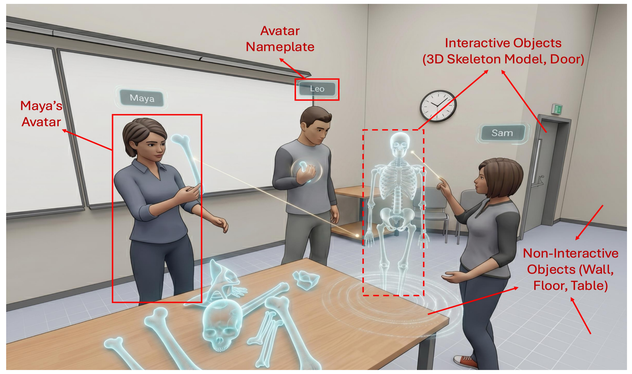
Relay and Betray: Exploiting Client-Side Authority in Multi-User Mixed Reality

Mutahar Ali, Habiba Farrukh

UC Irvine

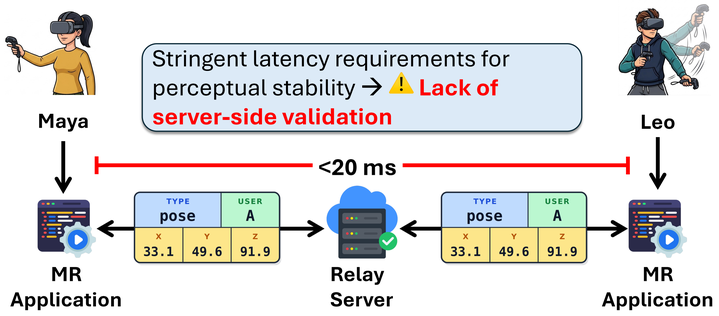
ARTIFACT EVALUATED USENIX AVAILABLE | ARTIFACT EVALUATED USENIX FUNCTIONAL | ARTIFACT EVALUATED USENIX REPRODUCED

Background

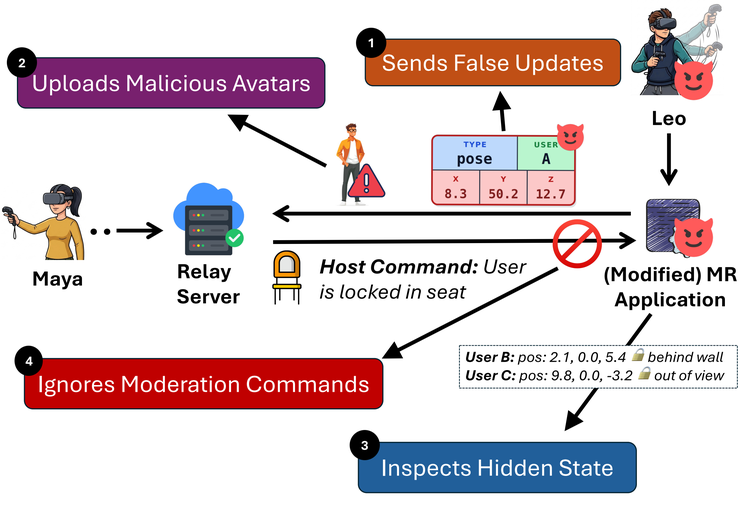


Multi-User Mixed Reality Apps

- Multi-user MR apps enable groups of users to experience the same 3D virtual space in real time.
- Users are represented by embodied avatars, and state (e.g., pose) is synchronized across users.
- Use cases include healthcare, education, training, design, social events, and gaming.



Threat Model



Methodology

Can a single malicious participant compromise the privacy, integrity, and safety of a shared MR session?

1 Data Collection

XR App Stores → Survey & filter → Multi-User MR Apps (APK) → Collect safety docs and guidelines → Safety Documentation

2 State Modeling

Recover Symbol Info

Recover symbols & metadata.
Outputs: classes, methods, variables, field names & offsets, metadata

Disassemble & Decompile

Disassemble native code and decompile to human readable form
Outputs: functions, call graph, xrefs, packet handlers, networking logic

Shared State Model

User	Object	World
• ID, avatar • score • token • safety bubble	• owner • transform • state • appearance	• access • host • other users • objects

3 Deriving Security Invariants

Invariant	Guidelines
Perceptual Secrecy	• Only see other users or objects when permitted by geometry and viewpoint. • A user should only hear others within a bounded spatial audio range.
State Integrity	• Only authorized users can update shared objects or world state. • Users can only take control of objects when permitted and must relinquish control when a valid transfer occurs. • Interactions must be proximity- and permission-bounded.
Safety Enforcement	• Blocking and safety features must suppress unwanted interaction. • Host controls (e.g., locking in seats, summoning users) must be enforced.
Availability	• A single user should not be able to degrade others' experience or render a session unusable through resource or state abuse.
Identity Integrity	• A user's visual and nominal identity should not be easily confused or cloned by another.

4 Identifying Invariant Violations

Trigger Security-Relevant Actions → Logging and Tracing → Analyzing Traffic → Security-Relevant Code → Runtime Instrumentation

Attacks

Surveillance
- Video surveillance
- Audio surveillance

Impersonation
- Avatar Cloning

Disruption
- Avatar Crashing
- Forced Disconnection

Tampering
- Object Manipulation

Safety Bypass
- Blocking Feature
- Host Controls

Video Demos of Attacks

Evaluation

Summary

- Analyzed 20 popular MR apps, including social, **healthcare**, **design**, **education**, and **gaming** apps.
- Attacks succeed across **all apps** on **all platforms**, except one app where attacks succeed on 2/4 platforms.

20 Apps

VRChat, Rec Room, Engage, Arkio, Gorilla Tag, VictoryXR, Spatial, Organon

150M+	6M+	280K+
installs	monthly active users	peak concurrent users

Use cases

4 Platforms

Meta, Android, SteamVR, VIVE

Video surveillance	20/20
Audio surveillance	20/20
Object manipulation	20/20
Block evasion	7/9 full, 2/9 partial
Host control evasion	10/10 apps with host controls
Avatar crashing	6/6 apps with fully custom avatars
Forcing disconnection	7/20 apps; rest are server-validated
Avatar cloning	6/6 apps with fully custom avatars

* **N<20** means the attack does not apply to some apps

Takeaways & Defenses

Takeaways

- Client-side trust creates a fundamental immersion-security tradeoff in multi-user MR.
- Apps rely on client authority for critical decisions, which leads to privacy, integrity, and safety risks.

Defenses

- Not all states are equally latency-sensitive; selective server-side validation for discrete high-impact actions
- State minimization and spatial filtering to avoid over-replicating sensitive perceptual data
- Client hardening; integrity checks and remote attestation to raise barrier for attacks

Findings responsibly disclosed to affected vendors; many acknowledged; some implemented hardening fixes.