Virtual Reality And Augmented Reality Headsets With External Cameras Are All Spying On You, And Those Around You, For Data Harvesting Scumbags

The research shows that hackers could use popular virtual reality (AR/VR) headsets with built in motion sensors to record subtle, speech-associated facial dynamics to steal sensitive information communicated via voice-command, including credit card data and passwords.

Government spy agencies pay Facebook and Google big bucks to listen in on what you do using their resources. You can assume that everything with a camera is 'bugged' Common AR/VR systems on the market include the popular brands Oculus Quest 2, HTC Vive Pro, and PlayStation VR.

To demonstrate the existence of security vulnerabilities, the researchers developed an eavesdropping attack targeting AR/VR headsets, known as "Face-Mic."

"Face-Mic is the first work that infers private and sensitive information by leveraging the facial dynamics associated with live human speech while using face-mounted AR/VR devices," says research leader Yingying "Jennifer" Chen, associate director of WINLAB and graduate director of electrical and computer engineering at Rutgers University-New Brunswick.

"Our research demonstrates that Face-Mic can derive the headset wearer's sensitive information with four mainstream AR/VR headsets, including the most popular ones: Oculus Quest and HTC Vive Pro." The researchers studied three types of vibrations captured by AR/VR headsets' motion sensors, including speech-associated facial movements, bone-borne vibrations and airborne vibrations. Chen notes that bone-borne vibrations in particular are richly encoded with detailed gender, identity, and speech information.

"By analyzing the facial dynamics captured with the motion sensors, we found that both cardboard headsets and high-end headsets suffer security vulnerabilities, revealing a user's sensitive speech and speaker information without permission," Chen says.

Although vendors usually have policies regarding utilizing the voice access function in headset microphones, Chen's research found that built-in motion sensors, such as an accelerometer and gyroscope within a VR headset, do not require any permission to access. This security vulnerability can be exploited by malicious actors intent on committing eavesdropping attacks.

Eavesdropping attackers can also derive simple speech content, including digits and words, to infer sensitive information, such as credit card numbers, Social Security numbers, phone numbers, PIN numbers, transactions, birth dates, and passwords. Exposing such information could lead to identity theft, credit card fraud, and confidential and health care information leakage.

Chen says once a user has been identified by a hacker, an eavesdropping attack can lead to further exposure of user's sensitive information and lifestyle, such as AR/VR travel histories, game/video preferences, and shopping preferences. Such tracking compromises users' privacy and can be lucrative for advertising companies.

Oculus Quest, for example, supports voice dictation for entering web addresses, controlling the headset, and exploring commercial products. The Face-Mic research shows that hackers may leverage these zero-permission sensors to capture sensitive information, leading to severe privacy leakages. Chen says she hopes these findings will raise awareness in the general public about AR/VR security vulnerabilities and encourage manufacturers to develop safer models.

“Given our findings, manufacturers of VR headsets should consider additional security measures, such as adding ductile materials in the foam replacement cover and the headband, which may attenuate the speech-associated facial vibrations that would be captured by the built-in accelerometer/gyroscope,” she says.

Chen and her colleagues are now examining how facial vibration information can authenticate users and improve security, and how AR/VR headsets can capture a user’s breathing and heart rate to measure well-being and mood states unobtrusively. To learn more about ongoing studies at WINLAB, click here.

The researchers will present their study at the annual International Conference on Mobile Computing and Networking in March. Additional collaborators are from Texas A&M University and University of Tennessee at Knoxville.The research shows that hackers could use popular virtual reality (AR/VR) headsets with built in motion sensors to record subtle, speech-associated facial dynamics to steal sensitive information communicated via voice-command, including credit card data and passwords. “…VR headsets should consider additional security measures…”

Common AR/VR systems on the market include the popular brands Oculus Quest 2, HTC Vive Pro, and PlayStation VR.

To demonstrate the existence of security vulnerabilities, the researchers developed an eavesdropping attack targeting AR/VR headsets, known as “Face-Mic.”

“Face-Mic is the first work that infers private and sensitive information by leveraging the facial dynamics associated with live human speech while using face-mounted AR/VR devices,” says research leader Yingying “Jennifer” Chen, associate director of WINLAB and graduate director of electrical and computer engineering at Rutgers University-New Brunswick.

“Our research demonstrates that Face-Mic can derive the headset wearer’s sensitive information with four mainstream AR/VR headsets, including the most popular ones: Oculus Quest and HTC Vive Pro.” The researchers studied three types of vibrations captured by AR/VR headsets’ motion sensors, including speech-associated facial movements, bone-borne vibrations and airborne vibrations. Chen notes that bone-borne vibrations in particular are richly encoded with detailed gender, identity, and speech information.

“By analyzing the facial dynamics captured with the motion sensors, we found that both cardboard headsets and high-end headsets suffer security vulnerabilities, revealing a user’s sensitive speech and speaker information without permission,” Chen says.

Although vendors usually have policies regarding utilizing the voice access function in headset microphones, Chen’s research found that built-in motion sensors, such as an accelerometer and gyroscope within a VR headset, do not require any permission to access. This security vulnerability can be exploited by malicious actors intent on committing eavesdropping attacks.

Eavesdropping attackers can also derive simple speech content, including digits and words, to infer sensitive information, such as credit card numbers, Social Security numbers, phone numbers, PIN numbers, transactions, birth dates, and passwords. Exposing such information could lead to identity theft, credit card fraud, and confidential and health care information leakage.

Chen says once a user has been identified by a hacker, an eavesdropping attack can lead to further exposure of user's sensitive information and lifestyle, such as AR/VR travel histories, game/video preferences, and shopping preferences. Such tracking compromises users' privacy and can be lucrative for advertising companies.

Oculus Quest, for example, supports voice dictation for entering web addresses, controlling the headset, and exploring commercial products. The Face-Mic research shows that hackers may leverage these zero-permission sensors to capture sensitive information, leading to severe privacy leakages. Chen says she hopes these findings will raise awareness in the general public about AR/VR security vulnerabilities and encourage manufacturers to develop safer models.

"Given our findings, manufacturers of VR headsets should consider additional security measures, such as adding ductile materials in the foam replacement cover and the headband, which may attenuate the speech-associated facial vibrations that would be captured by the built-in accelerometer/gyroscope," she says.

Chen and her colleagues are now examining how facial vibration information can authenticate users and improve security, and how AR/VR headsets can capture a user's breathing and heart rate to measure well-being and mood states unobtrusively. To learn more about ongoing studies at WINLAB, click here.

The researchers will present their study at the annual International Conference on Mobile Computing and Networking in March. Additional collaborators are from Texas A&M University and University of Tennessee at Knoxville.

Source: Rutgers University

But how worried should we be? To find out, we got in touch with UC-Davis researcher Oliver Kreylos, who discovered how to pull images from Oculus Rift sensors. He got back to us with very detailed answers not just about this particular hack, but VR tracking sensors in general. We found them to be very interesting, so we're sharing the email interview below in full.

Motherboard: Compared to the Vive's Lighthouse sensors, is the Oculus' method of position tracking comparatively sloppy, or would you say it has its benefits? As in, is it even worth it for Oculus to be taking these kinds of risks in its product?

Oliver Kreylos: This is a bit of a loaded question, so I'll have to establish some context to give it a fair answer. Cameras and tracking markers (such as Oculus' LEDs) are a long-established approach to 3-DOF (position only) and 6-DOF (position and orientation) tracking. [DOF stands for Degrees of Freedom, a system that mimics the visual sensation of looking around your body.] Almost all high-end motion capture systems are based on it. It was the foundation of the Wiimote's tracking, and it's been used for PC games in the form of NaturalPoint's TrackIR head tracker for many years.

The reason this method is so popular, especially among the hobbyist crowd, is that it delivers high quality results while only requiring little custom hardware. Cameras are ubiquitous and cheap, and the only other hardware component are rigid arrangements of tracking markers, which even hobbyists can build (see this old Wiimote project of mine). Everything else is done in software.

When Oculus needed a good and cheap 6-DOF tracking system for the Rift DK2 [Development Kit 2], cameras were the obvious and most appropriate choice. But instead of delivering a standard implementation, Oculus went the extra mile and delivered an ingenious variation which significantly reduced the computational load of the standard algorithm by designing tracking LEDs that could identify themselves to the camera (see here for all the gory details). As a result, the DK2 tracking system exceeded many people's expectations, including my own.

When Oculus moved from DK2 to CV1 [Consumer Version 1, or the first version used by buyers after the official launch], keeping the existing proven tracking system, and further improving it by employing better hardware (higher-resolution camera, global instead of rolling shutter), was a sound engineering decision. From a purely technical perspective, it was the correct decision. While there are some problems, Oculus' Constellation is an excellent 6-DOF tracking system. Due to Constellation's comparative closedness, I have not analyzed it as thoroughly as Valve's Lighthouse system, but I am expecting them to be about on par quality-wise.

That said, I rate Lighthouse as the more elegant of the two systems, in the sense that it collects less raw data, and has to do less processing, to deliver the same end result. The basic input for both systems' position calculation algorithm are three-dimensional rays starting at a central point— the camera focal point or the Lighthouse base station center—and pointing towards tracking LEDs or photodiodes in space. Constellation calculates these rays by capturing high-res images, streaming them to the host PC, finding blobs of bright pixels in those images, and calculating the (x, y) positions of their centers. Lighthouse does the same by timing when a sweeping laser hits a photodiode [a device that converts light into an electric current], converting times into angles based on the lasers' known angular velocity, and sending the resulting angles to the host PC. Constellation needs to send around 60 MBs of data per camera to the host, which puts severe stress on the host's USB subsystem, whereas Lighthouse sends so little data—I estimate tens of KBs—that it can do it wirelessly without causing issues.

But that does not mean Constellation is sloppy. It was a top-notch state-of-the-art system when it was developed, but Valve unveiled an unexpected and more elegant system at a time when it was too late for Oculus to change course. It was an ingenious bit of innovation that I don't believe anybody saw coming.

The main downside of Constellation is that it causes issues for some users due to its high USB bandwidth demands. The other, non-technical, downside is that it sends high-resolution images from several cameras to the host PC, and that those cameras, by necessity, have to be placed in almost ideal surveillance positions. I personally believe that the risk of some attacker gaining access to these images is minimal, but I cannot deny that it is theoretically possible.

This is the one aspect where Oculus could have done differently to avoid this issue entirely. I mentioned the Wiimote above, and that it also uses a camera (on the Wiimote) for tracking. But unlike Constellation, the Wiimote does not send images to the console for processing. The step that converts images to (x, y) LED positions is done inside the camera chip itself, by a custom piece of silicon. If Oculus had followed this approach and integrated such ASICs [application-specific integrated circuit, or a custom-designed circuit with a specific use] into the cameras themselves, they could have avoided

both of Constellation's issues. It would have reduced bandwidth from the cameras to the host by a factor of about 1000 (and fixed most users' issues), and it would have made it impossible to snoop images, because the images would never have been sent to the host PC in the first place.

I do not know whether this was "sloppy," in the sense that Oculus engineers overlooked the privacy concerns users might have, or if they weighed the risks and benefits and made an informed decision. There are benefits of sending images to the host: designing an ASIC takes time and money, software solutions are more flexible and easier to improve over time, and it is possible that Oculus are currently working on algorithms to use the camera images to track more than just LEDs, for example to bring user's hands or even full bodies into VR à la LeapMotion or Kinect.

As an aside, Oculus have consistently been referring to the CV1 cameras as "sensors" instead of cameras, and have insisted that they do not work like cameras. Initially, that made me think that they had indeed integrated image processing ASICs into the cameras, but that turned out to be false.

Do you know if it's possible to capture "images" with the Vive's sensors even without traditional cameras?

Lighthouse itself does not collect any data that would make it possible to reconstruct an image of the user's environment, or of the user itself. The only data sent from Lighthouse trackables to the host PC are timestamps at which individual photodiodes on the trackable are hit by sweeping lasers, and samples from the trackable's integrated inertial measurement unit, which together allow establishing the trackable's position and orientation in 3D space. Theoretically, by observing the position of all trackables over a long time period, one could reconstruct a rough 3D model of the user's environment (i.e., where walls or major obstacles are), but that is about it.

That said, the Vive headset does have a front-facing camera that is a standard webcam, and is advertised to the host PC's operating system as such. This camera is therefore exactly as vulnerable to exploits as any other webcam connected to a computer. Unlike Constellation's cameras, however, this one is not required for the overall VR system to function, and users concerned about privacy could disable it without negative consequences simply by covering it with a piece of tape.

Would you say the complications of the process to get a recognizable image (and the comparative limitations of the hardware's availability) is enough to dissuade the hacking of the sensors on a large scale?

I cannot make a judgment as I'm not a computer security expert. I know that there are real cases of remote attackers gaining access to webcams, but I do not know how those attacks were done, and whether those methods would work with the Constellation cameras. The approach I followed was simple, but would probably not work for a remote attacker as it required patching the Linux kernel's webcam driver to recognize the Constellation camera.

One thing I can say is that "standard" webcam attacks that might be out in the wild and be widely deployed will not succeed with the Constellation cameras, as they do not advertise themselves to the OS as standard webcams. If a generic webcam exploit were to run on an Oculus user's computer, it would not find them. That said, due to the Constellation cameras actually *being* standard webcams under the hood, exploits could be modified to target them nonetheless, but I do not know whether this would require small or large modifications.

Another complication for a would-be attacker is that the Constellation cameras are used by Oculus' run-time software while the headset is active. Existing webcam exploits might not be able to take over if another process is already using the cameras, or might shut down tracking in doing so, which would alert the user to shenanigans. In addition, each Constellation camera has an activity light on it. I do not know whether those are tied into the camera sensor's operation at an electric circuit level, but I do know that my camera's light turned on when I started recording images from it with my software.

While Augmented Reality (AR) and Virtual Reality (VR) are envisioned as the next iteration of the internet immersing us in new digital worlds, the associated headset hardware and virtual keyboard interfaces create new opportunities for hackers.

Such are the findings of computer scientists at the University of California, Riverside, which are detailed in two papers to be presented this week at the annual Usenix Security Symposium in Anaheim, a leading international conference on cyber security.

The emerging metaverse technology, now under intensive development by Facebook's Mark Zuckerberg and other tech titans, relies on headsets that interpret our bodily motions — reaches, nods, steps, and blinks — to navigate new worlds of AR and VR to play games, socialize, meet co-workers, and perhaps shop or conduct other forms of business.

A computer science team at UCR's Bourns College of Engineering led by professors Jiasi Chen and Nael Abu-Ghazaleh, however, has demonstrated that spyware can watch and record our every motion and then use artificial intelligence to translate those movements into words with 90 percent or better accuracy.

"Basically, we show that if you run multiple applications, and one of them is malicious, it can spy on the other applications," Abu-Ghazaleh said. "It can spy on the environment around you, for example showing people are around you and how far they are. And it can also expose to the attacker your interactions with the headset."

For instance, if you take a break from a virtual game to check your Facebook messages by air typing the password on a virtual keyboard generated by the headset, the spyware could capture your password. Similarly, spies could potentially interpret your body movements to gain access to your actions during a virtual meeting in which confidential information is disclosed and discussed.

The two papers to be presented at the cybersecurity conference are co-authored Abu-Ghazaleh and Chen toether with Yicheng Zhang, a UCR computer science doctoral student, and Carter Slocum, a visiting Assistant Professor at Harvey Mudd College who earned his docorate at UCR.

The first paper is titled "It's all in your head(set): Side-channel attacks on AR/VR systems." With Zhang as the lead author, it details how hackers can recover a victim's hand gestures, voice commands, and keystrokes on a virtual keyboard, with accuracy exceeding 90 percent. The paper further shows how spies can identify applications as they are launched and perceives other people standing near the headset user with a distance accuracy of about 4 inches (10.3 cm).

The second paper, "Going through the motions: AR/VR keylogging from user head motions," digs deeper into the security risk of using a virtual keyboard. With Slocum as lead author, it shows how subtle head movements made by users typing on virtual keyboards are sufficient for spies to infer the text that is being typed. The researchers then developed a system, dubbed TyPose, that uses machine

learning to extract these head motion signals to automatically infer words or characters that a user is typing.

Both papers are expected to inform the tech industry of their cyber security weaknesses.

“We demonstrate feasibilities of attacks, and then we do responsible disclosure,” Abu-Ghazaleh said. “We tell the companies that, hey, this is what we were able to do. And then we give them time to see if they want to fix it before we publish our findings.”