Who Bears the Burden? Technical Architectures for Age-Based Content Restriction

Dennis Jackson
Mozilla

Content restrictions of any kind raise many difficult questions. This paper does not argue for or against their introduction. Instead, it asks: if these restrictions are to be rolled out by legal fiat, who should be responsible for enforcing them at the technical level?

There are only two answers compatible with the end-to-end principle: either the device that displays the content to the user or the service that delivers the content to the device.

In a *service-enforced* system, services must not deliver age-restricted content to clients unless they first verify the user is old enough. These systems treat age restriction as a typical access control problem, resolved by requiring clients to authenticate. Such systems are already being implemented in several jurisdictions, including the United Kingdom, the European Union, Australia, France, and about half of US states.

In a *device-enforced* system, consumer devices must not display age-restricted content unless they have verified the user meets the age requirement. This shifts the enforcement layer from between the client and server to between the user and their device. This approach is supported by some child protection groups, adult content providers and major online platforms, but faces strong opposition from device makers such as Apple and Google.

This paper compares both approaches across a wide range of criteria, then assesses their overall suitability and proposes a pragmatic path forward.

Contents

Acceptability to Users

Even when the deployment and use of a system is a legal obligation, user acceptance is still important. If users perceive that a system causes excessive friction, invades their privacy, or conflicts with their values, they are strongly motivated to disable or circumvent it, rendering it ineffective.

Verifying a user's age is currently a high friction process. Users are confronted by dialogues asking them to sign in, turn on their webcam, or share personal information. While efforts are underway to create smoother flows through new APIs, these technologies will take years to arrive and only reduce friction to about that of cookie consent dialogues today.

How often users are going to be prompted is therefore central to acceptability. A service-enforced system requires users to verify their age at least once per age-restricted service they use. Given the wide range of adult content considered unsuitable for children and the ever-changing media landscape, this entails users having to verify their age frequently. Moreover, unless users create long-term accounts, they must verify their age again in each future interaction. In contrast, a device-enforced system requires only a single age verification per device.

The frequent interactions required by a service-enforced system also heighten users' perceived[1] loss of privacy. Many users are uncomfortable with document-based or biometric age verification, regardless of legal or cryptographic promises of privacy. After all, they have little reason to trust services with their identity and many reasons not to. In contrast, users already entrust sensitive data, including identity and biometrics, to their devices, making device-based age verification feel less invasive.

Service-enforced restrictions must follow legal definitions of restricted content and cannot reflect individual family preferences or age nuances. However, device-enforced systems support default policies that can be set out in regulation, while allowing families to tailor enforcement based on their values and child's maturity. This flexibility benefits both those seeking stricter controls and those who prefer looser settings, further reducing friction.

Overall, device-enforced systems reduce friction, improve perceived privacy, and better align with user values through configurable settings.

Cost of Compliance

Service-enforced and device-enforced systems differ in the obligations they place upon service providers and device manufacturers, leading to meaningful differences in compliance costs.

[1]We tackle the actual loss of privacy in Privacy Impacts.

Although neither can lawfully avoid complying, high costs can cause companies to delay adoption, engage in malicious compliance or exit a market entirely.

For service-enforced systems, each service communicating restricted content needs to establish an access control mechanism, establish or partner with an age verification service, identify which users should be subject to the access control (e.g. via geo-location), and assess which content should be restricted. Further, in order to minimize user friction, they are strongly incentivized to offer long-term user accounts, which requires its own technical investment and engages data protection regulations like the GDPR. These obligations place a substantial burden on each service that does not scale with audience size or revenue, disproportionately harming smaller players.

In a device-enforced system, the majority of the costs fall upon the device vendor, who must provide a mechanism for establishing the age of the user, an API which apps on the device can query for whether the user is subject to age restrictions and the necessary integrations in apps installed by default. Conveniently, suitable APIs already exist for some popular platforms[2] including MacOS and iOS's Declared Age Range API and Windows' Content Restrictions API.

Legislation mandating a device-enforced system would need to balance the need to comprehensively cover the range of devices which are available to children, against enforcing unnecessary requirements on devices which pose little risk due their limited distribution or niche uses.

Device-enforced systems also require investment from third-party apps containing age restricted content who would need to integrate support for content restriction APIs. This is far less burdensome than delivering a full age verification system as in the service-enforced model, but still requires the use of accurate content labels so that the device can enforce restrictions appropriately[3].

Comparing the two approaches, a device-enforced system offers substantial savings over a service-enforced system for most participants. The main exceptions are device manufacturers who bear greater responsibility, however, they already must comply with a wide range of existing regulations for electrical safety, usability and accessibility. In contrast, a service-enforced system imposes high costs on every compliant digital service, which will disproportionately burden smaller players.

[2]A notable exception amongst commercial consumer operating systems is Android; despite developer requests.
[3]Discussed further in Content Coverage

Content Coverage

Users engage with media through an ever-growing range of formats and technologies. Age-based content restriction systems must offer comprehensive coverage to effectively protect under-age users.

In the traditional media distribution model a server operated by a commercial entity communicates content to clients. Service-enforced systems work well in this setting, because the same entity is responsible for both distributing content and enforcing restrictions on accessing it, so there's no need for coordination. In contrast, device enforced systems split responsibility between the service who labels the content and the device who interprets those labels and restricts access accordingly.

Some labelling mechanisms already exist. A labeling solution called 'Restricted to Adults' has been available for the web for nearly two decades and has been widely deployed by adult content providers on a voluntary basis. Apple's App Store already requires app developers to provide age ratings for their apps and a similar system is used by the Google Play Store. Creating open and interoperable standards for content labelling is not a trivial task, but nor is it an unsolvable one.

An alternative to relying on services to label content is to employ on-device machine learning, as is used in Apple's Sensitive Content Warning, though this technology is still in its infancy, it shows considerable promise. There are also numerous rating services which offer labels for third party content which could be integrated into applications.

For some content, services may not have access to what they're communicating, or there maybe no responsible service. For example, it may be end-to-end encrypted (e.g. Signal, WhatsApp or iMessage) or shared peer-to-peer (e.g. Bluetooth or sneakernet) or even created locally with generative AI. In these contexts only the device has access to the content and so only a device-enforced approach can provide effective coverage.

While device-enforced systems require more coordination between content distributors and device vendors; these challenges are surmountable and workable solutions already exist. Contrastingly, though service-enforced systems can be cleanly applied to traditional media platforms, they are poorly aligned with modern forms of content consumption.

Adoption Dynamics

Device and service-enforced systems differ greatly in the adoption levels needed for effectiveness.

A naive view suggests the effectiveness of a service-enforced approach scales with the popularity of compliant services. But this fails to account for user behavior: demand for adult content is high, and users of any age can easily switch to non-compliant services when faced with friction or technical blocks. Legal action against non-compliant services is slow and often futile[4], especially when they operate from abroad. The long tail of less than legitimate adult sites who can conceal their funding sources, change their identity and swap domain names as-needed makes regulatory enforcement extremely challenging.

The shortcomings of this approach are illustrated by the fight against copyright infringement. Despite a global, long running campaign, it remains readily accessible to internet users of all ages. Ofcom's 2025 report on online copyright infringement found that one-third of UK internet users knowingly consumed pirated media in the three month study window, rising to over half of 12 to 15-year-olds.

Device-based enforcement is more robust. Users can switch websites or services easily, but not devices. If all of an underage user's devices implement the system, that user is effectively protected. In the UK, just 3 hardware vendors account for 85% of the mobile market and across the entire mobile market, only two operating systems are in popular use. The picture is similar for other types of devices such as tablets, game consoles, laptops and smart TVs. Even a device-enforced scheme whose deployment was limited to the very largest vendors would protect a high fraction of users. Further, these device vendors are necessarily subject to government jurisdiction and can be policed with existing market surveillance techniques for unsafe or non-compliant products.

However, there are two complicating factors. First, even if device vendors adopt the system, apps and services may not, meaning that users might only be partially protected. Labelling content with on-device machine learning, or the use of third-party rating services (discussed in Content Coverage) could mitigate this aspect.

Second, many unrestricted devices are already in use today and it is unlikely that restrictions can be retroactively deployed. Similarly, some vendors may continue to ship devices with absent or easily disabled controls which might appeal to under-age users looking to evade restrictions. As children have limited financial resources, the degree to which this is problematic for a device-enforced system's effectiveness is inversely proportional to adults' acceptance of the system. If well-implemented (i.e. low friction and perceived as effective by adults), then market forces will favour wider adoption of compliant devices and will minimize the availability of older or ineffective devices.

It seems unlikely that service-enforced restrictions can achieve the necessary adoption to effectively protect users. This is compounded by the high friction and costs of compliance noted

[4]At the time of writing, Ofcom (the UK state regulator) is providing a live feed of domains hosting adult content which do not perform age verification and have not been blocked.

earlier. On the other hand, device-enforced age restrictions need only be deployed by the most popular hardware vendors in order to protect a high fraction of users effectively.

Resistance to Circumvention

We also need to consider how effective these systems would be in the face of underage users determined to access restricted content. After all, if circumvention were not a concern, existing age dropdowns would suffice. This is a question of security analysis in which we need to consider: how might an attacker (the underage user) subvert or evade deployed access controls?

We first consider how under-age users might try to subvert an age verification challenge. Whilst device-enforced and server-enforced systems should offer comparable security for most popular challenge types, there are some notable exceptions.

Firstly, for biometric methods, a common bypass mechanism is to present a faked picture or video (created using a videogame, ID generator or generative AI). It is difficult, if not impossible, to prevent this in the service-enforced setting because the remote server has no effective way to verify whether the presented video is coming from a real camera, or is being emulated in software. Meanwhile, unless a child is able to interfere with the integrity of a device, the device's direct access to hardware can defeat these approaches, especially if robust liveness detection is used.

Secondly, methods based on inferring age from 3rd party systems are relatively easy for under-age users to compromise. Utility bills, credit cards, and adult's mobile phones are easily accessible to most under-age users and enable them to bypass these checks. Although device-enforced systems can leverage existing access control mechanisms (e.g. the use of passcodes or on-device biometrics like FaceID and TouchID) to protect parental control, as well as using more reliable (but higher-friction) age assurance methods.

An alternative to defeating age verification challenges directly, or choosing to visit unprotected sites (discussed in Adoption Dynamics) is for under-age users to avoid triggering challenges on otherwise protected services or devices.

In the service-enforced setting, content restrictions are triggered if the service detects the user is connecting from a country which requires age checks. However, users can easily spoof their location through the use of popular tools like VPNs or proxies, which are freely available, widely used and require no technical sophistication. Some age verification providers have claimed this could be prevented through the use of additional checks like querying a device's geolocation API. These claims are nonsense. Users can spoof the results of these APIs even more easily than masking their IP address with a VPN. Nor can access to VPNs be restricted. They have many legitimate users for both individuals looking to protect their privacy and corporate users looking for additional security.

In the device-enforced setting, the direct analogue is whether underage users can disable restrictions by tampering with their device. Although experts regularly discover new privilege escalation exploits, employing them is far beyond the technical capability of most under-age users and comparatively much harder than bypassing a service-enforced restriction by installing a VPN.

Alternatively, underage users in the device-enforced setting may try to evade checks by downloading applications which aren't aware of age restrictions and don't use the appropriate OS APIs. This is easily prevented by using existing OS mechanisms for restricting the installation of software without a suitable age rating or the approval of an administrative (adult) account.

Ultimately, circumvention is a challenge for any age-based content restriction system. In a service-enforced system, even unsophisticated underage users can bypass controls easily. In device-enforced systems, their success depends on both the device used and any settings which parents may have altered to improve (or reduce) the effectiveness of its enforcement.

Privacy Impact

We now examine how device or service-enforced content restriction systems could be abused to compromise the privacy of users.

Commercial actors may exploit age verification mechanisms to infer additional information about their users (e.g. to boost advertising revenues). Such abuse is difficult for users and regulators to detect , let alone prevent or punish. This is a challenge for both models, but far greater for a service-enforced system where users frequently need to verify their age with parties they have no prior relationship with. Whereas device-enforced systems can rely on the user having some level of trust in their device vendor, who are also more vulnerable to enforcement actions for data protection violations.

Some proponents of a service-enforced system have proposed designs which aim to prevent a user's age verification information from being linked to their interactions with restricted services. Unfortunately, many of the deployed systems rely on legal or technical promises which users have no way of verifying. In the future, new cryptographic approaches may enable strong mathematical guarantees of privacy to users, which may help address these concerns.

However, even if using such a cryptographic solution was mandatory, we have limited means to help users understand when it is (or isn't) being employed to protect their privacy. Users are notoriously bad at evaluating the trustworthiness of digital services and nothing prevents criminals from posing as age-verification providers and directly capturing the pictures, videos or documents provided by the user. Such scams are attractive to organised cybercriminals

because they enable identity fraud, phishing or even extortion over a user's interactions with adult services.

These privacy risks are largely limited to service-enforced systems. In device-enforced systems, users aren't faced with frequent age verification challenges which can leak their data or preferences, nor are they habituated to releasing sensitive documents, credentials or proof-of-age videos. In fact, in a well-designed device-enforced system, no information would need to leave the device at all.

Censorship Risks

Any system designed to restrict access to age-inappropriate content is vulnerable to scope creep and this is largely a question of policy. Here, we focus on how the two technical architectures provide different capabilities which might be abused.

In a service-enforced system, mechanisms must exist to prevent users from accessing non-compliant services that don't enforce age-based content restrictions, e.g. by through network filtering. Whilst these mechanisms would likely be ineffective in blocking access to widely sought content like adult media, they could still be used to effectively block more niche material, like journalism. Worse, the continued ineffectiveness of the service-enforced system at restricting access to adult content could be exploited to motivate the introduction of more invasive policies.

In a device-enforced system, there's no need to censor network traffic. However, it's easy to see how on-device restrictions aimed at children - which adults can disable - could be extended and repurposed to restrict categories of content for adults as well. By virtue of being on-device, the same mechanisms that make device-enforced content restriction more effective for blocking age-restricted content also make it more effective for censorship.

A key question is whether adults can ultimately disable the restrictions. Although technical controls to prevent children from accessing restricted content need not be sophisticated, analogous measures aimed at adults would need to prevent device owners from fully controlling their devices through the use of hardware-backed device attestation like Google Play Integrity and Apple App Attestations.

A further complication is that device vendors often have profit-driven incentives to push for the use of such mechanisms, which tighten their control over devices (at the expense of users). Reducing user agency allows vendors to advantage their own apps, prevent users from choosing alternatives or control how they consume content. These efforts align with the goals of censors looking to prevent content restrictions from being disabled, compounding the societal risk.

There is considerable prior art in this area. For example, the EU's controversial ChatControl, Apple's now-cancelled CSAM Detection and Google's now-cancelled Web Environment Integrity involve repurposing devices to work against their owners. Troublingly, a similar mechanism is being considered to secure the EU's service-enforced age verification system.

Abuse is a substantial risk of any content-restriction system. In the service-enforced case, effective censorship seems harder to achieve, but the ease of circumvention may drive the introduction of harsher and move invasive mechanisms. In device-enforced systems, the risk depends on how responsibly the legislation is written. If the solution preserves users' control over their own devices, an effective solution could be delivered with minimal risk. If not, legislation forcing devices to work against their owners' interests could be catastrophic.

Takeaways

If age-based content restrictions are to be mandated by law, there are strong arguments for enforcing them at the device level. Device-based systems offer clear advantages in user acceptability, cost effectiveness, and coverage across media types and technologies. They can protect a high fraction of users even if adoption is limited to the largest device vendors and resist circumvention much better than service-based restrictions. Importantly, they present fewer and more manageable risks to user privacy. However, there is a substantial danger that device-enforced content restriction is used to pass legislation which prevents users from being in full control of their own devices.

While service-based enforcement appeals to traditional notions of access control, it introduces substantial friction for users, high compliance burdens for services (especially smaller ones), and satisfies no one with ‘one-size-fits-all’ policies. Further, the ease of circumvention and the challenges for regulators trying to enforce the rules will likely drive the introduction of more intrusive measures, yet fail to improve efficacy, raising serious privacy and censorship concerns.

Given that some jurisdictions are already pursuing a service-enforced approach, regulators could take a pragmatic step by recognizing device-based enforcement as a suitable mechanism for enforcing age-based content restrictions in their guidance.

Such a mechanism would need little more than a signal from a device to indicate that it supports machine readable content labels and age restrictions, without revealing whether those restrictions are active or anything else about the device's settings or its user.

This would allow applications and services to adopt the approach and test its real world effectiveness. However, to achieve the high coverage that proponents of age-based content restrictions desire, it might be necessary to encourage device vendors to offer suitable APIs and standards to support widespread deployment.