Traceability and Cybersecurity

Experts’ Workshop Series on Encryption in India

November 2020

Traceability, or the ability track down the originator of a particular piece of content or message, is at the center of India’s debate around rules for online platforms and communications providers. In late 2018, the Indian Ministry of Electronics and Information Technology (MeiTy), proposed amendments to the Information Technology (Intermediaries Guidelines) Rules under the Information Technology Act.[1] Among the proposed changes is a demand for traceability, described as “enable tracing out of such originator[2] of information on its platform.” The amendment would make the online platform or provider liable for content posted by their users, if traceability is not provided. MeiTY invited public comments on the draft amendment in early 2019[3] and furthermore, nearly 30 cybersecurity and cryptographic experts in early 2020 sent an open letter to MeiTy expressing concerns about the proposed amendments.[4] Traceability is likewise at issue in an ongoing case at the Madras High Court between several large online platforms and the government concerning law enforcement access to user-generated content.[5]

Central to the ongoing debate are questions regarding:

- the achievability of traceability within end-to-end encrypted communications[6]
- what methods are available to enable traceability, and
- what are the ramifications of each?

Two techniques, the **use of digital signatures** and the **use of metadata**, have been proposed as methods to achieve traceability in end-to-end communications such as messaging apps (e.g. WhatsApp). Yet, to comply with traceability requirements, platforms may be forced to enable access to the contents of their users’ communications, **breaking end-to-end encryption** and considerably weakening the security and privacy of their product, in order to enable traceability.

[1] Information Technology [Intermediaries Guidelines (Amendment)] Rules.

[2] Under the Indian Information Technology Act, an originator is defined as the person who sends, generates, stores or transmits any electronic message or causes any electronic message to be sent, generated, stored or transmitted to any other person but does not include an intermediary”

[3] https://www.meity.gov.in/writereaddata/files/public_comments_draft_intermediary_guidelines_rules_2018.pdf

[4] https://www.internetsociety.org/open-letters/india-intermediary-guidelines/

[5] https://www.medianama.com/2019/09/223-sc-adjourns-hearing-on-facebook-transfer-petition-till-september-24/

[6] End-to-End (E2E) encryption — where the keys needed to unscramble an encrypted communication reside only on the devices communicating — provides the strongest level of security and trust, because by design, only the intended recipient holds the key to decrypt the message.

In a series of Chatham House Rule discussions held by the Internet Society in partnership with Medianama, an international group of cybersecurity security experts and policy experts examined the issue of message traceability in the Indian context. Experts had significant concerns around the two technical methods often proposed to enable traceability: the use of digital signatures and the use of metadata. Not only were these cited as threats to the privacy and security of users; there is doubt that they could be reliably used to attribute a message to its originator. Creating third party access to the contents of users' communications to enable traceability presents further concerns for security and privacy.

Digital Signatures

Some have recommended that the sender's digital signature[7] be added to messages, so that the originator of the message can be identified. For instance, in the case at the Madras High Court, Dr V. Kamakoti proposed the use of digital signatures to trace the originator of a forwarded message within WhatsApp.[8] In Kamakoti's proposal, the originator signature would either be visible to all in the message chain or be encrypted using a public key provided by WhatsApp. WhatsApp, using the corresponding private key, would be able to decrypt the originator information if there is a court order.

However, many, including experts in the Internet Society's discussions, raise concerns with the use of digital signatures to enable traceability:

- **Digital attribution is not absolute and vulnerable to impersonation:** To establish criminal liability, guilt has to be proven beyond reasonable doubt–a threshold that is hard to overcome, especially when impersonation online is so easy and pervasive. Proving that Person A, the user of Mobile/Computer A, actually sent those messages running a disinformation campaign is very difficult to prove even if law enforcement has the sender ID. Establishing whether the *use* of a person's device is necessarily evidence of the *person's use* of the device requires additional information.[9] Of even greater concern is that innocent users' may be implicated in illegal conduct by cyber criminals that impersonate their sender ID. This concern was mirrored in WhatsApp's own response to Komakoti's proposal, where they noted that "bad actors could use modified versions of the WhatsApp application to attribute a different phone number to a message."[10]

- **Digital signatures add vulnerabilities.** Private keys for digital signatures, particularly if held by a third party to the communication, like the communications service provider, would be a valuable target for bad actors. In the Kamokoti proposal, for instance, a compromised third party would have the potential to see when a particular user is sending a message – by receiving and decrypting the originator information. If put to the wrong use, the 'digital signature' approach gravely threatens citizens' freedom of speech, and can expose individuals (including the most vulnerable and the marginalized) to impersonation, harassment and persecution.

- **Cross-platform functionality would be unfeasible.** As different protocols govern different services and platforms, methods that would attempt to work across platforms would be unworkable. Inter-platform traceability is also difficult in federated systems like email or Internet Relay Chat (IRC). Methods to provide traceability across platforms, like using the same mandatory digital signature within the text of every message of every platform on a global basis, would be difficult to implement. A central registry of *every* device and *every* app client in the world would likely be required to authenticate the digital signatures. This would severely hamper innovation – forcing developers everywhere to coordinate their development with the operators of the central database.

[7] A digital signature is a mechanism to ensure that a message, or the information being sent, is authentic. It enables the sender of the message to attach a code that acts as a signature. Similar to a person's handwritten signature, it is unique to each signer, and can be likened to a tamper proof seal that guarantees that the information has not been altered in any way since it was sent

[8] https://www.medianama.com/2019/08/223-kamakoti-medianama-whatsapp-traceability-interview/

[9] For instance, someone else may be using their device.

[10] https://www.medianama.com/2019/08/223-exclusive-whatsapps-response-kamakotis-submission/

Additionally, a mandatory digital signature would put users at greater privacy and security risks as a single point of failure – or a single point for bad actors to target for compromise or to track a user's activities.[11] These digital signatures would also need to be easily revocable and re-issuable in case of compromise, which raises the question of how the associated private keys are secured. If this depends on users' biometrics, a further layer of technical and operational complexity is introduced.

Metadata

The use of metadata to enable traceability has also been raised by several entities in the encryption debate. Metadata, which contains information about the communication, but not the contents of the communication itself, can be used to determine things like the source, time, date, and destination of a communication, potentially the sender location, or even to estimate some attributes of the contents of the communication itself.

Some have highlighted that metadata around the size of the message, especially a media message, could potentially be used to check uncontrolled distribution of particular messages. For instance, WhatsApp keeps an encrypted log of media messages that are sent. It doesn't have access to the content of the media file, but it appears as a chunk of data of a particular size. Every time this media message is forwarded, the server has an estimate of how many times the chunk of data is forwarded – which it uses to fight spam on its platform. It may be possible to infer the message history with access to an unencrypted copy of the media file, thereby effectively removing the confidentiality of the communication.

Others contend that metadata can be analyzed to discern patterns such as who is talking to whom, how often and when. Basic social graphs can be created by platforms to create networks of interactions. Directory structures, which are fundamental to any communication platform, could be leveraged for investigation purposes without getting access to the content itself. These directories record who interacted with whom without intercepting the content itself.

However, experts, including those in the Internet Society's discussions, highlight concerns with the use of metadata to enable traceability:

- **Digital attribution is not absolute, particularly through metadata.** It is even harder to establish criminal liability based on metadata. Small changes to a message's contents could alter the metadata of a message, inhibiting the ability to follow a chain of similar metadata to an originator. Likewise, as with digital signatures, it is difficult to tie a user to a message when impersonation online is so easy and pervasive. Of even more concern is the potential for an innocent user to be implicated in criminal conduct purely based on spoofed metadata.

- **Undermine data minimization, privacy by design principles.** Dependence on metadata would harm platforms' moves towards data minimization and privacy by design, which an increasing number of data protection policies now require. This creates far more widespread risks to people's privacy and security by lowering the security standards for everyone. Metadata retained to help enable traceability would be a valuable target for bad actors. Criminals and foreign adversaries could use the stored metadata to develop social graphs of users or gather information that could enable attacks such as extortion, social engineering, or blackmail.

- **Risks of social profiling.** Where metadata is used to develop social graphs to help enable traceability, these social graphs are at risk of being compromised by criminals or foreign adversaries. There is also risk that these social graphs are monetized by the platforms themselves – which could expose sensitive details of government officials, elected officials, journalists, activists, lawyers, dissidents to data brokers and their customers.

[11] https://www.internetsociety.org/policybriefs/identity

- **Longer data retention periods create security risks:** Metadata retention rules often include requirements to retain the metadata for a certain length of time. If governments mandate longer metadata retention periods, it exacerbates privacy and security concerns by providing more metadata to be compromised in the case of a data breach. More data retained means more utility for criminals and foreign adversaries, and a more enticing target.

- **Not all platforms collect the same amount of metadata:** For example, Signal collects the bare minimum metadata that it needs to facilitate communication and doesn't collect any extra data.[12] Requirements to retain greater amounts of metadata could force platforms to significantly reconfigure their systems, incurring cost and raising the risks of creating new security vulnerabilities.

Breaking end-to-end encryption

As the utility of digital signatures and metadata to enable traceability is not clear, platforms may be forced to use methods to allow third parties to access the contents of communications, sometimes referred to as exceptional access, to comply with traceability requirements. By creating access to the contents of user messages, a platform or government entity could review messages sent by users – enabling them to flag objectionable content and identify the account that was sending the message.

Many techniques have been proposed to provide third party access to encrypted communications. This includes:

- *Key escrow*, where the keys used to decrypt messages are held (either partially or fully) by a third-party (like the platform provider) to enable access to contents of encrypted communications.
- *Ghost proposal*, where a third party is silently added as a participant to a communication.
- *Client-side scanning,* where communications, or hashes[13] created from the communications are reviewed for matches against a database of content before the message is sent to the intended recipient.

However, the consensus among experts, including those in the Internet Society's discussions, is that third party access methods would break end-to-end encryption by enabling third party access to content, and weaken the security and privacy protections for users.

- **Access for one is access for all.** By creating a way for a third party to access to users' encrypted communications, new vulnerabilities are effectively created in the system. Once found by bad actors, the same methods used to provide access for law enforcement or platforms could be used for nefarious activities. For example, if a bad actor were to gain access to the escrowed decryption keys, they would be able to decrypt all communications sent on a communications system. There is no way to ensure that the vulnerabilities created by an exceptional access method won't fall into the wrong hands.[14]

- **Exceptional access cannot be targeted and weakens security for all users.** When a system is modified to enable exceptional access, all users are put at greater risk. There is no way to deliver exceptional access to one user without creating a vulnerability for all users. For example, to implement the ghost proposal, the key distribution process must be altered by secretly distributing keys to people not in the group chat, and providers must suppress notifications to users that unauthorized third parties have access to their communications. By altering the key distribution and

[12] https://signal.org/blog/sealed-sender/
[13] A hash is a functionally unique digital "fingerprint" of user content
[14] https://www.schneier.com/academic/paperfiles/paper-keys-under-doormats-CSAIL.pdf

notification in a communications service, the platform introduces new vulnerabilities that could be used on all its users.[15]

- **Client-side scanning introduces vulnerabilities.** Some argue that client-side scanning is safe, particularly when the communications are hashed before being compared to a database of objectionable content. However, client-side scanning still introduces vulnerabilities that put the security and privacy of users at risk. Bad actors who gain access to a database of content could add new content to create false positives or monitor to whom, when, and where certain content was communicated.[16] Client-side scanning systems where the contents of a message are sent to a third party for manual review after matching to a database are particularly dangerous as they create a new way for bad actors to exploit to gain access to unencrypted communications.

- **National security concerns.** If one government or law enforcement agency were able to access to a user's communications, the same capability would be available to any other country in the world, including adversarial countries. Government officials and LEAs would not have access to secure communication channels either and risk becoming targets of surveillance by adversaries. Many government entities, including the European Commission[17] and the United States military,[18] have instructed their employees to use mass-market end-to-end encrypted communication services to protect their communications.

Conclusion

Traceability will likely continue to be a prominent issue in the debate in India around rules for digital platforms and communications service providers. However, there are credible concerns around the security, privacy and effectiveness of the two methods most often proposed to enable traceability, the **use of digital signatures** and the **use of metadata.** To comply with traceability requirements, communications service providers would be forced to access the contents of users' communications, greatly diminishing the security and privacy of a system for all users and putting national security at greater risk.

As policymakers, legislators, and judicial officials consider actions that would create traceability requirements, they must consider the severe implications of having content and service providers comply with these rules.

Acknowledgement
We are grateful for the participation of more than 50 technical and policy experts from Asia-Pacific, Africa, Europe, Latin America, and North America in the series of Global Technical Experts' Meeting held over the course of June to August 2020.

[15] https://www.internetsociety.org/resources/doc/2020/fact-sheet-ghost-proposals/
[16] https://www.internetsociety.org/resources/doc/2020/fact-sheet-client-side-scanning/
[17] https://www.politico.eu/article/eu-commission-to-staff-switch-to-signal-messaging-app/
[18] https://www.militarytimes.com/flashpoints/2020/01/23/deployed-82nd-airborne-unit-told-to-use-these-encrypted-messaging-apps-on-government-cellphones/