Hashing and matching is a key component of tackling child sexual abuse material (CSAM) at scale. It is a privacy forward way to share information of known and verified CSAM across law enforcement and industry.

Hashing technology is not new and has been used on CSAM for over a decade. Hashing is an algorithm that converts an image into a unique string of numbers. You cannot reverse engineer a hash into an image, it can only be matched to other hashes. Most images are hashed by known and trusted entities such as the National Center for Missing and Exploited Children (NCMEC), law enforcement, and some industry hash sets exist.

By converting images this way a company or law enforcement can share information about known CSAM in an anonymous manner. Hashes are most effective when shared across the ecosystem. A company can then set up a system to scan their images for known CSAM which is quick and effective.

In 2021, we know NCMEC received over 29 million reports from industry of CSAM which was over 84 million individual pieces of CSAM. The volume of this material is overwhelming. This increase represents both the real and alarming rise in CSAM dissemination online, and the improved ability and willingness of tech companies to be proactive in detecting, reporting, and removing this content from their platforms.

NCMEC and reviewing entities need the capabilities that hashing and matching provide in order to quickly determine what needs to be human reviewed and what is known CSAM that can immediately be reported to law enforcement. Companies can quickly remove known CSAM from their platform before it spreads further. With a problem at this scale privacy forward technology is necessary.

March 2022