frontiers in ICT

ORIGINAL RESEARCH
published: 09 April 2019
doi: 10.3389/fict.2019.00005

Cyber Security Threats and Challenges in Collaborative Mixed-Reality

Jassim Happa[1]*, *Mashhuda Glencross*[2] *and Anthony Steed*[3]

[1] *Department of Computer Science, University of Oxford, Oxford, United Kingdom,* [2] *Switch That Technologies Ltd., Oxford, United Kingdom,* [3] *Department for Computer Science, University College London, London, United Kingdom*

Collaborative Mixed-Reality (CMR) applications are gaining interest in a wide range of areas including games, social interaction, design and health-care. To date, the vast majority of published work has focused on display technology advancements, software, collaboration architectures and applications. However, the potential security concerns that affect collaborative platforms have received limited research attention. In this position paper, we investigate the challenges posed by cyber-security threats to CMR systems. We focus on how typical network architectures facilitating CMR and how their vulnerabilities can be exploited by attackers, and discuss the degree of potential social, monetary impacts, psychological and other harms that may result from such exploits. The main purpose of this paper is to provoke a discussion on CMR security concerns. We highlight insights from a cyber-security threat modelling perspective and also propose potential directions for research and development toward better mitigation strategies. We present a simple, systematic approach to understanding a CMR attack surface through an abstraction-based reasoning framework to identify potential attack vectors. Using this framework, security analysts, engineers, designers and users alike (stakeholders) can identify potential Indicators of Exposures (IoE) and Indicators of Compromise (IoC). Our framework allows stakeholders to reduce their CMR attack surface as well understand how Intrusion Detection System (IDS) approaches can be adopted for CMR systems. To demonstrate the validity to our framework, we illustrate several CMR attack surfaces through a set of use-cases. Finally, we also present a discussion on future directions this line of research should take.

Keywords: collaborative mixed reality, cyber security, attack modelling, attack surface, harm

OPEN ACCESS

Edited by:
Doug A. Bowman,
Virginia Tech, United States

Reviewed by:
Gang Wang,
Virginia Tech, United States
Blair MacIntyre,
Georgia Institute of Technology, United States

***Correspondence:**
Jassim Happa
jassim.happa@cs.ox.ac.uk

Specialty section:
This article was submitted to Virtual Environments, a section of the journal Frontiers in ICT

Received: *31 October 2018*
Accepted: *07 March 2019*
Published: *09 April 2019*

Citation:
Happa J, Glencross M and Steed A (2019) Cyber Security Threats and Challenges in Collaborative Mixed-Reality. Front. ICT 6:5. doi: 10.3389/fict.2019.00005

1. INTRODUCTION

Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) are receiving significant attention and the scale of deployment of the full spectrum of Extended Reality (XR) experiences is considerable. XR refers to all combinations of real-and-virtual environments and includes elements of Human-Computer Interaction (HCI). It is used as an umbrella term for VR, AR, MR and broader experiences incorporating sensors/wearables. Recent reports predict the global

market for XR hardware, software, and services will reach US $200 billion by 2021 (https://www.businesswire.com/news/home/20180227005719/en/Global-Augmented-Reality-Mixed-Reality-Market-Outlook and https://www.statista.com/statistics/591181/global-augmented-virtual-reality-market-size/). This rapid acceptance of XR technologies into a broad range of mainstream applications will result in applications being targeted by criminals. A benign, but creative, example entitled "*hello, we're from the internet*," demonstrates the ease by which application security can be breached. In this case, a group of artists curated an unauthorised virtual gallery consisting of work from eight artists drawn over notable digitised paintings displayed in an AR environment at the Museum of Modern Art (MoMA) (see https://next.reality.news/news/indie-artists-invade-moma-with-augmented-reality-reimagine-jackson-pollocks-works-0183271/).

Collaborative Mixed Reality (CMR) as described by Billinghurst and Kato (1999), extends Milgram and Kishino (1994)'s definition of MR to collaborative interfaces. The broad range of input/output (I/O) sensor technologies available today, means that CMRs can involve live tracking and/or sensor data incorporated into the Virtual Environment (VE) and represented (in some way) to the user. Consequently, users might not have a common set of interfaces to access a single shared VE. As such, the shared VE represents some combination of the real space of many users, combined with a synthetic model. Users can potentially collaborate using many different forms of sensors, data and interfaces. Recent advances suggest that within a few years time, hardware and software will render such CMR platforms more straightforward for mainstream use. For instance, Facebook recently released a social VR demo at Oculus Connect 2016 (https://www.youtube.com/watch?v=YuIgyKLPt3s), and Singh (see https://medium.com/@karansher/towards-stronger-human-connections-in-ar-vr-xr-53544bf6b10d) discusses how AR and VR might blur the distinction between our digital and physical reality. CMR research and development has also seen a renaissance with key actors such as *Oculus Rift* (https://www.oculus.com/) and *HTC Vive* (https://www.vive.com/uk/) in the VR space and highly popular AR games such as Pokemon GO (https://www.pokemongo.com/en-gb/) and Zombies Run! (https://zombiesrungame.com/).

In parallel to advances in CMR, there has been a substantial rise in research in cyber security research. This has focused on the Internet of Things (IoT), online games and Web applications. Most of this examines how new capabilities will affect *attack surfaces* (see Manadhata and Wing, 2010), as well as identify mechanisms to combat threats posed to these systems. Zhang et al. (2014) for instance outline challenges and research opportunities, Whitmore et al. (2015) provide a survey of topics and trends, Miessler (2015) discuss practitioner approaches to mapping out attack surfaces of IoT devices, and Happa et al. (2018) propose an ethics framework for research, development and deployment of heterogeneous systems that may not yet be fully defined or understood. Security research in the IoT space has progressed further than that in the CMR space, and significant overlaps exist (particularly on the attack surface consequences of interactions between heterogeneous sensors), thus we see it necessary to also review this literature.

In online virtual environments, the overall evolution of the world/model is both complex and distributed. There are many incentives and mechanisms to cheat (see Jeff Yan and Choi, 2002; Yan and Randell, 2005). These include; misplacing trust, collusion, abusing game procedure, manipulating virtual assets, exploiting machine intelligence, modifying client infrastructure, denying service to peer players, timing cheating, compromising authentication, exploiting bugs, compromising game servers, insider threats and social engineering. All these exploit the asymmetric availability of information within the game world (which takes this beyond normal problems of distributed applications, where the main attacks target the wire protocols, or central databases). Attacker objectives can vary and include bullying, hijacking online players' identities, obtaining virtual commodities, obtaining real money, achieving reputation damage and achieving political consequences. Bono et al. (2009) present an in-depth discussion on the importance of reducing the attack surface of Massively Multi-player Online Role-Playing Games (MMORPGs), given that a vulnerable game can leave many game instances compromised.

A wide range of collaboration platforms exist that use the Web or the Internet. These can be telepresence related, see Steuer (1992) and Szigeti et al. (2009), or Web platforms such as "its learnings" (https://itslearning.com/uk/) and "blackboard" (https://www.blackboard.com). Collaboration platforms are not limited to meetings or learning, and often aid users to achieve a common goal. Examples exist in the Intrusion Detection System (IDS) community, which has examined Collaborative Intrusion Detection Systems (CIDSs) for sharing Cyber Threat Intelligence (CTI) across different platforms (see Wagner et al., 2016; Giubilo et al., 2017; Happa, 2017; Nair et al., 2018). While these are driven by automation in sending event logs, the importance of making good use of the combination of automation and human analysts cognition cannot be understated. Together, these allow for enhanced decision making against detected cyber attacks. Collaborative platforms that allow for multi-user engagement through shared multi-modal experiences (see Axon et al., 2018) may enhance cyber security decision making through shared virtual environments, but these systems also need protecting.

1.1. Paper Contributions

A key motivation of this position paper, is to deliver a framework that assists stakeholders (e.g., security analysts, engineers, designers and users) to systematically identify and consider unknown, or poorly understood threats posed by attackers. This paper aims to formulate a starting point for understanding the security challenges of MR and CMR through provoking a discussion on CMR-security concerns, with the aim of helping stakeholders develop robust security solutions. We highlight insights from cyber-security threat modelling and propose key detection, mitigation, combat and deterrence strategies. Security concerns relate to how stakeholders can better understand the MR and CMR attack surface. We propose a general framework from which stakeholders can identify potential Indicators of Exposures (IoE) and Indicators of Compromise (IoC). Defense

strategies relate to how stakeholders can reduce the attack surface of applications, as well as how Intrusion Detection System-like (IDS) and Intrusion Prevention System-like (IPS) approaches (see Liao et al., 2013) might be both adapted and deployed for CMR environments.

The contributions of this paper are:

- **An outline of grand challenges in security for CMR applications**. This serves as a roadmap, highlighting key research, development and deployment challenges (in security) across various stakeholders.
- **An in-depth investigation of attacks on CMR systems**. Our investigation is informed by research in MR, as currently there is little investigation in CMR. We include identifying socio-technical properties of attacks (e.g., psychology, finance, reputation) as well as more traditional technology-centric perspectives (e.g., malware). We present this in the form of a framework, built from a handful of models and a novel and extendable CMR attack taxonomy. This taxonomy can be used to identify and quantify attack behaviors to predict subsequent harms and impacts.
- **An analysis of security use cases of CMR applications**. The aim is to illustrate why these grand challenges need to be addressed and develop our starting taxonomy. We specifically consider use cases of Intellectual Property theft, Virtual try on, a Virtual doctor and Gaming.
- **An outline of how existing and new defense approaches can defend against CMR attacks** through detection, mitigation, combat and deterrence.

We begin by discussing related work and detailing key challenge areas relevant to CMR applications. Then we present the security related properties of CMRs together with our threat modelling assumptions, to motivate our choice of framework models. We then detail our framework showing how each of our models work together to help stakeholders make more well-informed decisions about security issues, attacks and attack surfaces. We present our taxonomy and show how this is used to reason about the attack surface of a set of CMR use cases. Finally, we discuss specific defense mechanisms and how these may work within the context of CMR applications and finish with future directions of research and concluding remarks.

1.2. Related Work

Traditionally, when discussing security of systems, CIA (Confidentiality, Integrity, Availability) are considered the properties of systems that we wish to protect Cherdantseva and Hilton (2013). Confidentiality relates to data or systems only being disclosed to appropriate parties. Integrity relates to data or systems only being modifiable by appropriate parties. Availability relates to data and systems being accessible. Other properties exist that fall within the CIA triad, such as: Non-repudiation, Authorisation, Authentication, Anonymity and Unobservability among others. In CMR security, CIA can relate to, e.g., how compromising an application might result in leaking a user's personal data onto the CMR platform through its graphical interface, or some actions or behaviour by a CMR user which is signalled as being private, but is broadcast to a wide audience (confidentiality). These are two examples that could cause financial harm or reputational harm. Furthermore, words and actions by a user can be scrambled (integrity), or parts of the CMR application can been rendered unavailable (availability). For important meetings this may have business consequences. Security of social aspects and harms are still poorly understood, as discussed by Agrafiotis et al. (2016) and worthy of detailed consideration.

CMR security literature is very limited. There has been some prior work on security and privacy challenges and approaches in MR such as the work by De Guzman et al. (2018). The authors survey the landscape of the published literature in the field and report that <2% of AR/MR literature investigates topics of security and privacy. In contrast to De Guzman et al. (2018) who provide a data-centric categorisation-approach based on a historical understanding of input/data access/output protection strategies, particularly focusing on target technical security and privacy properties, we consider a much wider gamut for the attack surface. Our approach includes non-technical aspects of the attack surface (psychological, financial, reputational harms) and we provide an attack modelling tactic and guidelines to consider present day as well as likely future types of attacks in the CMR landscape. This wider look at the threat landscape enables us to identify socio-technical harms of attacks. By socio-technical harms, we mean psychological, emotional, financial, reputation or other non-technical damage to users, in addition to the technical harms typically discussed in cyber security works. We believe a socio-technical approach to attack modelling in the CMR space will be required to comprehensively prepare for future types of attacks.

Early research into MR focused on categorising the technologies themselves. Milgram and Kishino (1994) presented a taxonomy of MR displays on a reality-virtuality contiunuum. Benford et al. (1996) described the idea of a shared space and CMR, breaking down concepts such as transportation, arteficiality and spatiality. Many early works, as highlighted by De Guzman et al. (2018) focus on challenges related to performance, alignment, interfacing, visualisation, mobility of MR.

Ethics considerations and value-sensitive design, have also been stressed to push the boundaries of research into topics such as data ownership, privacy, secrecy and integrity protection of AR systems, as discussed by Roesner et al. (2014a). They define a roadmap for protecting computer security and privacy of AR systems by considering new security and privacy challenges, such as the complex set of always-on input sensors. These include cameras, GPS, microphones, platforms that can run multiple applications simultaneously, including interfacing with other AR systems and output devices such as displays, and earphones.

Billinghurst and Kato (1999) review MR techniques for developing Computer Supported Collaborative Work (CSCW) interfaces and describe lessons learned from developing a variety of collaborative MR interfaces. The authors develop several examples, demonstrating how user experiences with their interfaces facilitate collaboration in a natural manner, enabling people to use normal gestures and non-verbal behaviour in face-to-face and remote collaboration.

Stevens (2006) discusses CopyBots in Second Life. Copybots enable attackers to copy and export (by re-uploading) locally stored content, including animations, clothing, gestures, meshes and sounds. Being able to copy virtual items, through the uses of scripts, poses challenges with regards to Intellectual Property and end-user rights.

Lofgren and Fefferman (2007) reviews one of the first VE epidemics, in World of Warcraft (https://worldofwarcraft.com/). A release from Sept 2005, granted access to an area known as "Zul'Gurub." In this area, higher-level players avatars were affected by a virtual disease known as "corrupted blood." This infection was an inconvenience, designed to make combat in this area more challenging and slowly drain life from players. Once the high-level characters returned to the main areas of the game, their pets could still be afflicted by the disease. Due to a bug, the disease was able to unintentionally leave the high-level area, and quickly kill lower-level characters in other parts of the game. Balicer (2007) studied how we can model infectious diseases through online role-playing games, and "Corrupted blood" is mentioned as an example of how a potential attack might affect a shared VE.

Mennecke et al. (2007) presents a roadmap for research on virtual worlds and highlights some security and trust concerns. The security concerns focus on e-commerce, demonstrating a need for clarity and transparency on when actions are secure, and privacy is assured.

Oliveira et al. (2009) discuss interoperability between different solutions, despite several solutions sharing great functional overlap. The paper presents a broad domain analysis on shared virtual environments, providing readers with a sense of how the parts are connected in a greater whole. The authors also propose solutions to resolve the fragmentation in this space, which they label the *Analysis Domain Model*. This model loosely inspires our view on the attack surface.

D'Antoni et al. (2013) argues that AR applications and non-mouse/keyboard user inputs need to be natively supported by operating systems. The authors propose that sensitive data may be mixed in with user inputs, and discuss research directions to solve these privacy concerns by having multiple applications share one instance of the augmented reality. From a resource perspective, we argue this is a reasonable approach in that it limits a user's digital footprint and unifies inputs, processing and outputs of any AR system. It is unclear whether such a system could lead to a single point of failure, in which the adversary is only required to have elevated privileges to manipulate the entirety of the shared AR space, unless additional access controls are implemented to support this paradigm (such as sandboxing).

Jana et al. (2013) expands on D'Antoni et al.'s work, proposing a new operating system abstraction: the recogniser. The recogniser takes raw sensor data and provides higher-level objects such as skeletons or faces to applications. The authors propose a permissions-based system at different levels of granularity, and demonstrate this on Windows with 87 AR applications and four recognisers. They also demonstrate a visualisation to show users how data (related to privacy) is exposed, and survey 462 participants to establish a privacy ordering over recognisers.

Vilk et al. (2015) discuss privacy concerns from the raw data from novel input devices, by examining how they may emerge on the Web. They demonstrate a proof of concept that projects Web page content across multiple surfaces in a room, and runs across multiple mobile platforms taking natural user inputs. The browser is designed with privacy in mind, taking the principle of least privilege into account when projecting and managing content displayed. The authors conduct a user survey to demonstrate the revealed information was acceptable.

Madary and Metzinger (2016) discuss the potential for deep behaviour manipulation, in the context of VEs that can be quickly modified with the goal of influencing behaviour. This point is quite critical to consider, as the risk of harm to users of MR applications can be magnified through hackers manipulating users. With these factors of plasticity and potential for deep behaviour manipulation being a specific concern in rich MR environments. The harm possible with cyber threats can be significantly greater than the monetary costs alone from a data security breach.

Lebeck et al. (2017, 2018) investigate how to secure AR output by tackling risks associated with malicious or buggy AR output. These risks can compromise safety of users, if an application accidentally or intentionally obscures output of other AR applications. The authors demonstrate a novel platform (Arya) to tackle this issue. Arya is an AR platform that controls application output according to a set of policies specified in their framework (see Roesner et al., 2014b). The authors add support for recognisers (from Jana et al., 2013) to detect objects. Arya exists between applications and the output drivers.

Sluganovic et al. (2017) discuss how existing methods to secure device pairing are not directly applicable to AR headsets as they: (1) assume single-user controls on two devices (when AR systems involve two users, each with their own device, users should not be required to take them off), and (2) they do not assume that an adversary can fully eavesdrop the out-of-band communication. The authors demonstrate the design and evaluation of HoloPair: a system for secure and usable pairing of two AR headset. This pairing protocol, along with its implementation, is evaluated with 22 participants. The study demonstrates a need for multi-device pairing protocols and shows how security for such devices is sometimes assumed to be sufficient for AR platforms. We suspect security protocols for CMR systems, such as the one in this paper will play a significantly larger role in the future.

Han et al. (2018) discusses how heterogeneous devices will be required to have a shared reference point, and for certain environments, different modalities will make this challenging to achieve. The authors propose a context-based pairing mechanism that uses time as the common factor across differing sensor types. Their system creates event fingerprints, matched across different IoT devices, for autonomous pairing purposes.

Alrawi et al. (2019) proposes a methodology to analyse security properties of IoT devices. The authors systematise the literature for smart home IoT devices to summarise key attack techniques, mitigation and stakeholders. They demonstrate these for 45 devices to identify neglected areas of research. While not directly relevant to this body of work, IoT devices are

likely to share many capabilities that can be attacked in a similar way.

Other ethics considerations relating to heterogeneous systems have been considered in IoT research (see van den Hoven, 2012; Wachtel, 2012; Baldini et al., 2018). We believe many of these ethical concerns will affect CMR applications as well:

- **Incentives for businesses and other threat actors** so they do not manipulate users or users' data, see for instance the Cambridge Analytica case (see Persily, 2017) (Note that throughout this paper we use the term *users* to indicate potential victims of attacks. While it is more precise to use the term *data subjects* (EU General Data Protection Regulation (GDPR) terminology), we employ the term users to avoid confusion for non-privacy/security readers).
- **Increasing user awareness of data processing**. A user needs to see the complete set of data as well as understanding what that data can be used for, to make a rational choice about how their data can be treated and the level of security they need.
- **Psychological biases**, Users have different perceptions of risk, and risk appetites. Reducing psychological biases will be vital to increase robustness and resilience against attacks.
- **Accountability of applications regarding users privacy**. This is less of an issue for EU citizens with the introduction of GDPR.
- **Miniaturisation and invisibility**. It is important to keep technology visible for inspection, audit, quality control and accountability. If users are unaware of their existence, they cannot know how their attack surface is increased.
- **Ambiguity and ontology**. It is necessary to deal with unclear criteria of identity and system boundaries, as the distinctions between objects artefacts and human beings blur together.
- **Identification**. More and more seemingly insignificant devices are being assigned identities. What should the rules be for assigning, administering and managing these identities? This concern, combined with the miniaturisation of devices, means that systems can perform background processing related to identity, without users being aware of how they are being monitored.
- **Mediation and autonomous agency**. Heterogeneous devices provide ways of extending and augmenting human agency. We need to consider how to ensure appropriate automated data-processing and decision-making.
- **Embedded "smart" capabilities**. Smart objects may embed intelligence/knowledge function as devices become external extensions to the human body. In the context of CMR, we envisage that devices with remote capabilities can predict what we recognise, choices we make and patterns of behaviour, facilitated by XR capabilities. For instance, AR can produce names of people on screen on your behalf linking them to other databases.
- **Seamless transfers invoking unpredictability and uncertainty**. Information flow may become invisible to end-users, which may raise concerns about whether people understand how their data is being handled.

A recent overlap between VR and security focuses on how to facilitate cyber threat training scenarios. Kabil et al. (2018) describes the use of CVEs to improve users cyber situational awareness. Piskozub et al. (2017) also describe ways to augment decision-makers' understanding of cyber-physical situational awareness, through VR simulations. This prior work focuses on how VR/MR can enhance security capabilities. In contrast, we discuss the security of the devices and the CMR environment itself.

2. GRAND CHALLENGES

When deriving grand challenges, we need to explore the asymmetric relationship between attackers and defenders in the context of CMR systems in a systematic manner. Through analyzing the roles, vulnerabilities and exploits of CMR systems we can contribute to security standards and provide meaningful use cases. We need to investigate a plethora of issues, examples include; socio-technical manipulation, breach of confidentiality, or rendering a system unusable through a denial of service attack. De Guzman et al. (2018) present a list of open challenges. This includes discussions on topics pertaining to: security and privacy of existing devices and platforms; regulation of MR applications; native support between applications; targeted data protection; user welfare, society, policy, and ethics; profitability and interoperability; networking challenges; smart futures. We extend these further:

- **Challenge Area 1: CMR attacks and attack surfaces**. This includes modelling the behaviours of attackers in the context of CMR devices, how users interact with them and how attackers can exploit this to their advantage. We can do so, by developing in-depth use cases of attacks, but also explore different techniques to express behaviour of systems, users and attackers, including for instance UML, Message Sequence Charts, Flow Charts, Attack Graphs or Formal Methods (including model checking and protocol analysis).
- **Challenge Area 2: CMR attack impacts and harms**. Measuring direct and indirect socio-technical consequences of CMR attacks, including the complex dependencies of harms and their relationships with one another is a challenge. While we can use tools such as the Common Vulnerability Score System (see Scarfone and Mell, 2009) to compute a quantified measurement of a vulnerability's technical impact, being able to measure its follow-on impacts or harms at the socio-technical level is a far from trivial task.
- **Challenge Area 3: CMR ethics and norms**. Within the context of CMRs, what represents good and bad (bad here relating to nefarious and inappropriate, but not malicious) behaviour? How might the boundaries and definitions behaviour evolve over time? For example, when might teasing in a CMR cross the boundary into bullying.
- **Challenge Area 4: Effective tools and techniques to**:
  - **Detect CMR attacks**, including proactive and reactive measures. Proactive measures can include identifying how the education of stakeholders can improve resilience and robustness to attacks, also allowing engineers, programmers and designers to make better informed decisions w.r.t.

developing CMR applications. Key challenges exist in education in cyber security as people need be able and willing to apply their understanding. This may require changes in attitudes and intentions (see Bada and Sasse, 2014) when developing and/or using systems. Reactive means include uses of IDS-like approaches, such as misuse detection or anomaly detection systems (see Liao et al., 2013 for a state-of-the-art literature review on the subject, host-based and network-based solutions for detection at optimal locations on a system and architecture or infrastructure to best identify attacks.

– **Mitigate CMR attacks**. By exploring robustness and resilience of systems, we can withstand attacks and also recover faster. An aspect of this challenge, should perhaps cover IPS-like systems in which we develop systems that automate defensive actions in the interests of the integrity of the CMR system.
– **Combat CMR attacks**. We need to explore the incident responses that are the most efficient, effective and appropriate. Can we combat attacks (offensively), or should defenses remain passive like in most current defense systems?
– **Deter CMR attacks**. Can we identify means to prevent attacks from occurring in the first place? Once approach might be to harden hardware and software (minimise software installed and hardware used, while also making sure these are always up to date with the latest security patches) to reduce the attack surface with the aim of reducing the attacker's appetite for targeting the system or users in the first place.

3. CMR PROPERTIES RELATING TO SECURITY

Many collaborative mixed reality systems employ a client-server architecture. This is partially due to the complexity of services required, but also because a central point provides easy discovery and network packet reflection across heterogeneous connections. For example, one current CMR system, High Fidelity, has a fairly complex network architecture because of the services it provides: content streaming, audio mixing, currency and marketplace services and naming (https://blog.highfidelity.com/high-fidelity-system-architecture-f30a7ba89f80).

Tools to build centralised servers are well supported in common development toolkits such as Unity and Unreal. Many smaller scale systems use peer to peer networking strategies. For example, the Oculus Platform SDK has examples for achieving this with avatar support (https://developer.oculus.com/documentation/platform/latest/concepts/dg-rooms/). Such technologies have evolved from similar tools for desktop, mobile and console games, so the technological and architectural issues are necessarily similar (Jeff Yan and Choi, 2002; Yan, 2003; McGraw and CTO, 2008).

A number of security-related properties are essential to consider in CMR applications, these include:

- **Network communication architecture**. The topology of the network can affect where malicious attacks may be targeted. For example in a classical client-server architecture, attacking a server with a denial of service attack is an obvious choice.
- **Social interaction**. Social interaction, spatial constraints, facilitating collaborative access (see Pettifer and Marsh, 2001; Tolone et al., 2005), proximity to other users, the use of space as access control (see Bullock and Benford, 1999) and programmed access controls (see Wright and Madey, 2010) can all potentially be exploited by an attacker.
- **Impact of immersion**. The impact of having users engaged with a system that is immersive, can cause psychological harm (see Stanney et al., 1998; Lawson, 2015; Madary and Metzinger, 2016). An attacker could manipulate a CMR to cause further psychological harm. There are parallels emerging in studies of social media addiction and depression Tandoc et al. (2015)
- **Software considerations**. Downloading code to run on a local device also has an impact. Similar issues exist with app models, signed code, etc. It is necessary for security considerations to understand the limitations of this kind of model. Extensible run-time is generally not a well-researched topic outside the Web and server centre domain (e.g., Docker https://www.docker.com/), see Sallés et al. (2002).

Untethered AR/MR devices bring some interesting implications for MR network communication. In particular, the interest in network communication over 5G cellular infrastructure will necessitate smart ways to deal with balancing of local/remote computation, local compositing and network access to resources for maintaining synthetic environments. The network architectural requirements and pros and cons of various distribution architectures are outside the scope of this work so we refer the reader to Bastug et al. (2017).

There is a considerable body of research on network architectures specifically for CVEs, CMEs, and networked games (see Steed and Oliveira, 2009). A number of early works in VR proposed various ways to maintain persistent VE state over varied network architectures (see for example DIVE Benford et al., 1995 MASSIVE Greenhalgh et al., 2000, DEVA Pettifer et al., 2000, and NPSNET Macedonia et al., 1994). More recently, data and state models have also been explored in networked gaming (see for instance Hawkins et al., 1999; Hu et al., 2006).

We classify client server as an architecture in which there is one locus of control, this may be on a cloud server, virtualised (see Marsh et al., 2006) or may exist on a single server. We classify broadly peer-to-peer as device to device communication where only minimal state information is held by two independent loci of control that are kept in synchrony through message passing of some sort (see Steed and Oliveira, 2009 for further details on network architectures). This is important from the perspective of threat modelling. For this reason, we will consider use case applications that fit into these network categories.

3.1. Adversary Model and Assumptions

We focus on CMR applications particularly because of the (as yet) largely unexplored socio-technical aspects of harms in this space, including consequences in human

perception, immersion and presence. We also consider financial, reputational, political, cultural, psychological (and health more broadly), see Agrafiotis et al. (2016). De Guzman et al. (2018), as discussed, presents a data-centric approach to MR security, we argue that a wider, socio-technical perspective is necessary to be comprehensive. In contrast to Madary and Metzinger (2016) we do not discuss the broader ethical considerations of these technologies, beyond how they relate to attacks.

Attacks are actuated threats, with the explicit intent to cause harm. Outcomes of attacks can be successful, unsuccessful or have degrees of impact and harms. We assume the adversary to be a perfect and fully potent attacker, like in Roscoe and Goldsmith (1997)'s model. If the attacker can be successful, then they are always (perfect), and the maximum harm/impact is always achieved (fully potent). This methodology gives a simple worst-case approach to reasoning about attacker behaviour and enables us to consider the security of CMR systems deterministically. The attacker requires no prior knowledge of potential flaws in the system; instead, we take an exploratory approach in the attempt to exhaustively investigate socio-technical harms (e.g., CIA, nausea, trauma, manipulation, reputational, financial etc.) that can arise from an MR attack.

Since our approach is open-ended, we can in principle model the attack surface and goals (of attackers), and behaviour based on known and reasonable assumptions. This open-endedness leaves room for future attacks to be detailed, once we know more about the attack surface in future years. Adversaries may have a copy of the CMR system to hand, but hardware access may not be a requirement for all kinds of attacks. The CMR system is assumed to connect to a service as part of a client-server architecture. We assume three types of adversaries:

- **Non-users of the CMR system**. Adversaries are not users of the CMR system, but instead attempt to hack the system for nefarious purposes.
- **Legitimate users of a CMR system** are assumed to have to authenticate in CMR systems. We make no assumptions about whether identity verification is required, only that activities in a CMR application, will be tied to a user account.
- **Legitimate administrators of a CMR system** have administration privileges. These can be abused (insider threat). Adversaries may also wish to obtain escalated privileges in order to reach a large audience or target a particular user, and some attacks may require this type of capability in order to execute the attack.

We do not consider attackers without the intention to cause harm, this means that we will not be covering *acts of god* (natural hazards outside human control), *benign attacks/threats* (those that cannot cause harm) or *accidental threats*. We also exclude accidental attacks, (accidental consequences, but the intention of the attack is still there). For an in-depth discussion on our starting points, please see Cohen (1997a,b) which both discuss properties of attacks and defenses, respectively. Agrafiotis et al. (2016) discusses harms, and De Guzman et al. (2018) presents a literature survey of security in the MR space.

4. A FRAMEWORK TO UNDERSTAND THE CMR ATTACK SURFACE

Our extendable framework uses three models. These are: the Environment Attack Model, the Data and State Attack Model and the I/O Attack model. Each sub-part of the framework looks at the attack surface through a different lens enabling reasoning at various different levels of abstraction.

4.1. Overview

We can combine three models to form the basis of our framework, which can be reasoned about through an attack taxonomy. Each model works as a *lens* (see **Figure 1**) by providing a different representation of understanding attackers and attacks. These include:

- **Environment Attack Model**: considers how the attacker uses their environment (*Threat Environment*, TE) to compromise or otherwise influence the *System* and *Virtual Environment* (SE, VE respectively) (see **Figure 2**).
- **Data and State Attack Model**: considers how the attacker can use lateral movement across different states of systems as well as subsystems to achieve their goals (see **Figure 3**).
- **I/O Attack Model**: considers how the attacker can compromise the SE and VE through exposures of vulnerabilities in the Inputs, Boxes, Outputs and Integration (see **Figure 4**).

By considering how attackers relate to their environments; how they may navigate in a stateful publish/subscriber-like system; and finally: how they can target the attack surface of the SE in question, we can systematically consider a set of building blocks to reason about attacks themselves. This should allow analysts to form a comprehensive understanding of how attacks can affect the attack surface and predict what risk-owners need to do in response. This is by no means a complete solution, and we anticipate the need for other models to extend ours. The following sections provide more information about each model in the framework as well as the taxonomy itself.

4.2. Environment Attack Model

When discussing computer graphics and attack modelling, there is an overlap in vocabulary. We recognise it is necessary to decouple the terms in order to discuss "environments" in more meaningful ways. In attack modelling and model checking, the term *environment* is often used to describe the set of external factors that can affect a system. In computer graphics, *environment* can both refer to the real environment, virtual environments and combinations of both found in AR systems. The environment can also refer to the physical system's topology (e.g., a network topology) in which a device or threat exist.

For the purposes of this paper, we distinguish these terms as follows. VE denotes any synthetic or imaginary environment (software and models see Ellis, 1989). We use SE (typically often referred to as simply *System*) as the technology landscape in on which the VE is generated. Finally, we use *Threat Environment*

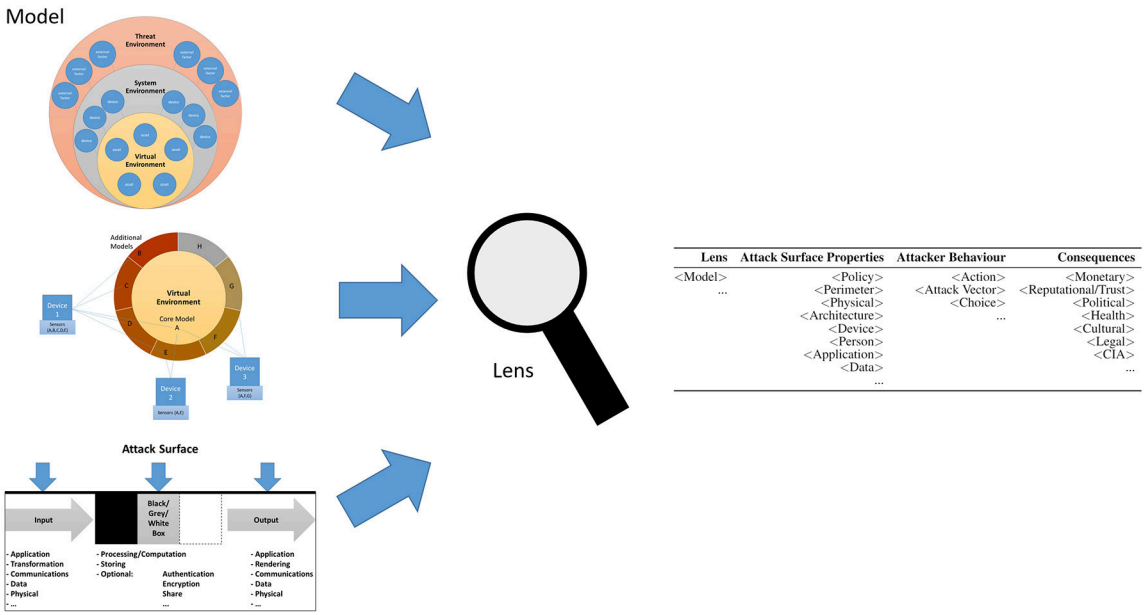

FIGURE 1 | An overview of how to use our framework. Each model acts as a lens for our taxonomy. Our framework assumes that other models can be used to reason about the attack. The purpose of the framework is to use the taxonomy from different perspectives to achieve a more holistic view of CMR applications, in order to identify possible vulnerabilities and mitigation strategies. Note this figure is an overview, bigger images can be found in **Figures 2–4** and **Table 1**.

(TE) to denote the space in which attackers operate. This is where attackers make use of software, hardware and even knowledge resources outside of IT administrators' control. We consider these environments having direct relationships with each other, as illustrated in **Figure 2**. In a CMR application, we can say that the attacker can operate outside the realm of the CMR space. In the TE, attackers intrude with the intention to directly or indirectly engage with the SE in order to compromise it or the VE, as shown in **Figure 2**. In the figure, we see that external factors or behaviour of the attacker can affect the system.

Reducing the attack surface of the SE and VE becomes challenging since for security analysts to understand the attack surface, they require rights to potentially privileged information on devices in order to map their capabilities, capacities and ability to withstand an attack. However, for CMR applications, especially those in which any device can connect to the SE and VE, IT administrators of such a system are unlikely to be able to easily obtain this information. This means that it is nigh impossible for stakeholders to obtain a comprehensive view of the attack surface. Thus, suggesting that the (already worrisome) asymmetric relationship between attacker and defender (in the attacker's favour, i.e., the attacker needs only to be successful once to compromise a system) is particularly problematic and needs to be considered in-depth by CMR application stakeholders.

In the following subsections we describe the necessary threat model assumptions, along with the attack surface of each of the constituent components: *Input*, (black/grey/white) *Box*, *Output* and finally the *Integration* of the previous components. Note: as the TE exists outside the bounds of what analysts are privy to, any assumption we make about the TE will be speculation or based on heuristics from attacker history. We assume that any third-party dependencies to a CMR application, is a part of that SE.

4.3. Data and State Attack Model

We envisage CMR systems making greater use of sensors, cameras, on-body/off-body networks, wearable, implants etc. The inherent implication of having rich variation in sensors and interfaces that each connected user might have at their disposal, is that I/O to/from the VE from users will substantially differ between users. If we see sensor I/O to/from a CMR as a local contribution to a shared VE's state, then the state of the whole CMR itself will be dependent on these contributions. In other words, each user contributes information to the complete state of the VE, and each user obtains their own (device dependent) perception of the complete state. This means that each CMR effectively has a fuzzy layer of states defined by the combination of sensors which could be vulnerable to various attacks. More dangerously, it is also entirely possible that no-one has a complete picture of the VEs state at all times meaning it is challenging to verify correct and secure behaviour.

We can more formally define the complexities of state in CMRs by borrowing terms from Marsh et al. (2006) *subjective* and *objective* reality concepts from the Deva Architecture (also see Pettifer et al., 2000) for shared VEs. Marsh et al. (2006) characterise behaviour of entities in shared VEs in terms of their objective (that is, part of the synchronised, semantic state of the environment) or subjective state (part of what is necessary to best-portray the correct interpretation of the objective state

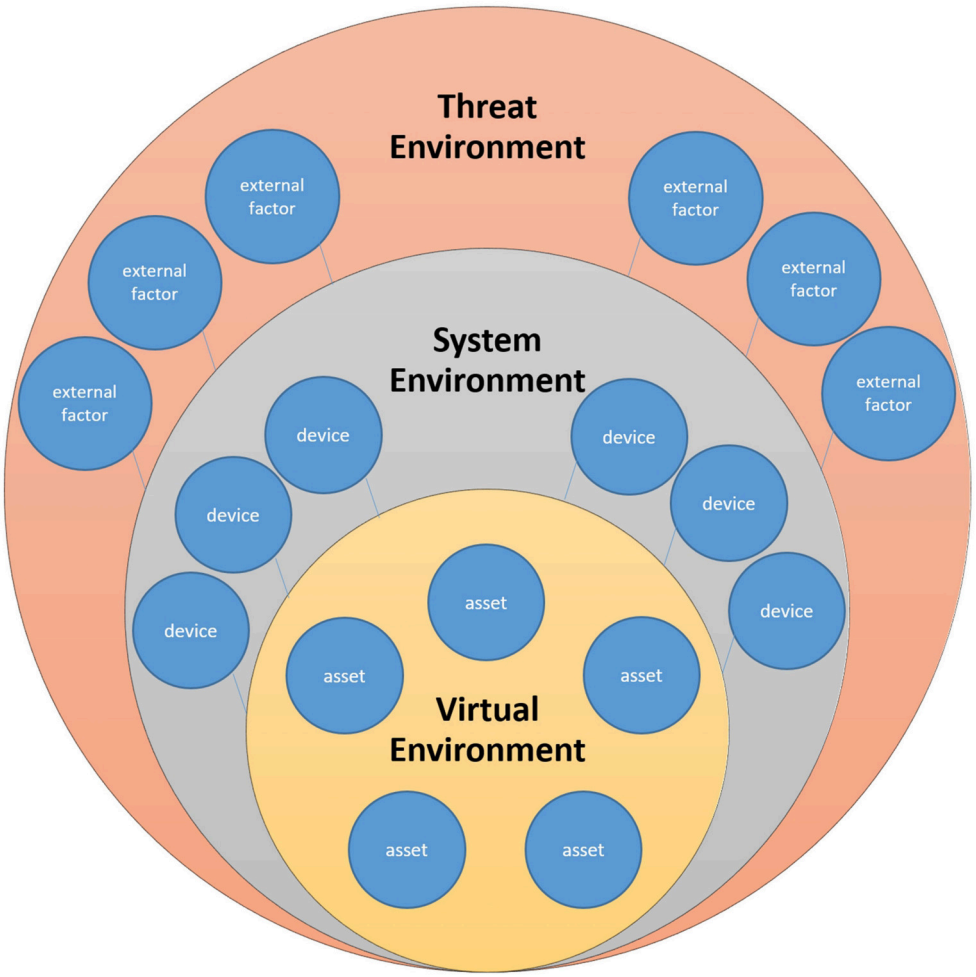

FIGURE 2 | Attackers can operate in a Threat Environment (external forces to a system). Threat actors can interact with a System Environment (architecture/topology) directly or indirectly, and subsequently (once compromised) attack the Virtual Environment (graphics).

to the user). This is akin to differentiating between the world around us and our actual perception of it. These ideas are useful to encapsulate the situation in CMRs, particularly where users contribute information to the overall state from a broad and diverse range of devices and sensors. For the purposes of this paper, we will call them Core Model (objective state) and Additional Models (subjective states). We add to these definitions, the *unified* Model (the sum of the core and all additional models, here represented as a doughnut circle) of the CMR which is the sum of all potential contributions from every participant in the CMR. This unified CMR depends on whom and what is connected at any given time and the capabilities of their individual I/O hardware, sensors and states. The Core Model is that which any participant with a minimum viable setup can access (see **Figure 3**).

Some examples include:

- **Render one or more component of a CMR unusable, or have it misbehave**, either by bricking (rendering permanently unusable) the component, temporarily flooding it so it cannot perform as expected (denial of service), or misbehave (affecting the integrity of the system).
- **Identity theft on CMR platforms**. Compromising a CMR system can enable an attacker to pose as a legitimate user and manipulate others to reveal other sensitive data or manipulate them to behave in a manner in which they normally would not. Such behaviour can have serious financial and reputational consequences.
- **Health risks**. Griefing (see Chesney et al., 2009), trolling and cyber bullying on the Internet are not new concepts, but it translates straightforwardly into CMR applications. Online activities of this sort have also been linked to cases of suicide (see Hinduja and Patchin, 2010). Another health risk can be to compromise a system intended for simulation and training. For example, an application for studying and learning to treat schizophrenia using VR, (see Freeman, 2008), or other mental health issues, (see Gregg and Tarrier, 2007), could be compromised to teach bad practice or to have detrimental effects on victims (e.g., cause nausea or exacerbate the victim's mental health disease).

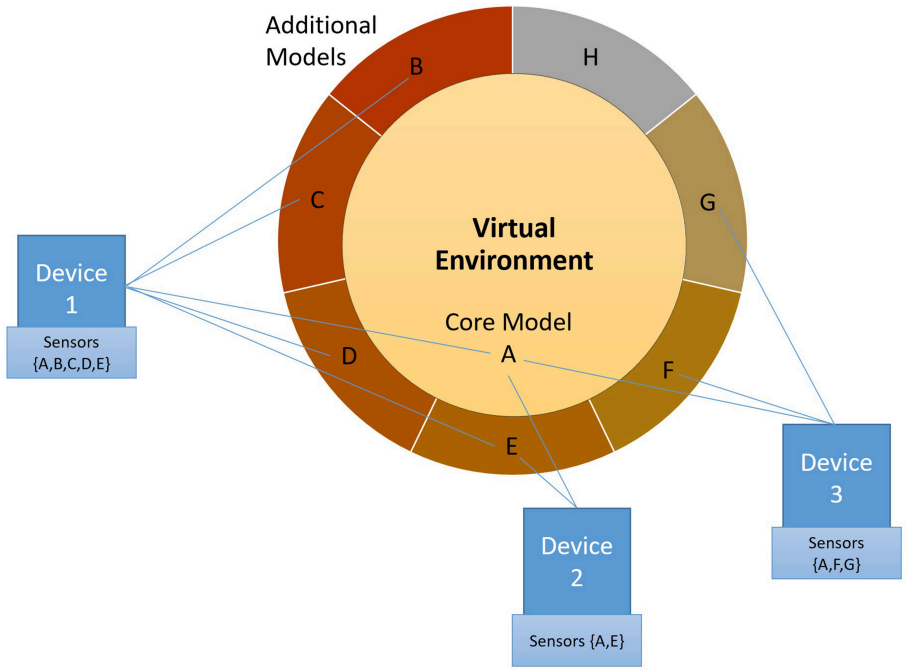

FIGURE 3 | A conceptualisation of the threat landscape in which attack actors can operate. Despite the attacker being confined to a Core Model, they have the potential to intrude and compromise the Unified CMR (all models) because of the relationship between models that is often readily available in CMR applications. This approach is akin to a publish/subscribe model, in which the expected areas in the circle are only accessible to those devices with the right sensors and that subscribe to that part of the circle. Of note: this does not preclude any attacker from attacking other models (even if those sensors that make up those models are not accessible to the device), as the attack can still go through the core model.

- **Personal data utilisation for social manipulation** (rather than persuasion). We have seen how utilisation of personal data on social media can help manipulate large audiences through fake news and online propaganda, which is argued to have had a dramatic effect on the US election (see Persily, 2017). Content displayed to users can be projected based on psychological profiling (eye-gaze data can for instance expose a range of proclivities). With MR and CMR systems evolving, it will be necessary to consider how perception and manipulation of VE content through such social manipulation techniques can affect end-users. A particular factor that magnifies the risks and harm possible with security violations in CMRs is the plasticity of the human mind (see Madary and Metzinger, 2016 for a detailed discussion).

4.4. I/O Attack Model

In this section we outline our input/box/output/integration exposure (I/O Attack) model. For our attack surface, similar to De Guzman et al. (2018), we consider the whole system that is necessary for the CMR (see **Figure 4**). The attack surface can be considered to be like a data-flow pipeline. This model is loosely inspired by Oliveira et al. (2009), the key difference being we look at the analysis domain from the attacker's perspective. We assume that every device, and every sub-component of the device can have this model applied to it. We break down the attack surface into: **Input** concerns all interfaces that consume data from the real world (including device sensors, end-user inputs, physical wires, etc.). **Box** is a shorthand notation for a black/grey or white box, a common term in software engineering. It is any subsystem that a stakeholder can directly or indirectly access. Inside it, some processing takes place to modify the input according to the requirements and specification of the developer. Output concerns the processed data passed from one subsystem to another. **Output** concerns the outgoing data from a box. Output may be input for another subsystem, or it may also be the final destination for the processed data (e.g., a display). **Integration** (the black outline of the figure) concerns the interplay between all components of the system. While as separate units, they may behave within acceptable thresholds, it may be that the system as a whole, when viewed altogether, may misbehave as a consequence of an attack, which may not be detectable by any single component.

4.4.1. Input

Given the Data and State model as described in **Figure 3**, we assume that multiple subsystems can interface with each other in sequences (like a daisy chain). Stakeholders may be in a position in which they have no control of this sequence, one subsystem's input will be another subsystem's output (and vice versa). We consider the input and output attack surface to consist of components pertaining to: raw data; data transfer; manipulation of real-world information; and manipulation of the

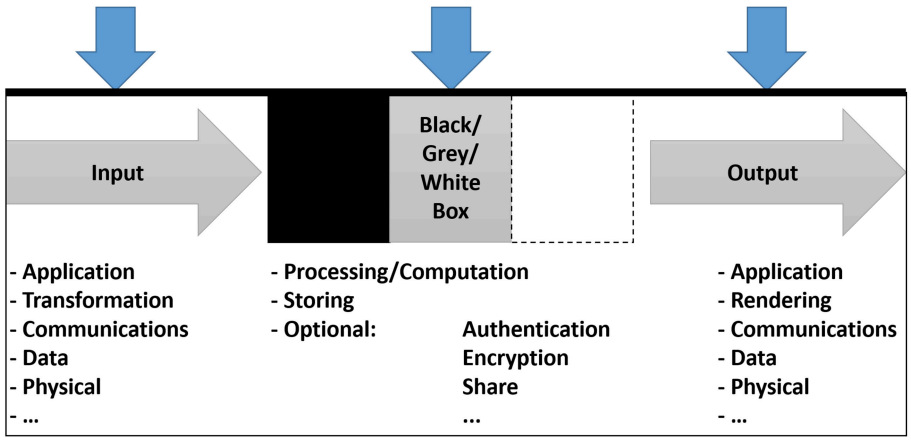

FIGURE 4 | Input/box/output/integration (I/O Attack Model) describes the constituent components that are exposed to (and can be compromised by) the attacker.

semantic meaning of the user-level application of information in question.

- **Application** is comprised of a combination of rendering to form different varieties of realism that the user interprets (see, for instance Ferwerda, 2003, although we expand the idea of realism to other senses than vision alone). Attackers may for instance manipulate the interpreted meaning in the content of the VE.
- **Conversion** is the part of the pipeline that converts one type of signal into another through sensing (transformation, input) or projection (rendering, output).
  - **Transformation** (input) includes the low level processes of converting between real world, digital and synthetic signals. Inputs considers converting real world and synthetic signals into digital data to be communicated/transported.
  - **Rendering** (output) includes the sensory information that can be attacked, such as for instance: visual, audio, tactile, olfactory. We regard this as being the information that is projected in some form (e.g., display, speaker, haptic device or digital scent emitter). Attackers may for instance manipulate the content of the VE.
- **Communications** includes protocols necessary to transmit/transfer the data. Attackers may for instance manipulate the content of the VE.
- **Data** includes the data buffers themselves.
- **Physical** includes the hardware itself, such as cables and monitors necessary to receive (input) or transmit/transfer (output) data.

We cautiously observe that the attack surface of the input/output is in similar to that of the OSI model. While a comparison may be useful conceptually, we note that these are analogous rather than "equivalent to" comparisons, specifically because wireless inputs/outputs will be equivalent to the OSI model, whereas transmissions and transfers over cables will not be (architecturally speaking).

4.4.2. Box

In our work, we assume that there are many types of boxes that all follow the black/grey/white box paradigm. Black boxes has the subsystem not visible to engineers and developers who did not design the system. Grey boxes have parts of the system visible and modifiable to other (including users), and white boxes have the entire architecture visible and modifiable. The attack surface is dependent on box functionalities, software and hardware they incorporate. As CMR applications are still immature and evolving, we break down the criteria for being considered a box into mandatory and optional criteria, as well as what these may mean for the attack surface:

- **Mandatory criteria**:
  - **Processing/computation**. There must be some processing or computation involved (other than signal transformations).
  - Access to **Persistent storage**. There must exist a unified, additional and core models.
  - **Sharing/networking capabilities** aiming at sharing processed or stored data.
- **Optional criteria**, i.e., a system built with security in mind is likely to also have:
  - **Authentication** considers access controls, so the users only have access to what they should be seeing at any time.
  - **Encryption** aims at ensuring privacy for the user.

Each of these boxes, each a standalone unit in their own right, dependent on hardware, software and network topology configurations will dictate how open to attack exposure a box is. The attacker can modify a box to for instance:

misbehave, leak information, prevent authorised users from accessing their systems, or render the box unusable. Examples of boxes include:

- **Desktop computers, mobile phones, laptops and servers**. These boxes will contain a number of subboxes, such as GPUs, CPUs, RAM etc.
- **Head-mounted displays** such as Oculus, Vive or other AR/MR glasses.
- **IoT devices** that use sensors, but that also have additional processing capabilities to augment the sensed data.
- **Wearables** such as smart watches, fitbits, clothes.
- **Assistants** such as Google Assistant, Alexa and Siri.
- **Software processes** such as tracker drivers or the VR compositor (e.g., WebVR https://webvr.info/).

While the link between MR and IoT is tenuous, we believe this will strengthen as more IoT devices become available, and MR devices become common. More recently for instance, consumer untethered VR headsets have become available.

4.4.3. Integration

The integration (part of the attack surface) considers that the interplay between the individual components can be manipulated by the attacker by considering the system *as a whole* (rather than constituent components). Human-level manipulation plays a significantly larger role than technical ones in the integration as some attacks are disguised as regular usage of the system. Example attack vectors may include:

- **Reconnaissance**. An attacker wishing to identify how their node/device on the CMR application can influence the rest of the SE and VE. Attackers may use reconnaissance activities as part of a larger attack Hutchins et al. (2011).
- **Social engineering**. An attacker can influence users to do their bidding, for instance because the other users have legitimate access, or to frame other users or manipulate them into nefarious activities.
- **Identity theft**. An attacker can hijack and pose as another user in a CMR application. As different levels of abstraction can play a role in CMR applications through the uses of avatars, other users may not notice that the victim's identity has been compromised before the attacker has acted on their objectives.
- **Imbued behaviour** Pettifer and Marsh (2001) highlight for instance how functionality can be imbued upon an entity by the VE in which it is created. Such functions may be "enforced" by an environment (thus being a "law of the world") whilst others are optional. This escalation of functionality (such as privileges) could give attackers unfair advantages, especially in social contexts or VEs where real world benefits may result. For instance, the attacker winning a game of "CMR poker" because they are privy to other users' cards.

Again, as the CMR application space is still evolving and not particularly mature, we suspect the types of attack vectors are likely to grow over time. The purpose of our I/O Attack model is to demonstrate which parts of a pipeline the attacker can exploit. We provide further examples of integration attacks in section 5.

4.5. Attack Taxonomy

Each model provides a different lens for viewing the different attacks, and describes varying scopes of attacks that enable more straightforward reasoning about potential kinds of attacks. We envisage that other models can also be incorporated into the taxonomy. Since this taxonomy is general and provides a way to reason in the context of specific applications we outline in **Table 1** some high level consequences. These could range from "Users stop using the systems," "Loss of Trust," "Personal Data Leak" and so on. However, these consequences are application dependent. An excellent article discussing hate in social media outlines some potential consequences in the context of social VR (https://www.adl.org/resources/reports/hate-in-social-virtual-reality#hopes-and-fears-for-the-future-of-social-virtual-reality). For example, "harassment," "racism," "anti-semitism," "bullying" etc. To make the high-level consequences of our taxonomy more concrete, the reader is referred to the case studies in section 5 which list some more context specific consequences based on the CMR application types.

Our taxonomy builds on **assimilating three key categories of (observable) properties** about the attack surface, attacker and consequences (impacts) to allow analysts to formulate their own descriptions of attacks:

- **Lens**: *from which perspective to view the attack?* No model is fully accurate or precisely describes all attacks. We assume that any model used, provides a different perspective on an attack (i.e., uses a different lens). In our paper, we have developed three different models, specifically designed with CMR application use cases in mind. However, we envisage that any attack model that applies to CMR applications can be applied here.
- **Attack Surface Property**: *what properties must be present prior to the attack?* These can be technological or human-centred, such as policies, vulnerabilities, perimeter (e.g., firewall), physical, (network) architecture paradigm (e.g., client-server), device (host), person, application and data. This is like mapping the CMR system to that of *defense in depth/perimeter defense*, see Smith (2003).

TABLE 1 | Taxonomy of the CMR application attack surface.

Lens	Attack surface property	Attacker behaviour	Consequence
<Model>	<Policy>	<Action>	<Monetary>
<TBE>	<Perimeter>	<Attack Vector>	<Reputational/Trust>
	<Physical>	<Choice>	<Political>
	<Architecture>	<TBE>	<Health>
	<Device>		<Cultural>
	<Person>		<Legal>
	<Application>		<CIA>
	<Data>		<TBE>
	<TBE>		

- **Attacker Behaviour**: *what does the attacker need to do to compromise the CMR application?* This includes attack vectors and actions necessary to execute attacks. To facilitate the identification of these properties, analysts may also make good use of UML, message sequence charts, flow charts, attack graphs or formal methods (including model checking), w.r.t. the attack surface properties. From the behaviours we can identify the part of the attack surface the attacker needs to exploit.
- **Consequence**: *what are the indirect and direct outcomes (impact) of the attack?* Note that the term consequence is specifically used here to denote a *reaction* or *effect* of the attack. This can be inconsequential, i.e., without any impact or harm. Net detrimental effects is harm, consequences could also be net positive (e.g., lessons learnt from an attack meant that some users who previously regarded security as not important, now do take it seriously and were able to withstand a more significant attack not that far ahead in the future), or no positive or negative effect may be measured at all.

Combined, each category paints an accurate picture of the building blocks that make up an entire attack surface. Our taxonomy accommodates future directions of CMR applications research, and can also be used in the IoT space and in traditional network defense research (see **Table 1**). Where *<TBE>* is present, we explicitly mean *To Be Extended* with other future models, properties, behaviours or consequences.

In essence, we consider the taxonomy also able to predict outcomes by considering how:

Models (Lenses) give different perspectives of: *Attack Surface Properties* + *Attacker Behaviour* → *Consequences*. Note here that consequences is used as a generic term to denote both impact (typically used to describe effects of attacks in the digital space) and harms more broadly.

We take a simple approach of describing attack surface properties, attacker behaviours and consequences in short sentences or single words using different perspectives. These different models (e.g., attack graphs or attack trees) can be used to describe attributes in the taxonomy. Viewing different attacker behaviour descriptions using different models, for instance, may enable enhanced cognition among security practitioners and researchers allowing for lateral thinking among users of our taxonomy.

Attacks against CMR applications are not well-known, we consider it highly important to explore threats to CMR applications and their users. As a catalyst to derive likely attacks in our taxonomy, we describe threat awareness of attacks using the basic classification scheme, as described by Chismon and Ruks (2015):

- **Known knowns**: a threat that is well-understood and we have techniques to prevent or mitigate the threat.
- **Known unknowns**: a threat that has been identified (i.e., we know of its existence), but that we do not know any details about it.
- **Unknown unknowns**: a threat we know nothing about, not even its existence.

A key motivation of this position paper is to deliver a framework that allows stakeholders to systematically identify and consider threats posed by attackers. We can do so by considering how threats are first "unknown unknowns," then "known unknowns," and then eventually "known knowns." By scoping threats through our framework, CMR application stakeholders can straightforwardly reason about threats that may actuate into attacks on their particular CMR attack surface, and better defend against them.

5. CMR ATTACK USE-CASE STUDIES

We briefly analyse four use cases reasoned using our framework, presenting these in the form of our attack taxonomy. These case studies are sourced from typical CVE and CMR applications implemented by other researchers. Specifically, a range of collaborative *virtual prototyping* for engineering design (see Gomes De Sá and Gabriel, 1999; Kan et al., 2001; Glencross et al., 2007), *virtual try on* (see Kartsounis et al., 2003; Kim and Forsythe, 2008), and *health-care* applications (see Rizzo and Kim, 2005; Fujita et al., 2010) and *gaming* have been proposed over the last few decades. Although, we focus on just four illustrative use cases, there are of course many more application scenarios in social, entertainment, interior design, training etc. For brevity, we consider limited example attack behaviours in our instantiation of the taxonomy for each use case. It is not possible to perform an exhaustive analysis of the entire attack surface for each use case within the scope of this paper, instead we show the reasoning is extendable. We include in supplementary materials a more thorough analysis for one use case.

5.1. Use Case 1: Industrial Espionage

The first scenario we examine is that of a targeted industrial espionage attack of a CMR application used for prototyping or engineering design. In this situation, the CMR has some information relating to valuable intellectual property (IP), perhaps in the form of computer-aided design (CAD, for example a hypothetical shared version of VR Rolls Royce Engine design https://www.youtube.com/watch?v=CyMzFenEuNI) models, simulation or industrial know how. We consider this scenario to adopt a broadly client-server architecture and list some considerations in **Table 2**. The potential for physical consequences can arise from denial of service attacks that could reduce framerate on target devices inducing nausea and/or frustration with the usability of the application. In CMR engineering applications, exploiting vulnerabilities in connected cameras to spy on users of the system can also lead to psychological harms, for instance: a colleague may blackmail or coerce another as a consequence of spying.

5.2. Use Case 2: Virtual Try-On

In this scenario, we examine the considerations involved with a virtual try-on application. Many such applications are emerging

TABLE 2 | Use Case 1: Industrial Espionage.

Lens	Attack surface property	Attacker behaviour	Consequence
Environment	Valuable IP	Theft	Reputation/Monetary
model	User Acceptance	DDoS/DoS	Users stop using the systems
	Insider Access Control	Escalated Privileges	System Misuse
	<TBE>	<TBE>	<TBE>
Data and	Simulation	Theft	CIA of Sim. Data Affected
State model	3D Model Design	Theft	IP theft/Monetary
	Collaboration Model	DDoS/Dos	Collaboration Impossible
	Multiple Users	Spoofing	Loss of Trust
	Administrative Data	Theft	Personal Data Leak
	<TBE>	<TBE>	<TBE>
I/O Attack	Central Server	DDoS/DoS	Access Limitation
model	Central Server	DDoS/DoS	Induced Nausea
	Simulation	Theft	Monetary
	3D Model Design	Theft	Monetary/IP
	Camera Data	Cyber Spying	Psychological Harm
	Administrative Data	Theft	Monetary Cost
	Collaboration Model	DDoS/DoS	Usability Limitation
	<TBE>	<TBE>	<TBE>

that offer users the ability to try on anything from eyewear to clothing and to visualise these before buying. Indeed Amazon has recently patented a blended reality virtual mirror which aims to allow users to try on virtual clothes (https://www.theverge.com/circuitbreaker/2018/1/3/16844300/amazon-patent-mirror-virtual-clothes-fashion). Again, in this scenario, we consider this type of application to adopt a broadly client-server architecture and list considerations in the **Table 3**. This application scenario has the potential for a range of privacy threats and personal psychological harms that can arise from access to and misuse of personal data. The degree of harm done to targets, is also far from uniform as the psychological impact on an individual will depend on their own prior experience, state of mind, vulnerabilities, personal body image as well as what the attacker actually does with the data. For example, an attacker might use the data to body shame, cyber stalk and bully the target of the attack. A parallel can be drawn with the celebrity hacking scandal, in which Apple's iCloud server was compromised and a collection of almost 500 private photos (some containing nudity) of various celebrities, mostly women, were posted on the imageboard 4chan (https://www.theguardian.com/technology/2014/sep/01/naked-celebrity-hack-icloud-backup-jennifer-lawrence).

5.3. Use Case 3: Health-Care: Virtual Doctor

In this scenario, we consider a health-care application which enables consultation with a virtual doctor. A number of different network architectures have been proposed for such applications, but we consider specifically the one-to-one consultation scenario. For this reason, we cast it in a broadly peer to peer network architecture, even though some application data could well be held in a central location. Like the previous use case, this one also has potential for very serious privacy threats, personal, psychological and even medical harms that might arise from an attack. A particular threat could be someone masquerading as a doctor and manipulating the user to cause them physical harms. Another potential threat is one where a third person enters the consultation uninvited, which could cause significant feelings of violation and loss of trust in the application.

Much like the previous use case, the level of harm possible to an individual is very dependent on their own mental, physical, medical condition as well as the specifics of how the attacker exploits the system. There is a further key requirement of a high level of confidentiality of communication in the system to engender trust and acceptance by users. If users lose this trust, they may discontinue an essential therapy. We outline these in **Table 4**.

5.4. Use Case 4: Shared Gaming

In this scenario, we examine the considerations involved in gaming applications (see also **Table 5**). Gaming is possibly the most mature of the use cases provided in this paper, as many implementations of CMR-like solutions already exist, so this is by no means an exhaustive list. Gaming allows for many connected devices to collaborate to a shared goal, often bound by the rules of the game in question.

Social attacks: Gaming can be manipulated at the social level through social engineering. Example user can manipulate other players into believing they are another player by masquerading as them (identify theft is often mitigated in several games, esp. MMORPGs, today by providing a unique identifier that other players can examine.), or bully or discriminated against other players, or manipulate other players in-game to conduct real-world actions they normally would not do. For instance. virtual assets can be confiscated and held hostage by another player.

TABLE 3 | Use Case 2: Virtual Try-On.

Lens	Attack Surface Property	Attacker Behaviour	Consequence
Environment	Body Image Sensitivity	Social Engineering	Psychological Harm
model	Demographic Vulnerability	Stalking/Bullying	Psychological/Physical Harm
	Predisposition to sharing	Shaming Online	Reputation Harm
	<TBE>	<TBE>	<TBE>
Data and	Body Model/Avatar	Deep Fakes & Theft	Reputation Harm
State model	Garment Design	Theft	Copying/Fashion Fakes
	Fabric Simulation	Theft/Manipulation	Retail Monetary Loss
	User Credential Data	Theft	Access/Monetary Loss
	<TBE>	<TBE>	<TBE>
I/O Attack	Central Server	DDoS/DoS	Access Limitation
model	Central Server	DDoS/DoS	Induced Nausea
	Central Server	DDoS/DoS	Retail Monetary Loss
	Personal Camera Data	Shaming online	Reputation Harm
	<TBE>	<TBE>	<TBE>

Technical attacks: Attackers can manipulate bugs, vulnerabilities or weaknesses of the game system to cheat, whether this be for competitive advantage or for the amusement of the attacker. Latency-issues during gameplay is often addressed by the server to give players an even playing field. This well-meaning latency handicap can be misused by the attacker to gain a competitive advantage. Sophisticated attackers might also be able to manipulate what is rendered on screen to other players, or steal bank details from their accounts.

Socio-technical attacks: Players can be stalked by other players through IP address retrieval, leading to scenarios in which one player fears for their well-being. Attackers can also manipulate the rules of the game for their own advantage, whether this is financial gain (e.g., exploiting a bug to gain significant amount of in-game currency), cause reputational harm to a player or game developer. The mix of heterogeneous sensors that CMRs are likely to include in the future, many of which borrowed from the IoT space, are likely to increase the attack surface substantially. Game players may collude, sharing resources to flood the connection of other players.

6. DISCUSSION: DEFENSE STRATEGIES

6.1. Recommendations

6.1.1. Address Low-Hanging Fruit

Many security practices exist in other systems today, esp. those involving sanitation of inputs and outputs. De Guzman et al. (2018) has already pointed out a number of these efforts, including:

- **intrinsic input sanitisation** such as user inputs or well-formedness checks.
- **context-based sanitisation** such as checking for the contextual meaning of assets, or this could also mean malicious behavior in the CMR application itself (even if the input is clean, the higher level interpretation of the input may not be).
- **video/output sanitisation** to check well-formedness of output.
- **extrinsic input sanitization** such as those that relate to privacy preferences by end users.
- **protection through abstraction** such as revoking untrusted applications' access to any raw input feeds.
- **protection through encryption** can keep information confidential between parties if protected with basic end-to-end encryption.
- **secret sharing or secure multi-party computation** (i.e., splitting the secret across multiple trusted parties). Data can be split among untrusted parties such that information can only be inferred when the distributed parts are together.
- **gesture- and physiological-based authentication** (including biometrics) such as continuous authentication and device attestation techniques can provide assurances that the items in questions and end-users have not been hampered with.

Other sanitization or run-time checking methods may be implemented to ensure players behave according to the rules of the system, check expected throughput and volume of data fall within acceptable thresholds (e.g., spammers and content being sent across the wire). Monitoring of user patterns may be enough to identify anomalous behaviour, however, this may be used in an attack if an adversary if they escalate their privileges.

6.1.2. Educate Users About Attack Surfaces and Defenses

In section 2, we discuss that educating users is a proactive measure to detect attacks, as well as increase victims' threshold for malicious behaviour (to not be "as affected" by attacker harms). Both are non-trivial tasks. CMR attacks are poorly understood, because many have yet to be conducted in real systems. Furthermore, with the attacker-defender also being an arms race, some victims eventually might become attackers after they have been compromised. If security literacy increases overall, this

TABLE 4 | Use Case 3: Health-care: Virtual Doctor.

Lens	Attack Surface Property	Attacker Behaviour	Consequence
Environment	Confidentiality	Spoofing/Appearing	Loss of Trust
model	Mental Health	Manipulation	Psychological Harm
	Physical Condition	Manipulation	Physical Harm
	Suggestivity	Manipulation	Legal Consequences
	<TBE>	<TBE>	<TBE>
Data and	Treatment Simulator	Theft	Physical/Psychological Harm
State model	Treatment History	Theft	Loss of Trust/Privacy
	Treatment Plan	Theft/Manipulation	Loss of Trust/Physical Harm
	Patient Credentials	Theft Spoofing	Physical/Psychological Harm
	Practicioner Credentials	Theft/Spoofing	Legal Consequences
	Medical Data	Theft	Loss of Trust/Privacy
	Medical Data	Sale	Legal Consequences
	Medical Data	Manipulation	Incorrect Diagnosis
	<TBE>	<TBE>	<TBE>
I/O Attack	Peer-to-Peer	DoS	Access Limitation
model	Central server	DDoS/DoS	Induced Nausea
	Camera Data	Theft/DoS	Loss of Trust/Usability
	Microphone Data	Theft/DoS	Loss of Trust/Usability
	Medical Sensors	Theft/DoS/Manipulation	Physical Harm
	Real-time Response	DoS	Efficacy of treatment
	interface modifications	DoS	Access Limitation
<TBE>	<TBE>	<TBE>	<TBE>

TABLE 5 | Use Case 4: Shared Gaming.

Lens	Attack Surface Property	Attacker Behaviour	Consequence
Environment	Vulnerable Demographic	Bullying/Discrimination	Stop Playing
model	Socially Dependent	Bullying/Manipulation	Monetary/Psychological harm
	Personal Insecurity	Manipulation	Player Disadvantaged
	Body Image	Bullying	Legal/Psychological Harm
	Mental Health	Bullying	Depression/Suicide
	Suggestivity	Manipulation	Player Disadvantaged
	Location	Stalking	Fear/Physical Harm
	Group Dynamics	Spoofing	Loss of Reputation
	<TBE>	<TBE>	<TBE>
Data and	Game Simulation	DDoS/Manipulation	Monetary Loss
State model	Joining Credentials	Theft	Access Limitation
	Group/Team Membership	Escalate Privileges	Unfair Advantage
	Player Identity	Spoofing/Theft	Loss of Trust/Privacy
	Multiple Players	DDoS	Loss of Causality
	Player Assets	Theft	Monetary Cost
	Game Levels	Cheating	Unfair Advantage
	<TBE>	<TBE>	<TBE>
I/O Attack	Peer-to-Peer	DoS	Flooding/Access Denial
model	Central server	DDoS/DoS	Induced Nausea
	Camera Data	Theft/Sale	Loss of Trust/Usability
	Microphone Data	Theft/Sale/DoS	Loss of Trust/Usability
	GPS/location Data	Stalking/Harassment	Physical/Psychological harm
	Camera/Microphone/GPS	Theft/Sale/Harassment	Legal Consequences
<TBE>	<TBE>	<TBE>	<TBE>

may help defenders in the short term, but the asymmetric relationship between attackers and defenders is an arms race.

6.1.3. Explore Applications of Detection Systems

In traditional network security: IDSs/IPSs are hardware and software sensors dedicated to monitor and analyse network traffic data, log files and other system call traces to detect attacks. We postulate it is possible to employ IDSs and IPSs to detect attacks posed to CMR systems. Specifically, we believe we can deploy intrusion detection sensors at the various parts of the attack surface: input, box, outputs and integration of the SE.

Additional sensors may be deployed to identify semantic issues in VEs (integration). These types of detection systems would be far more complex than detection systems today, in that they must be able to take human context (e.g., to what degree a hijacking is a joke) into consideration as well. Data sent for analysis to detect malicious activities or policy violations needs to produce reports to a management station (e.g., a security operation centre). To the best of our knowledge, neither of these types of systems have been implemented for CMR applications.

Two classes of detection methods exist: *misuse detection* and *anomaly detection* Liao et al. (2013). Misuse detection works by identifying actions that match patterns (signatures) of known attacks. The severity of misuse signatures is determined by how much the analyst who set the rule believes that the attack pattern in question matters (intrinsically). Misuse detection rules range from simple well-formedness and validation checks in data to keeping track of system states. Anomaly detection on the other hand look for actions that significantly deviate from "*normal behaviour*." Normal behaviour is often defined by statistical deviations of features being monitored. These systems will look to statistically-based (no prior knowledge about the system necessary), knowledge-based (Prior knowledge about system necessary) and machine-learning based (algorithm learns normal behaviour over time) normal behaviour. IPSs may also provide measures for automating fixing-of-abuse. However, both IDSs and IPSs are prone to generate a plethora of false positives, and should therefore be used experimentally to detect potential intrusions.

6.1.4. Develop CMR Applications Aimed at Collaborative Intrusion Detection

Zhou et al. (2010) describe various architectures for collaboration in intrusion detection, including: a centralised correlation unit, hierarchical approaches and fully distributed architectures. Vasilomanolakis et al. (2015) describe a taxonomy that is broken down into: Local Monitoring; Membership Management; Correlation and Aggregation; Data Dissemination; Global Monitoring. CMR applications may play a larger role in collaborations across different cyber security analysis teams in the future. Systems could be developed to support sharing of CTI. We should consider how such systems can be best leveraged by security analysts, but also consider how these systems can be attacked. The attacker may wish to exploit each of these aspects in the Collaborative IDS (CIDS) taxonomy, which we would need to consider how to protect.

6.1.5. Develop Metrics or Assessments for Harm

We characterise the perceived distance of a security breach from individuals in reasoning about the level of harm and potential long term psychological impact. For example, victims of burglary, display long term psychological impacts from the very personal violation of their home (see Beaton et al., 2000), similarly victims of identity theft suffer long term psychological harm (see Roberts et al., 2013). In contrast, we speculate that victims of security threats such as denial of service attacks, loss of National Health Service data and loss of password/personal data stored on company servers are potentially less harmed psychologically as the crime is perceived to be distant i.e., affecting the companies/organisations rather than the individuals whose data has been breached. Stealing an avatar, or masquerading as another player is also another example of identity theft.

6.1.6. Use Distributed Ledgers Critically

Li et al. (2017) present a survey of the security of blockchain systems. Distributed ledgers such as blockchain may provide assurances that a sequence of events have been executed in order. This method could provide reassurance that the current core model of the system (e.g., SE and VE) is correct, but should any CMR system wish to employ a smart contract architecture, this system will also be susceptible to smart contract attacks and misuse and stakeholders should follow common best-practices in security (https://consensys.github.io/smart-contract-best-practices/known_attacks/). In an environment in which much transient information (users dropping in and out, i.e., overlapping states), there may be an unmanageable problem of maintaining verifiable subjective and unified states even with a blockchain solution.

DuPont (2017) describes a form of algorithmic governance: a short-lived experiment to create a "Decentralised Autonomous Organisation" (The DAO) on the Ethereum blockchain. Within days the DAO's code was exploited by an attacker who used unintended behaviour of the code's logic to rapidly drain funds worth 3.7m in Ethereum tokens. The DAO, demonstrates how the use of smart contracts can automate large portions of a system. CMR applications could expand to include automation in governance, but any automation should be continually peer-reviewed by and managed by a human for the foreseeable future to prevent similar attacks occurring in CMR applications.

6.2. Benefits and Disadvantages of our Framework

We believe the key benefits of our approach are:

- **It comprehensively considers the grand challenges of the future** in CMR to a degree that no previous work has previously investigated.
- **It is an abstract-based reasoning approach** which allows for the benefit of making use of different levels of abstractions to protect a CMR architecture from different types of cyber attacks with potentially socio-technical harm (in particular psychological and health-related harms). It is an aggressive approach to identify and address security concerns, first and foremost aimed at enabling stakeholders to more

straightforwardly understand and reason about the MR and CMR attack surface, by considering the inputs, processing, outputs and integration of existing systems.
- **It is a starting point for research** into CMR attack surfaces, intended to provoke discussions and research.

Key limitations of our approach are:

- **It is largely conceptual**. Presently, our framework is very conceptual as it stands, with little empirical or experimental evidence to support the assertion that our approach will be successful for real, and in particular large scale CMR applications.
- **It lacks specificity**. Our approach aims to be generic enough to fit most conceivably directions that CMR applications may take in the future, at the cost of specificity. This also opens up the possibility of each research project using the framework differently, making it challenging to ensure that projects apply as intended.

6.3. Future Directions of Research

Roesner et al. (2014a), Sluganovic et al. (2017), and De Guzman et al. (2018) are examples that demonstrate how use cases need to be developed to consider security and safety issues in CMR devices. In particular, we see a need to exhaustively examine how CMR systems are likely to be used, develop secure protocols to support both usability as well as security and safety aspects of such systems, while also being able to develop techniques to detect compromises and nefarious activities will be key issues moving forward.

The CMR landscape today is likely to look very different 5 years from now. In preparation for this, we argue it is important to advance research by expanding our work in these key areas:

- **Empirical and experimental evidence**. A number of use cases have to be tested and executed in production environments to identify how well our model can reflect reality. We believe that our abstraction-by-design approach gives analysts and researchers the flexibility to develop new use cases. Use cases should be tested with concrete research questions and hypothesis testing. For instance, Marsh et al. (2006)'s idea of imbuing credentials: *can this be done legitimately or by the attacker*?
- **Development of further use cases of attacks**. Further use cases will enable other analysts and security researchers to start from templated examples of attacks to form ones that are more relevant for their own environments.
- **Extending the taxonomy**. We have already discussed that the taxonomy is not exhaustive, and that further use-cases will inform the value of the taxonomy and its ability to predict consequences. Formalisation of certain properties may also be a possibility. Other models can also be incorporated into the framework, and used in the taxonomy to describe behaviour and attack surface properties. Earlier, we described how attack graphs may be one approach to describe attacker behaviours. This does however mean that our taxonomy is incomplete, this is by design to future-proof the approach. Eventually, we aim to provide a detailed resource for the research community: a template of building blocks that researchers use to reason about, and model likely CMR threats based on known characteristics about the attack surface and attacker. For instance, if we know the CMR application uses avatars, we may then provide the building blocks to start reasoning about topics such as identify theft.
- **Metrics and function development**. Further work should investigate socio-technical aspects of CMR security that will be required to understand the relationships between taxonomy properties in order to accurately and precisely predict consequences, perhaps through simulation or post-incident analysis.
- **Attack detection tool development** IDS and IPS like tools can and should be deployed for CMRs.

7. CONCLUSION

In this paper, we have proposed and investigated grand challenges areas in CMR security, and discussed the degree to which potential harms may result from attacks. We have discussed mitigation strategies and IDS approaches to detect possible indicators of compromise, and how to begin to reduce the CMR attack surface. Use cases presented in this paper are not exhaustive and the taxonomy proposed is extendable by design. We only regard them as a starting point for future research. We assume that many attack vectors are not immediately obvious from our present-day understanding of CMR security. In preparation to tackle these challenges, we therefore deem it necessary to propose an extendable taxonomy which can be used as an abstraction-based reasoning approach to detect, mitigate, combat and deter CMR attacks. This work has been largely exploratory from desktop analysis, chiefly to provoke a discussion on the state of security in CMR today, or rather, lack thereof. We hope that results from this research will help establish a more fundamental, scientific foundation for security of CMR applications. CMR applications are likely to integrate with heterogeneous sensors. We anticipate that in the following decades there will be an convergence toward a *Smart Extended Reality* (SXR). What this means at this stage is currently uncertain, however this framework shows the need to consider how attacks can be executed and the harms that can come from them in order to develop better defense strategies.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

AS acknowledges the UK EPSRC-funded project Context Aware network architectures for Sending Multiple Senses (EP/P004016/1) for funding his work in this area.

REFERENCES

Agrafiotis, I., Bada, M., Cornish, P., Creese, S., Goldsmith, M., Ignatuschtschenko, E., et al. (2016). *Cyber Harm: Concepts, Taxonomy and Measurement*. Technical Report, Sad Business School WP 2016-23.

Alrawi, O., Lever, C., Antonakakis, M., and Monrose, F. (2019). "Sok: security evaluation of home-based iot deployments," in *SoK: Security Evaluation of Home-Based IoT Deployments* (San Francisco, CA: IEEE).

Axon, L., Alahmadi, B., Nurse, J. R., Goldsmith, M., and Creese, S. (2018). "Sonification in security operations centres: what do security practitioners think?" in *Workshop on Usable Security (USEC) at the Network and Distributed System Security (NDSS) Symposium 2018*.

Bada, M., and Sasse, A. (2014). *Cyber Security Awareness Campaigns: Why do They Fail to Change Behaviour?* Technical Report, University of Oxford.

Baldini, G., Botterman, M., Neisse, R., and Tallacchini, M. (2018). Ethical design in the internet of things. *Sci. Eng. Ethics* 24, 905–925. doi: 10.1007/s11948-016-9754-5

Balicer, R. D. (2007). Modeling infectious diseases dissemination through online role-playing games. *Epidemiology* 18, 260–261. doi: 10.1097/01.ede.0000254692.80550.60

Bastug, E., Bennis, M., Médard, M., and Debbah, M. (2017). Toward interconnected virtual reality: opportunities, challenges, and enablers. *IEEE Commun. Mag.* 55, 110–117. doi: 10.1109/MCOM.2017.1601089

Beaton, A., Cook, M., Kavanagh, M., and Herrington, C. (2000). The psychological impact of burglary. *Psychol. Crime Law* 6, 33–43. doi: 10.1080/10683160008410830

Benford, S., Bowers, J., Fahlén, L. E., Greenhalgh, C., and Snowdon, D. (1995). "User embodiment in collaborative virtual environments," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, CO: ACM SIGCHI), 242–249.

Benford, S., Brown, C., Reynard, G., and Greenhalgh, C. (1996). "Shared spaces: transportation, artificiality, and spatiality," in *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work* (ACM), 77–86.

Billinghurst, M., and Kato, H. (1999). "Collaborative mixed reality," in *Proceedings of the First International Symposium on Mixed Reality (ISMR 99). Mixed Reality Merging Real and Virtual Worlds* (Verlag: Springer), 261–284.

Bono, S., Caselden, D., Landau, G., and Miller, C. (2009). Reducing the attack surface in massively multiplayer online role-playing games. *iEEE Secur. Priv.* 7, 13–19. doi: 10.1109/MSP.2009.75

Bullock, A., and Benford, S. (1999). "An access control framework for multi-user collaborative environments," in *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work* (ACM), 140–149.

Cherdantseva, Y., and Hilton, J. (2013). "A reference model of information assurance & security," in *2013 International Conference on Availability, Reliability and Security* (Regensburg, Bavaria: IEEE), 546–555.

Chesney, T., Coyne, I., Logan, B., and Madden, N. (2009). Griefing in virtual worlds: causes, casualties and coping strategies. *Inf. Syst. J.* 19, 525–548. doi: 10.1111/j.1365-2575.2009.00330.x

Chismon, D., and Ruks, M. (2015). *Threat Intelligence: Collecting, Analysing, Evaluating*. MWR InfoSecurity Ltd.

Cohen, F. (1997a). Information system attacks: a preliminary classification scheme. *Comput. Secur.* 16, 29–46.

Cohen, F. (1997b). Information system defences: a preliminary classification scheme. *Comput. Secur.* 16, 94–114.

D'Antoni, L., Dunn, A. M., Jana, S., Kohno, T., Livshits, B., Molnar, D., et al. (2013). "Operating system support for augmented reality applications," in *HotOS*, Vol. 13 (Santa Ana Pueblo, NM), 21–21.

De Guzman, J. A., Thilakarathna, K., and Seneviratne, A. (2018). Security and privacy approaches in mixed reality: a literature survey. *arXiv preprint arXiv:1802.05797*. Available online at: https://arxiv.org/abs/1802.05797

DuPont, Q. (2017). "Experiments in algorithmic governance: a history and ethnography of "the dao," a failed decentralized autonomous organization," in *Bitcoin and Beyond*, 157–177.

Ellis, S. R. (1989). Nature and origins of virtual environments: a bibliographical essay. *Comput. Syst. Eng.* 2, 321–347. doi: 10.1016/0956-0521(91)90001-L

Ferwerda, J. A. (2003). "Three varieties of realism in computer graphics," in *Human Vision and Electronic Imaging VIII*, Vol. 5007, (International Society for Optics and Photonics), 290–298.

Freeman, D. (2008). Studying and treating schizophrenia using virtual reality: a new paradigm. *Schizophr. Bull.* 34, 605–610. doi: 10.1093/schbul/sbn020

Fujita, H., Hakura, J., and Kurematsu, M. (2010). "Virtual doctor system (vds): Framework on reasoning issues," in *Proceedings of the 2010 Conference on New Trends in Software Methodologies, Tools and Techniques* (Yokohama: IOS Press), 481–489.

Giubilo, F., Sajjad, A., Shackleton, M., Chadwick, D. W., Fan, W., and de Lemos, R. (2017). "An architecture for privacy-preserving sharing of cti with 3rd party analysis services," in *International Conference for Internet Technology and Secured Transactions (ICITST)* (Cambridge).

Glencross, M., Jay, C., Feasel, J., Luv, K., Mary, W., and Hubbold, R. (2007). "Effective cooperative haptic interaction over the internet," in *In Proceedings of IEEE Virtual Reality* (Charlotte, NC: IEEE), 115–122.

Gomes De Sá, A., and Gabriel, Z. (1999). Virtual reality as a tool for verification of assembly and maintenance processes. *Comput. Graph.* 23, 389–403.

Greenhalgh, C., Purbrick, J., and Snowdon, D. (2000). "Inside massive-3: flexible support for data consistency and world structuring," in *Proceedings of the Third International Conference on Collaborative Virtual Environments* (San Francisco, CA: Association for Computing Machinery), 119–127.

Gregg, L., and Tarrier, N. (2007). Virtual reality in mental health. *Soc. Psychiatry Psychiatr. Epidemiol.* 42, 343–354. doi: 10.1007/s00127-007-0173-4

Han, J., Chung, A. J., Sinha, M. K., Harishankar, M., Pan, S., Noh, H. Y., et al. (2018). "Do you feel what i hear? enabling autonomous iot device pairing using different sensor types," in *Do You Feel What I Hear? Enabling Autonomous IoT Device Pairing using Different Sensor Types* (San Francisco, CA: IEEE).

Happa, J. (2017). "PROTECTIVE: a European-wide NREN cyber threat intelligence sharing platform–lessons learnt to date," in *OASIS/FIRST Borderless Cyber Conference and Technical Symposium* (Prague).

Happa, J., Nurse, J. R. C., Goldsmith, M., Creese, S., and Williams, R. (2018). "An ethics framework for research into heterogeneous systems," in *Living in the Internet of Things: Cybersecurity of the IoT-2018* (London), 1–8.

Hawkins, W. M., Tversky, O. J., Robins, N., and Hester, S. K. (1999). *Networked Computer Game System With Persistent Playing Objects*. U.S. Patent 6,009,458.

Hinduja, S., and Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. *Arch. Suicide Res.* 14, 206–221. doi: 10.1080/13811118.2010.494133

Hu, S.-Y., Chen, J.-F., and Chen, T.-H. (2006). Von: a scalable peer-to-peer network for virtual environments. *IEEE Netw.* 20, 22–31. doi: 10.1109/MNET.2006.1668400

Hutchins, E. M., Cloppert, M. J., and Amin, R. M. (2011). "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," in *Leading Issues in Information Warfare & Security Research*, vol. 1, 80.

Jana, S., Molnar, D., Moshchuk, A., Dunn, A. M., Livshits, B., Wang, H. J., et al. (2013). "Enabling fine-grained permissions for augmented reality applications with recognizers," in *USENIX Security Symposium* (Washington, DC), 415–430.

Jeff Yan, J., and Choi, H.-J. (2002). Security issues in online games. *Electr. Libr.* 20, 125–133. doi: 10.1063/1.5033679

Kabil, A., Duval, T., Cuppens, N., Le Comte, G., Halgand, Y., and Ponchel, C. (2018). "Why should we use 3d collaborative virtual environments for cyber security?" in *IEEE Fourth VR International Workshop on Collaborative Virtual Environments (IEEEVR 2018)* (Reutlingen: IEEE), 1–8.

Kan, H., Duffy, V. G., and Su, C.-J. (2001). An internet virtual reality collaborative environment for effective product design. *Comput. Ind.* 45, 197–213. doi: 10.1016/S0166-3615(01)00093-8

Kartsounis, G., Magnenat-Thalmann, N., and Rodrian, H.-C. (2003). "E-tailor: Integration of 3d scanners, cad and virtual-try-on technologies for online retailing of made-to-measure garments," in *E-Business Applications* (Springer), 137–152.

Kim, J., and Forsythe, S. (2008). Adoption of virtual try on technology for online apparel shopping. *J. Interact. Market.* 22, 45–59. doi: 10.1002/dir.20113

Lawson, B. D. (2015). "Motion sickness symptomatology and origins," in *Handbook of Virtual Environments: Design, Implementation, and Applications*, eds K. S. Hale and K. M. Stanney (CRC Press), 531–600.

Lebeck, K., Ruth, K., Kohno, T., and Roesner, F. (2017). "Securing augmented reality output," in *Security and Privacy (SP), 2017 IEEE Symposium on* (St. Augustine, FL; IEEE), 320–337.

Lebeck, K., Ruth, K., Kohno, T., and Roesner, F. (2018). Arya: Operating system support for securely augmenting reality. *IEEE Secur. Priv.* 16, 44–53. doi: 10.1109/MSP.2018.1331020

Li, X., Jiang, P., Chen, T., Luo, X., and Wen, Q. (2017). A survey on the security of blockchain systems. *Future Generat. Comput. Syst.* doi: 10.1016/j.future.2017.08.020

Liao, H.-J., Lin, C.-H. R., Lin, Y.-C., and Tung, K.-Y. (2013). Intrusion detection system: a comprehensive review. *J. Netw. Comput. Appl.* 36, 16–24. doi: 10.1016/j.jnca.2012.09.004

Lofgren, E. T., and Fefferman, N. H. (2007). The untapped potential of virtual game worlds to shed light on real world epidemics. *Lancet Infect. Dis.* 7, 625–629. doi: 10.1016/S1473-3099(07)70212-8

Macedonia, M. R., Zyda, M. J., Pratt, D. R., Barham, P. T., and Zeswitz, S. (1994). Npsnet: a network software architecture for large scale virtual environments. *Presence* 3, 265–287. doi: 10.1162/pres.1994.3.4.265

Madary, M., and Metzinger, T. K. (2016). Real virtuality: a code of ethical conduct. recommendations for good scientific practice and the consumers of vr-technology. *Front. Rob. AI* 3:3. doi: 10.3389/frobt.2016.00003

Manadhata, P. K., and Wing, J. M. (2010). An attack surface metric. *IEEE Trans. Softw. Eng.* 37, 371–386. doi: 10.1109/TSE.2010.60

Marsh, J., Glencross, M., Pettifer, S., and Hubbold, R. (2006). A network architecture supporting consistent rich behavior in collaborative interactive applications. *IEEE Trans. Vis. Comput. Graph.* 12, 405–415. doi: 10.1109/TVCG.2006.40

McGraw, G., and CTO, C. (2008). *Exploiting Online Games: Cheating Massively Distributed Systems*. Addison-Wesley.

Mennecke, B., Roche, E., Bray, D., Konsynski, B., Lester, J., Rowe, M., et al. (2007). "Second life and other virtual worlds: a roadmap for research," in *Communications of the Association for Information Systems*.

Miessler, D. (2015). *Iot Attack Surface Mapping*. Presentation at DEFCON. (accessed Sept 29, 2018).

Milgram, P., and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Trans. Inform. Syst.* E77-D, 1321–1329.

Nair, S., Ganesan, A., Joshi, K. P., Oates, T., Joshi, A., and Finin, T. (2018). "Early detection of cybersecurity threats using collaborative cognition," in *Conference: 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)* (Philadelphia, PA).

Oliveira, M., Jordan, J., Pereira, J., Jorge, J., and Steed, A. (2009). Analysis domain model for shared virtual environments. *Int. J. Virtual Real.* 8, 1–30. Available online at: http://discovery.ucl.ac.uk/8266/

Persily, N. (2017). The 2016 us election: can democracy survive the internet? *J. Democracy* 28, 63–76. doi: 10.1353/jod.2017.0025

Pettifer, S., Cook, J., Marsh, J., and West, A. (2000). "Deva3: architecture for a large-scale distributed virtual reality system," in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (Seoul: Association for Computing Machinery), 33–40.

Pettifer, S., and Marsh, J. (2001). "A collaborative access model for shared virtual environments," in *Wetice* (Cambridge, MA: IEEE), 257.

Piskozub, M., Creese, S., and Happa, J. (2017). "Dynamic re-planning for cyber-physical situational awareness," in *CPS Conference on Computational Science and Computational Intelligence* (Las Vegas, NV), 1–6.

Rizzo, A. S., and Kim, G. J. (2005). A swot analysis of the field of virtual reality rehabilitation and therapy. *Presence* 14, 119–146. doi: 10.1162/1054746053967094

Roberts, L. D., Indermaur, D., and Spiranovic, C. (2013). Fear of cyber-identity theft and related fraudulent activity. *Psychiatry Psychol. Law* 20, 315–328. doi: 10.1080/13218719.2012.672275

Roesner, F., Kohno, T., and Molnar, D. (2014a). Security and privacy for augmented reality systems. *Commun. ACM* 57, 88–96. doi: 10.1145/2580723.2580730

Roesner, F., Molnar, D., Moshchuk, A., Kohno, T., and Wang, H. J. (2014b). "World-driven access control for continuous sensing," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (ACM), 1169–1181.

Roscoe, B., and Goldsmith, M. (1997). *The Perfect Spy for Model-Checking Crypto-Protocols*. Technical Report, University of Oxford.

Sallés, E. J., Michael, J. B., Capps, M., McGregor, D., and Kapolka, A. (2002). "Security of runtime extensible virtual environments," in *Proceedings of the 4th International Conference on collaborative Virtual Environments* (Bonn: ACM), 97–104.

Scarfone, K., and Mell, P. (2009). "An analysis of cvss version 2 vulnerability scoring," in *Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement* (IEEE Computer Society; Buena Vista), 516–525.

Sluganovic, I., Serbec, M., Derek, A., and Martinovic, I. (2017). "Holopair: Securing shared augmented reality using microsoft hololens," in *Proceedings of the 33rd Annual Computer Security Applications Conference* (San Juan: ACM), 250–261.

Smith, C. L. (2003). "Understanding concepts in the defence in depth strategy," in *Security Technology, 2003. Proceedings. IEEE 37th Annual 2003 International Carnahan Conference on* (Taipei: IEEE), 8–16.

Stanney, K. M., Mourant, R. R., and Kennedy, R. S. (1998). Human factors issues in virtual environments: a review of the literature. *Presence* 7, 327–351.

Steed, A., and Oliveira, M. F. (2009). *Networked Graphics: Building Networked Games and Virtual Environments*. Elsevier.

Steuer, J. (1992). Defining virtual reality: dimensions determining telepresence. *J. Commun.* 42, 73–93.

Stevens, V. (2006). Second life in education and language learning. *TESL-EJ* 10, 1–4. Available online at: http://www.cc.kyoto-su.ac.jp/information/tesl-ej/ej39/int

Szigeti, T., McMenamy, K., Saville, R., and Glowacki, A. (2009). *Cisco Telepresence Fundamentals*. Cisco Press.

Tandoc, E. C., Ferrucci, P., and Duffy, M. (2015). Facebook use, envy, and depression among college students. *Comput. Hum. Behav.* 43, 139–146. doi: 10.1016/j.chb.2014.10.053

Tolone, W., Ahn, G.-J., Pai, T., and Hong, S.-P. (2005). Access control in collaborative systems. *ACM Comput. Surv.* 37, 29–41. doi: 10.1145/1057977.1057979

van den Hoven, J. (2012). *Fact Sheet-Ethics Subgroup iot–Version 4.0*. Technical Report.

Vasilomanolakis, E., Karuppayah, S., Mühlhäuser, M., and Fischer, M. (2015). Taxonomy and survey of collaborative intrusion detection. *ACM Comput. Surv.* 47, 55. doi: 10.1145/2716260

Vilk, J., Molnar, D., Livshits, B., Ofek, E., Rossbach, C., Moshchuk, A., et al. (2015). "Surroundweb: Mitigating privacy concerns in a 3d web browser," in *2015 IEEE Symposium on Security and Privacy (SP)* (San Francisco, CA: IEEE), 431–446.

Wachtel, T. (2012). *10th Meeting of the Internet of Things Expert Group*. Technical Report.

Wagner, C., Dulaunoy, A., Wagener, G., and Iklody, A. (2016). "Misp: the design and implementation of a collaborative threat intelligence sharing platform," in *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security* (Vienna: ACM), 49–56.

Whitmore, A., Agarwal, A., and Da Xu, L. (2015). The internet of things a survey of topics and trends. *Inf. Syst. Front.* 17, 261–274. doi: 10.1007/s10796-014-9489-2

Wright, T., and Madey, G. (2010). Discretionary access controls for a collaborative virtual environment. *Int. J. Virtual Real.* 9, 61. Available online at: https://hal.archives-ouvertes.fr/hal-01530498/

Yan, J. (2003). "Security design in online games," in *Computer Security Applications Conference, 2003. Proceedings. 19th Annual* (Las Vegas, NV: IEEE), 286–295.

Yan, J., and Randell, B. (2005). "A systematic classification of cheating in online games," in *Proceedings of 4th ACM SIGCOMM Workshop on Network and System Support for Games* (Hawthorne, NY: ACM), 1–9.

Zhang, Z.-K., Cho, M. C. Y., Wang, C.-W., Hsu, C.-W., Chen, C.-K., and Shieh, S. (2014). "Iot security: ongoing challenges and research opportunities," in *Service-Oriented Computing and Applications (SOCA), 2014 IEEE 7th International Conference on* (Matsue: IEEE), 230–234.

Zhou, C. V., Leckie, C., and Karunasekera, S. (2010). A survey of coordinated attacks and collaborative intrusion detection. *Comput. Secur.* 29, 124–140. doi: 10.1016/j.cose.2009.06.008

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.