Internet Impact Brief

2021 Indian Intermediary Guidelines and the Internet Experience in India

Neeti Biyani, Internet Society
Amrita Choudhury, CCAOI

Version 1.0 (8 November 2021)

Abstract

In February 2021, the Indian Government introduced the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (hereafter, 'the Intermediary Liability Guidelines 2021' or 'the Guidelines') under the parent legislation, the Information Technology Act 2000. One of the primary aims of the Intermediary Liability Guidelines 2021 is to regulate social media intermediaries, such as messaging services and media-related intermediaries such as digital media houses, thereby recasting what intermediary liability protection means in the Indian context. This report uses the Internet Impact Assessment Toolkit[1] (IIAT) to assess how the Guidelines may affect the global Internet by impacting what the Internet needs to thrive as an open, globally connected, secure and trustworthy resource for all.

Methodology

The Internet owes its strength and success to a foundation of critical properties that, when combined, represent the Internet Way of Networking (IWN). This includes: an accessible Infrastructure with a common protocol, a layered architecture of interoperable building blocks, decentralized management and distributed routing, a common global identifier system, and a technology neutral, general-purpose network.

To assess whether the present proposal has an impact on the Internet, this report will examine its impact on the IWN foundation the Internet needs to exist, and what it needs to thrive as an open, globally-connected, secure and trustworthy resource.

Context

The Intermediary Liability Guidelines 2021[2] supersede the Information Technology (Intermediaries Guidelines) Rules, 2011.[3] Both

[1] https://www.internetsociety.org/issues/internet-way-of-networking/internet-impact-assessment-toolkit/ The IIAT was developed by the Internet Society to be used by anyone who wants to check if a particular policy, development, or trend affects the critical properties of the Internet Way of Networking (IWN).

[2] The Information Technology (Intermediary Guideliens and Digital Media Ethics Code) Rules, 2021: <https://www.meity.gov.in/writereaddata/files/Intermediary_Guidelines_and_Digital_Media_Ethics_Code_Rules-2021.pdf>

[3] The Information Technology (Intermediaries Guidelines) Rules, 2011: <https://www.meity.gov.in/writereaddata/files/GSR314E_10511%281%29_0.pdf>

versions of the Intermediary Guidelines are mainly focused on Section 79 of the Information Technology Act 2000 (hereafter, 'IT Act'), which is a 'safe harbour' clause exempting intermediaries from liability for the content users create and share on the intermediaries' platforms, as long as the intermediaries comply with the IT Act. The Intermediary Liability Guidelines 2021 impose various compliance requirements on intermediaries that limit their ability to claim protection from liability under the safe harbour clause.

In determining the requirements social media intermediaries must comply with, the Guidelines make a distinction between "social media intermediaries" and "significant social media intermediaries". While both are defined as "an intermediary which primarily or solely enables online interaction between two or more users and allows them to create, upload, share, disseminate, modify or access information using its services", significant social media intermediaries are platforms with a current threshold of 5 million or more registered users in India. This is not a high threshold for a country like India, which has a population of 1.3 billion – making most platforms qualify as significant social media intermediaries. Significant social media intermediaries also include any intermediary that the government deems as presenting "a material risk of harm to state security or the sovereignty and integrity" of the country.

The Intermediary Liability Guidelines 2021 require social media and significant social media intermediaries to adhere to a number of new compliance mechanisms, including content takedown within tight deadlines [Rule 3(1)(d)], automated content filtering [Rule 4(4)] and voluntary identification of users on social media intermediaries [Rule 4(7)]. But perhaps the most concerning requirement contained within the Guidelines is for a significant social media intermediary providing services primarily in the nature of end-to-end encrypted messaging – such as Signal or WhatsApp – to enable the identification of the "first originator" of any information they carry on their platforms. [Rule 4(2)].

End-to-end encryption allows only the sender and the receiver of the information to read the message, and ensures the information remains scrambled while it is in transit. End-to-end encryption offers the best protection of confidentiality between two endpoints, as only the sender and receiver can see their messages in decrypted form. There is no intermediary point at which a third party such as an Internet Service Provider (ISP), mobile service provider or messaging platform can eavesdrop on their conversation.

In the case of the Indian Intermediary Liability Guidelines 2021, the first originator is assumed to mean the very first person who introduces the message on the messaging platform in question. Further, the Guidelines provide that where the first originator of any information of an intermediary is based outside the territory of India, the first originator of that information within the territory of India shall be deemed to be the first originator of that information.

An important feature of the traceability requirement in the Indian Intermediary Liability Guidelines 2021 is that it requires end-to-end encrypted messaging services to enable traceability of *all* the messages exchanged over their platforms.[4]

The Government of India is proposing two models of implementing this traceability requirement, claiming neither will undermine end-to-end encryption:

(a) The identity of the originator will be tagged and included in an encrypted form with each message. Originally formulated by Professor V. Kamakoti of the Indian Institute of Technology, Madras,[5] this proposal will require an intermediary to hold in escrow the key to decrypt the originator's information, which may be used to reveal the same for a particular message in response to an order from an authorized body. Further, the proposal suggests that users mark a message 'forwardable' or 'non-forwardable' in order to indicate assumption of responsibility as an originator. If a user originates a message and marks it as 'forwardable', their information gets linked with the message. However, if a sender marks a message 'non-forwardable' and the recipient forwards it nonetheless, the recipient becomes the originator and their information is linked with the message.

(b) A library of alpha-numeric hashes for every message will be maintained, against which the hash of a message subject to a government order can be compared to enable traceability.[6] Hashing is the practice of using an algorithm to map data to a fixed length. An end-to-end encrypted messaging service would therefore have to maintain a library of numerous hashes to assist the government trace the originator of a message in case of an authorized order.

This Internet impact brief analyzes how the Intermediary Liability Guidelines 2021 could impact what the Internet needs to exist and thrive. Specifically, the report analyzes the ways in which the various compliance requirements mentioned above, and the traceability proposals being considered by the Government of India, would impact the Internet as an open, globally connected, secure and trustworthy resource.

---

[4] News18, New IT Rules: Why WhatsApp Is Worried Over Tracking Billions of Users, 1 July 2021: <https://www.news18.com/news/explainers/why-whatsapp-is-worried-over-tracking-billions-of-users-3778013.html>

[5] Prof. Kamakoti's proposal on identifying originator information on Whatsapp: <https://www.medianama.com/wp-content/uploads/Dr-Kamakoti-submission-for-WhatsApp-traceability-case-1.pdf>

[6] Surabhi Agarwal, 'India proposes alpha-numeric hash to track WhatsApp chat', Economic Times, 23 March 2021: <https://economictimes.indiatimes.com/tech/technology/govt-proposes-alpha-numeric-hash-to-track-whatsapp-chat/articleshow/81638939.cms>

How do India's Intermediary Liability Guidelines 2021 Affect the Realization of the Full Potential of the Internet

The Internet is built on a number of unique foundational properties that make it work and thrive for everyone. To assess how the Indian Intermediary Liability Guidelines 2021 might impact what we need for an Internet that is open, globally connected, secure and trustworthy, we will look at through the lens of the enablers of these goals.

Unrestricted reachability

"Internet users have access to all resources and technologies made available on the Internet and are able to make resources available themselves. Once a resource has been made available in some way by its owner, there is no blocking of legitimate use and access to that resource by third parties."[7]

Rule 3(1)(d) of the Guidelines lays out procedures that social media intermediaries and significant social media intermediaries must follow while addressing content takedown notices. An intermediary is required to remove access to content within 36 hours on receiving "actual knowledge" about allegedly objectionable content. Such notice may come from a court or the "appropriate government or its agency". Rule 3(1)(d) is incompatible with similar provisions in the Intermediary Liability Guidelines' parent legislation, the IT Act. In the event of a content takedown notice, the IT Act allows the intermediary or the originator of information an opportunity to respond to the request for blocking. No such provision exists in the Intermediary Liability Guidelines 2021.

The scope of the objectionable content laid out in the Guidelines is wide and loosely defined – it includes any unlawful information related to the interest of the country's sovereignty and integrity, national security, external relations, public order, decency, morality, defamation, contempt of court, etc. This makes abuse of Rule 3(1)(d) possible, and may result in an overreach of the government's powers to request content takedowns. Together, these measures could result in the frequent, erroneous blocking of legal content, restricting access to resources on the Internet and depriving users of global knowledge production, thereby hindering the global, connected Internet.

Data confidentiality of information, devices, and applications

"Data confidentiality, usually accomplished with tools such as encryption, allows end users to send sensitive information across the Internet so that eavesdroppers and attackers cannot see the content or know who is communicating. Allowing the transfer of sensitive information helps create a secure Internet. Data confidentiality also extends to data-at-rest in applications. (N.B., "confidentiality" also contributes to privacy, which is part of a trustworthy Internet)."[8]

[7] https://www.internetsociety.org/resources/doc/2021/enablers-of-open-globally-connected-secure-trustworthy-internet/
[8] Ibid.

As we have noted, Rule 4(2) of the Intermediary Liability Guidelines 2021 requires significant social media intermediaries that provide services primarily in the nature of messaging to enable the identification of the 'first originator' of the information. This rule ostensibly targets popular end-to-end encrypted messaging services, but may be extended to social media platforms that provide 'direct messaging' capabilities to their users.

The rule clarifies that significant social media intermediaries are not "required to disclose the contents of the electronic message", which suggests that law enforcement agencies aim to identify the first originator of an encrypted message that they already have access to through other means. This could happen by gaining access to a device involved in a messaging chain, or in case of a message going viral with potentially thousands of recipients, one of whom is reporting the message.

End-to-end encrypted messaging services do not have the capability to read messages shared on their platform or identify the first originator of messages. As argued by Meheshwari and Nojeim, the proposals being considered by the Government of India currently may not overtly require intermediaries to *disclose* a message, but that does not mean the intermediary will never need to *access* the contents of the message.[9] In most cases when the government is looking to seek originator information, it may be on the basis of already knowing the contents of a particular message. To comply with the law and shield themselves against liability, intermediaries may need to link the known content of a message with the originator. They could only do so by accessing each message in a forwarding thread, thus breaking end-to-end encryption.

However, Rule 4(2) goes against the parent legislation of the Intermediary Liability Guidelines – the IT Act. The IT Act does not grant the government the power to demand traceability. Rule 4(2), in demanding traceability, breaks end-to-end encryption systems, undermines user privacy, threatens data confidentiality, goes against the principle of data minimization, and may create room for government overreach and censorship. Altogether, these proposal requirements seriously threaten the goal of a secure Internet.

End-to-end encryption ensures that data and communication is confidential between the sender and receiver. Any attempt to undermine end-to-end encryption systems endangers individuals and businesses. In mandating traceability of end-to-end encrypted messages, Rule 4(2) will result in associated costs for storing all the data needed for tracing every message ever exchanged on such a platform, impacting affordability and preventing new entrants from easily and affordably offering end-to-end encrypted messaging services in the country. There are innumerable marginalized groups, such

[9] Maheshwari & Nojeim, 'Part 2: New Intermediary Rules in India Imperil Free Expression, Privacy and Security', Center for Democracy & Technology, June 2021: <https://cdt.org/insights/part-2-new-intermediary-rules-in-india-imperil-free-expression-privacy-and-security/>

as journalists, victims of domestic abuse and members of the LGBTQ+ community, who would be in particular danger if end-to-end encryption is threatened.

Further, Rule 4(4) of the Intermediary Liability Guidelines 2021 mandates significant social media intermediaries to "deploy technology-based measures, including automated tools or other mechanisms" to proactively detect and filter content depicting rape, child sexual abuse material (CSAM), or content removed earlier in response to a takedown notice. This will require social media intermediaries to set up client-side scanning, which refers to systems that scan message contents — i.e., text, images, videos, files — for matches against a database of objectionable content before the message is sent to the intended recipient.[10] This compromises data confidentiality as this rule requires intermediaries offering end-to-end encrypted messaging to monitor and filter all the content a user sends, thus defeating the purpose of end-to-end encryption.

Reliability, resilience, and availability

"The Internet is reliable when technology and processes are in place that permit the delivery of services as promised. If, for example, an Internet service's availability is unpredictable, then users will observe this as unreliable. This can reduce trust not just in one single service, but in the Internet itself. Resilience is related to reliability: a resilient Internet maintains an acceptable level of service even in the face of errors, malicious behavior, and other challenges to its normal operations."[11]

Rule 4(2) of the Guidelines, which mandates traceability on end-to-end encrypted messaging services, breaks the crucial properties of and fundamental expectations from end-to-end encryption on messaging platforms, thus making such services unreliable. It is in fact better to not encrypt a service at all, rather than linking every user to the content they generate, as then they are vulnerable to the new risks posed by traceability.

Rule 4(4) of the Guidelines mandates client-side scanning with the ostensible aim of tracking down CSAM and content depicting violence against women. While removal of such content is a justifiable aim, automated content filtering systems are error-prone, have inherent biases and prejudices, and threaten free and legitimate speech. They can filter out unrestricted content, and more often than not result in over-removal of content.[12] This will erode the reliability of end-to-end encrypted messaging services. Further, any endeavour to filter, monitor and block such content should not be designed solely by private social media intermediaries, as it leads them to assume an adjudicatory role.[13]

---

[10] Internet Society, 'Fact Sheet: Client-Side Scanning', March 2020, <https://www.internetsociety.org/resources/doc/2020/fact-sheet-client-side-scanning/>

[11] https://www.internetsociety.org/resources/doc/2021/enablers-of-open-globally-connected-secure-trustworthy-internet/

[12] Article 19, 'Freedom of Expression Unfiltered: How blocking and filtering affect free speech', December 2016, <https://www.article19.org/data/files/medialibrary/38586/Blocking_and_filtering_final.pdf>

[13] Praveen Arimbrathodiyil vs. Union of India: <https://sflc.in/praveen-arimbrathodiyil-vs-union-india-sflcin-assists-challenging-part-ii-intermediary-rules-2021>

Accountability

"Accountability on the Internet gives users the assurance that organizations and institutions they interact with are directly or indirectly acting in a transparent and fair way. In an accountable Internet, entities, services, and information can be identified and the organizations involved will be held responsible for their actions."[14]

Rule 3(1)(d) of the Guidelines, which lays out the criteria for content takedowns, mounts unfair pressure on social media intermediaries to adhere to potentially arbitrary takedown notices, some of which may even come from non-judicial government bodies – thus impacting the transparency and accountability of online services and undermining trust in the Internet.

Rule 4(4) of the Guidelines, mandating automated content filtering of CSAM and content depicting violence against women may result in disparate standards of offensive content by different social media platforms. Most automated content filters are error-prone, have inherent biases and prejudices and threaten free and legitimate speech. They can filter out unrestricted content, and more often than not result in over-removal of content, leading to unnecessary takedowns. Additionally, Rule 4(4) needs to balance the threats to users' rights against the use of automated tools. Instead of leaving the development and implementation of automated tools to social media intermediaries, the tools should be developed in a collaborative manner between the technological industry and child rights organizations, and contain multistakeholder oversight requirements. It is also alarming that the Guidelines 2021 do not establish mechanisms allowing users to appeal against takedown of legal content, a provision that is contained within the parent legislation, the IT Act.

Privacy

"Privacy on the Internet is the ability of individuals and groups to be able to understand what information about them is being collected and how, and to control how this is used and shared. Privacy often includes anonymity."[15]

Rule 4(2) of the Guidelines, mandating traceability on end-to-end encrypted messaging platforms, does not conform with the parent legislation – the IT Act. End-to-end encrypted services provide users with a secure platform to communicate freely without the fear of being watched or surveilled by third parties, including the state. Traceability as proposed in the Guidelines will lead to self-censorship as users will be reticent in their communication for the fear of being watched. This also extends to fear of trade secrets, commercial and financial communication and privileged communication being surveilled. This is a threat to citizens' right to privacy. Traceability of the first originator of certain information as well as users involved in a particular chain of messages infringes upon users' right to privacy and their right to anonymity.

[14] https://www.internetsociety.org/resources/doc/2021/enablers-of-open-globally-connected-secure-trustworthy-internet/

[15] Ibid.

Rule 4(2) also gives access to law enforcement agencies, but is at risk of being exploited by state authorities to further surveillance, tracking political dissent and so on. This weakens security and privacy for all users who use end-to-end encrypted services and platforms.

Finally, Rule 4(4) mandating client-side scanning compromises user privacy on end-to-end encrypted messaging services, as adding client-side functionality creates additional ways to interfere with communication by manipulating the database of objectionable content. Adversaries with the ability to add digital fingerprints to the database and receive alerts when matches to those fingerprints are found, would be able to monitor select user content before it is encrypted and sent. This would allow them to track to whom, when, and where certain content was communicated. These fingerprints could include commonly used passwords or other information to enable attacks such as social engineering, extortion, or blackmail.[16] It would also require minimal effort to reconfigure client-side scanning from CSAM to include any targeted content, enabling authorized and unauthorized parties to search for arbitrary content in the absence of warrants.[17]

Summary and Conclusions

This Internet Impact Brief on the Indian Intermediary Liability Guidelines 2021 has revealed several ways in which the rules hinder what the Internet needs to thrive. In particular, it impacts the enablers of an Open, Globally Connected, Secure and Trustworthy Internet by undermining ease, reachability, data confidentiality, reliability, accountability, and privacy online. The following conclusions should help the Government of India revisit its Intermediary Liability Guidelines and reconsider rules to mitigate harming the Internet.

In trying to regulate the Internet in the country in this manner, India is impacting many of the things the Internet needs to thrive as a unique and ubiquitous resource.

Many rules laid out in the Intermediary Liability Guidelines 2021 contravene the openness, interoperability, reliability and trustworthiness of the Internet. We must distinguish between the Internet as an infrastructure and the services and platforms built on top of it. While the aim to address the unique challenges that social media intermediaries present as well as prevent CSAM and content portraying violence against women is justifiable, many of the issues the Guidelines wants to curb are in fact social problems that existed long before the Internet.

[16] Internet Society, 'Fact Sheet: Client-Side Scanning', March 2020, <https://www.internetsociety.org/resources/doc/2020/fact-sheet-client-side-scanning/>

[17] Abelson et al., 'Bugs in our Pockets: The Risks of Client-Side Scanning', October 15, 2021, <https://arxiv.org/abs/2110.07450>

Conclusions:

- Rule 3(1)(d) of the Intermediary Liability Guidelines, which mandates content take down on social media intermediaries, hinders the unrestricted reachability of the Internet. The vague definition of the content that may be subject to takedown also makes predictability hard for users as well as social media intermediaries.
- Rule 4(2) of the Guidelines mandating traceability on end-to-end encrypted messaging services should be withdrawn. Traceability is incompatible with end-to-end encryption, impacts data confidentiality and applications and threatens users' privacy.
- Rule 4(4) mandates client-side scanning by way of automatic filtering and blocking of CSAM and content portraying violence against women. While this is a worthy goal, it should not be implemented in the current design of the Guidelines as it poses significant harm to both the Internet and its users.
- Rule 4(7) of the Guidelines, requiring users to voluntarily identify themselves to social media intermediaries, restricts easy access to the Internet and should be withdrawn.