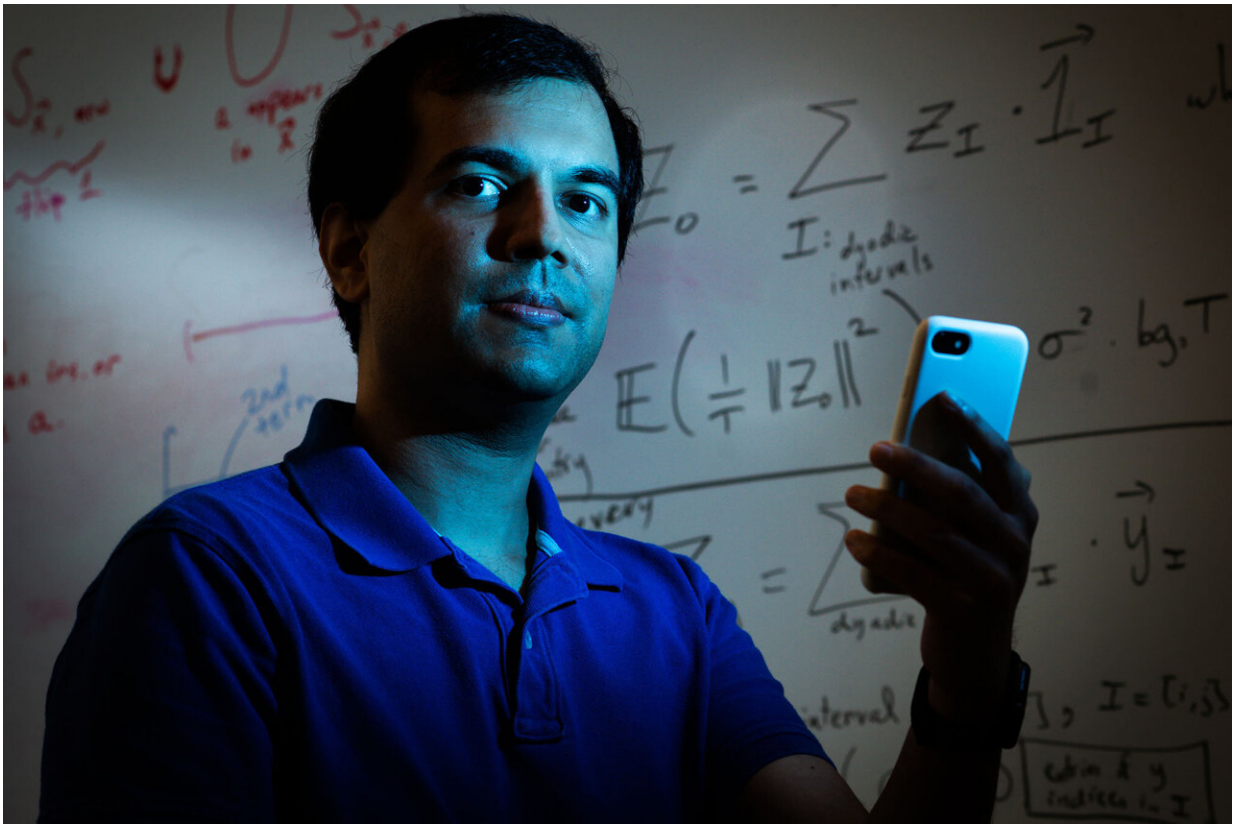
Can WhatsApp messages be secure and encrypted—but traceable at same time?

August 10 2022, by Andrew Thurston



Cryptographer and computer scientist Mayank Varia, a Boston University Faculty of Computing & Data Sciences associate professor, says his work can help balance privacy and human rights with online trust and safety. Credit: Jackie Ricciardi, Boston University

Cryptographers love an enigma, a problem to solve—and this one has it

all. Indestructible codes, secret notes, encryption and decryption.

Here's the puzzle: Someone wants to send a secure message online. It has to be so private, so secret, that they can deny they ever sent it. If someone leaks the message, it can never be traced back to the sender. It's all very Mission: Impossible. But there's a kicker: if that message peddles abuse or misinformation, maybe threatens violence, then anonymity may need to go out the window—the sender needs to be held to account.

And that's the challenge: is there a way to allow people to send confidential, secure, untraceable messages, but still track any menacing ones?

Mayank Varia might have cracked the conundrum. A cryptographer and computer scientist, Varia is an expert on the societal impact of algorithms and programs, developing systems that balance privacy and security with transparency and social justice. Working with a team of Boston University computer scientists, he's designed a program called Hecate—fittingly named after the ancient Greek goddess of magic and spells—that can be bolted onto a secure messaging app to beef up its confidentiality, while also allowing moderators to crack down on abuse. The team is presenting its findings at the 31st USENIX Security Symposium.

"Our goal in cryptography is to build tools and systems that allow people to get things done safely in the digital world," says Varia, a BU Faculty of Computing & Data Sciences associate professor. "The question at play in our paper is what is the most effective way to build a mechanism for reporting abuse—the fastest, most efficient way to provide the strongest security guarantees and provide the weakest possible puncturing of that?"

It's an approach he's also applying beyond messaging apps, building online tools that allow local governments to track gender wage gaps—without accessing private salary data—and enable sexual assault victims to more safely report their attackers.

Everything is deniable

When two people chat in a private room, what they talk about is just between them—there's no paper trail, no recording; the conversation lives on in memory alone. Put the same conversation online—Twitter, Facebook, email—and it's a different story. Every word is preserved for history. Sometimes that's good, but just as often it's not. An activist in an authoritarian state trying to get word to a journalist or a patient seeking help for a private health issue might not want their words broadcast to the world or held in an archive.

That's where end-to-end encryption comes in. Popularized by apps like WhatsApp and Signal, it scrambles sent messages into an unreadable format, only decrypting them when they land on the recipient's phone. It also ensures messages sent from one person to another can't be traced back to the sender; just like that private in-person chat, it's a conversation without a trail or record—everything is deniable.

"The goal of these deniable messaging systems is that even if my phone is compromised after we've had an encrypted messaging conversation, there are no digital breadcrumbs that will allow an external person to know for sure what we sent or even who said it," says Varia.

Amnesty International calls encryption a human right, arguing it's "an essential protection of [everyone's] rights to privacy and free speech," and especially vital for those countering corruption or challenging governments. Like much in the online world though, that privacy can be exploited or bent to more sinister ends. "There are specific times where

this can be a bad thing," says Varia. "Suppose the messages someone is sending are harassing and abusive and you want to go seek help, you want to be able to prove to the moderator what the message contents were and who said them to you."

A study of elementary, middle, and high school students in Israel, where more than 97 percent of kids reportedly use WhatsApp, found 30 percent had been bullied on the app, while UK prosecutors have said end-to-end encryption could harm their ability to catch and stop child abusers. Extremist groups, from Islamic State to domestic terrorists, have leaned on encrypted apps like Telegram and Signal to spread their calls for violence.

The task for tech companies is finding a way to support the right to privacy with the need for accountability. Hecate offers a way to do both—it allows app users to deny they ever sent a message, but to also be reported if they say something abusive.

A message in invisible ink

Developed by Varia and doctoral students Rawane Issa and Nicolas Alhaddad, Hecate starts with the accountability side of that contradictory deniable and traceable combination. Using the program, an app's moderator creates a unique batch of electronic signatures—or tokens—for each user. When that user sends a message, a hidden token goes along for the ride. If the recipient decides to report that message, the moderator will be able to verify the sender's token and take action. It's called asymmetric message franking.

The fail-safe, says Varia, the part that allows for deniability, is that the token is only useful to the moderator.

"The token is an encrypted statement that only the moderator knows how

to read—it's like they wrote a message in invisible ink to their future self," says Varia. "The moderator is the one who builds these tokens. That's the nifty part about our system: even if the moderator goes rogue, they can't show and convince the rest of the world—they have no digital proof, no breadcrumbs they can show to anyone else."

The user can maintain deniability—at least publicly.

Similar message franking systems already exist—Facebook parent Meta uses one on WhatsApp—but Varia says Hecate is faster, more secure, and futureproof in a way current programs are not.

"Hecate is the first message franking scheme that simultaneously achieves fast execution on a phone and for the moderator server, support for message forwarding, and compatibility with anonymous communication networks like Signal's sealed sender," says Varia. "Previous constructions achieved at most two of these three objectives."

The civic impact of algorithms

The team says Hecate could be ready for implementation on apps like Signal and WhatsApp with just a few months of custom development and testing. But despite its technological advantages, Varia suggests companies approach Hecate with caution until they've fully investigated its potential societal impact.

"There's a question of can we build this, there's also a question of should we build this?" says Varia. "We can try to design these tools that provide safety benefits, but there might be longer dialogues and discussions with affected communities. Are we achieving the right notion of security for, say, the journalist, the dissident, the people being harassed online?"

As head of CDS' Hub for Civic Tech Impact, Varia is used to

considering the societal and policy implications of his research. The hub's aim is to develop software and algorithms that advance public interest, whether they help to fight misinformation or foster increased government transparency. A theme through recent projects is the creation of programs that, like Hecate, straddle the line between privacy and accountability.

During a recent partnership with the Boston Women's Workforce Council, for example, BU computer scientists built a gender wage gap calculator that enables companies to share salaries with the city without letting sensitive pay data leave their servers.

"We're designing tools that allow people—it sounds counterintuitive—to compute data that they cannot see," says Varia, who's a member of the federal government's Advisory Committee on Data for Evidence Building. "Maybe I want to send you a message, but I don't want you to read it; it's weird, but maybe a bunch of us are sending information and we want you to be able to do some computation over it."

That's caught the interest of the Defense Advanced Research Projects Agency and Naval Information Warfare Center, which both funded the work that led to Hecate and have an interest in asking computer experts to crunch data without ever seeing the secrets hidden within it.

Varia's approach to encryption could also benefit survivors of sexual abuse. He recently partnered with San Francisco–based nonprofit Callisto to develop a new secure sexual assault reporting system. Inspired by the #MeToo movement, its goal is to help assault victims who are frightened of coming forward.

"They report their instance of sexual assault into our system and that report kind of vanishes into the ether," says Varia. "But if somebody else reports also being assaulted by the same perpetrator, then—and only

then—does the system identify the existence of this match."

That information goes to a volunteer attorney—bound by attorney-client privilege—who can then work with the victims and survivors on next steps. Just like Hecate, Varia says it finds a balance between privacy and openness, between deniability and traceability.

"When we talk about trade-offs between privacy, digital civil liberties, and other rights, sometimes there is a natural tension," says Varia. "But we can do both: we don't have to build a system that allows for bulk surveillance, wide-scale attribution of metadata of who's talking to who; we can provide strong personal privacy and human rights, while also providing online trust and safety, and helping people who need it."

More information: Hecate: Abuse Reporting in Secure Messengers with Sealed Sender

Provided by Boston University