Yet another work has focused on a learning method that ensures a certified error upper bound, which no adversarial attack can cause to go beyond (Raghunathan et al. 2018). This method can also be used during training as a regularizer that encourages robustness against attacks.

3) Adversarial training includes adversarial examples during training to make the model robust against them (Goodfellow et al. 2014). Another method for training is the use of distillation within the DNN to improve its generalization and, hence, enhance its robustness against adversarial perturbations (Papernot et al. 2016). However, adversarial training on one model may not be adept over adversarial perturbations produced from another NN model; likewise, the cross-model adaptability of detection- and modification-based defenses may also not be ensured. To address that, a recent work proposed to train using an ensemble of training data with adversarial augmentations transferred from various models (Tramèr et al. 2017).

As we have mentioned earlier, most (if not all) of these defenses have received counter-attacks that defeat the defenses. Nevertheless, regardless of the approach, the intention of the defense is to (1) provide integrity to the detection process (e.g., classification), (2) ensure its availability, and (3) ensure that only authentic inputs are detected. Consequently, adversarial machine learning has
also been used as a privacy-preserving method to confuse an adversarial data collector; we present these techniques in the following subsection (i.e., Section 3.2.2).

### 3.2 Data Protection

Most of the data protection techniques proposed and discussed in the wider literature were implemented on generic systems and not necessarily MR systems. Furthermore, most of the approaches we will discuss have been aimed at traditional visual media, i.e., images and video. While MR still relies heavily on visual data, 3D spatial data is now primarily utilized to represent spatial understanding. Nevertheless, as MR data moves away from the MR platform, say to auxiliary cloud- or web-based support services, the similar protection techniques may be used for MR data. Thus, we present an exposition of these techniques on the following data flow aspects: (1) data aggregation, (2) privacy-preserving data processing, and (3) data storage, particularly, access to it, has to be protected as well.

Generally, the aim of these data protection approaches is to allow services or third-party applications to learn without leaking unnecessary and/or personally identifiable information. Usually, these protection approaches use privacy definitions such as $k$-anonymity, and differential privacy. $k$-anonymity (Samarati 2001; Samarati and Sweeney 1998) ensures that records are unidentifiable from at least $k-1$ other records. It usually involves data generalization techniques to ensure privacy, but suffers from scaling problems, i.e., larger data dimensions, which can be expected from MR platforms and devices with a high number of sensors or input data sources. Differentially private algorithms (Dwork et al. 2014), however, inserts randomness to data to provide plausible deniability and unlinkability. The guaranteed privacy of differentially private algorithms is well-studied (McSherry and Talwar 2007). In the following subsections, we now focus on the different threats and approaches to data aggregation, processing, and storage.

#### 3.2.1 Protected Data Collection and Aggregation

Essentially, data collection also falls under the input category, but we focus on data after sensing and how systems, applications, and services handle data as a whole. Protected data collection and aggregation approaches are also implemented as an intermediate layer as in Figure 5. Usually, data manipulation or similar mechanisms are run on this intermediary layer to provide a privacy guarantee, e.g., differential privacy or $k$-anonymity, among released data. RAPPOR or randomized response (Erlingsson et al. 2014) is an example of a differentially private data collection and aggregation algorithm. It is primarily applied for privacy-preserving crowd-sourced information such as those collected by Google for their Maps services. Privacy-preserving data aggregation (PDA) has also been adopted for information collection systems (He et al. 2007, 2011) with multiple data collection or sensor points, such as wireless sensor networks or body area networks. Overall, the goal of privacy-preserving data collection and aggregation is to get aggregate statistics or information without divulging individual information; thus providing anonymity, unlinkability, and plausible deniability between the aggregate information (as well as its derivative processes and further resulting information) and the data source entity, i.e., a user.

#### 3.2.2 Protected Data Processing

After collection, most services will have to process the data immediately to deliver outputs in real-time. Thus, similar to data collection, the same privacy threats of information disclosure, linkability, detectability, and identifiability holds. During processing, third-party applications or services can directly access user data, which may contain sensitive or personal information if no protection measures are implemented. The subsequent exposition of protection approaches presents a collection of MR-related work particularly on privacy-preserving and secure image and video processing.
Fig. 7. Generic block diagrams of two example data protection approaches: (1) cryptographic technique using secure multi-party computation where two or more parties exchange secrets (1.1 and 1.3) to extract combined knowledge (1.2 and 1.4) without the need for divulging or decrypting each others data share; and (2) personal data stores with “trusted” applets.

(1) **Encryption-based techniques.** Homomorphic encryption (HE) allows queries or computations over encrypted data. In visual data processing, this has been used for image feature extraction and matching for various uses such as image search, and object detection. HE-Sift (Hsu et al. 2011) performs bit-reversing and local encryption to the raw image before feature description using SIFT (Lowe 2004). The goal was to make dominant features, which can be used for context inference, recessive. As a result, feature extraction, description, and matching are all performed in the encrypted domain. A major drawback is the very slow computation time due to the near full-homomorphism used as well as the approach being algorithm-specific. Using leveled HE can reduce the computation time of HE-Sift (Jiang et al. 2017).

SecSift (Qin et al. 2014, 2016) improves on the computation time of HE-Sift by instead using a somewhat homomorphic encryption, i.e., order-preserving encryption. They split or distribute the SIFT feature computation tasks among a set of “independent, co-operative cloud servers to keep the outsourced computation procedures as simple as possible and avoid utilizing homomorphic encryption.” Other improvements utilized big data computation techniques to expedite secure image processing such as the use of a combination of MapReduce and ciphertext-policy attribute-based encryption (Zhang et al. 2014), or the use of Google’s Encrypted BigQuery Client for Paillier HE computations (Ziad et al. 2016).

(2) **Secret Sharing or Secure Multi-party Computation.** Data can be split among untrusted parties assuming that information can only be inferred when the distributed parts are combined (Huang et al. 2011; Yao 1986). Secure multi-party computation (SMC) or secret sharing allows computation of data from two or more sources without necessarily knowing about the actual data each source has. The diagram labelled 1 in Figure 7 shows a possible SMC setup. For example, App 1 requires data from another party (could be another application) to provide a certain service. It encrypts its share of the data (step 1.0) and sends it to the other party (1.1). The other party then encrypts its other share (1.2) and sends it to App 1 (1.3). Both can compute the results over the combined encrypted shares without the need

\[\text{SIFT} \text{ or Scale-invariant Feature Transform} \text{ is a popular image feature extraction and description algorithm.}\]
to decrypt their shares. For example, P3 (Ra et al. 2013) uses two-party secret sharing “by splitting a photo into a public part, which contains most of the volume (in bytes) of the original, and a secret part, which contains most of the original’s information.” It uses AES-based symmetric keys to encrypt the secret part and allows the use of a tunable parameter between storage/bandwidth and privacy. This approach, however, is JPEG-format specific.

A virtual cloth try-on (Sekhavat 2017) service used secret sharing and secure two-party computation. The anthropometric information of the user is split between the user’s mobile device and the server, and are both encrypted. The server has a database of clothing information. The server can then compute a 3D model of the user wearing the piece of clothing by combining the anthropometric information and the clothing information to generate an encrypted output, which is sent to the user device. The user device decrypts the result and combines it with the local secret to reveal the 3D model of the user “wearing” the piece of clothing.

(3) Data Manipulation and Perturbations. Other MR-related privacy-preserving techniques have focused on facial de-identification using image manipulation to achieve k-anonymity for providing identity privacy (Gross et al. 2006, 2008; Newton et al. 2005). Succeeding face de-identification work has focused on balancing utility and privacy (Du et al. 2014); much recent work have utilized generative adversarial networks, using the same adversarial idea discussed in Section 3.1.3 but leveraging it for deceiving a potentially adversarial data collector, to de-identify faces but ensuring high utility of the resulting de-identified face (Brkic et al. 2017; Wu et al. 2018).

The same adversarial manipulation can be extended over 3D spatial data that is utilized in current AR/MR systems. Instead of providing complete 3D spatial data, a sanitized or “salted” virtual reconstruction of the physical space can be provided to third-party applications. For example, instead of showing the 3D capture of a table in the scene with all 3D data of the objects on the table, a generalized horizontal platform or surface can be provided. The potentially sensitive objects on the table are thus kept confidential. A tunable parameter provides the balance between sanitization and utility. Using this tunability, similar notions of privacy guarantee to differential privacy and k-anonymity can be provided. However, this approach is yet to be realized but virtual reconstruction has been used to address delayed alignment issues in AR (Waegel 2014). This approach can work well with other detection (Section 3.1.2) and rendering (Section 3.3.2) strategies of sanitization and abstraction as well as in privacy-centred collaborative interactions (Section 3.4.1). It also opens the possibility to have an active defence strategy where “salted” reconstructions are offered as a honeypot to adversaries.

Overall these privacy-preserving data processing techniques aim to provide the privacy properties of unlinkability, unobservability, and plausible deniability between the process (as well as its results) and the data source. Furthermore, the encryption- and secret sharing-based techniques further provide security properties of integrity and authorization as only the authorized parties can process the data while ensuring confidentiality through homomorphism. All these techniques complement each other and may be used simultaneously. Thus, any or all of these techniques can be applied to MR and it will only be a matter of whether the technique is appropriate for the amount of data and level of sensitivity of data that is tackled in MR environments.

3.2.3 Protecting Data Storage. After collection and aggregation, applications store user data on separate databases in which users have minimal or no control over. Privacy concerns on how

---

6Anthropometric information are body measurements of an individual that capture size, shape, and so on.
these applications use user data beyond the expected utility to the user have been posed (Felt et al. 2012; Lee et al. 2015; Ren et al. 2016). Aside from these privacy threats, there are inherent security threats such as tampering, unauthorized access, and spoofing. To provide security against such threats, the Advanced Encryption Standard (or AES) has been specified as the industry standard.

When trustworthiness of third-party applications and services are not ensured, protected data storage solutions, such as personal data stores (PDS), with managed application access permission control is necessary. PDSs allows the users to have control over their data and which applications have access to it. Figure 7 shows a generic block diagram (labelled 2) of how a PDS protects the user data by running it in a protected sand-box machine that can monitor the data that is provided to the applications. Usually, applet versions of the applications run within the sand-box. Various PDS implementations have been proposed such as the personal data vaults (PDV) (Mun et al. 2010), OpenPDS (de Montjoye et al. 2014), and the DABOX (Crabtree et al. 2016). Other generic protection approaches focused on encrypted fragmented data storage (Ciriani et al. 2010) or decentralized storage using blockchains (Zyskind et al. 2015). As a result, PDS provides accountability and subsequently the non-repudiation security property as applications cannot deny that they have accessed the stored data. Privacy-preserving aggregation can also be implemented within the PDS to provide privacy properties of anonymity, unlinkability and plausible deniability between the released aggregate data and the user as a data source. For example, OpenPDS releases private aggregates or answers through its SafeAnswers interface.

3.2.4 Remaining Challenges in Data Protection. There are necessary modifications that applications have to partake to implement these data protection strategies. Aside from implementation complexity additional resources may be necessary such as the inherent need of memory, and compute capacity when employing encryption. There are attempts to eliminate the necessity of code modification, such as in GUPT (Mohan et al. 2012), which focuses on the sampling and aggregation process to ensure distribution of the differential privacy budget and eliminating the need for costly encryption. Also, combining these techniques with protected sensor management and data storage to provide confidentiality through sanitization and authorized access control is promising.

3.3 Output Protection

The prime objective of MR is to deliver immersive experiences. To achieve that, applications ship services and experiences in the form of rendered outputs. In general, there are three possible types of outputs in MR systems: real-world anchored outputs, non-anchored outputs, and outputs of external displays. The first two types are both augmented outputs. The last type refers to outputs of other external displays that can be utilized by MR systems, and vice versa. Protecting these outputs is of paramount importance aside from ensuring input and data protection. As a result, there are three aspects when it comes to output protection: output control, protected rendering, and protecting external displays.

3.3.1 Output Reliability and User Safety. Current MR systems have loose output access control. As a result, adversaries can potentially tamper or spoof outputs that can compromise user safety. Further threats of denial of service, and policy and consent non-compliance is also present.

Output control policies can be used as a guiding framework on how MR devices will handle outputs from third-party applications. This includes the management of rendering priority, which could be in terms of synthetic object transparency, arrangement, occlusion, and other possible spatial attributes to combat attacks such as clickjacking. An output access control framework (Lebeck et al. 2016) with an object-level granularity have been proposed to make output handling enforcement easier. It can be implemented as an intermediary layer, as in Figure 5, and follows a set of output policies. In a follow-up work, they presented a design framework (Lebeck et al. 2017) for
output policy specification and enforcement, which combined output policies from Microsoft’s HoloLens Developer guidelines, and the U.S. Department of Transportation’s National Highway Traffic Safety Administration (NHTSA) for user safety in automobile-installed AR. They designed a prototype platform called ARYA that will implement the application output control based on the output policies specified, and evaluated ARYA on various simulated scenarios. Aside from policy compliance, ARYA ensures security properties of integrity, non-repudiation, availability, and authorization; that is, correct outputs are always available, an output’s originator cannot be denied, and only authorized applications can produce such outputs. Succeeding work builds up on ARYA’s weakness when it comes to dynamic and complex environments especially when various, differently sourced policies are required (Ahn et al. 2018); they utilised reinforcement learning to determine the optimal policy enforcement assisted by fog-based servers. This approach reinforces the properties of integrity and availability in complex and dynamic environments, and further provides confidentiality by doing processing at the edge instead of the cloud.

3.3.2 Threats During Rendering. Other MR environments incorporates any surface or medium as a possible output display medium. For example, when a wall is used as a display surface in an MR environment, the applications that use it can potentially capture the objects or other latent and/or sensitive information within the wall during the detection process. This specific case intersects very well with the input category, because what is compromised here is the sensitive information that can be captured in trying to determine the possible surfaces for displaying.

Privacy-preserving Rendering. Applications that requires such displays do not need to know what the contents on the wall are. It only has to know that there is a surface that can be used as a display. Protected output rendering protects the medium and, by extension, whatever is in the medium. Least privilege has been used in this context (Vilk et al. 2014). For example, in a room-scale MR environment, only the skeletal information of the room, and the location and orientation of the detected surfaces (or display devices) is made known to the applications that wish to display content on these display surfaces (Vilk et al. 2015). This utilizes the same abstraction strategy, which has been used both in input and data protection to provide unobservability and undetectability and authorized access. This example of room-scale MR environments is usually used for collaborations.

3.3.3 Threats to Output Displays. Output displays are vulnerable to physical inference threats or visual channel exploits such as shoulder-surfing attacks. These are the same threats to user inputs (Section 3.1.2) especially when the input and output interfaces are on the same medium or are integrated together such as on touch screens.

Protecting Outputs from External Inference. To provide secrecy and privacy on certain sensitive contexts, which requires output confidentiality (e.g., ATM bank transactions), MR can be leveraged. This time, MR capabilities are leveraged for output defense strategies.

(1) Content hiding. EYEGUIDE (Eaddy et al. 2004) used a near-eye HMD to provide a navigation service that delivers secret and private navigation information augmented on a public map display. Because the EYEGUIDE display is practically secret, shoulder surfing is prevented.

Other approaches involve the actual hiding of content. For example, VR CODES (Woo et al. 2012) takes advantage of rolling shutter to hide codes from human eyes but can be detected by cameras at a specific frame rate. Shutter glasses can also be used to similarly hide displayed content (Yerazunis and Carbone 2001). A similar approach has been used

---

7Here are two example descriptions of the policies: (1) “Don’t obscure pedestrians or road signs” is inspired from the NHTSA; (2) “Don’t allow AR objects to occlude other AR objects” is inspired from the HoloLens’s guidelines.

ACM Computing Surveys, Vol. 52, No. 6, Article 110. Publication date: October 2019.
to hide AR tags in video (Lin et al. 2017). This type of technique can hide content from human attackers but is still vulnerable to machine-aided inference or capture.

(2) **Visual cryptography.** Secret display approaches have also been used in visual cryptographic techniques such as visual secret sharing (VSS) schemes. VSS allows the “mechanical” decryption of secrets by overlaying the visual cipher with the visual key. However, classical VSS was aimed at printed content (Chang et al. 2010) and requires strict alignment, which is difficult in AR and MR displays, particularly handhelds and HMDs. The VSS technique can be relaxed to use code-based secret sharing, e.g., barcodes, QR codes, and 2D barcodes. The ciphers are publicly viewable while the key is kept secret. An AR-device can then be used to read the cipher and augment the decrypted content over the cipher. This type of visual cryptography have been applied to both print (Simkin et al. 2014) and electronic displays (Andrabi et al. 2015; Lantz et al. 2015).

Electronic displays are, however, prone to attacks from malicious applications that have access to the display. One of these possible attacks is cipher rearrangement for multiple ciphers. To prevent such in untrusted electronic displays, a visual ordinal cue (Fang and Chang 2010) can be combined with the ciphers to provide the users immediate signal if ciphers have been rearranged.

These techniques can also be used to provide protection for sensitive content on displays during input sensing. Instead of providing privacy protection through post-capture sanitization, the captured ciphers will remain secure as long as the secret shares or keys are kept secure. Thus, even if the ciphers are captured during input sensing, the content stays secure. In general, these visual cryptography and content-hiding methods provide visual access control, i.e., *authorization* and *confidentiality* in shared or public resources. More device-level examples of this technique are discussed in Section 3.5.2.

### 3.3.4 Remaining Challenges in Output Protection

Similar to input protection, output protection strategies can use the same abstraction approach applied as an intermediary access control layer between applications and output interfaces or rendering resources. To enforce these output abstractions, a reference policy framework has to exist through which the abstraction is applied. As a result, perhaps, the biggest challenge is the specification and enforcement of these policies, particularly who will specify them and how they will be effectively enforced. In the output side, risks and dangers are more imminent, because adversaries are about to actuate or have already actuated the malicious response or output. Thus, these access control strategies and policies are necessary for output protection.

Malicious inference or capture of outputs present the same threats as input inference. Section 3.5.2 will focus on device-level protection approaches to output interfaces and displays.

### 3.4 Protecting User Interactions

When it comes to collaboration, technologies such as audio-visual teleconferencing, and *computer-supported collaborative work* (or CSCW) enable live sharing of information among multiple users. These are called *shared space* technologies as more than one user interacts in the same shared space, whether physically or virtually, as shown in Figure 8(a). And MR offers a much more immersive sharing space for collaborations.

Concerns on the boundaries between physical and virtual spaces (Figure 3) in MR and on the directionality of these boundaries have been raised (Benford et al. 1998). The directionality can influence the balance of power, mutuality and privacy between users in shared spaces. For example, the boundary (labelled 1) in Figure 8(b) allows User 2 to receive full information (solid arrow labelled 2) from User 1 while User 1 receives partial information (broken arrow labelled 3) from
User 2. The boundary enables an “imbalance of power,” which can have potential privacy and ethical effects on the users. For example, early observation work shows territoriality in collaborative tabletop workspaces (Scott et al. 2004).

Early attempts on ensuring user privacy in a collaborative MR context provided users with a private space, as shown in Figure 8(c), that displays outputs only for that specific user while having a shared space for collaboration. For example, the PIP or personal interaction panel was designed to serve as a private interface for actions and tasks that the user does not want to share with the collaborative virtual space (Szalavári and Gervautz 1997). It is composed of a tracked “dumb” panel and a pen. It was used as a gaming console to evaluate its performance. In the following subsection, we first look at some other early work on collaborative interactions, and, then, proceed with examples of how to manage privacy in shared spaces.

3.4.1 Threats During Collaborative Interactions. Early collaborative platform prototypes demonstrated fully three-dimensional collaboration in MR (Billinghurst and Kato 1999; Grasset and Gascuel 2002; Hua et al. 2004; Regenbrecht et al. 2002; Schmalstieg and Hesina 2002). However, none have addressed the concerns raised from information sharing due to the boundaries created by shared spaces. An adversarial user can potentially tamper, spoof, or repudiate malicious actions during these interactions. As a result, legitimate users may suffer denial of service and may be unaware that their personal data may have been captured and then leaked.

Protecting Collaborative Interactions. Most of the following approaches ensures the privacy properties of content awareness and policy and consent compliance. We look at the specific strategies and targets of protection these approaches are aimed at.

1) Enabling user-originated policies. EMMIE (Environmental Management for Multi-user Information Environments) (Butz et al. 1999) is a hybrid multi-interface collaborative environment that uses AR as a 3D interface at the same time allowing users to specify privacy of certain information or objects through privacy lamps and vampire mirrors (Butz et al. 1998). EMMIE’s privacy lamps are virtual lamps that “emit” a light cone in which users can put objects within the light cone to mark these objects as private. However, the vampire mirrors are used to determine privacy of objects by showing full reflections of public objects while private objects are either invisible or transparent. However, the privacy lamps and vampire mirrors only protect virtual or synthetic content and does not provide protection to real-world objects. Similar user-enabled privacy has been demonstrated in ROOMPLANNER using hand gestures to enforce privacy, through private spaces and outputs, in a digital tabletop (Wu and Balakrishnan 2003).
Kinected Conference (DeVincenzi et al. 2011) allows the participants to use gestures to impose a temporary private session during a video conference. Aside from that, they implemented synthetic focusing using Microsoft Kinect’s depth sensing capability where other participants are blurred to direct focus on a participant who is speaking, and augmented graphics hovering above the user’s heads to show their information such as name, shared documents, and speaking time. The augmented graphics serve as feed-through information to deliver signals that would have been available in a shared physical space but is not readily cross-conveyed between remote physical spaces.

(2) Multi-user coordination policies. Early work on mediating conflicts in digital workspaces explored the use of multi-user coordination policies (Morris et al. 2006a). For example, to increase group awareness, they employed cooperative gestures, which requires gesture contributions from more than one user to enforce a single command, such as clearing the entire screen when users do the erase gesture together (Morris et al. 2006b).

(3) Feed-through signalling. SecSpace (Reilly et al. 2014) explores a feed-through mechanism to allow a more natural approach to user management of privacy in a collaborative MR environment. In contrast to Kinected Conference’s gesture-based privacy session, and Emmie’s privacy lamps and vampire mirrors, users in SecSpace are provided feed-through information that would allow them to negotiate their privacy preferences. Figure 8(b) shows an example situation in which User n enters the shared space (labelled 4) on the same physical space as User 2, which triggers an alarm (labelled 5) or notification for User 1. The notification serves as a feed-through signalling that crosses over the MR boundary. By informing participants of such information, an imbalance of power can be rebalanced through negotiations.

Non-AR Feed-through signalling have also been used in a non-shared space context like the candid interactions (Ens et al. 2015), which uses wearable bands that lights up in different colors depending on the smart-phone activity of the user, or other wearable icons that change shape, again, depending on which application the icon is associated to. However, the pervasive nature of these feed-through mechanisms can still pose security and privacy risks, thus, these mechanisms should be regulated and properly managed. In addition, the necessary infrastructure, especially for SecSpace, to enable this pervasive feed-through system may be a barrier to adaptation. A careful balance between the users’ privacy in a shared space and the utility of the space as a communication medium is sought to be sought.

(4) Private and public space interactions. Competitive gaming demands secrecy and privacy to make strategies while performing other tasks in a shared environment. Thus, it is a very apt use case for implementing user protection in a shared space. Private Interaction Panels (or PIPs) demonstrates a gaming console functionality where a region that is defined within the PIP panel serves as a private region (Szalavári et al. 1998). For example, in a game of Mah-jongg, the PIP panel serves as the user’s space for secret tiles while all users can see the public tiles through their HMDs. The PIP pen is used to pick-and-drop tiles between the private space and the public space. However, TouchSpace implements a larger room-scale MR game. It uses an HMD that can switch between see-through AR and full VR, an entire floor as shared game space with markers, and a wand for user interactions with virtual objects (Cheok et al. 2002). Essentially, Emmie’s privacy lamps and mirrors also act as private spaces.

BragFish (Xu et al. 2008) implements a similar idea on privacy to that of the PIP with the use of a handheld AR device, i.e., Gizmondo. In BragFish, a game table with markers serves as the shared space, while each user has the handheld AR device that serves as the
private space for each user. The handheld AR device has a camera that is used to “read” the markers associated to a certain game setting, and it frees the user from the bulky HMDs as in PIP and TouchSPACE. The Gizmondo handheld device has also been used in another room-scale AR game (Mulloni et al. 2008). Similarly, camera phones have been used as a handheld AR device in a tabletop marker-based setup for collaborative gaming (Henrysson et al. 2005).

Overall, the aim of these protected collaborative interactions in MR is to provide confidentiality for relevant information that users may deem sensitive in a shared context, while a few also provided non-repudiation so that an action that affects other users’ activity cannot be denied by the actor or subsequently used to identify them. Other approaches, such as cooperative gestures, also leads to ensuring availability of the shared task. Perhaps, an important aspect that has risen is the utilization of different portions of space with different functions during interactions, namely, a public portion for shared objects or activities, and a private portion for user-sensitive objects or tasks. However, shared space platforms assume that user can freely share and exchange content or information through an existing interaction channel. In the next section, we focus on how to protect the sharing channel in an MR context.

3.4.2 Threats to Sharing Initialization. All those shared space systems that were previously discussed rely on a unified architecture to enable interactions and sharing on the same channel. However, there might be cases that sharing is necessary but no pre-existing channel exists, or an entire architecture, just like in SecSPACE or EMMIE, to support sharing is not readily available. Thus, a sharing channel needs to be initialized. The same threats of spoofing and unauthorized access from Personal Area Networks, such as in ZigBee or Bluetooth PAN arises.

Securing Sharing Channels. Similar techniques to out-of-band channels can be used to achieve a secure channel initialization. LOOKSGOODToME is an authentication protocol for device-to-device sharing (Gaebel et al. 2016). It leverages on the camera/s and wireless capabilities of existing AR HMDs. Specifically, it uses the combination of distance information through wireless localization and facial recognition information to cross-authenticate users. In other words, using the AR HMD that has a camera and wireless connectivity, users can simply look at each other to authenticate and initiate sharing. HOLOPAIR, however, avoids the use of wireless localization, which may be unavailable and inefficient in devices, and instead utilizes exchange of visual cues between users to confirm the shared secret (Slugarovic et al. 2017). Both uses the visual channel as an out-of-band channel.

3.4.3 Remaining Challenges in Sharing and Interactions. The most apparent challenge is the varying use cases in which users interact or share. A recent work exposes the various user concerns, such as technological misuse and access negotiation, that can arise on a multi-user MR environment (Lebeck et al. 2018). Depending on the context or situation, privacy and security concerns, as well as the degree of concern, can vary. For example, feed-through signalling may be necessary in classroom scenarios to inform teachers when students enter and leave the classroom. However, there would also be occasions that it could be perceived to be too invasive or counter-intuitive, for example during military negotiations in the field. Thus, there is a great deal of subjectivity to determine what is the most effective protection mechanism during sharing or interactions. Perhaps, before everything else, we should ask, “Who or what are we protecting?”

3.5 Device Protection

Given the capabilities of MR devices, their privacy and security risks have been raised. Various data protection approaches have also been proposed as described in the previous subsections. To
complement these approaches, the devices themselves have to be protected as well. There are two general aspects that needs protection at the device level: device access, and input/output interfaces.

3.5.1 Threats to Device Access. The primary threats to device access are identity spoofing and unauthorized access. All approaches described below aim to provide protection against such threats. Some approaches, particularly those using physiological or biometric information, also ensures identifiability of users in addition to authorization and authentication.

Novel Authentication Strategies. Device access control ensures that authorized users are provided access while unauthorized ones are barred. Password still remains as the most utilized method for authentication (Dickinson 2016). To enhance protection, multi-factor authentication (MFA) is now being adopted, which uses two or more independent methods for authentication. It usually involves the use of the traditional password method coupled with, say, a dynamic key that can be sent to the user via SMS, email, or voice call. The two-factor variant has been recommended as a security enhancement, particularly for on-line services like email, cloud storage, e-commerce, banking, and social networks.

Aside from passwords are PIN- and pattern-based methods that are popular as mobile device authentication methods. A recent study (George et al. 2017) evaluated the usability and security of these established pin- and pattern-based authentication methods in virtual interfaces and showed comparable results in terms of execution time compared to the original non-virtual interface. The following sections look at other novel authentication methods that leverages existing and potential capabilities of MR devices.

1) Gesture- and Active Physiological-based Authentication. We look at the various possible gestures that can easily be captured by MR devices, specifically finger, hand, and head gestures. Mid-air finger and hand gestures have been shown to achieve an accuracy between 86–91% (based on corresponding accuracy from the equal error rate or EER) using a 3D camera-based motion controller over a test population of 200 users (Aslan et al. 2014). A combination of head gestures and blinking gestures triggered by a series of images shown through the AR HMD have also been evaluated and promises an approximately 94% of balanced accuracy rate in user identification over a population of 20 users (Rogers et al. 2015). However, HEADBANGER uses head-movements triggered by an auditory cue (i.e., music) and achieved a True Acceptance Rate (TAR) of 95.7% over a test population of 30 users (Li et al. 2016a). Other possible gestures or active physiological signals, such as breathing (Chauhan et al. 2017), are also potential methods.

2) Passive Physiological-based Authentication. Passive methods include physiological or biometric signals. Physiological-signal-based key agreement (PSKA) (Venkatasubramanian et al. 2010) used PPG features locked in a fuzzy-vault for secure inter-sensor communications for body area networks or BAN. Despite existing MR devices not having PPG sensing capabilities, the PSKA method can be utilized for specific use cases when MR devices need to communicate with other devices in a BAN, which can potentially be PPG sensing capable. However, SKULLCONDUCT (Schneegass et al. 2016) uses the bone conduction capability of the Google Glass for user identification (with TAR of 97%) and authentication (EER of 6.9%). All these novel methods show promise on how latent gestures, physiological signals, and device capabilities can be leveraged for user identification and authentication.

3) Multi-modal and/or Biometric Authentication combines two or more modes in a singular method instead of involving other methods or bands of communication is called multi-modal authentication. One multi-modal method combines facial, iris, and periorcular information for user authentication and has an EER of 0.68% (Raja et al. 2015).
Fig. 9. Sample interface and display protection strategies: (1) inserting a polarizer to prevent or block display leakage; and (2) visual cryptography, e.g., using secret augmentations (2.2) through decryption (2.1) of encrypted public interfaces (2.0). All elements to the left of the optical display element are considered vulnerable to external inference or capture.

...combines gaze gestures and touch keys as a singular pass-key for smart phones to counter shoulder-surfing attacks on touch-based pass keys (Khamis et al. 2016). These types of authentication methods can readily be applied to MR devices that has gaze tracking and other near-eye sensors.

3.5.2 Threats to Physical Interfaces. As discussed in Sections 3.1.2 and 3.3.3, MR interfaces are vulnerable from malicious inference, which leads to disclosure of input activity, and/or output display information. Currently available personal AR or MR see-through HMDs project or display content through lenses. The displayed content on the see-through lenses can leak private information and be observed externally. Visual capture devices can be used to capture and extract information from the display leakage. External input interfaces suffer from the same inference and side-channel attacks such as shoulder-surfing.

Protection Approaches. There are optical and visual strategies that can be used to provide interface and activity confidentiality and unobservability. Figure 9 shows example strategies of optical blocking and visual cryptography.

1) Optical strategies have been proposed, such as the use of polarization on the outer layer (as in Figure 9, labelled 1), use of narrowband illumination, or a combination of the two to maximize display transmission while minimizing leakage (Kohno et al. 2016). As of yet, this is the only work on MR display leakage protection using optical strategies.

There are other capture protection strategies that have been tested on non-MR devices, which allows objects to inherently or actively protect themselves. For example, the TaPS widgets use optical reflective properties of a scattering foil to only show content at a certain viewing angle (Möllers and Borchers 2011). Active camouflaging techniques have also been used, particularly on mobile phones, which allows the screen to blend with its surrounding just like a chameleon (Pearson et al. 2017). Both TaPS widgets and the chameleon-inspired camouflaging are physically hiding sensitive objects or information from visual capture. The content-hiding methods discussed in Section 3.3.3 to hide outputs are also optical strategies.
Visual cryptography and scrambling techniques for display protection have also been discussed in Section 3.3.3. The same can also be used for protecting sensitive input interfaces. EyeDecrypt (Forte et al. 2014) uses visual cryptography technique to protect input/output interfaces as shown in Figure 9, labelled 2. The publicly viewable input interface is encrypted (Figure 9, step 2.0), and the secret key is kept or known by the user. The user uses an AR device to view the encrypted public interface and, through the secret key, is visually decrypted (Figure 9, step 2.1). As a result, only the user can see the actual input interface through the AR display (Figure 9, step 2.2). It utilizes out-of-band channels to securely transmit the cryptographic keys between two parties (i.e., the client, through the ATM interface, and the bank). It also provides defense if the viewing device, i.e., the AR HMD, is untrusted by performing the visual decryption in a secure server rather than on the AR device itself.

Another AR-based approach secretly scrambles keyboard keys to hide typing activity from external inference (Maiti et al. 2017). Only the user through the AR device can see the actual key arrangement of the keyboard. However, these techniques greatly suffer from visual alignment issues, i.e., aligning the physical interface with the AR rendered objects.

Remaining Challenges in Device Interface Protection. Despite the use-cases with visual cryptography using AR or MR displays, the usability of this technique is still confined to specific sensitive use cases due to the requirements of alignment. Furthermore, this type of protection is only applicable to secrets that are pre-determined, specifically, information or activities that are known to be sensitive, such as password input or ATM PIN input. These techniques are helpful in providing security and privacy during such activities in shared or public space due to the secrecy provided by the near-eye displays, which can perform the decryption and visual augmentation. Evidently, it only protects the output or displayed content of external displays but not the actual content, which are displayed through the AR or MR device.

We have presented both defensive and offensive, as well as active and passive, strategies to device protection. Nonetheless, there are still numerous efforts on improving the input and output interfaces for these devices and it is opportune to consider in parallel the security and privacy implications of these new interfaces.

3.6 Summary of Security and Privacy Approaches in Mixed Reality

Finally, we present in Table 2 an overall comparison of these approaches based on which security and privacy properties they are addressing and to which elements these properties are provided for using the same symbols used in Figure 3, namely, \( \Diamond \) data flow, \( \bigcirc \) process, \( \square \) storage, and/or \( \triangle \) entity.

Generalizations and Gaps. Unsurprisingly, majority of the approaches are targeting confidentiality, i.e., preventing information disclosure, except for some non-MR-but-related work such as adversarial example defenses (Section 3.1.3). Furthermore, the categorization, to some extent, localized the targeted properties. It is also clear that some properties are rather specific to certain approaches, e.g., authentication is of course targeted by authentication approaches. Nonetheless, trends and clustering of target properties among the categories are evident. The first two major categories roughly target properties that are more privacy leaning. However, the last three categories were fairly balanced. Moreover, after confidentiality, the next most targeted properties are authorization, undetectability and unobservability, and policy and consent compliance.

Consequently, it is evident that MR-targeted protection approaches, particularly those approaches under input protection, still primarily lack provisions for plausible deniability. It is of no surprise that data protection approaches (which are mostly generic and non-MR targeted, while a few are proto-MR or MR-related) are the ones that primarily target this property. Furthermore,
Table 2. Summary of MR Approaches that Have Been Discussed and Which Security and Privacy Properties They Provide to Which Data Flow Element: ◇ data flow, ○ process, □ storage, and/or △ entity

| Section | Approach | Integrity | Non-Repudiation | Availability | Authorization | Confidentiality | Anonymity | Unlinkability | Undetectability | Deniability | Awareness | Compliance |
|---------|----------|-----------|-----------------|--------------|---------------|-----------------|-----------|--------------|----------------|------------|-----------|------------|
| 3.1.1   | DARKLY (Jana et al. 2013b)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.1   | Context-based sanitization (Zarepour et al. 2016)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.1   | PLACEAVOIDER (Templeman et al. 2014)† |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.1   | 3D humanoids replace humans (Szczuko 2014)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.1   | OPENFACE/RTFACER (Wang et al. 2017)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.1   | Capture-resistant spaces (Truong et al. 2005)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.1   | See-through vision (Hayashi et al. 2010)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.1   | World-driven access control (Roessner et al. 2014c‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.1   | I-fit (Aditya et al. 2016)† |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.1   | PRIVACYCAMERA (Li et al. 2016b)† |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.1   | CARDEA (Shu et al. 2016)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.1   | MARKIT (Raval et al. 2014, 2016)† |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.2   | PrePose (Figueiredo et al. 2016)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.2   | RECOGNIZERS (Jana et al. 2013a)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.2   | SEMABROID(Xu and Zhu 2015)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.3   | Adversarial Detection (Carlini et al. 2016; Meng and Chen 2017)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.3   | Model Modification (Gu and Rigazio 2014; Raghunathan et al. 2018; Zantedeschi et al. 2017)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.1.3   | Adversarial Training (Goodfellow et al. 2014; Papernot et al. 2016; Tramèr et al. 2017)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |

Data Protection Approaches

| Section | Approach | Integrity | Non-Repudiation | Availability | Authorization | Confidentiality | Anonymity | Unlinkability | Undetectability | Deniability | Awareness | Compliance |
|---------|----------|-----------|-----------------|--------------|---------------|-----------------|-----------|--------------|----------------|------------|-----------|------------|
| 3.2.1   | Randomized Response (Erlingsson et al. 2014)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.2.2   | HE-SIFT (Hsu et al. 2011)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.2.2   | LevelLED HE-SIFT(Jiang et al. 2017)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.2.2   | CRYPTO-M (Ziad et al. 2016)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.2.2   | SEC-SIFT (Qin et al. 2014, 2016)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.2.2   | P3 (Ra et al. 2013)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.2.2   | Cloth Try-on (Sekhavat 2017)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.2.2   | k-anonymous faces (Gross et al. 2006, 2008; Newton et al. 2005)† |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.2.2   | GAN-based face de-identification (Brikic et al. 2017; Du et al. 2014; Wu et al. 2018)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.2.3   | PDV (Mun et al. 2010)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.2.3   | OPENPDS (de Montjoye et al. 2014)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.2.3   | DATABASE(Crabtree et al. 2016)§ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |

Output Protection Approaches

| Section | Approach | Integrity | Non-Repudiation | Availability | Authorization | Confidentiality | Anonymity | Unlinkability | Undetectability | Deniability | Awareness | Compliance |
|---------|----------|-----------|-----------------|--------------|---------------|-----------------|-----------|--------------|----------------|------------|-----------|------------|
| 3.3.1   | ARIA (Lebeck et al. 2016, 2017)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.3.1   | Fog-based output security (Ahn et al. 2018)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.3.2   | SURROUNDWEB (Vilk et al. 2015, 2014)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.3.3   | EYEGUIDE (Eaddy et al. 2004)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.3.3   | VR CODES (Woo et al. 2012)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |
| 3.3.3   | Psycho-visual modulation (Lin et al. 2017)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |

(Yerazunis and Carbone 2001)‡ |          | ◇               | △            | △             | □              | △         | △            | ◇              | △          |           |            |

(Continued)
The entity can be the data itself, the user as the originator of the data, or the adversary (say, identifiability of an adversary as a security provision).

The approaches have been applied to either an § MR context, an † MR-related (or proto-MR) context, or a ◦ non-MR context.

For the target elements, we only focus the top two elements if the approaches target more than two.

majority of the data and input protection approaches were applied over traditional media such as images and video, while current widely utilized MR platforms are capturing 3D spatial data to represent the physical space. Thus, there is a huge necessity to design and evaluate data protection approaches aimed at current and upcoming types of data utilized by MR systems.

4 OPEN CHALLENGES

The previous section has elaborated on the different security and privacy approaches in the research literature and highlighted the remaining challenges for each category. Now, we present the
high-level open challenges brought about by existing and upcoming MR devices, and our potential future with MR.

**Security and Privacy of Existing Devices and Platforms.** Risks over MR outputs (Baldassi et al. 2018) and during interactions (Lebeck et al. 2018) have recently been presented. A similar systematic security and privacy analysis of MR applications, devices, and platforms has to be performed to identify other potential and latent risks particularly on its input capabilities. For example, scanning capability of current devices, such as the Hololens, should be investigated whether it can potentially be used to detect heartbeats or other physiological signals of bystanders, or how these 3D spatial data can be further abused by adversaries. Then, we can evaluate these MR systems against the security and privacy requirements we have specified. Upon consideration of these requirements, we can, then, come up with protection mechanisms targeted for these devices and platforms.

**Native Support with Fine-grained Access Permissions.** Furthermore, the current MR processing pipeline is performed mostly by the applications and in a very monolithic manner. They would need direct access to sensing interfaces (via APIs) to perform detection, and to output interfaces for rendering. Once provided permission, AR applications, inevitably, have indefinite access to these resources and their data. Native, operating system-level support for detection and rendering can provide an access control framework to ensure security and privacy of user data. OS-level abstractions have been proposed to “expose only the data required to [third-party] applications” (D’Antoni et al. 2013). Apple’s ARKit and Google’s ARCore APIs made the move to provide such support to MR application development on their platforms, but the granularity of their access control is still coarse and similar risks are still present.

**Society, Policy, and Ethics.** Possible social consequences (Feiner 1999), and, further, implications on bystander privacy, which have already been brought up several times, may arise from the user’s appearance while wearing the device. Existing universal, ethical, and legal concepts on privacy and policy may need to be revisited to catch up with MR becoming mainstream. Ultimately, as these technologies will be delivering information overlaid on the physical space, the correctness, safety, and legality of this information has to be ensured (Roesner et al. 2014a). This is now more evident with demonstrations such as Google’s Duplex and Visual Positioning System projects, which they demonstrated during Google I/O 2018, as well as the now prevalent work on deepfakes.

**Best Practices for MR Security and Privacy.** Despite the existing security and privacy best practices guidelines recommended by government, industry, and non-government institutions, we argue that a more specific set of guidelines targeted for MR is necessary.\(^8\) First, we emphasize on how access to various data sources (i.e., sensors) should elicit permission requests to the user to disaggregate access privileges. Second, APIs should allow separate access to raw spatial data to that of the released spatial data, which could be a generalized version of the raw one. This allows for the implementation of tunable transformations to produce a privacy-preserving version of the spatial data. Last, we recommend runtime access permission requests with visualizations informing users of the content (or version of the content) applications are desiring access to. MR’s freedom from physical displays can be leveraged to present such visualizations in a more

\(^8\)For example, the Australian government has published guidelines for mobile app developers [https://www.oaic.gov.au/agencies-and-organisationsguides/guide-for-mobile-app-developers]; other agencies or institutions providing guidelines for security and privacy best practices include NIST [https://www.nist.gov/programs-projects/nist-cybersecurity-iot-program], and the IoT Security Foundation [https://www.iotsecurityfoundation.org/best-practice-guidelines/].
informative and immersive manner. Practicing such guidelines that emphasize privilege separation among data flows in MR remains as a big challenge.

5 CONCLUSION

This is the first survey to take in the endeavour of collecting, categorizing, and reviewing various security and privacy approaches in MR. We have raised various known and latent security and privacy risks associated with the functionalities of MR and gathered a comprehensive collection of security and privacy approaches on MR and related technologies. We have identified five aspects of protection, namely, input, data access, output, interactivity, and device integrity, and we categorized the approaches according to these five aspects. Furthermore, we identify which security and privacy properties are targeted by the approaches using a list of thirteen known properties. We, then, used these properties to present a high-level description of the approaches and use it to compare them. Among the properties, confidentiality, authorization, and undetectability were the most targeted, while there is a considerable lack of provision for other properties. Furthermore, there is also a lack of data protection strategies that are targeted toward 3D MR data, specifically those that are employed by now widely used MR platforms, i.e., Google ARCore, Apple ARKit, and Windows MR API. Therefore, it is opportune to design, investigate, and implement security and privacy mechanisms that can be integrated with existing and upcoming MR systems, while their utilization and adoption are still not widespread.

REFERENCES

A. Acquisti, R. Gross, and F. Stutzman. 2011. Privacy in the age of augmented reality. Proc. National Academy of Sciences.

Alessandro Acquisti, Curtis R. Taylor, and Liad Wagman. 2016. The economics of privacy. Journal of Economic Literature 54, 2 (2016), 442–92.

Paarijat Aditya, Rijurekha Sen, Peter Druschel, Seong Joon Oh, Rodrigo Benenson, Mario Fritz, Bernt Schiele, Bobby Bhattacharjee, and Tong Tong Wu. 2016. I-pic: A platform for privacy-compliant image capture. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services. ACM, 235–248.

Surin Ahn, Maria Gorlatova, Parinaz Naghizadeh, Mung Chiang, and Prateek Mittal. 2018. Adaptive fog-based output security for augmented reality. In Proceedings of the Morning Workshop on Virtual Reality and Augmented Reality Network. ACM, 1–6.

Naveed Akhtar and Ajmal Mian. 2018. Threat of adversarial attacks on deep learning in computer vision: A survey. IEEE Access 6 (2018), 14410–14430.

Sarah J. Andrabj, Michael K. Reiter, and Cynthia Stanton. 2015. Usability of augmented reality for revealing secret messages to users but not their devices. In Proceedings of the Symposium on Usable Privacy and Security (SOUPS’15). 89–102.

Ilhan Aslan, Andreas Uhl, Alexander Meschtscherjakov, and Manfred Tscheligi. 2014. Mid-air authentication gestures: An exploration of authentication based on palm and finger motions. In Proceedings of the 16th International Conference on Multimodal Interaction. ACM, 311–318.

Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent advances in augmented reality. IEEE Comput. Graph. Appl. 21, 6 (2001), 34–47.

Ronald T. Azuma. 1997. A survey of augmented reality. Presence: Teleoper. Virtual Environ. 6, 4 (1997), 355–385.

Stefano Baldassi, Tadayoshi Kohno, Franziska Roesner, and Moqian Tian. 2018. Challenges and new directions in augmented reality, computer security, and neuroscience—part 1: Risks to sensation and perception. Arxiv Preprint Arxiv:1806.10557.

Steve Benford, Chris Brown, Gail Reynard, and Chris Greenhalgh. 1996. Shared spaces: Transportation, artificiality, and spatiality. In Proceedings of the ACM Conference on Computer Supported Cooperative Work. ACM, 77–86.

Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva. 1998. Understanding and constructing shared spaces with mixed-reality boundaries. ACM Trans. Comput.-Hum. Interact. 5, 3 (1998), 185–223.

Mark Billinghurst and Hirokazu Kato. 1999. Collaborative mixed reality. In Proceedings of the 1st International Symposium on Mixed Reality. 261–284.

K. Brkic, I. Sikiric, T. Hrakc, and Z. Kalafatic. 2017. I know that person: Generative full body and face de-identification of people in images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW’17). 1319–1328. DOI: https://doi.org/10.1109/CVPRW.2017.173

ACM Computing Surveys, Vol. 52, No. 6, Article 110. Publication date: October 2019.
Andreas Butz, Clifford Beshers, and Steven Feiner. 1998. Of vampire mirrors and privacy lamps: Privacy management in multi-user augmented environments. In Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology. ACM, 171–172.

Andreas Butz, Tobias Höllerer, Steven Feiner, Blair MacIntyre, and Clifford Beshers. 1999. Enveloping users and computers in a collaborative 3D augmented reality. In Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR’99). IEEE Computer Society, Washington, DC, 35.

Nicholas Carlini, Pratyush Mishra, Tavish Vaidya, Yuankai Zhang, Micah Sherr, Clay Shields, David Wagner, and Wenchao Zhou. 2016. Hidden voice commands. In Proceedings of the 25th USENIX Security Symposium (USENIXSecurity’16). 513–530.

Krzysztof Chalupka, Pietro Perona, and Frederick Eberhardt. 2014. Visual causal feature learning. Arxiv Preprint Arxiv:1412.2309.

Joy Jo-Yi Chang, Ming-Jheng Li, Yi-Chun Wang, and Justie Su-Tzu Juan. 2010. Two-image encryption by random grids. In Proceedings of the International Symposium on Communications and Information Technologies (ISCI'T10). IEEE, 458–463.

Dimitris Chatzopoulos, Carlos Bermejo, Zhanpeng Huang, and Pan Hui. 2017. Mobile augmented reality survey: From where we are to where we go. IEEE Access 5 (2017), 6917–6950.

Jagmohan Chauhan, Yining Hu, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. 2017. BreathPrint: Breathing acoustics-based user authentication. In Proceedings of the 15th International Conference on Mobile Systems, Applications, and Services. ACM, 278–291.

Adrian David Cheok, Xubo Yang, Zhou Zhi Ying, Mark Billinghurst, and Hirokazu Kato. 2002. Touch-space: Mixed reality game space based on ubiquitous, tangible, and social computing. Person. Ubiqu. Comput. 6, 5-6 (2002), 430–442.

Valentina Ciriani, Sabrina De Capitani Di Vimercati, Sara Foresti, Sushil Jajodia, Stefano Paraboschi, and Pierangela Samarati. 2010. Combining fragmentation and encryption to protect privacy in data storage. ACM Trans. Info. Syst. Secur. 13, 3 (2010), 22.

Andy Crabtree, Tom Lodge, James Colley, Chris Greenhalgh, Richard Mortier, and Hamed Haddadi. 2016. Enabling the new economic actor: Data protection, the digital economy, and the databox. Person. Ubiqu. Comput. 20, 6 (2016), 947–957.

Keith Curtin. 2017. Mixed Reality will be most important tech of 2017. Retrieved from https://thenextweb.com/insider/2017/01/07/mixed-reality-will-be-most-important-tech-of-2017.

Loris D’Antoni, Alan M. Dunn, Suman Jana, Tadayoshi Kohno, Benjamin Livshits, David Molnar, Alexander Moshchuk, Eyal Ofek, Franziska Roesner, T. Scott Saponas et al. 2013. Operating system support for augmented reality applications. In Proceedings of the Workshop on Hot Topics in Operating Systems (HotOS’13). Vol. 13. 21–21.

Yves-Alexandre de Montjoye, Erez Shmueli, Samuel S. Wang, and Alex Sandy Pentland. 2014. openpmd: Protecting the privacy of metadata through safeanswrs. PloS One 9, 7 (2014), e98790.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In Proceedings of the Conference on Computer Vision and Pattern Recognition. Ieee, 248–255.

Mina Deng, Kim Wuys, Riccardo Scandariato, Bart Preeneel, and Wouter Joosen. 2011. A privacy threat analysis framework: Supporting the elicitation and fulfillment of privacy requirements. Require. Eng. 16, 1 (2011), 3–32.

Anthony DeVincenzi, Lining Yao, Hiroshi Ishii, and Ramesh Raskar. 2011. Kinected conference: Augmenting video imaging with calibrated depth and audio. In Proceedings of the ACM Conference on Computer Supported Cooperative Work. ACM, 621–624.

Ben Dickinson. 2016. 5 authentication methods putting passwords to shame. Retrieved from https://thenextweb.com/insider/2016/03/31/5-technologies-will-flip-world-authentication-head.

Klaus Dorfmueller-Ulhaas and Dieter Schmalstieg. 2001. Finger tracking for interaction in augmented environments. In Proceedings of the IEEE and ACM International Symposium on Augmented Reality. IEEE, 55–64.

Liang Du, Meng Yi, Erik Blasch, and Haibin Ling. 2014. GARP-face: Balancing privacy protection and utility preservation in face de-identification. In Proceedings of the IEEE International Joint Conference on Biometrics (IJCB’14). IEEE, 1–8.

Cynthia Dwork, Aaron Roth et al. 2014. The algorithmic foundations of differential privacy. Found Trends Theor. Comput. Sci. 9, 3–4 (2014), 211–407.

Marc Eaddy, Gabor Blasko, Jason Babcock, and Steven Feiner. 2004. My own private kiosk: Privacy-preserving public displays. In Proceedings of the 8th International Symposium on Wearable Computers (ISWC’04). Vol. 1. IEEE, 132–135.

Barrett Ens, Tovi Grossman, Fraser Anderson, Justin Matejka, and George Fitzmaurice. 2015. Candid interaction: Revealing hidden mobile and wearable computing activities. In Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology. ACM, 467–476.

Ülfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. Rappor: Randomized aggregatable privacy-preserving ordinal response. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. ACM, 1054–1067.

Chengfang Fang and Ee-Chien Chang. 2010. Securing interactive sessions using mobile device through visual channel and visual inspection. In Proceedings of the 26th Annual Computer Security Applications Conference. ACM, 69–78.
Academic Computing Surveys, Vol. 52, No. 6, Article 110. Publication date: October 2019.
Anindya Maiti, Murtuza Jadliwala, and Chase Weber. 2017. Preventing shoulder surfing using randomized augmented reality.

Anindya Maiti, Oscar Armbruster, Murtuza Jadliwala, and Jibo He. 2016. Smartwatch-based keystroke inference attacks.

David G. Lowe. 2004. Distinctive image features from scale-invariant keypoints.

Pei-Yu Lin, Bin You, and Xiaoyong Lu. 2017. Video exhibition with adjustable augmented reality system based on temporal psycho-visual modulation. EURASIP J. Image Video Process. 2017, 1 (2017), 7.

Sugang Li, Ashwin Ashok, Yanyong Zhang, Chenren Xu, Janne Lindqvist, and Macro Gruteser. 2016a. Whose move is it anyway? authenticating smart wearable devices using unique head movement patterns.

Kiron Lebeck, Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. 2018. Towards security and privacy for multi-user augmented reality: Foundations with end users.

Ang Li, Qinghua Li, and Wei Gao. 2016b. PrivacyCamera: Cooperative privacy-aware photographing with mobile phones.

Christos Kalloniatis, Evangelia Kavaki, and Stefanos Gritzalis. 2008. Addressing privacy requirements in system design: The PriS method. Require. Engineer. 13, 3 (2008), 241–255.

Mohamed Khamis, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, and Andreas Bulling. 2016. Gazetouchpass: Multimodal authentication using gaze and touch on mobile devices. In Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM, 2156–2164.

Tadayoshi Kohno, Joel Kollin, David Molnar, and Franziska Roesner. 2016. Display Leakage and Transparent Wearable Displays: Investigation of Risk, Root Causes, and Defenses. Technical Report. Microsoft Research. https://www.microsoft.com/en-us/research/publication/display-leakage-and-transparent-wearable-displays-investigation-of-risk-root-causes-and-defenses.

Bernard Kress and Thad Starner. 2013. A review of head-mounted displays (HMD) technologies and applications for consumer electronics. In Proceedings of the International Society for Optics and Photonics (SPIE’13), Vol. 8720. 87200A.

Alexey Kurakin, Ian Goodfellow, and Samy Bengio. 2016. Adversarial examples in the physical world. Arxiv Preprint Arxiv:1607.02533.

Patrik Lantz, Bjorn Johansson, Martin Hell, and Ben Smeets. 2015. Visual cryptography and obfuscation: A use-case for decrypting and deobfuscating information using augmented reality. In Proceedings of the International Conference on Financial Cryptography and Data Security. Springer, 261–273.

Tadayoshi Kohno, Joel Kollin, and Franziska Roesner. 2017. Securing augmented reality output. In Proceedings of the IEEE Symposium on Security and Privacy (SP’17). IEEE, 320–337.

Kiron Lebeck, Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. 2018. Towards security and privacy for multi-user augmented reality: Foundations with end users. In Proceedings of the IEEE Symposium on Security and Privacy (SP’18). IEEE, 392–408.

Yann LeCun, Corinna Cortes, and C. J. Burges. 2010. MNIST handwritten digit database. AT&T Labs. Retrieved from http://yann.lecun.com/exdb/mnist.

Patrik Lantz, Bjorn Johansson, Martin Hell, and Ben Smeets. 2015. Visual cryptography and obfuscation: A use-case for decrypting and deobfuscating information using augmented reality. In Proceedings of the International Conference on Financial Cryptography and Data Security. Springer, 261–273.

Frank McSherry and Kunal Talwar. 2007. Mechanism design via differential privacy. In Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS’07). IEEE, 94–103.

Dongyu Meng and Hao Chen. 2017. Magnet: A two-pronged defense against adversarial examples. In Proceedings of the ACM SIGSAC Conference on Computer and Communications Security. ACM, 135–147.

Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. IEICE Trans. Info. Syst. 77, 12 (1994), 1321–1329.

Paul Milgram, Haruo Takemura, Akira Utsumi, Fumio Kishino, et al. 1994. Augmented reality: A class of displays on the reality-virtuality continuum. In Telemmanipulator and Telepresence Technologies, Vol. 2351. SPIE, 282–292.

Frashanth Mohan, Abhradeep Thakurta, Elaine Shi, Dawn Song, and David Culler. 2012. GUPT: Privacy preserving data analysis made easy. In Proceedings of the ACM SIGMOD International Conference on Management of Data. ACM, 349–360.

Meredith Ringel Morris, Anthony Cassanego, Andreas Paepcke, Terry Winograd, Ann Marie Piper, and Anqi Huang. 2006a. Mediating group dynamics through tabletop interface design. IEEE Comput. Graph. Appl. 26, 5 (2006), 65–73.

ACM Computing Surveys, Vol. 52, No. 6, Article 110. Publication date: October 2019.
Meredith Ringel Morris, Anqi Huang, Andreas Paepcke, and Terry Winograd. 2006b. Cooperative gestures: Multi-user gestural interactions for co-located groupware. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 1201–1210.

Alessandro Mulloni, Daniel Wagner, and Dieter Schmalstieg. 2008. Mobility and social interaction as core gameplay elements in multi-player augmented reality. In Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts. ACM, 472–478.

Min Mun, Shuai Hao, Nilesh Mishra, Katie Shilton, Jeff Burke, Deborah Estrin, Mark Hansen, and Ramesh Govindan. 2010. Personal data vaults: A locus of control for personal data streams. In Proceedings of the 6th International Conference on Emerging Networking Experiments and Technologies (CoNEXT’10). ACM, 17.

Elaine M. Newton, Latanya Sweeney, and Bradley Malin. 2005. Preserving privacy by de-identifying face images. IEEE Trans. Knowl. Data Engineer. 17, 2 (2005), 232–243.

Nicolas Papernot, Patrick McDaniel, Xi Wu, Somesh Jha, and Ananthram Swami. 2016. Distillation as a defense to adversarial perturbations against deep neural networks. In Proceedings of the IEEE Symposium on Security and Privacy (SP’16). IEEE, 582–597.

Jennifer Pearson, Simon Robinson, Matt Jones, Anirudha Joshi, Shashank Sahoo, and Srimath Subramanian. 2017. Chameleon devices: Investigating more secure and discreet mobile interactions via active camouflaging. In Proceedings of the CHI Conference on Human Factors in Computing Systems. ACM, 5184–5196.

Zhao Qin, Jingbo Yan, Kui Ren, Chang Wen Chen, and Cong Wang. 2014. Towards efficient privacy-preserving image feature extraction in cloud computing. In Proceedings of the 22nd ACM International Conference on Multimedia. ACM, 497–506.

Zhao Qin, Jingbo Yan, Kui Ren, Chang Wen Chen, and Cong Wang. 2016. SecSIFT: Secure image SIFT feature extraction in cloud computing. ACM Trans. Multimedia Comput. Commun. Appl. 12, 4s (2016), 65.

Moo-Ryong Ra, Ramesh Govindan, and Antonio Ortega. 2013. P3: Toward privacy-preserving photo sharing. In Proceedings of the USENIX Conference on Networked Systems Design and Implementation (NSDI’13). 515–528.

Ihsan Rabbi and Sehat Ullah. 2013. A survey on augmented reality challenges and tracking. Acta Graphica 24, 1–2 (2013), 29–46.

Aditi Raghunathan, Jacob Steinhardt, and Percy Liang. 2018. Certified defenses against adversarial examples. Arxiv Preprint Arxiv:1801.09344.

Kiran B. Raja, Ramachandra Raghavendra, Martin Stokkenes, and Christoph Busch. 2015. Multi-modal authentication system for smartphones using face, iris and periorcular. In Proceedings of the International Conference on Biometrics (ICB’15). IEEE, 143–150.

Nisarg Raval, Animesh Srivastava, Kiron Lebeck, Landon Cox, and Ashwin Machanavajjhala. 2014. Markit: Privacy markers for protecting visual secrets. In Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. ACM, 1289–1295.

Nisarg Raval, Animesh Srivastava, Ali Razeen, Kiron Lebeck, Ashwin Machanavajjhala, and Lanond P. Cox. 2016. What you mark is what apps see. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services. ACM, 249–261.

H. T. Regenbrecht, M. Wagner, and G. Baratoff. 2002. MagicMeeting: A collaborative tangible augmented reality system. Virtual Reality 6, 3 (Oct. 2002), 151–166.

Derek Reilly, Mohamad Salimian, Bonnie MacKay, Niels Mathiasen, W. Keith Edwards, and Juliano Franz. 2014. SecSpace: Prototyping usable privacy and security for mixed reality collaborative environments. In Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems. ACM, 273–282.

Jingjing Ren, Ashwin Rao, Martina Lindorfer, Arnaud Legout, and David Choffnes. 2016. Recon: Revealing and controlling pii leaks in mobile network traffic. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services. ACM, 361–374.

Franziska Roesner, Tamara Denning, Bryce Clayton Newell, Tadayoshi Kohno, and Ryan Calo. 2014a. Augmented reality: Hard problems of law and policy. In Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. ACM, 1283–1288.

Franziska Roesner, Tadayoshi Kohno, and David Molnar. 2014b. Security and privacy for augmented reality systems. Commun. ACM 57, 4 (Apr. 2014), 88–96. DOI: https://doi.org/10.1145/2580723.2580730

Franziska Roesner, David Molnar, Alexander Moschchuk, Tadayoshi Kohno, and Helen J. Wang. 2014c. World-driven access control for continuous sensing. In Proceedings of the ACM SIGSAC Conference on Computer and Communications Security. ACM, 1169–1181.

Cynthia E. Rogers, Alexander W. Witt, Alexander D. Solomon, and Krishna K. Venkatasubramanian. 2015. An approach for user identification for head-mounted displays. In Proceedings of the ACM International Symposium on Wearable Computers. ACM, 143–146.
Pierangela Samarati. 2001. Protecting respondents identities in microdata release. *IEEE Trans. Knowl. Data Engineer.* 13, 6 (2001), 1010–1027.

Pierangela Samarati and Latanya Sweeney. 1998. *Protecting Privacy when Disclosing Information: k-Anonymity and Its Enforcement Through Generalization and Suppression.* Technical Report. SRI International, 19.

Dieter Schmalstieg and Gerd Hesina. 2002. Distributed applications for collaborative augmented reality. In *Proceedings of the Conference on Virtual Reality.* IEEE, 59–66.

Stefan Schneegass, Youssef Oualil, and Andreas Bulling. 2016. SkullConduct: Biometric user identification on eyewear computers using bone conduction through the skull. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, 1379–1388.

Stacey D. Scott, M. Scheelagh T. Carpendale, and Kori M. Inkpen. 2004. Territoriality in collaborative tabletop workspaces. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work.* ACM, 294–303.

Yoones A. Sekhavat. 2017. Privacy preserving cloth try-on using mobile augmented reality. *IEEE Trans. Multimedia* 19, 5 (2017), 1041–1049.

Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K. Reiter. 2016. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security.* ACM, 1528–1540.

Prakash Shrestha and Nitesh Saxena. 2017. An offensive and defensive exposition of wearable computing. *ACM Comput. Surveys* 50, 6 (2017), 92.

Jiayu Shu, Rui Zheng, and Pan Hui. 2016. Cardea: Context-aware visual privacy protection from pervasive cameras. *Arxiv Preprint Arxiv:1610.00889*.

Mark Simkin, Dominique Schröder, Andreas Bulling, and Mario Fritz. 2014. Ubic: Bridging the gap between digital cryptography and the physical world. In *Proceedings of the European Symposium on Research in Computer Security.* Springer, 56–75.

Ivo Slugačević, Matej Serbec, Ante Derez, and Ivan Martinovic. 2017. HoloPair: Securing shared augmented reality using microsoft hololens. In *Proceedings of the 33rd Annual Computer Security Applications Conference (ACSAC’17).* ACM, New York, NY, 250–261. DOI: https://doi.org/10.1145/3134600.3134625

Zsolt Szalavári, Erik Eckstein, and Michael Gervautz. 1998. Collaborative gaming in augmented reality. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology.* ACM, 195–204.

Zsolt Szalavári and Michael Gervautz. 1997. The personal interaction panel—a two-handed interface for augmented reality. In *Proceedings of the Computer Graphics Forum*, Vol. 16. Wiley Online Library.

Piotr Szczukowski. 2014. Augmented reality for privacy-sensitive visual monitoring. In *Proceedings of the International Conference on Multimedia Communications, Services and Security.* Springer, 229–241.

Robert Templeman, Mohammed Korayem, David J. Crandall, and Apy Kapadia. 2014. PlaceAvoider: Steering first-person cameras away from sensitive spaces. In *Proceedings of the Network and Distributed System Security Symposium (NDSS’14).*

Bruce H. Thomas and Wayne Piekarski. 2002. Glove-based user interaction techniques for augmented reality in an outdoor environment. *Virtual Reality* 6, 3 (2002), 167–180.

Grace Woo, Andrew Lippman, and Ramesh Raskar. 2012. VRCodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR’12).* IEEE, 59–64.
Mike Wu and Ravin Balakrishnan. 2003. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology. ACM, 193–202.

Yifan Wu, Fan Yang, and Haibin Ling. 2018. Privacy-protective-gan for face de-identification. Arxiv Preprint Arxiv:1806.08906.

Chao Xu, Pathak, and Prasant Mohapatra. 2015. Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch. In Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications. ACM, 9–14.

Yan Xu, Maribeth Gandy, Sami Deen, Brian Schrank, Kim Spreen, Michael Gorbsky, Timothy White, Evan Barba, Iulian Radu, Jay Bolter et al. 2008. BragFish: Exploring physical and social interaction in co-located handheld augmented reality games. In Proceedings of the International Conference on Advances in Computer Entertainment Technology. ACM, 276–283.

Zhi Xu and Sencun Zhu. 2015. Semadroid: A privacy-aware sensor management framework for smartphones. In Proceedings of the 5th ACM Conference on Data and Application Security and Privacy. ACM, 61–72.

Andrew Chi-Chih Yao. 1986. How to generate and exchange secrets. In Proceedings of the 27th Annual Symposium on Foundations of Computer Science. IEEE, 162–167.

William S. Yerazunis and Marco Carbone. 2001. Privacy-enhanced displays by time-masking images. In Australian Conference on Computer-Human Interaction.

Xiaoyong Yuan, Pan He, Qile Zhu, and Xiaolin Li. 2019. Adversarial examples: Attacks and defenses for deep learning. IEEE Trans. Neural Netw. Learn. Syst. 30, 9 (2019), 2805–2824.

Valentina Zantedeschi, Maria-Irina Nicolae, and Ambrish Rawat. 2017. Efficient defenses against adversarial attacks. In Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security. ACM, 39–49.

Eisa Zarepour, Mohammadreza Hosseini, Salil S. Kanhere, and Arcot Sowmya. 2016. A context-based privacy preserving framework for wearable visual lifeloggers. In Proceedings of the IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops’16). IEEE, 1–4.

Lan Zhang, Taeho Jung, Puchun Feng, Xiang-Yang Li, and Yunhao Liu. 2014. Cloud-based privacy preserving image storage, sharing and search. Arxiv Preprint Arxiv:1410.6593.

Juncheng Zhao and Hock Soon Seah. 2016. Interaction in marker-less augmented reality based on hand detection using leap motion. In Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry. ACM, 147–150.

M. Tarek Ibn Ziad, Amr Alanwar, Moustafa Alzantot, and Mani Srivastava. 2016. Cryptoing: Privacy preserving processing over encrypted images. In Proceedings of the IEEE Conference on Communications and Network Security (CNS’16). IEEE, 570–575.

Guy Zyskind, Oz Nathan et al. 2015. Decentralizing privacy: Using blockchain to protect personal data. In Proceedings of the Security and Privacy Workshops (SPW’15). IEEE, 180–184.

Received July 2018; revised June 2019; accepted July 2019