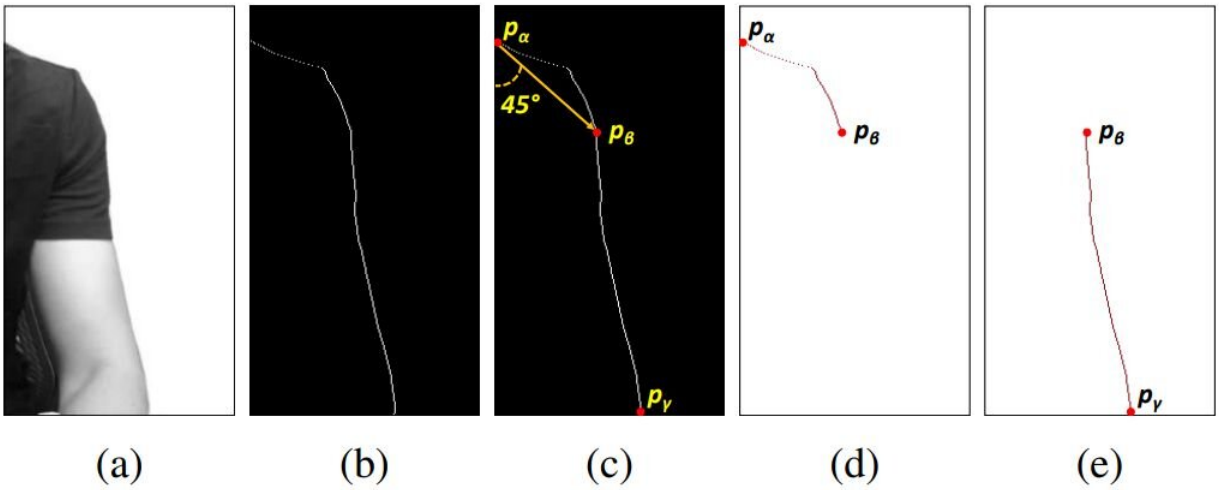
Zoom hack reveals text contents by viewing shoulder movement

November 3 2020, by Peter Grad



(a) A keystroke frame segment, (b) Outer contour (OC), (c) 45° projection from pα that intersects OC at p_β, (d) Shoulder contour (SC), and (e) Arm contour (AC). Credit: arXiv:2010.12078 [cs.CR]

Countless companies around the world see the wildly popular video-conferencing tool Zoom as a savior during this era of quarantine and work-at-home routines. Zoom estimates it has more than 300 million meeting participants daily.

But progress in the digital world is often accompanied by problems. Some are minor, such as participants complaining about their

unflattering appearance due to poor lighting or high-resolution images detailing zits; unplanned appearances by infants, pets or parents; and distracting noises and sounds from one's household. More embarrassing are those comments made by participants mistakenly believing their microphones were off.

More serious issues arose as "Zoom-bombing" troublemakers have sneaked into meetings and uttered rude comments or displayed inappropriate images. Malicious animated GIFs placed in Zoom chats carried harmful code into meetings. Worse—there's no official record for how many times this has happened—one hapless well-known American network TV commentator last month ended his Zoom conference call by touching himself in places one doesn't normally display in public—before realizing he had not shut his camera off.

This week, a new problem has been added to the list. Researchers at the University of Texas discovered they could determine what Zoom participants are typing in private side chats during Zoom meetings.

Murtuza Jadiwala, a computer science professor heading the research project, said his team was able to identify the contents of texts by examining body movement of the participants. Specifically, they focused on the movement of their shoulders and arms to extrapolate the actions of their fingers as they typed.

Given the widespread use of high-resolution web cams during conference calls, Jadiwala was able to record and analyze slight pixel shifts around users' shoulders to determine if they were moving left or right, forward or backward. He then created a software program that linked the movements to a list of commonly used words. He says the "text inference framework that uses the keystrokes detected from the video … predict[s] words that were most likely typed by the target user. We then comprehensively evaluate[d] both the keystroke/typing

detection and text inference frameworks using data collected from a large number of participants.

In a controlled setting, with specific chairs, keyboards and webcam, Jadiwala said he achieved an accuracy rate of 75 percent. However, in uncontrolled environments, accuracy dropped to only one out of every five words being correctly identified.

Other factors contribute to lower accuracy levels, he said, including whether long sleeve or short sleeve shirts were worn, and the length of a user's hair. With long hair obstructing a clear view of the shoulders, accuracy plummeted.

He noted that a user's typing style also can affect the results.
"The joint movements associated with keystrokes following the initial keystroke depends primarily on the user's typing style, e.g., hunt-and-peck, touch-typing, or hybrid. Certain typing styles, such as hunt-and-peck, result in significant upper hand movements (not just fingers or wrist) between keystrokes, than others," making it easier to discern the contents of texts.

He suggested that blurring of body or shoulder contours in Zoom videos could disrupt the ability of hackers to determine the contents of messages.

Snooping by examining one's posture joins a long list of eavesdropping techniques digital technology has brought us in recent years. Snooping on smartphones' accelerometer and gyroscope readings can reveal PIN codes entered for credit card purchases. Israeli researchers were able to recreate speech and music sounds by using a telescope to scan a lightbulb in a room where a meeting was being held; barely perceptible bulb vibrations caused by the sounds were analyzed and interpreted with a stunning degree of accuracy. Years earlier, MIT, Microsoft, and Adobe

achieved similar results by examining micro-vibrations from a bag of potato chips. And decades earlier the Soviet Union tapped into infrared waves bouncing off windows to eavesdrop on conversations.

For now, users are cautioned to use a strong password system for meeting participation, avoid side chats and lock the door to your room so Fido doesn't hop onto your lap mid-meeting.

And keep your pants on.

More information: Zoom on the Keystrokes: Exploiting Video Calls for Keystroke Inference Attacks, arXiv:2010.12078 [cs.CR] arxiv.org/abs/2010.12078