Spatiotemporal Security in Mixed Reality Systems

Yasra Chandio
University of Massachusetts, Amherst
ychandio@umass.edu

Fatima M. Anwar
University of Massachusetts, Amherst
fanwar@umass.edu

ABSTRACT

This paper exhaustively explores the threat landscape of coordinated spatiotemporal attacks in mixed reality systems. Novel device-level and cross-device time translation and spatial shift attacks are launched, and their impact on deep learning based sensor fusion is evaluated. A major focus of this work is to establish stealthiness in the presence of sophisticated security mechanisms with an added constraint that mixed reality systems allow minimal time durations for covert operation. The efficacy of proposed attacks is evaluated through a preliminary study on inertial and visual data streams.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**.

ACM Reference Format:
Yasra Chandio and Fatima M. Anwar. 2020. Spatiotemporal Security in Mixed Reality Systems. In *The 18th ACM Conference on Embedded Networked Sensor Systems (SenSys'20), November 16–19, 2020, Virtual Event, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3384419.3430401

1 INTRODUCTION

The immersiveness of Mixed Reality (MR) applications relies on multi-modal sensing of user activities and surrounding environment via a range of commodity sensing devices such as Head Mounted Displays (HMD), hand-held controllers, and external cameras. One can envision future applications that use smartphones, wearable,and body area networks with existing MR platforms to deliver improved user experience. Widespread adoption of MR technology in critical applications such as surgery, medical therapies, and neurorehabilitation has made it a lucrative target for malicious activities. MR peripherals open door for novel and stealthy attacks, due to their multimodal attack surface and extensive reliance on spatiotemporal services both at the device and data level. Though prior research has focused on device authentication, authorization and data integrity [1, 4], we present attacks that maliciously induce temporal and spatial variations among data points on one device or across devices in a coordinated fashion leading to undesirable consequences.

Spatiotemporal attacks have various dimensions. `Data-level` temporal and spatial semantics can be attacked by manipulating sampling rate, frame rate, and pixel distribution. For example, MEMS inertial sensors in HMDs are susceptible to resonant acoustic interference that induces drifts in sampling rates to shift the inertial signal in time [5]. Similarly, an attacker with image processing capabilities can put together a sequence of frames to change the spatial landscape.`Device-level` manipulation of timing and location services via delaying timestamps and distance enlargement is also prevalent.

In contrast to other smart spaces, MR systems are more susceptible to spatiotemporal attacks, particularly relying on fusing heterogeneous sensing modalities in real-time to track human movements and gestures that are contingent upon device-level and data-level coordination at temporal and spatial scale.

In this work, we make a case for *coordinated spatiotemporal attacks* on MR systems. We demonstrate that while an attack on a single sensor can divert the user from its path, a coordinated attack on multiple sensors provides a more precise control over the attack and makes it hard for the system to detect abnormal trajectory. By manipulating the timing and spatial properties of devices and data, we show the adverse impact on critical applications. First we present `Device-level` temporal and spatial attacks and `cross-device` attacks. Then we demonstrate how these attacks are coordinated for stealthiness.

2 COORDINATED SPATIOTEMPORAL ATTACKS

We illustrate the effects of coordinated spatiotemporal attacks in Figure 1 on MR based surgical usecases. D represents a sensor producing data. Let's assume D_1 is a camera with a given frame rate and inter-frame correlation(IFC) metric. The solid yellow rectangle around consecutive frames f_1, f_2, and f_3 depicts (I) `device-level time translation attack` that changes not only the sampling rate but also the ordering of frames. This attack affects a surgeon performing pedicle screw surgery using HMD tracking to calculate a precise needle angle, and a few degrees of error will result in dire consequences. The green rectangle around frames fr_3 and fr_4 represents a (II) `device-level spatial shift attack` manipulating IFC at a fixed sampling frequency. For instance, this attack translates into misaligned spatial features of the images from Computed Tomography Angiography (CTA), and results in wrong renderings of patient's overlayed arteries, thus causing misplaced perforations. Red dotted rectangles around f_3, fr_3, g_3, and a_3 show data misalignment among four devices due to (III) `cross-device time translation attack`. This attack can be launched by maliciously drifting device clocks in the desired direction. As organs keep changing during surgical interactions, timely and tightly synchronized data-streams are crucial for hazard-free surgical procedure. Finally, the blue rectangle around fr_2 and g_2 depicts a (IV) `cross-device spatial shift attack`, where the attacker maliciously introduces relative rotations or spatial translation on selected samples from different sensing modalities. Any spatial changes across devices will impair surgeon of necessary information like position of instrument and/or blurred

SenSys '20, November 16–19, 2020, Virtual Event, Japan
© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-7590-0/20/11…$15.00
https://doi.org/10.1145/3384419.3430401

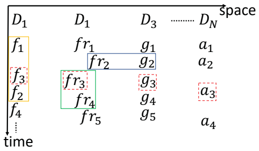

Figure 1: Illustration of device-level and cross-device time translation and spatial shift attacks.

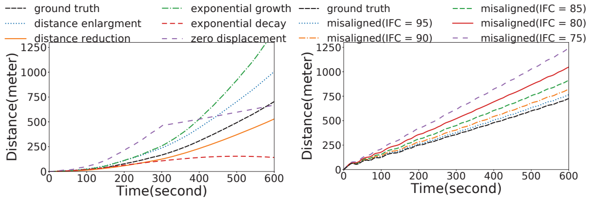

Figure 2: Device-level attacks: (a) Temporal attacks on inertial datastream. (b) Spatial attacks on visual stream.

virtual rendition of patient's organs. The goal of a stealthy attacker is to coordinate attacks to remain hidden and slowly diverge system from its true state. For instance, if a verifier keeps track of a fixed sampling rate to counter device-level attacks, attacker selectively drops CTA frames with coordinated time dilation of inertial data to give a false illusion of constant rate. An attacker can also insert partially forged frames and synchronize it with constricted time for covert operation. In addition, cross-device coordinated attacks misalign frames to hide variations in IFC due to frame forgery.

MR applications depend on multimodal data and are increasingly using deep learning based approaches for sensor fusion. These approaches are sensitive to the spatial and temporal properties of data, exposing them to spatiotemporal attacks. Our preliminary study [3] suggests that data misalignment from various modalities results in reduced model accuracy. Also, spatial and temporal misalignment distorts the fundamental characteristics of latent features from different sensing modalities. This presents new challenges for sensor fusion techniques and motivates further exploration on spatiotemporal attacks which is the focus of our work.

3 EVALUATION

Our preliminary results test the efficacy of spatiotemporal attacks on OxIOD dataset [2]. We evaluate a scenario where a user is tracked by two sensors on a HMD (inertial sensor and camera), while slowly moving away from its initial position. The user's position is tracked at every time step i.e. 10 millisecond via inertial and visual relative pose estimations. Ground truth distance traversed by the user is obtained separately from the inertial and visual positions at each time step. Finally, both streams are fused using Kalman filter to determine ground truth distance for sensor fusion. To detract the user from its intended path, we simulate temporal attacks on inertial data and spatial attacks on visual data. Figure 2a shows one kind of device-level time translation attack on inertial data. This attack modulates the sampling rate to achieve desired deviation from ground truth distance. Thus providing the attacker fine-grained control over distance enlargement and reduction at a constant or exponential rate. The attacker would also achieve zero-displacement in a chosen time interval to remain hidden from sanity checks of the verification system. However it is hard to avoid large deviations from the intended outcome through one sensing modality, thus jeopardizing the attacker's ability to be stealthy and cause more damage. In Figure 2b, we demonstrate spatial attack on visual data. This attack misaligns pixel distribution of consecutive frames such that IFC stays within a certain detection threshold. Lower IFC between successive frames results in quick and large deviations from the expected outcome that can alarm the verifier. Even 80% IFC significantly impacts the outcome in Figure **??**. We suggest that an attacker can exercise fine control over device parameters and craft a coordinated attack across various sensing modalities to cause lasting damage without getting detected.

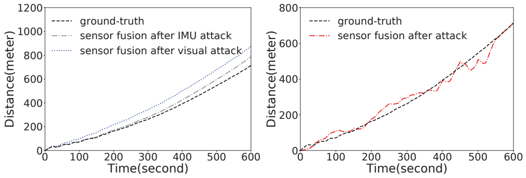

Figure 3: (a) Coordinated attack on either inertial or visual stream at a time. (b) Zero displacement coordinated attack.

In the previous experiments, we demonstrated how an individual sensor stream can be manipulated for an attack. In practice, data from more than one sensors is fused together to increase tracking accuracy by removing noise present in any single sensor's data. We first evaluate the efficacy of an attack on one sensor's stream to determine if it can divert the user trajectory or the attack is filtered as noise. Figure 3a shows the individual attacks, on either visual stream or inertial data, both affecting the fused stream. However, the ability to only manipulate a single sensor at a time restricts the attacker to simple attacks as other streams can be used to detect abnormal patterns of sensor under attack. To solve this problem, we launch coordinated attacks on both sensor's stream. The goal of the attack is to keep diverting the user from original path to arbitrary location such that at the end user should reach its original destination (zero displacement strategy). Figure 3b shows that the attacker is able to launch a precisely controlled attack on the fused data stream. It is hard to detect this attack as both sensors have reciprocated.

4 CONCLUSION AND FUTURE WORK

This work presents an unexplored yet a large attack surface on MR systems. We have shown that attacks across temporal and spatial scale, when coordinated, cause undetected damage. Our next steps would explore new attack parameters, experiment with distributed sensing modalities, and evaluate over real MR infrastructure including Hololens2 headset, external cameras, and wearables.

REFERENCES

[1] P. Casey, I. Baggili, and A. Yarramreddy. 2019. Immersive Virtual Reality Attacks and the Human Joystick. *IEEE TDSC* (2019).

[2] C. Chen, P. Zhao, C.X. Lu, W. Wang, A. Markham, and N. Trigoni. 2018. OxIOD: The Dataset for Deep Inertial Odometry. arXiv:1809.07491 [cs.RO]

[3] S. S. Sandha, J. Noor, F. M. Anwar, and M. Srivastava. 2020. Time Awareness in Deep Learning-Based Multimodal Fusion Across Smartphone Platforms. In *IEEE/ACM IoTDI*.

[4] I. Sluganovic, M. Serbec, A. Derek, and I. Martinovic. 2017. HoloPair: Securing Shared Augmented Reality Using Microsoft HoloLens. In *ACSAC*.

[5] Y. Tu, Z. Lin, I. Lee, and X. Hei. 2018. Injected and Delivered: Fabricating Implicit Control over Actuation Systems by Spoofing Inertial Sensors. In *USENIX Security*.