Augmented Reality: Challenges and Opportunities for Security and Privacy

Franziska Roesner  
University of Washington

Tadayoshi Kohno  
University of Washington

David Molnar  
Microsoft Research

Abstract

Augmented reality (AR) technologies overlay computer-generated visual, audio, and haptic feedback onto an individual’s perception of the world. Early-generation AR technologies are shipping today on mobile phones, including seamless translation and information overlays. We anticipate significant advances in coming years, so it is crucial to consider security and privacy issues today. In this paper, we identify new challenges posed by AR technologies, including issues arising from always-on interfaces. We discuss existing risks that increase significantly in the AR context, such as the tension between privacy and cloud services. These, in turn, raise new research questions. We close with opportunities created by AR for improved security and privacy.

1 Introduction

Augmented reality (AR) technologies promise to enhance our perception of and interaction with the real world. Unlike virtual reality, in which the real world is replaced by a simulated one, augmented reality senses properties of the physical world and overlays real-world feedback with computer-generated visual, audio, and haptic signals.

Early-generation AR technologies are shipping on widely available mobile platforms. Examples include the Word Lens iPhone application — an application that overlays translated text on the camera’s view of foreign text — and Layar — a geolocation-based AR platform that allows developers to create augmented reality layers for the world (e.g., for game playing). The recent advent of 1 GHz processors, location sensors, and high resolution, autofocusing cameras in mobile phones has made these applications possible.

Beyond the mobile phone, devices are becoming available that enhance sensing, display, and data sharing. For example, Looxcie — an over-the-ear, always-on video camera — recently announced a feature enabling wearers to share their live video feed with anyone else in the world. Transparent, wearable displays are now available for research purposes from several companies. Figure 1 breaks out four categories of AR technologies: sensors, feedback devices, cloud services for storage and processing, and data sharing. Many of these are shipping today, while others are still experimental.

These technologies are at the cusp of significant innovation, so now is the time to consider security and privacy issues. We ask the following questions: (1) what new security and privacy research challenges arise with AR technologies? (2) what “classic” security and privacy goals and challenges become more serious and ubiquitous in the context of AR technologies? (3) what new opportunities do AR technologies create for improving our digital and physical security and privacy?

We find that AR technologies form an important, new, and fertile playground for computer security and privacy research. Of course, AR technologies should leverage standard security best practices, such as on-device and network encryption. Nevertheless, we find unique obstacles — such as overcoming security risks arising from a complex feedback loop between physical-world and digital-world inputs and outputs — that are simultane-

2http://www.looxcie.com
ously intellectually challenging yet surmountable. Other challenges, such as giving users control of their data, are well known in other arenas but become even more important for AR technologies with their always-on, always-sensing nature. Given the future importance of AR technologies, researchers already tackling these issues in other domains can find value in refocusing their attention on AR applications.

Finally, AR technologies can improve security and privacy, both digitally and physically. For example, we can provide personal digital views of content on personal displays. Imagine a password manager that superimposes highlights over the correct keys for a complex password when a user looks at a keyboard, or an application that alerts the user when someone is lying!

To the best of our knowledge, we are the first to consider all of these issues together in the context of augmented reality. We begin in Section 2 with a collection of near-future scenarios that highlight the capabilities of AR technologies, their usage scenarios, and the spectrum of security and privacy threats that might arise.

2 Scenarios and Attacks

To make our discussion concrete, we introduce three scenarios that use the technology represented in Figure 1. Each of these scenarios is possible to some extent today. The advancement of AR technology will deepen each scenario, as we discuss. We then describe attacks possible in each scenario.

Scenario 1: Translation. Alice has recently purchased an augmented reality system that includes a body-worn camera, an invisible Bluetooth earpiece, and a transparent heads-up display. Because she is soon traveling to a foreign country and does not speak the language, she visits the “App Store” associated with her augmented reality system and downloads a translation application. During her visit, the system provides Alice with visual translations of both spoken and written words via her heads-up display. When others speak to Alice, she hears audio translation through her Bluetooth earpiece. When the translator encounters a word or sentence that it cannot translate, it sends the fragment to a crowdsourcing system, where a human quickly translates the word and the result is returned to Alice’s application.

We see precursors of this scenario today in smartphone applications like the Word Lens translator for the iPhone, which instantaneously translates words viewed through the iPhone’s camera and displays them on the iPhone’s screen in place of the original words. In the context of this scenario, we consider several attacks:

Malicious Real-World Input Attack. On the second day of her trip, Alice sees an advertisement on a bus. Unbeknownst to Alice, the text in the advertisement is maliciously crafted to exploit a vulnerability in the optical character recognition program used by the translation application. When the translation application attempts to process the input, the vulnerability is exploited and the attacker takes control.

Deception Attack. With complete control over the system, the attacker uses a deception attack in which he changes the reality that Alice perceives. He pastes false street signs and translations over the real signs, leading Alice into a dangerous neighborhood. The attacker eliminates from Alice’s heads-up display and earpiece any indication of a mysterious man coming up from behind.

Feedback Overload Attack. The attacker uses his control over Alice’s feedback devices to blind her through her display and deafen her through her earpiece. Alice is incapacitated. The approaching man, a conspirator of the attacker, can now easily take Alice’s wallet and other belongings.

Extortion Attack. The attacker speaks to Alice through the earpiece and refuses to re-enable Alice’s AR system until she transfers an additional sum of money. She complies because she is so dependent on the translation application and the rest of the system that she would not be able to function without it.

Crowdsource Poisoning Attack. Alice’s arch-rival Mallory knows which translation service is used by Alice’s system. She arranges for people loyal to her to perform all of Alice’s translations. At the crucial moment, Mallory ensures that while Alice thinks she is asking how to buy tobacco, she in fact has declared, “My hovercraft is full of eels.” Alice is arrested for the insult.

Scenario 2: Local Directions. Bob wants to find a good sandwich shop at which to eat lunch. He wanders around the neighborhood, pausing at places that his heads-up display (or earpiece) tells him might match his tastes, as determined based on his previous experience with sandwiches. When he approaches each establishment, his augmented reality feedback devices give him information about the menu, ratings, and personalized recommendations for a place. Because none of the restaurants match his mood today, his system gives him a recommendation for a pizza place in another neighborhood. He consents and follows the visual, audio, and haptic (vibrations telling him to turn left or right) instructions provided by his system to take a bus to the selected location.

This scenario represents the AR version of services that exist today, such as Yelp for restaurant reviewing

---

4http://www.questvisual.com

5Reference to the “Dirty Hungarian Phrasebook” Monty Python sketch, first aired in 1970.
<table>
<thead>
<tr>
<th>Commercially Available Today</th>
<th>Experimentally Only</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensors</strong></td>
<td></td>
</tr>
<tr>
<td>Body-worn RGB cameras</td>
<td>Haptic sensors</td>
</tr>
<tr>
<td>GPS (error of 5 meters or more)</td>
<td></td>
</tr>
<tr>
<td><strong>Feedback</strong></td>
<td></td>
</tr>
<tr>
<td>Opaque near-eye display</td>
<td>Transparent near-eye display</td>
</tr>
<tr>
<td>Phone display/speaker</td>
<td>Embedded displays (e.g., contact lenses [13])</td>
</tr>
<tr>
<td>Invisible Bluetooth earpiece</td>
<td>Haptic feedback [9]</td>
</tr>
<tr>
<td><strong>Services</strong></td>
<td></td>
</tr>
<tr>
<td>Simple cloud services (e.g., photo gallery)</td>
<td>Complex cloud services (e.g., object recognition)</td>
</tr>
<tr>
<td>Marker-based tracking [17]</td>
<td>Markerless tracking</td>
</tr>
<tr>
<td>Good face detection (not recognition) [16]</td>
<td>Good face recognition</td>
</tr>
<tr>
<td>Expensive or cheap but inaccurate transcription</td>
<td>Cheap accurate transcription</td>
</tr>
<tr>
<td><strong>Sharing</strong></td>
<td></td>
</tr>
<tr>
<td>Selective sharing (photos, videos, location)</td>
<td>Automatic sharing [4]</td>
</tr>
</tbody>
</table>

Figure 1: Augmented Reality Technology. We categorize technologies based on their availability. Some technologies are commercially available “off the shelf” today. Others are under development and exist on an experimental basis.

and Google Maps for directions and navigation. In the AR context, this scenario raises a number of additional attack vectors:

**Ubiquitous Spam and Phishing.** As Bob wanders the neighborhood, his display is flooded with advertisements for other restaurants, stores, and products. At times there are so many advertisements flooding Bob’s display that he can’t distinguish advertisements from real signs for restaurants he passes. At one point, he almost enters Todd’s Tacos because a sign from Paul’s Pizza has been pasted over the sign in his display — paid for by Todd, who hopes that once people walk into his restaurant they will stay even upon discovering he doesn’t serve pizza. The prevalence and intrusiveness of these attacks is different in degree from what we see today.

**Physical Search Engine Optimization Attack.** Bob’s camera pans over Mallory’s Subway Sandwiches. Unbeknownst to Bob, Mallory’s has a small sign in the corner of its shop window that happens to look to Bob’s camera just like the sign in front of Alice’s Excellent Subs. Bob’s service is confused and tells Bob to go into Mallory’s instead of giving him directions to Alice’s Subs. This attack is related to existing concerns about location control, or geoslavery [5].

**Scenario 3: Augmented Memories.** Alice’s husband Charlie is going on a ski vacation and wishes to vividly remember his experiences. Because he does not have an augmented reality system himself, he borrows Alice’s for the weekend. He uses the body-worn camera and microphone to record the entire trip. Charlie does not need to worry about recording and can fully enjoy skiing: the wearable sensors are unobtrusive and invisible, and the vast amount of data that they collect are uploaded immediately to a cloud service. There the memories are cataloged and indexed for search, making it easy for Charlie to call up information about his trip later when he wishes to remember which runs he skied, with whom he had conversations and about what they spoke, etc. The system also lets Alice share in her husband’s adventures remotely by hooking into his video and audio feeds.

Today, the Looxcie wearable video camera, which constantly buffers thirty seconds of video to allow users to begin recording with a thirty second lag at any time, represents an early commercial version of such a scenario. The Looxcie camera will also soon allow users to grant others access to their live video and audio feeds. Relatedly, Bell and Gemmell have explored the challenges for lifetime data storage in the MyLifeBits project [2]. Here, we consider a number of data privacy issues in the context of this scenario:

**Data Ownership.** When Charlie’s camera captures a video of a fellow skier, Steve, who takes an embarrassing fall, Steve asks that Charlie delete the recording. Charlie refuses, arguing that he owns the recorded data and does not wish to introduce gaps in his recording of this trip. Steve argues that he should have a say in the data stored about him and points out that he never consented to being recorded.

**Proof of Harmlessness.** People listening nearby now realize that Charlie has been surreptitiously recording everything and also become angry. He tries to calm them by telling them that his recording system has a special filter to blur out the faces of strangers. This is true, but the bystanders don’t believe him and he has no way to demonstrate otherwise.

**Lending AR Devices.** When Charlie borrows Alice’s AR system, he logs in with his own credentials so that the recorded data is stored in his account with the cloud provider. He believes that Alice can only view his video and audio feeds when he explicitly authorizes her to hook into the feeds. However, Alice is suspicious and uses a program that recovers data resident in the system’s cache.

---

7http://www.looxcie.com
when the system is returned to her. She discovers a video of Charlie flirting with someone; Charlie and Alice don’t speak for days. This problem is heightened in the AR context due to the always-on nature of recording devices.

**Data Access Policies.** The vast amount of data that Charlie records includes sensitive information that he wishes to protect. For example, the system records video while he enters his debit card PIN and while he reads a controversial publication. The cloud provider that stores these data encrypts them, so Charlie isn’t worried. Unfortunately, when Alice divorces Charlie because of his ongoing affair, the cloud provider is subpoenaed for his data, which then become public record. Though it is unrelated to the divorce case, the information about his reading of the controversial publication prevents him from winning the bid for state senator the following year.

3 Challenges

We distill several challenges from our scenarios. With these challenges come specific research opportunities.

**Challenge 1: Real-World Interfaces.** Augmented reality systems interface directly with the real world, unlike traditional systems. Sensors take constant input from the user’s real-world surroundings, and feedback devices transmit sensory input directly to the user. These circumstances allow for a number of the attacks introduced in the translation scenario: real-world malicious input, feedback overload, and deception attacks. This challenge presents the security community with a number of concrete research opportunities:

*Aggregate Input Validation:* Sensors and services handling sensor data will need to be robust to real-world malicious input attacks. In the case of a maliciously crafted sign, traditional input validation techniques are likely to apply. However, a qualitatively new threat is sensor input that passes validation in isolation but thwarts the receiving system in aggregate. To take an example in the collaborative sensing space, a review site might leverage location tracking to measure a restaurant’s popularity by noting the average number of people present during the day. A canny restauranteur may then pay people to stand in the restaurant without buying anything. The restaurant’s measured popularity rises but has no relationship to its quality. AR devices that constantly collect data will drive the adoption of collaborative sensing applications such as this; thus, these security concerns will increase in importance. Trusted sensors [14]—while important to prevent other attacks—do not help in these cases, as real-world conditions are manipulated.

*Output validation:* Feedback devices must be designed with feedback overload threats in mind. For example, consider a telepresence system that allows Carol to “stand in” for a remote Alice: when Alice moves her facial muscles, Carol feels a stimulus that assists her in moving her face to match Alice’s expression. When Alice moves her arm, Carol feels a tug on her arm. Without proper protection, an attacker with control of the system could use it to physically hurt Carol. Such systems must distinguish malicious from benign but intense outputs.

“Get me out of here”: Deception attacks rely on the fundamental integration of AR feedback into a user’s perception of reality. To address both this and attacks on feedback devices, users must be able to reliably turn off the AR system (input and output) and verify that it has been turned off. In the near term, removing the system is a simple way to achieve this. Another approach here is that every AR system should have a secure attention sequence, analogous to Ctrl-Alt-Del on Windows computers. Still another approach is to reserve a trusted region of the display that always shows the real world. Determining the best such sequence, or the right input mode (e.g., gestures or speech), requires research for each AR system.

**Challenge 2: Tensions between Privacy and Functionality.** Because augmented reality needs to process information about the physical world, the technology stresses the classic tension between privacy and functionality. To support rich AR applications, sensor and other input data will fundamentally need to be stored and processed remotely. The spread of AR technology and its applications will drive users to create, store, and share a larger amount of more sensitive data than ever before. The scenarios that we introduced in the previous section raised issues like data ownership, consent to record, and trust in cloud service providers; here we outline several concrete research directions to address these issues.

*Bystander Privacy Protection:* Bystanders should be able to opt out of or be anonymized in the recordings of others; prior work has examined such issues [7, 15]. In Section 2, Charlie was unable to prove to bystanders that his system blurred their faces. Without a way to prove that such safeguards are in place, it does not matter that they exist. To further complicate the issue, users must first be made aware that they are being recorded; ensuring this is another research challenge.

The CVDazzle project pursues a different approach: using makeup to confuse face detection algorithms. This provides privacy without needing compliant cameras. The key limitation is that CVDazzle is painstakingly hand-tuned for one particular face detection algorithm. Is it possible to find a general algorithm for synthesizing makeup that fools face detection?

**Crowdsourcing Privacy Protection:** Hard algorithmic problems that may be solved by computers in the future are today sometimes solved by crowdsourcing to hu-
mans, as in the translation scenario or for object recognition (as in [3]). For some applications (e.g., remote instruction by an expert), humans may always be needed. Both the user and the crowdsourced worker must be protected: the user’s situation and data may be sensitive, and the worker should not be subjected to offensive or disturbing content.

A natural approach to protect privacy is to break tasks into pieces that are small enough that little or no private data is leaked to an individual worker. We examined a random sample out of 576 transcription tasks on Amazon’s Mechanical Turk crowdsourcing service and found that none follow this practice. Each task we examined reveals an entire video or conversation. Some tasks, however, did ask workers to sign an NDA before sending the complete audio file for transcription.

Others are actively working on how to best break large amounts of data into smaller pieces and return useful results. We are not aware, however, of work on the security of these mechanisms against an adversary that can create and coordinate multiple worker accounts. A basic question: given a method for recruiting workers, how many worker accounts need an attacker control before gaining the ability to reconstruct a large fraction of the original private data?

Cloud Data Privacy: Security and privacy concerns with respect to cloud service providers are not new to the AR space. Similarly, concern about the secrecy of data resident on a user’s devices (e.g., after device loss or confiscation) is not new. The difference in degree of concern, however, gives new context, which we hope will inform researchers tackling these problems today. AR technology and applications will drive users to produce, store, process, and share orders of magnitude more content, of greater sensitivity, than they do today. Rather than recording data selectively, users will be constantly recording everything in their experiences, causing a qualitative shift in the importance of these issues.

Challenge 3: Understanding and Control. As augmented reality technology leads users to record, store, and share vast amounts of sensitive data, the mental models that the systems ask users to grasp will become increasingly complex. We outline a set of concrete research questions for usable security and system design:

Mental Models of Access Settings: Today, users already have difficulty forming mental models of their privacy settings on services like Facebook because of the complexity of relationships between people and data items [10]. With the spread of AR systems, it will become infeasible for users to manually manage the access settings for every person and every data item that they have ever encountered. Researchers will need to develop methods and interfaces that allow users to intuitively grasp and manage these settings.

Application Permissions: In the translation scenario, we alluded to the existence of an App Store for AR systems. Users will need to grant applications permissions, as they do today via manifests (e.g., on Android and Windows phones) or prompts (e.g., on the iPhone). In the AR context, it is unclear whether these are the right models and what permissions should be available to applications. Should applications that modify the display have access to the entire display?

While this model may be acceptable for today’s primitive phone-based AR systems, it seems too permissive for more aggressive AR systems with heads-up or even embedded displays; a better model may be to grant an application restricted permissions to a portion of the display when a particular event occurs or in a particular context (e.g., at home). The problem becomes more pressing when multiple applications are running simultaneously: how do users link a specific augmentation to an application? What happens when five applications all want to subtitle the same object? Should we restrict applications from viewing the raw camera feed and instead only expose that certain objects are present? As difficult as the permission problem is today [6], AR will make it still more critical, as applications will not only read data but modify users’ perception of reality.

Cross-cutting Challenge: Crossing the Chasm. If we consider the lifetime of technologies, then we must realize that their usage scenarios will change over time. As a technology “crosses the chasm” from early adopters to mainstream users, these shifts will cause new security and privacy issues to arise. These issues are difficult to face because they do not surface during the initial deployment among highly technical, mostly trusting users. Here we highlight three examples of “crossing the chasm” issues that might arise if not considered sufficiently in advance.

Legitimate Advertising: AR technologies will create rich opportunities for advertising. Researchers must find ways to enable this advertising while preventing spam, deception, and physical search engine manipulation. One possibility we envision is the notion of a “verified place,” in which a service like Google or Bing Maps provides physically unforgeable objects [12] for establishments to display, which can then be detected by AR systems. This would prevent attacks like the one mounted by Todd’s Tacos in the second scenario of Section 2.

Loss and Lending: Before AR systems are ubiquitous,

---

9http://www.crowdflower.com


11Consider the transition of the Internet from an academic network to a commercial network and the resulting security and privacy issues.
users may borrow the AR devices of others (as in the third scenario in Section 2). However, AR designers may not consider device loss and lending until there are more adopters. Researchers must design appropriate models and semantics that protect data in the case of device loss while allowing for device lending under appropriate restrictions. AR systems will be more integrated with our reality than systems today, potentially complicating existing lending semantics.

Interfaces and Frameworks: Security vulnerabilities often arise at the interfaces between components. Widespread adoption of AR technologies will create explicit or implicit frameworks for communication among AR devices on one person and between people (e.g., Alice’s camera communicates wirelessly to Bob’s display). Security researchers must anticipate these developments and ensure that the interfaces between devices and services are designed for security. What is the best way to design “an operating system for AR”?

4 Opportunities

We introduce concrete ideas to use AR for improving security and privacy. Our list is undoubtedly incomplete; we hope to see rich future work in the area.

Leveraging Personal Views. Personal displays present a strong defense against shoulder surfing, as users may interact with applications visible in their own view only. For example, someone using a laptop on an airplane today exposes everything they view and type to their seat neighbors; a personal heads-up display combined with a haptic sensor for discreet input would allow for greatly improved privacy.

Personal displays further enable encrypted content in the real world that can only be decrypted by the AR systems of the intended recipients. For example, a company can post encrypted notices on a bulletin board that employees can read through their company-issued AR systems, but that visitors to the company’s building cannot. Precursors of such a system are possible today using 2D barcodes that encode URLs to data with appropriate access control.

Other methods of granting access to specific individuals can also leverage the personal view. For example, if Alice wishes to grant Bob access to her home while she is away, today she must give the physical key to Bob ahead of time or leave it in a publicly accessible place (e.g., under the doormat), relying on the fact that only Bob knows that and where she hid the key. With an AR system (and an electronic door lock), Alice could leave a virtual “key” at any time, which Bob (but no one else) can retrieve when his AR system detects that he is in the right place at the right time (i.e., an AR version of [1]).

Leveraging Information Overlays. AR systems can overlay information on a user’s view of the world, presenting alerts to improve privacy, security, and safety.

For example, an AR system could alert users when it detects camera lenses pointed at them, using (for instance) computer vision to detect the glint of light reflecting off a lens. Alternatively, legislation or market forces could lead to cameras that announce their status (as in [11]), which in turn could inform the AR system. Of course, not all cameras would be compliant, but if many are, this would significantly raise the bar. More broadly, AR systems could be used to detect eavesdropping of any kind (e.g., a laser pointed at a window).

Such systems could also detect deception attempts, or provide warnings of possible physical harm, noting that a person is lying, that someone is hiding in the bushes, or that an ATM is outfitted with a skimmer. One of our colleagues refers to this application as “spidey sense.”

Easier Authentication. An AR system can present password hints to users. For example, an AR system’s heads-up display could outline the appropriate characters that the user must enter on legacy devices like ATM PIN pads. Users could then be assigned strong passwords, as they would never need to actually remember the password. This application requires markerless tracking and a system design that properly protects the stored passwords.

Beyond storing passwords, AR systems can be used for implicit authentication of their users. The plethora of sensors attached to people can be used to authenticate them with biometric and behavioral characteristics. Prior work has examined the possibility of such mechanisms on mobile phones [8]; AR systems would provide far more powerful authentication.

Beyond the sensors attached to an individual (e.g., Alice), the sensors of bystanders could also be used to authenticate her by providing the authentication system with third-party visual, audio, and other sensory views of Alice. This third-party authentication system would distribute trust to systems and persons with no incentive to falsely authenticate Alice.

5 Conclusion

We identified four major challenges—real-world interfaces, tensions between privacy and functionality, user understanding of AR, and crossing the chasm—and the corresponding opportunities for security and privacy research. As the technology is deployed, it will be difficult to retroactively add security and privacy protections. The time is now to consider security and privacy as important values in the design of nascent AR technologies.

References


