Image DePO: Towards Gradual Decentralization of Online Social Networks using Decentralized Privacy Overlays

JACOB LOGAS, Georgia Institute of Technology, USA
ARI SCHLESINGER, Georgia Institute of Technology, USA
ZHOUYU LI, Georgia Institute of Technology, USA
SAUVIK DAS, Georgia Institute of Technology, USA

Centralized online social networks — e.g., Facebook, Twitter and TikTok — help drive social connection on the Internet, but have nigh unfettered access to monitor and monetize the personal data of their users. This centralization can especially undermine the use of the social internet by minority populations, who disproportionately bear the costs of institutional surveillance. We introduce a new class of privacy-enhancing technology — decentralized privacy overlays (DePOs) — that helps cOSN users regain some control over their personal data by allowing them to selectively share secret content on cOSNs through decentralized content distribution networks. As a first step, we present an implementation and user evaluation of Image DePO, a proof-of-concept design probe that allows users to upload and share secret photos on Facebook through the Interplanetary File System peer-to-peer protocol. We qualitatively evaluated Image DePO in a controlled, test environment with 19 queer and Black, Indigenous, (and) Person of Color (BIPOC) participants. We found that while Image DePO could help address the institutional threats with which our participants expressed concern, interpersonal threats were the more salient concern in their decisions to share content. Accordingly, we argue that in order to see widespread use, DePOs must align protection against abstract institutional threats with protection against the more salient interpersonal threats users consider when making specific sharing decisions.

CCS Concepts: • **General and reference** → *Design*; • **Computer systems organization** → *Peer-to-peer architectures*; • **Networks** → *Peer-to-peer protocols*; **Network privacy and anonymity**; **Online social networks**; • **Software and its engineering** → *Middleware*; **Walkthroughs**; • **Information systems** → **Social networks**; • **Human-centered computing** → *User studies*; **Social content sharing**.

Additional Key Words and Phrases: decentralized; steganography; online social network; privacy; queer; LGBTQ; BIPOC

ACM Reference Format:
Jacob Logas, Ari Schlesinger, Zhouyu Li, and Sauvik Das. 2022. Image DePO: Towards Gradual Decentralization of Online Social Networks using Decentralized Privacy Overlays. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 60 (April 2022), 28 pages. https://doi.org/10.1145/3512907

1 INTRODUCTION

The early years of social networking facilitated massive decentralized social movements like the Arab Spring [21], the Occupy movement [101] and the Indignados movement [81]. At their best, online social networks (OSNs) offer people unprecedented abilities to connect, coordinate, and collectively

Authors' addresses: Jacob Logas, logasja@gatech.edu, Georgia Institute of Technology, Atlanta, Georgia, USA, 30332; Ari Schlesinger, a.schlesinger@gatech.edu, Georgia Institute of Technology, Atlanta, Georgia, USA, 30332; Zhouyu Li, zli853@gatech.edu, Georgia Institute of Technology, Atlanta, Georgia, USA, 30332; Sauvik Das, Georgia Institute of Technology, New York, USA, sauvik@gatech.edu.

2573-0142/2022/4-ART60 $15.00
https://doi.org/10.1145/3512907

act. However, as these platforms have matured and the monetary incentives of collecting, processing, and monetizing personal data drive nigh trillion dollar market capitalizations [2, 45], we now see social networking platforms as powerful institutions with disturbingly unilateral control over users' personal data and correspondences with others. We have now entered what privacy scholars have termed the "Golden Age of Surveillance" [98] and "The Age of Surveillance Capitalism" [108].

Two common technological mechanisms used to resist the surveillance enabled by centralized OSNs (cOSNs) are (1) data masking and (2) decentralization. (1) Data masking allows for sharing secret data over untrusted lines of communication; one example is FaceCloak, in which users share "real" text content with a trusted third party and fake text content with Facebook [66]. The trusted third party only delivers the "real" data to viewers with the FaceCloak app installed, bypassing Facebook's data processes. (2) Decentralization allows for sharing data from a network wherein there is no single control point. Decentralized online social networks (dOSNs), like mastodon.social, allow users to connect with each other through independent, federated servers that compete on user-amenable privacy policies. The data masking approach has the benefit of allowing users to continue using a platform they know and in which they are invested, but offload trust from one centralized institution to another. The decentralization approach has the benefit of distributing trust, but can only succeed if a critical mass of users migrate to the dOSN. However, neither data masking nor decentralization have seen success in attracting a user-base outside of the most privacy conscious users. What is needed is an approachable alternative that allows users to distribute content in a decentralized manner without requiring them to de-platform themselves.

In this paper, we introduce a new technical approach, decentralized privacy overlays (DePOs): client-side technologies that allow users to share secret content on cOSNs through decentralized content distribution networks. DePOs combine the benefits of data masking, in allowing users to remain on their preferred platforms, with the benefits of decentralized content delivery, in eliminating the need to share personal data with cOSNs. We propose DePOs as a mechanism to help reduce the impact of institutional surveillance through the *gradual* decentralization of cOSNs.

To unpack how DePOs might be received by end-users, we implemented a proof-of-concept design probe, Image DePO: a browser extension that enables sharing secret images through Facebook. Design probes are exemplary artifacts that can help elicit concrete, grounded reactions to a concept technology and are commonly employed in HCI design research [107]. When sharing a secret image with Image DePO, users click on an injected button in the Facebook composer interface (fig. 3) and select an image to share. Image DePO adds the secret image to the InterPlanetary File System (IPFS) — a decentralized file system in which files can be located through the cryptographic hash of its content [11]. We encode the unique hash of the secret image into a generated cover image using steganography, a method of hiding secret data within publicly view-able data (e.g., image, video, audio). This cover image is then uploaded onto Facebook instead of the secret image. Viewers with Image DePO installed and who have access to the cover image through Facebook's standard access control settings will have the secret image dynamically hot-swapped with the cover image when it comes into their view; but, Facebook and other viewers will see only the cover image.

We qualitatively evaluated Image DePO with 19 participants from the queer and Black, Indigenous, (and) People of Color (BIPOC) communities [1], recruited through the Prolific online participation platform. We focused on recruiting participants from minority populations because, as surveillance scholar Simone Browne notes, the costs of institutional surveillance are disproportionately borne by these populations [20]. Thus, by centering the voices of those who stand the most to gain from DePOs and other counter-surveillance technologies, the insights we derive should improve DePO design for those with the most at stake. Given that the primary goal of our evaluation was to extract design

[1] We note several authors of this paper belong to these communities (section 7).

insights about how to design DePOs to better address lived concerns over institutional surveillance, rather than to suggest that Image DePO is an immediately deployable solution, we focused on answering the following research questions in our evaluation:

RQ1 What threats, and mitigating strategies thereof, do queer and BIPOC participants consider when sharing photos on cOSNs?
RQ2 How might a decentralized privacy overlay (DePO) help address the concerns identified in RQ1?
RQ3 What new concerns might a DePO introduce for queer and BIPOC populations when sharing photos on centralized OSNs?

We found that our participants focused on three central threats when considering sharing photos on cOSNs: **threats within social circle**, or specific people with whom they were connected on a cOSN; **threats outside social circle**, or strangers with whom they were not directly connected; and, **institutional threats**, like Facebook itself or state agencies. Participants associated a number of concerns with these threats, ranging from fear over leaking personal information, wrongful moderation, data permanence and searchability. We also found that participants were generally positive towards Image DePO; several stated that it helped address concerns over inappropriate access to the content they shared by people outside of their intended audience, that it helped prevent strangers from searching for content shared in the past, and that it strengthened their sense of control over the content they shared on Facebook. Some stated that Image DePO could help when, e.g., sharing photos of participation in protest or for organizing demonstrations.

In general, however, when deciding whether to share a photo, we found that participants were more concerned about interpersonal social threats — a person or group of people who participants did not want to see a photo — than the institutional threats Image DePO was specifically designed to address. Therein lies an important disconnect between we who design anti-surveillance technologies and the general cOSN-using public: protection against institutional harms like surveillance and censorship in sharing original content online is a *secondary* concern, much like all security and privacy concerns [38]. One may value the properties of security and privacy when sharing a photo, but one's primary concern in sharing that photo is not to be secure or private but to relay a message or solicit social support, etc. Accordingly, we argue that in order to see more widespread use and facilitate gradual decentralization, DePOs must be designed to align protection against secondary institutional threats with protection against primary social threats.

In short, we make the following contributions in this work:

(1) We developed of a proof-of-concept DePO, Image DePO, that allows participants to share secret images through a decentralized CDN over Facebook.
(2) We identified the lived threats that our queer and BIPOC participants considered when posting images to cOSNs like Facebook, as well as the mitigating behaviors they took to protect their content from those threats.
(3) We evaluated Image DePO and synthesized design considerations for the development of future DePOs to help reduce the impact of surveillance through gradual decentralization of online social networks.

2 RELATED WORK

Casemajor *et al.* defines active non-participation as the "politically willful engagement in a platform in order to slow it down or disrupt it, or exiting the platform entirely due to active decision making." In this definition, there are three methods of active non-participation: *obfuscation*, *sabotage*, and *exodus* [24]. Here, we briefly survey related work in a form of active non-participation called *obfuscation*.

We also survey prior work on minority perspectives and analyses to inform the design and evaluation of Image DePO—centering the voices of queer and BIPOC folks in its design and evaluation.

2.1 Active Non-Participation through Obfuscation

Obfuscation is the deliberate addition of ambiguous, confusing, or misleading information to interfere with surveillance and data collection with the goals of buying time, providing cover or deniability, evading observation, interfering with profiling, or expressing protest [22]. Obfuscating tactics have been used throughout human history by marginalized populations to communicate under the gaze of powerful centralized institutions — so much so that Brunton and Nissenbaum have called obfuscation the "weapon of the weak." Obfuscation can take many forms but here we focus on prior work in image obfuscation.

2.1.1 Stateful Image Obfuscation. Stateful image obfuscation involves splitting an image the user desires to be protected–the secret image–into different observable "states". Ra *et al.* obfuscate sensitive images by splitting the image into two parts, one public which is posted to social media and the other which is saved in a cloud provider (e.g. Dropbox) with the two pieces only combined if a viewer has to appropriate permissions [83]. He *et al.* similarly encrypted only portions of an image (e.g., face) to be decrypted by the key-holder [51]. These approaches allow for protection of shared images, though they are simple to detect by a casual observer as the images are highly perturbed and thus simple to disrupt by a central authority.

2.1.2 Stateless Image Obfuscation. Stateless image obfuscation is a WYSIWYG form of information hiding that redacts sensitive parts of an image while maintaining its overall cohesion. Hassan *et al.* presented a tool to detect sensitive portions of an image and overlay a cartoon image to censor it [50]. In a qualitative evaluation of their tool, Hasan *et al.* found that users make a trade-off between the aesthetics of an obfuscated image and the protection it provides [49]. One major drawback of the "stateless" approach is the large amount of perturbation to the original image, effecting its aesthetic and making it simple to detect by a casual observer. With Image DePO, we were interested in covert obfuscation that does not affect the aesthetics of the secret image and is non-obvious to a cOSN.

2.1.3 Steganographic Obfuscation. Steganography is a method of obfuscation in which data is hidden within a "cover" medium (e.g., photo, document, video). The effectiveness of steganography is determined by three heuristics: data retention, data detect-ability, and data capacity. As with any adversarial method, work into making data less detectable animated a forensic effort to detect and uncover steganographic messages. This kind of "arms race" is familiar to those within the security community, and is an ongoing relationship in the steganography community [62]. However it is important to recall the goals of obfuscation are transient in nature—evade, obstruct, or buy time and not necessarily to *permanently* protect. With this said, there is a rich set of literature for obfuscation schemes that target OSNs. With FaceCloak, Luo *et al.* obfuscated all text fields on Facebook, hot-swapping the fake values with real ones from a proprietary server [66]. Directly related to FaceCloak, NOYB protects Facebook user profile data by shuffling "atoms" of data among NOYB users and disentangling them by reference via a public dictionary file – salted with non-user data to hinder mining by adversary [47]. Hummingbird by Cristofaro *et al.* provided privacy on Twitter by running a parallel service to encrypt a users tweets and decrypt them for approved followers [32]. Lucas and Borisov introduced FlyByNight, a Facebook application for the encryption of direct messages [65]; accomplished by registering as a Facebook app and running a proprietary server. Beato *et al.* presents a tool for hiding links to data on a publicly addressable storage service; however, this scheme still relies on third-party "hashmap services" (e.g., TinyURL) for content delivery [10]. Image DePO is unique from other OSN obfuscation approaches in that it does not require the user to trust or rely on a separate third party institution — IPFS is a P2P protocol, not a service — and Image DePO's use is hidden from casual observers.

2.2 Decentralized Online Social Networks

Beyond masking and obfuscating personal data shared through cOSNs, other prior work attempted to spur an *exodus* from cOSNs to dOSNs. Paul *et al.* surveyed 16 dOSNs in the literature and provided insights to the possible implications of dOSNs if widely adopted including problems around *resource provisioning*, *profile availability*, and *functionality* that cOSNs are currently better suited for [78]. This analysis covered exploratory work proposing new dOSNs like Cachet [73], DECENT [57], Persona [3], Safebook [33], and Vis-a-Vis [89]. Each of these exploratory platforms presented a blueprint for dOSNs, though their approaches and evaluations were mainly technical and did consider human factors. In a critical analysis of the failings of dOSNs, however, Narayanan *et al.* argue that pure technological innovation is not enough for adoption and human factors must be considered [72].

There have been several attempts to bring dOSNs to market with Mastodon[2] as one of the better known solutions—a decentralized Twitter-like OSN. However, Mastodon has failed to gain mainstream appeal mirroring the fate of other commercial dOSNs (e.g., Diaspora[3]). These commercial failures can be attributed to the requirement for technical knowledge that cOSNs abstract away (e.g., hosting a server) [78]. This complexity in usability is compounded by the high cost users feel about leaving an OSN they have invested in—the so-called cold start problem that all new OSNs encounter be they centralized or decentralized [6, 8, 9, 53, 63, 79]. DePOs curtail the cold start problem by allowing users to remain on their cOSN of choice; they add the ability to share select content with their existing social network on the cOSN through a decentralized CDN.

2.3 Queer and BIPOC folks' use of and experiences with technology

While OSNs have been invaluable for the community building, organizing, and celebration of BIPOC and queer users, the processes cOSNs subject their users to have caused out-sized harm to these communities. Prior work has documented said harms, pushed for more equitable practices, and analyzed the sociotechnical effects of technology on marginal groups [76, 94]. Technology, including cOSNs, codifies discrimination in the way they are built and function by centering design around "normal" users (read cisgender, heterosexual, and white) while still being considered "value-neutral" and "objective" [12, 19, 88].

The disproportionate surveillance of black and brown individuals aided by cOSN data collection and aggregation reflect how supposedly "value-neutral" systems can harm minority populations [20, 75]. Noble's "Algorithms of Oppression" discusses how the outcomes of aggregated data analysis often reflect societal attitudes as opposed to truths (e.g., criminalized black men and sexualization of black girls in Google search) [74]. Similarly, queer users often find themselves more heavily moderated, scoring as more "toxic" than racist content due to the use of "mock impoliteness" in queer communities [36]. Prior work has also shown queer folk — especially transgender individuals — feel at risk of harassment even in safe spaces on public platforms [87]. Compounding the issue, queer users still "in the closet" are at risk of "being outed" to people online and offline hostile to their identity because of normative nudges towards full disclosure on OSNs [30].

However, even with these online struggles both queer and BIPOC users have been able to carve out portions of OSNs for themselves and even make the systems work for them [60, 70], a tradition not unfamiliar to the marginalized [13]. Queer users have found social technology useful in exploring

[2] mastodon.social
[3] joindiaspora.com

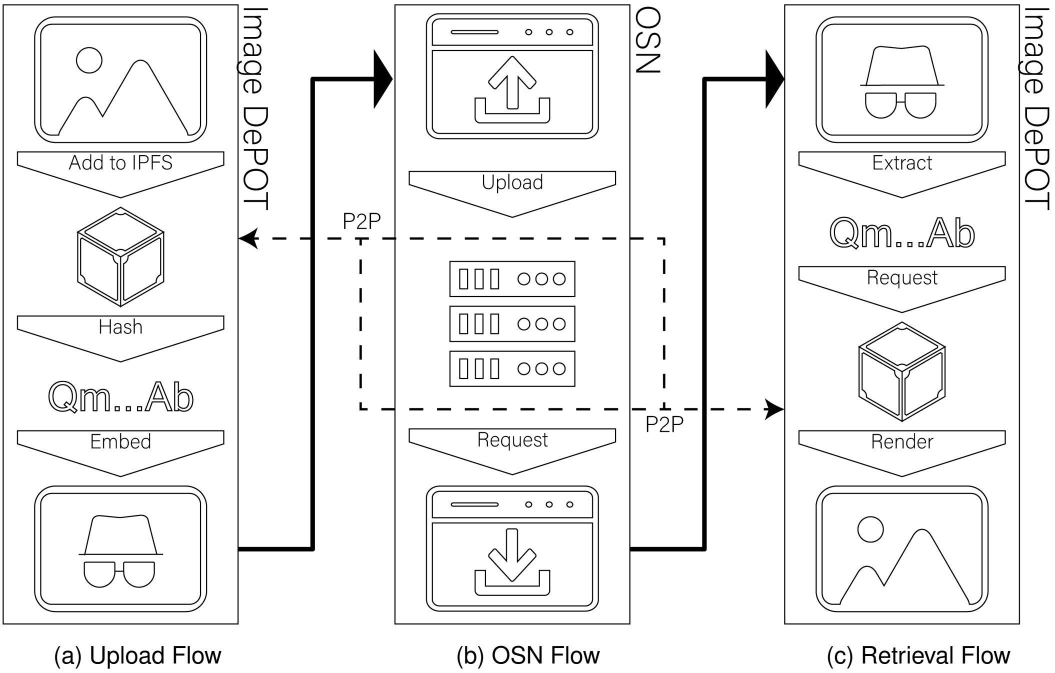


Fig. 1. System diagram with the three stages of Image DePO

and developing their sexual identity while living in a heteronormative or anti-queer environment [48, 67, 68]. Similarly, through OSNs, BIPOC users can interact and exchange culturally relevant content as Brock illustrates with "Black Twitter" as an example of technology use within systems that often reinforce long-standing racial practices [19]. Given that marginalized folks — like those who belong to queer and BIPOC communities — disproportionately bear the costs of institutional surveillance, we centered their voices in the design and evaluation of Image DePO.

3 SYSTEM DESIGN AND IMPLEMENTATION

We implemented Image DePO as a proof-of-concept decentralized private overlay (DePO) for sharing secret images on Facebook. We focus on Facebook because it remains the largest and most popular cOSN on which photo sharing is a popular use case. Importantly, our goal in this work was to use our implementation of Image DePO as a design probe through which we might explore and evaluate the utility of DePOs from the perspective of end-users; as such, we have not yet released it for public use nor do we claim that it is, in its present form, a reliable means of countering surveillance.

Image DePO is a browser extension that allows users to share secret images on Facebook through IPFS — a peer-to-peer data delivery protocol. To minimize disruption to the end-user experience of Facebook, Image DePO injects interface elements into the Facebook web user interface. When users decide to share a secret image, they click on a button injected into Facebook's web composer element titled: "Share with Image DePO". Users select the image they would like to secretly share, and then this image is served with a locally running IPFS node. We then steganographically encode the secret image's address into an automatically generated cover image and this cover image is uploaded onto Facebook's servers. Viewers with Image DePO installed can see the secret image — it will be dynamically hot-swapped for the cover image by Image DePO, which will look for an encoded address in any image that Facebook loads onto the user's browser. Facebook and viewers who do not have Image DePO installed, however, will see only the cover image. Note that we do not add an additional layer of audience selection onto Image DePO; any viewer who has Image DePO installed and who can see the cover image on Facebook through the viewer's chosen privacy settings will be able to see the secret image.

The core threat model we aim to address with Image DePO is a cOSN that hosts and distributes content generated by end-users to their connections and other users of the platform. Here, we assume the target cOSN is a *passive warden* – a steganography adversary who passively notices and disrupts a hidden message [26]. We assume the cOSN can modify the images that are uploaded onto its servers and indeed do compress them, but has an incentive to keep images aesthetically similar to the original — failure to do so might result in end-users migrating off the platform to one that does not modify their content, for example. Given this threat, then, two key design goals for Image DePO included resilience against naïve detection and corruption of secret content in uploaded cover images. Our implementation of Image DePO focused on Facebook as the oppositional cOSN — but there is nothing specific about the broad approach unique to Facebook.

In short, in designing Image DePO, we were driven by the following three design goals:

- Secret images should be distributed via Facebook in a decentralized manner.
- Cover images should be resilient against naïve detection and corruption of secret content.
- There should be minimal disruption to the user experience of uploading and viewing images on Facebook.

3.1 Decentralized Content Delivery

This significantly lowers the technical knowledge necessary for the average user to use the decentralized web, allowing them to easily "pin" files to their node for serving on the network.

To avoid reliance on a new trusted, centralized server for hosting and distributing secret images we used a distributed file sharing protocol called IPFS. IPFS is a peer to peer file-sharing protocol that consists of a swarm of interconnected nodes. While traditionally content is requested with a HTTP address indicating *where* to find it, IPFS makes requests using a 46 character string representing the cryptographic hash of the content. This hash is known as the "content identifier" (CID) and is both unique and immutable for the data it represents. Any node hosting a piece of content can respond to a request, delivering content through a peer-to-peer connection with the requester. Each IPFS node has both ephemeral and persistent data it serves. A node that received content from a given CID in turn begins to serve shards of the content until the garbage collector removes it from the data store. On the other hand, content that one wishes to keep alive can be "pinned" to the node, serving the whole file and keeping the garbage collector from removing it. Using IPFS is participatory and collaborative with no economic incentive built in, if nobody using IPFS is hosting content identified by a CID it is inaccessible. Fortunately, IPFS nodes can run on a wide range of devices like phones, personal computers, servers, or even within a web-page using JavaScript. IPFS was chosen over other distributed file delivery systems as it requires little technical knowledge for users install, self perpetuates content such that availability increases as it is viewed by more people, and costs users nothing beyond some hard-drive use.

3.2 Resilience Against Detection and Corruption

Given that Image DePO is reliant on Facebook to host and distribute the cover image, it is imperative that these images are viewed as no different than any other image that may be uploaded on to Facebook. Thus, in encoding secret content into the uploaded cover image, we must ensure that there are no obviously discernible artifacts either visually or at the byte level. These requirements led us to explore steganography — methods to conceal secret data within publicly view-able content. We chose the YASS steganographic method — based on the F5 algorithm [92, 103] — which encodes secret data into the discrete cosine transform (DCT) coefficients of images in the JPEG format. Newer steganographic methods are more resistant to steganalysis as they use deep learning [28]. However, implementing a more advanced method needlessly complicates the design probe as Facebook is considered a passive warden.

We steganographically encode the CID of the secret image. Given our use of steganography, one might ask: why not skip IPFS altogether and steganographically store the secret image in the cover image? In steganography, there is an inverse relationship between visual degradation and the number of bits hidden in a cover photo. As such, embedding a secret image into a cover image requires it to be substantially larger than the secret image. This constraint poses a practical challenge — Facebook compresses images that are above a certain size (section 3.4). Thus, the secret image would need to be significantly scaled down to be directly encoded in a cover image without causing obviously discernable artifacts in the image.

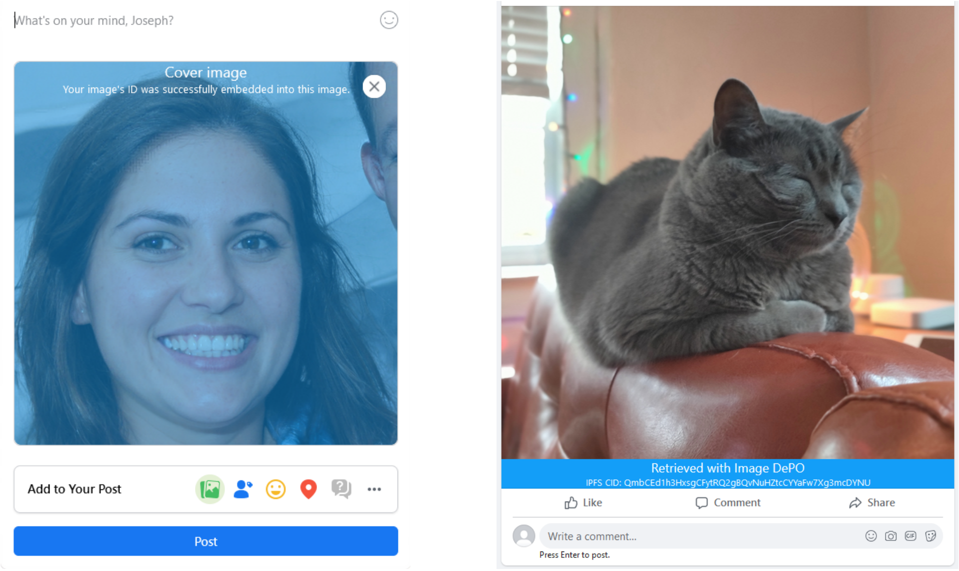
By storing a fixed-length address, instead, we evade this problem by reducing the amount of data to be hidden and fixing the size of the hidden message. Moreover, doing so affords us space for redundancy, to help ensure the encoded secret information can still be recovered even through the compression process, as explained in the next section.

3.2.1 Avoiding Secret Corruption. When an image is uploaded onto Facebook, it is transformed, compressed and manipulated which can corrupt an encoded secret. To avoid this corruption, we conducted a forensic analysis of Facebook's image processing pipeline and designed the secret embedding process to minimize the risk. As Facebook's image processing pipeline is not public, we employed the use of image forensic analyses as demonstrated in prior work [1, 23, 25]. We built a corpus of images with different sizes and shapes we uploaded to Facebook and retrieved to determine changes made to the platform since Amsden and Chen's analysis in 2015 [1]. We combined knowledge from prior work and our own empirical experiments to profile the optimal meta-attributes (e.g., size, shape, compression) of a cover image (section 3.4).

In our design, we also focus on avoiding naïve detection and corruption of the encoded secret. Although YASS has well documented methods of detection that could be used against Image DePO cover images to actively detect their upload by Facebook [59], recall we consider Facebook to be a passive actor. The simplest form of steganalysis is a byte-by-byte comparison of a cover image against an original, a function that can be made client-side by the Facebook interface (e.g., block upload of *the* Image DePO cover image). With this in mind, we took the approach of using a freely available Generative Adversarial Network service, https://thisxdoesnotexist.com. This approach ensures each cover photo from Image DePO is unique and cannot be compared against an original.

3.3 Minimizing Disruption to the User Experience

Finally, given that users often struggle with and reject security and privacy solutions that encumber or complicate their use of computing systems [15, 38, 43, 43, 86, 104], a core design goal for Image DePO was to minimize disruption to the user experience of Facebook. Accordingly, we directly integrated all end-user touch-points for Image DePO within Facebook's web interface. To facilitate use of Image DePO for *uploaders*, we inserted a "Share with Image DePO" button into Facebook's new post composer element — ensuring that using Image DePO would require little additional effort once the decision to share a photo on Facebook is made. Conversely, to simplify the *viewer* experience, Image DePO dynamically swaps the cover image (delivered by Facebook) with the secret image (delivered by IPFS). Thus, after its initial installation, Image DePO requires little-to-no



Cover image

Your image's ID was successfully embedded into this image.

Retrieved with Image DePO

IPFS CID: QmbCEd1h3HxsgCFytRQ2gBQvNuHZtcCYYaFw7Xg3mcDYNU

Press Enter to post.

(a) (b)

Fig. 2. (a) Image upload process result 1a. (b) Image retrieval process result 1c.

additional effort on the part of the user. We offer a more detailed explanation of the user experience in the sections below.

3.3.1 Image Owner. The steps to share a secret image with a cover image uploaded to Facebook are as follows (fig. 1a):

(1) The user clicks the "Share with Image DePO" button injected into the new post form (fig. 3).
(2) The user chooses the image they wish to share from the file explorer.
(3) The user is presented with the post creation dialogue with their cover image and can tag, comment, etc., as normal (fig. 2a).
(4) The user clicks "Post", sharing the cover image.

The inserted Image DePO button when selected acts as an HTML file input element. Upon the selection of a file, it is "pinned" to the locally running IPFS node. The extension retrieves a randomly generated photo from the site https://thisxdoesnotexist.com/—a repository of hosted Generative Adversarial Networks. Using a modified JPEG encoder that implements YASS, the CID is embedded into the generated cover image. The resulting cover image is then programmatically added to Facebook's input form. This triggers Facebook's proprietary events presenting the user with Figure

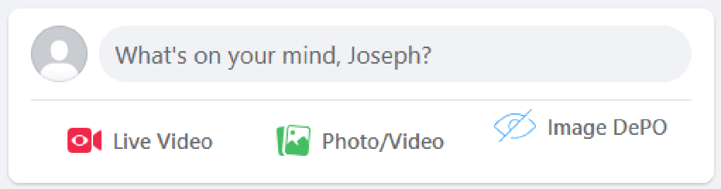


Fig. 3. Facebook posting controls with Image DePO installed.

2a with an injected image overlay. From this point, sharing is like any other image where users can provide a caption, tag a user, and so on.

3.3.2 Image Viewer. Image DePO requires no direct action on the part of the viewer. Images from Image DePO show as native to the platform with only an added overlay element to indicate they were retrieved from Image DePO. The full process of image retrieval is summarized in Figure 1c. When the Facebook interface attempts to fetch an image from their servers, the extension intercepts this request. The full-sized image is instead requested and using the modified JPEG encoder the image is scanned for an embedded CID. If a CID is not found, the retrieved image from Facebook is forwarded to the interface. Otherwise, IPFS is queried with the retrieved CID and the resulting image is inserted in place of the cover image. The end result of this process is seen in Figure 2b.

3.4 Forensic Analysis of Facebook's Image Processing Pipeline

In order to encode secret content into cover images in a manner that is unlikely to be corrupted by Facebook's image processing pipeline, we required an understanding of how Facebook's image processing pipeline manipulated the byte-content of uploaded images. To acquire this understanding, we conducted a forensic analysis on images uploaded onto Facebook.

Hiney *et al.* performed a similar forensic analysis in 2015 [55]. However as Facebook is an ever-evolving platform, it was deemed likely that the image upload pipeline had changed. Facebook, as with most other platforms, makes use of the JPEG format—a lossy compression format based on discrete cosine transforms (DCTs) to convert an image to the frequency domain and Huffman encoding. Our forensic analysis aimed to find the image size, compression ratio, and quantization tables that minimize perturbations to the DCT coefficients of an uploaded JPEG image. Minimal DCT coefficient perturbation is necessary as YASS encodes data into the DCT coefficients. To perform our analysis, we used a subset of 500 randomly selected images from the Flickr 8K dataset [56]. The goal of our analysis was to develop a compression function $C_{ImageDePO}$ for an image x such that:

$$C_{Fb}(x) \approx C_{ImageDePO}(x) \tag{1}$$

JPEG images are stored as DCT coefficients in three different channels: Y, U, and V. The Y channel represents image's luminance while the U and V channels indicate the chrominance of the image. It is common for JPEG image steganography techniques, and indeed YASS, to hide data into Y channel. In our study, images were compared with a suite of tools including ImageMagick's "Identify" tool [100], the "findimagedupes" linux package, and a Matlab script that provided the diff between two JPEG image DCT coefficients.

Intuitively, $C_{ImageDePO}$ should closely match C_{Fb} to avoid perturbation. To confirm this intuition we needed to examine if an image re-uploaded to Facebook is approximately the same after re-compression:

$$\begin{aligned} C_{Fb}(x) &= x' \\ C_{Fb}(x') &= x'' \\ x' &\approx x'' \end{aligned} \tag{2}$$

Thus, all 500 images of varying sizes were uploaded to Facebook and their full-size server counterparts were downloaded. A random set of 20 images were then re-uploaded and downloaded on Facebook. Using the aforementioned tools, we compared x' and x'' and found only minimal differences with DCT coefficients at maximum perturbed ±1.

Additionally, we experimented to find the range of x with respect to its pixel dimensions, finding:

$$\max height(x), width(x) \leq 2048\text{px} \tag{3}$$

With the confirmation that $x' \approx x''$ and knowledge of the size constraint, we developed $C_{ImageDePO}$ to closely mirror C_{Fb} while embedding a secret message.

JPEG compression consists of quantization and sampling steps and we found all images uploaded to Facebook used the same quantization table. Similarly, a sampling of 1x2x2 is applied to all images uploaded to Facebook. This means no sampling is applied to Y channel DCT coefficients — the channel we are interested in. Hence, we developed a native JavaScript YASS [92] library that embeds and extracts messages while compressing with Facebook's quantization table and sampling scheme.

4 USER EVALUATION

We qualitatively evaluated our working proof-of-concept of Image DePO by allowing participants to use the prototype in a controlled, test-environment and conducting follow-up semi-structured interviews and free-response surveys. Our goal in evaluating Image DePO was not to prove its immediate effectiveness as a tool to subvert cOSNs but to gain insights into how DePOs, more broadly, might be designed to address the lived threats that participants consider when sharing photos on cOSNs. As noted in prior design research in HCI, the act of building and designing a working system used by real people can help foreground design insights that are otherwise difficult to attain [107]. As a reminder, our goal was to answer the following research questions:

RQ1 What threats, and mitigating strategies thereof, do queer and BIPOC participants consider when sharing photos on cOSNs?

RQ2 How might a decentralized privacy overlay (DePO) help address the concerns identified in RQ1?

RQ3 What new concerns might a DePO introduce for queer and BIPOC populations when sharing photos on centralized OSNs?

4.1 Participants

ID	*Age*	*Ethnicity*	*Gender*	*Queer*
1	32	Asian	Man	-
2	34	Black	Woman	✓
3	31	Asian	Woman	-
4	26	White	Woman	✓
7	49	Black	Man	-
8	21	Asian	Woman	✓
9	50	White	Man	✓
10	38	Asian	Man	-
11	30	Asian	Non-binary	✓
12	44	White	Woman	✓
13	34	White	Woman	✓
14	51	White	Man	✓
15	27	White	Woman	✓
16	23	Asian, White	Woman	-
17	35	Asian	Man	-
18	29	White	Woman	✓
19	32	Asian	Man	-
20	31	White	Woman	✓
21	31	Black	Woman	-

Fig. 4. Participant demographics

Prior work suggests that the costs of surveillance are disproportionately borne by those from minority, marginalized and non-dominant groups [20, 60, 69]. By centering the voices of those who stand to gain the most out of subverting the surveillance and censorship capabilities of cOSNs, we should uncover insights that will help the design of DePOs for all users. To that end, we recruited participants who identified as queer or BIPOC from the Prolific platform — two communities with whom several of the authors of this paper identify. Prospective participants were only told that they would be evaluating a browser extension, but were not given details about what the extension did. Eliding that a study is specifically about privacy in recruiting participants is a common strategy in the usable privacy literature to avoid a self-selection bias in recruiting (see, e.g., [18, 40]). In total, 19 participants were recruited: an even split of 11 queer and 11 BIPOC with some overlap. Participants 5 and 6 were disqualified from the study as they identified as white and queer in prescreening but identified as non-queer during the interview.

A majority of the respondents were between the ages of 21 and 35 with a range of 21 to 50 years old. A single participant identified as non-binary while 11 identified as women and 7 identified as men. Eight participants identified as white, seven identified as Asian, three as Black, and a single participant identified as both Asian and white. Participants self reported a mean computer literacy of 3.95 on a linear scale from 1 to 5. Participants were familiar with browser extensions, with only one participant indicated having 0 and the rest reporting between 1 and 6 installed. Additionally, the majority of our participants indicated that they used the Google Chrome browser (a statistic that mirrors wider adoption statistics of browser usage).

4.2 Procedure

We carried out our evaluation in three parts during a one-hour session: an intake survey, a semi-structured interview that involved participants trialing Image DePO in a controlled context, and an exit interview.

4.2.1 Intake survey. At a scheduled time, each participant was asked to complete a web survey recalling at least three instances in which they wished to share a photo but chose not to or reconsidered and deleted. To protect the privacy of our participants and allow for a wider set of situations the participant may consider reporting, we elected to ask for an image description as opposed to a file upload.

For each photo, the participant was asked which site they wished to share it on and why they decided not to share it. Participants were also asked if they shared the photo some other way (i.e., via SMS, Email, Messaging App, Encrypted Messaging App, and Other).

4.2.2 Trialing Image DePO and semi-structured interview. After the intake survey, the participant was connected to a researcher who carried out a semi-structured interview while having the participant use the extension. To create a controlled environment compatible with remote participation, we employed a pre-configured virtual machine. The interviewer greeted participants and gave a short introduction, explaining what the extension does and its decentralized nature. Following this, the interviewer engaged the participant in a discussion about considerations they had in their decision to not share the images they described earlier, the level of control they feel over their data online, and what steps they take to protect themselves online.

Participants were then connected to the virtual machine via remote desktop software. The interviewer asked the participant to download an image of their choice (e.g., dog, cat, flower) and then share the image using Image DePO on Facebook. We employed a study-specific test Facebook account to protect our participants' privacy and abide by the recruitment platform's terms of service. Next, the interviewer asked the participant to express any confusions, frustrations, or other thoughts they had using Image DePO. Finally, the interviewer observed the use of the tool, only directing when requested.

When finished with the extension walk-through, we gathered participants' immediate reaction to Image DePO. The interviewer investigated Image DePO's usability to identify concerns missed in development. Next, the interviewer brought the discussion to considerations and concerns mentioned earlier by the participant and whether Image DePO addressed them. Interviewers attempted to gain concrete reasons behind participant responses as opposed to simply accepting affirmations. After this, the discussion was driven to how the participant felt about the decentralized component of Image

DePO. In the end, participants had the opportunity to express any feedback about Image DePO not covered by the discussion.

4.2.3 Exit survey. The exit survey revisits the photos described in the intake survey after using Image DePO. Participants are asked directly if they think they would use Image DePO for these images and questioned about for what purpose they would and would not see themselves using the extension's features. Finally, a text box for general feedback is provided.

4.3 Thematic analysis

Interviews were transcribed using the https://otter.ai speech to text tool with the resulting transcripts manually corrected. The lead researcher read the resulting transcripts and short answers, extracting excerpts relating to the research questions. The excerpts were then classified into categories using open coding by the lead researcher which were then discussed and focused in cooperation with the secondary researchers. This process was repeated until a focused set of excerpts and codes categorizing the excerpts was achieved. Next, the lead researcher identified patterns in the codes and grouped them according to emergent narrative themes (i.e., thematic analysis [17]). These themes were again discussed and focused in cooperation with the secondary researcher over multiple iterations ensuring cohesion. Finally, the resulting codes and themes were taken back to the data to evaluate their validity when applied in the wider context of the transcript or free response. This whole process was iterated upon until the themes and codes were stable between iterations.

5 FINDINGS

We next present our findings on the threats considered by our participants at the time of posting a photo, their current strategies for mitigating said threats, how Image DePO might address these threats, and new concerns Image DePO might raise. Overall, we find: 1) our participants' lived threats and mitigation strategies largely align with prior work, 2) Image DePO addresses much of the threats and concerns to give participants a greater sense of control over the data they share, and 3) Image DePO is missing key features for users to consider using it regularly such as customizable cover images and access control.

5.1 Threats that inhibit photo sharing on online social networks

We started by asking participants to reflect on salient threats that inhibited their sharing of specific photos on a cOSN. This understanding was necessary in order to contextualize participants' responses to Image DePO. Our findings in this section help confirm and extend many of the findings from prior work examining sharing behavior, boundary regulation, context collapse, regrets and self-censorship on social media.

We found that participants primarily considered three threats that inhibited photo sharing on cOSNs: **threats within their social circle**, **threats outside of their social circle**, and **institutional threats**. In turn, each threat was associated with a number of concerns, mostly centered around who might access an image that might be shared. Finally, to mitigate these concerns in sharing photos on cOSNs, participants employed a number of ad-hoc strategies including: leveraging inbuilt privacy controls, managing multiple accounts, employing judicious selection of friends and followers, using direct messaging, and practicing self-censorship.

Table 1 lists concerns raised by participants who were reluctant to share photos.

5.1.1 Threats Within Social Circle. In deciding whether or not to post an image, participants revealed how consideration of people within their social circle influenced the decision. Specifically, the decision to post a photo largely relied on a consideration of a subset of the sharer's social circle (e.g., work contacts, extended family) — i.e., the imagined audience [64].

Example Photo	*Reason*	*Shared*	*Concern*
My cat sitting on my lap and me trying to work on my computer.	Too many cat photos, didn't want to look like I'm not taking work seriously.	-	Disclosure
My daughter in a dance pose with her leg in the air.	The dance moms are competitive, I didn't want them judging her.	-	Disclosure
A friend and me dressed up for Mardi Gras.	My friend's wife didn't know that she was visiting me.	SMS	Disclosure
A horse I ride.	A friend is afraid of horses and the horse's owner is friends with me.	SMS	Disclosure
A pie our neighbor dropped off.	I didn't want neighbors to be jealous.	SMS	Disclosure
My son at the beach.	Location tagged and maskless.	-	Searchability
Boyfriend and me on the couch together.	Discomfort sharing life with people I'm not close friends with.	SMS	Disclosure
A group of friends taken post-dinner, in the living room.	I think people would assume things about violating COVID rules.	-	Disclosure
Henna on my hand.	Hard to share aspects of my culture.	-	Disclosure
A group selfie with friends.	People might assume I don't follow restrictions despite wearing masks.	SMS	Disclosure
My son holding a turkey drumstick this Thanksgiving.	I worry that it would give the wrong message about social gatherings.	-	Disclosure
My new condo.	Don't want people to find my home.	-	Searchability
My friend and me after we finished an escape room.	Taken during COVID, and I thought there might be a negative reaction.	SMS	Disclosure
Selfie while hiking a mountain.	People may find where I was going.	SMS	Searchability
A selfie with my new sunglasses	Afraid of this picture becoming public, I don't trust Instagram.	-	Searchability
Rayshard Brooks mural	Felt risky after what happened to protesters.	-	Searchability

Table 1. Photo examples, the reason they weren't shared on a cOSN, the alternate method if applicable, and the concern it aligns with. Threats separated in the table (top: inside social circle, middle: outside social circle, bottom: institutional).

The strongest concern associated with within social circle threats was a fear of *disclosure* to the "the wrong person" leading to conflict or judgement. Several participants had coworkers as followers on an OSN, causing the concern an image they want to share could make them look lazy or unproductive. Others worried a subset of their friends or connections may cast unfair judgement on them or the subjects of the content. Some had the concern a specific acquaintance would see a post or series of posts deemed inappropriate for the person.

In short, **within social circle threats** were often the result of context collapse [29, 35] — the conflation of multiple social networks into one on cOSNs — leading to challenges with boundary regulation [97] and impression management [82] as has been documented in prior literature studying behavioral inhibitions on social media.

5.1.2 Threats Outside Social Circle. Many participants also reported concerns about persons on the periphery or outside of their social circle (e.g., friends of friends, random internet users, hackers).

Participants were concerned about their *disclosure* of information to the public. Participants expressed concern that images they post may allow people outside their social circles to infer or make judgements on their location, beliefs, relationships, or attitude toward COVID restrictions. This is specifically attributed to outside social circle concerns as several participants mentioned the broad user base of cOSNs like Facebook and Instagram as a factor: " I'm hesitant about posting family photos on my Facebook, I'm not comfortable with strangers seeing those photos of me. I don't want to get messages or people commenting that I don't know. " (P4)

Participants were also concerned the content they hoped to keep semi-public might be found through careful *searching*. " I'm uncomfortable that someone can Google my name and find a photo of me as I have an uncommon last name. " (P4) Some participants revealed being harassed as they had a publicly findable account: " I made it public to try and get more followers. Then I kept getting a lot of weird direct messages from random guys. " (P16) However, one participant desired to be found on the cOSN but not on Google. " [I have settings so] if you're on Google, you can't search for me, but on Facebook you can find me. " (P21)

Our findings add to a long tradition of scholarship in privacy and social computing that foregrounds the tension between privacy and publicity on the social internet — our participants, like most people, both wanted privacy and protection while keeping the benefits of community, connection and fame that OSNs afford [42, 77, 95].

5.1.3 Institutional Threats. Institutional threats (e.g., corporations, governments) provided some anxiety to participants but were not at the forefront of their concerns when making photo sharing decisions.

Many participants indicated a concern about *data permanence*—the idea that uploaded content can never truly be deleted from a cOSN. These participants placed the blame on the platforms themselves, with distrust over data being truly deleted. " I can delete it but I believe once everything is on the internet, it is staying there and regardless backups are created at the server side even if deleted locally. " (P10) This uncertainty and distrust led participants to feel a sense of powerlessness over their data. " I don't think there is a lot of control [over what I upload]. When I put stuff on Facebook, people can see it for a long time. I can delete it, but it can be found if somebody really wanted to find it. I really don't trust [the OSNs]. " (P21)

Participants also expressed concern about the *searchability* of personal data, foregrounding fears around algorithmic analysis, indexing, and reprisal. One instance was the automatic person tagging by OSNs with facial recognition models of users. " I remember when that spooky stuff started happening, where it was like, do you want to tag Sherry? And I'm like, yeah, wait a second. " (P9) Also, fears around reprisal from the state due to the content of a shared photo. " My knowledge of data security affects what I take photos of. I went to a protest and left my phone in the car so I didn't even accidentally take a photo in my pocket and have that metadata. " (P13) This concern that institutions can search participants' prior posts combined with the distrust around the removal of uploaded content helps explain why many choose to self-censor.

Some participants revealed a concern around *wrongful moderation* — content a user shared is unfairly removed from the platform and restrictions placed on the poster. One participant who expressed playful rudeness or vulgarity faced consequences that made them reconsider what they could post. " One time I uploaded an [Instagram] story and I had my middle finger [up] and I got shadowbanned for two weeks. [The ban] was super quick, I was surprised I was banned over that. " (P16) Another participant was concerned about the removal of content considered to be socially important. " I've seen a lot of videos that were flagged for violence, just because someone was arrested. There was no violence... maybe some strong words here and there but the kind of thing that you'd see in all kinds of videos. " (P9)

The combined concerns of data permanence, searchability, and wrongful moderation from institutional threats appeared to have had a chilling effect on our participants use of online social networks — participants either posted uncontroversial content (e.g., nature, pets, or selfies) or not at all. While we do not necessarily advocate sharing *more* content on OSNs, it is important to note that if OSNs are only seen as platforms for sharing "uncontroversial" content, marginalized communities who are often considered controversial will be pushed out.

5.1.4 Current Mitigation Strategies. Our participants employed strategies to avoid threats and allay concerns while still unlocking some of the benefits of cOSNs.

Privacy Controls on cOSNs aided participants in directing who could access content while maintaining a large following or friends list. Where available, participants created or joined specialized groups for targeted sharing (e.g., close friends, family). " I set up the website with very strict privacy settings and when I upload a picture, I check it's shared with [only] the people I want. " (P7)

Multiple Accounts allowed participants to compartmentalize parts of their social circle with multiple accounts either on a single cOSN or across multiple cOSNs. " I'm more inclined to post [images of family and me] on Instagram [as I have less followers there than on Facebook]. " (P4)

Judicious friends/following list management goes beyond privacy control use, keeping some acquaintances out of an online social circle entirely. " I never friend current colleagues, because I like to bitch about work. I didn't want it to get back to anybody if I crossed the wrong person. Not that I'm a super contentious human. There are just people out there who like to stir up trouble. " (P12)

Direct Messaging a small, specific audience allowed some participants to avoid concerns with cOSNs altogether. Participants mentioned wanting to restrict "access", avoid moderation, the threat of search-ability and the presence of personal details in the photo as rationales for engaging in this behavior. " I prefer using platforms with... features like encrypted messages. I want to make sure that my message cannot get caught by someone else. " (P7)

Self-censorship gave participants peace of mind by avoiding engaging the risks at all. Distrust of inbuilt privacy tools, in turn, caused some participants to avoid sharing altogether. " Even if [I use encrypted messaging] I choose not to share [a photo] whatsoever because I have this mindset that my privacy can be breached or invaded and I get anxious. " (P15) Self-censorship was by far the most prevalent way participants mitigated their fears with 13 of the 19 participants explicitly mentioning it. Participants *wanted* to share personal information, but the perceived institutional and social privacy threats outweighed their desire to share.

Note many of these mitigation strategies are were found in prior work as well. Privacy control use, friend list management, maintaining multiple accounts, and direct messaging all fall are strategies of audience selection. Audience selection allowed participants to share posts to curated audiences. These results echo prior work on channel selection and audience separation as a boundary regulation strategy for sharing on social media [39, 91, 97]. Similarly, self-censorship results mirrored prior literature in social computing [34, 90].

5.2 Threats addressed by Image DePO

Understanding participants' privacy concerns with sharing photos on cOSNs, we next focused on understanding participants reactions to and experiences with Image DePO (section 4).

Many participants reported having a favorable impression towards Image DePO, mentioning that they would use it frequently to share content through Facebook if made publicly available. Note, however, that such statements cannot be trusted at face value owing to the Hawthorne effect. Our goal with this analysis, instead, was to use Image DePO as a design probe through which we could understand how a DePO might address the lived threats that queer and BIPOC folks grapple with in making day-to-day decisions about what content to share over cOSNs. As such, we primarily analyzed the rationale for *why* participants reported a favorable or unfavorable impression towards Image DePO in the sections to follow, organized by the threats participants reported as being most salient to their decision making process for sharing photos on cOSNs.

5.2.1 Threats Within Social Circle. With respect to the threats within their social circle and the concerns they entail, participants reported Image DePO would be a welcome addition to their *audience selection* toolset. Several participants mention that their understanding of Image DePO's approach simplified the sharing of semi-private content: " I recently eloped, I would probably [use Image DePO to] share photos from that event with close friends and loved ones but wouldn't want to take the time to send to them one by one. " (P1)

Other participants felt Image DePO provided increased assurance that shared content would not be distributed beyond the groups specified: " I feel more comfortable sharing with a close group of friends... I can be confident that the photo is protected and only [a] selected group of people can see it... I know that only people who have access to this group can see it. " (P7)

Improving audience selection was not a design goal for Image DePO; indeed, any one of a user's Facebook friends who has Image DePO installed would be able to see the secret shared image if they were in the original audience specified for the cover image on the cOSN. For example, if Alice shared a secret image via Image DePO on Facebook with the visibility of the cover image set to all of her friends, then any of her friends with Image DePO installed would be able to see the secret image.

That participants saw the utility of Image DePO as a means through which audience selection might be improved points to two upshots: first, the messaging of what is being protected from whom should be made explicit and clear every time a user attempts to share content through a DePO to avoid unexpected interpersonal privacy violations; second, it appears that the more salient threat to these participants, when navigating decisions about whether to share photos on cOSNs, are social and not institutional. Accordingly, focusing strictly on protecting against institutional threats may fail to attract users.

5.2.2 Threats Outside Social Circle. The concerns participants associated with outside social circle threats included leaking *personal details*, and the *searchability* of their personal data.

P16 indicated that Image DePO could help address concerns about the *searchability* of data, especially older data: " You can post personal things, but not have to worry about if you get too big, and then [your posts] are really public and people can just track you down. I feel like something like this could make me want to go public again. " (P16) The temporal challenges of setting appropriate access control policies to content shared on cOSNs is a longstanding challenge that has been well studied in the usable privacy and social computing literature [4, 77]. While not an original design goal, Image DePO can at least partially protect against data against automated searching by strangers — as P21 noted: "since Facebook can't see it, [strangers] can't see it either."

Indeed, many of the concerns related to *searchability* that our participants raised were, in large part, enabled or facilitated by the cOSN through which the photos were shared — e.g., search functionalities and public-by-default content timelines. If the cOSN does not have direct access to the shared image, it cannot facilitate searches that lead to that image.

5.2.3 Institutional Threats.

The concerns that comprised institutional threats included *surveillance*, *data permanence*, *searchability*, and *wrongful moderation*. Despite the fact that Image DePO was explicitly designed to address these threats, fewer participants mentioned protection from institutional threats when discussing their impressions of Image DePO.

Participants who *did* discuss institutional threats felt a lack of control when posting to Facebook, and stated that Image DePO increased their control over their data due to its decentralized distribution model: "with Image DePO, I own my data." (P19) and "this gets around the platform in a way that private postings don't." (P9)

P11 felt that Image DePO helped address concerns of *surveillance* specifically: "I would use Image DePO for practically all of the photos I share on social media websites as long as [OSNs] are unable to steal/gather info from the images I share." (P11)

Several participants referenced using Image DePO for sharing images depicting " civil unrest or controversial police activity. " (P9) Additionally, some participants felt Image DePO would be useful for "planning an event like a protest or rally." (P18) Image DePO, thus, could help people spur and/or participate in collective action with reduced fear of institutional surveillance.

P9 felt Image DePO sufficiently addressed the issue of *wrongful moderation*. " [Videos are] getting taken down for what seemed like spurious reasons and that could be a result of an algorithm that wouldn't see it if shared with Image DePO. " (P9) This participant generalized Image DePO for sharing video content on a platform like YouTube. While beyond the scope of Image DePO, one could imagine a Video DePO that enables such functionality — allowing YouTube users to watch de-centrally distributed videos instead of needing to immediately migrate to a less active, decentralized video streaming platform like DTube.

Participants considered concerns over *data permanence* and *searchability* in tandem when assessing the utility of Image DePO. In explaining why they would use Image DePO, P4 said: " I feel like once Facebook can see the image, the ownership of it goes to them. They can save it forever and share with other parties. For example, an image I posted on Twitter is on Google now forever from a different website and I cannot take it down. " (P4) P20 was concerned what the permanence of data shared on a cOSN could mean for children and felt Image DePO could help address it: " Is it ethical to be posting pictures of kids? In 20 years they'll have to reckon with those pictures since they're findable on the internet forever. " (20) Participants felt that Image DePO could combat against both of these threats: "it feels like it's not just putting images out there forever." (P21)

Some participants (like P16) felt that Image DePO would allow them to engage in active non-participation to protest against institutional threats. " I definitely think privacy breaches and having our data sold is such a real problem. I'm into combating it. I think being aware of what they do is the first step but then I get unfortunate feeling like I have to limit myself in what I can post. But something like Image DePO would help. " (P16)

In summary, while the cOSN was not at the forefront of participants' threats when making decisions about sharing photos on social media, participants generally found Image DePO to be a useful tool that afforded them a greater sense of control over their personal data. Other participants mistakenly believed that Image DePO could provide protection against interpersonal threats, suggesting that these threats may be more salient for those participants at the moment of decision making. However, this mistaken assumption could result in privacy violations if Image DePO were used without appropriate expectation setting about what data is being protected from whom. We discuss the implication of these results for the design of future DePOs.

5.3 Threats resulting from Image DePO

Novel technologies — even those designed with good intentions — are rarely unilaterally good. We next explored if and how Image DePO raised new concerns for our participants.

Several participants were concerned about access control in Image DePO and the risks they perceive of decentralization. Others raised concerns about the availability of the images protected by Image DePO. Some worried that the use of generated images would affect their public-facing persona. Finally, a few participants were concerned about the ethics of a tool that allow users to bypass "legitimate" forms of moderation.

5.3.1 Data and Device Access. While some participants assumed Image DePO would afford enhanced audience selection capabilities, others recognized the lack of direct support for enhanced interpersonal audience selection as a short-coming: " I think the reasons I don't post to Facebook are general privacy and Facebook's algorithm, yes, but also who among my Facebook friends can see things? I am the daughter of a police officer, I would not want him to be able to access images on Image DePO. " (P13) This quote explicitly illustrates the mismatch between Image DePO's focus on protection against institutional threats versus the day-to-day interpersonal threats that were more salient to some of our participants when making decisions to share photos on cOSNs.

Some participants voiced concerns about the security of the IPFS protocol and the safety of their device. For example, P15 felt positively about Image DePO's capabilities "as long as [content retrieval from personal device] doesn't affect me in any way negatively". This sentiment is echoed by P16, who voiced a more specific fear that content delivery using IPFS could provide "friends & family (or harmful third party)" with "access to my device". Using Image DePO requires users to join the IPFS network, which does entail distributing data to other nodes on the network for content that a user elects to "pin" — but being a node on the network does not allow other users direct access to one's device. A top-level upshot, then, is that DePOs must clearly communicate to end-users what exactly what content is available for peer-nodes on the IPFS network to access.

5.3.2 Content Availability. Image DePO was developed for desktop browsers but cOSNs provide access to shared content on a multitude of devices. Participants expressed concerned that a decentralized approach to content delivery would reduce the availability of the content they shared to specific browsers and devices, or require re-sharing on multiple devices. P12, for example, wondered how their content might be distributed if their device was offline:"what if my device is not active?" (P12) Another concern was compatibility between different versions of Facebook or other cOSN. As there is no companion mobile application to Image DePO, yet, mobile users would only see cover images. Some worried about visibility, that people "will see the wrong thing if I want to show someone on my phone." (P16)

Many of these concerns are not necessarily fundamental limitations of DePOs but a matter of engineering effort — indeed, our implementation of Image DePO was a proof-of-concept. Nevertheless, these concerns suggests that some participants may not be ready to use DePOs until they are reliably supported on a broad range of devices.

5.3.3 Control over cover images. Many participants voiced a concern about the content of the cover image: " I am just a little thrown that one of the cover photos is a kid instead of like an object." (P20) The GAN used for cover photos generated human portraits, meaning a prolific user of Image DePO would have " a bunch of incongruous photographs of people" making them worried their account would “not make sense to someone” (P18) or worry that people would “think I’m hacked, a bot or a troll” (P14).

Participants indicated a greater comfort with generated images “without people in them, I know they have GANs for cats, landscapes, there’s all sorts of different ones.” (P14) Some participants indicated a desire for the cover image to be customizable, allowing them to “curate the cover images to reflect myself.” (P14)

More broadly, participants expressed a desire to have stronger control over the cover image selection process so that they could manage their impression towards non-Image DePO users.

5.3.4 Moderation. Some participants expressed concern that the ability for Image DePO to circumvent content moderation could be a “double edged sword,” with malicious actors able to use it too. “ I certainly want to protect those who protest. ...On the other hand...I very much want law enforcement to be able to look at [dangerous content] with a warrant. ” (P13)

That privacy-enhancing technologies might be used for anti-social ends is a common critique (e.g., encryption and the “going dark” problem [98]); we respond in more detail in section 6. It should be noted, however, that while Image DePO might prevent law enforcement from subponeaing Facebook for a given user’s secretly shared content, it will not prevent them from acquiring a warrant to search through a given suspect’s personal device. Moreover, any viewer of the secret content will also be able to save and report the content to authorities if deemed dangerous.

6 DISCUSSION

Although we envisioned DePOs, and Image DePO specifically, as curtailing the monitoring and monetizing of personal data by institutional actors, our findings suggest that while institutional actors might cause or exacerbate sharing concerns, the institutional actors themselves were not at the forefront of participants’ mind when making sharing decisions. Rather, the salient threats at the moment of decision making tended to be more social: e.g., participants were concerned about other people they knew or strangers being able to find and view their images. Through our human-centered evaluation of Image DePO, we uncovered the threats and concerns users consider when sharing photos on cOSNs like Facebook and how a DePO might or might not address these concerns. In interviewing queer and BIPOC folks about their considerations and concerns about posting photos online, we found that our participants were concerned about three categories of threats: **within social network**, **outside social network**, and **institutional actors**. Based on these findings, we distill a series of design implications for DePOs and other third-party technologies designed to lessen the impact of surveillance.

6.1 Gradual decentralization of online social networks

Shortly after the emergence of social media platforms, technology Utopians predicted a move toward decentralization [106]. However, while decentralized alternatives like Mastodon have emerged, they have thus far failed to see widespread adoption and use. Part of the challenge is that competing social networks face a cold-start problem [41]: people will only join if they can connect with friends already on the platform. In The Master Switch, legal scholar Tim Wu argues that disruptive innovations in information and communication technologies — like social networking platforms — go through a predictable cycle [105]. They often start out open, democratized, and used in favor of counter-cultural challenges to entrenched power structures (e.g., with broadcast radio). Gradually, however, as their power is more well understood, they centralize and amalgamate power into the hands of new centralized giants (e.g., AT&T’s dominance over telecommunications in the early-mid 20th century).

We can see the structure of this cycle taking form with online social networking. Indeed, over the past decade, social networking has become increasingly centralized; today, the vast majority of social media users use platforms owned by a few large corporations [53, 93].

DePOs allow for a gradual transitioning away from cOSNs: users can continue to use cOSNs, but can choose to share select secret content through dCDNs with their social networks instead. Gradually, as increasingly many users install gateways to the dCDN in order to unlock the secret content shared by their early-adopter friends, it may be possible to do away with the cOSN altogether. In short, DePOs allow for early adopters to access the benefits provided by decentralization with minimal disruption to their existing communities. Our design probe — Image DePO — is a proof of concept for realizing this vision for secret image sharing on Facebook.

6.2 Aligning Protection Against Social and Institutional Threats

One key design implication from our findings is that, in order to see widespread use and facilitate the gradual decentralization of cOSNs, DePOs like Image DePO must align protection against abstract institutional actors with protection against the more salient social threats users consider when making specific sharing decisions on cOSNs. For example, participants expressed interest in using Image DePO to selectively hide content from certain groups of friends and connections, but were sometimes less concerned about hiding data from Facebook itself. The results of our analysis recall that of Dourish et al. [38], who found that security concerns tend to be secondary — desired, but peripheral to one's primary task and any given moment.

The burgeoning public interest in Virtual Private Network (VPN)s may provide an illustrative example to follow. Both VPNs and DePOs can be thought of as client-side overlay layers that help circumvent institutional privacy threats. However, while VPNs have been in the security and privacy minded user's toolkit for decades, they have only recently seen mainstream use. This is in no small part due to the realities of institutional harms that more users are becoming aware of [71]. Even with this awareness, however, VPNs needed to become "one-click" solutions that emphasized benefits more directly aligned with users' primary concerns (i.e., access to country-specific streaming offerings on Netflix). Similarly, although the technologies underlying Image DePO — decentralization and obfuscation — are well studied and developed, DePOs must be designed in a manner in which protection against institutional threats are *aligned with or peripheral to* other, more primary concerns for end-users before they are likely to see widespread adoption.

6.3 Institutional Security vs Privacy from Institutions

A common critique of privacy-enhancing technologies is that they may be used by criminals and malicious individuals to evade law enforcement — a conundrum sometimes referred to as the "going dark" problem [98]. The concern is that criminals can avoid "legitimate" means of moderation or surveillance that keep us as a society safe with their use of privacy enhancing tools [44].

It is important to note, however, that while the threat of criminal use of privacy-enhancing technology can be harmful, the threat of institutional privacy invasions are already causing immediate real-world harm. The cycle described by Wu seems fateful with the once open, democratized, and counter-cultural internet [105] becoming a useful tool for repression. Companies who own cOSNs can be served with warrants for a user's information and are under no obligation to notify the user [84]. Repressive regimes use the centralized internet for mass surveillance and control [27, 46, 52, 58, 80, 85, 96] with the targets of surveillance and repression by institutions often being those most marginalized [14, 16, 20, 31, 37, 54, 61, 99, 102].

The ethical conundrum of privacy is and always has been that by affording an individual greater freedom from surveillance, we increase the ability for any given individual to engage in anti-social activities. However, this argument presents a false choice between privacy and security. As many privacy scholars have noted, law enforcement and intelligence institutions have many other methods outside of surveilling cOSNs to detect and counter criminal activities. For better or for worse, today, there is no meaningful way for anyone to "go dark" [44].

6.4 Limitations

Here we address limitations to our work. First, we asked participants to recall past instances when they *wanted* to share a photo on a cOSN but decided against it. People often struggle with recalling the exact reasons behind past actions. To address this concern we had participants to focus on specific images they wanted to share and grounding the discussion specifically on those images.

Second, COVID restrictions were still in place while the evaluation was taking place. Thus, participation was done remotely over video call and the walk-through was carried out via a remote machine. While remote participation is less rich than in-person participation, remote participation is now common in many forms of user-testing. Third, in our analysis we noticed many concerns voiced by participants about judgement from others had to do with breaking COVID restrictions — e.g., breaking physical distancing guidelines. Re-running the study at a different time may result in a greater diversity of censored content and, in turn, reveal new threats and concerns associated with sharing photos on cOSNs.

Fourth, the decision to focus on desktop / laptop computers was a conscious one to simplify the development of our design probe. However, if DePOs are to see widespread use, it will be necessary to implement a service that works on a broad range of devices, including mobile devices.

Fifth, we recognize that — especially in HCI — technical applications like Image DePO are often presented as a panacea to problems that are sociotechnical in nature [5, 7]. We are not arguing that Image DePO, alone, is capable of fully subverting centralized, institutional surveillance of social activity on the Internet. Rather, we present DePOs as a new addition to a toolset of privacy-enhancing technologies that can empower users to reclaim some control over the personal data they share online.

Finally, we anticipate criticisms of the decision to recruit queer and BIPOC participants without specifically centering the conversation around identity. It is important to understand a wide range of experiences and concerns for people in marginalized groups, to focus on more than just their identities. Just as whiteness, maleness, or straightness do not make up the entirety of a human's experience; sexuality, race, or gender expression are only facets of much larger systems. Examining more holistically allows us to gain a better understanding on how larger systems of identity, experience, and categorization impact interactions with and choices around technology. By actively centering the voices of people from marginalized groups in this work, we are able to understand the harms that need mitigation without reducing our participants experiences simply to their identity.

7 CONCLUSION

We introduced a new class of privacy-enhancing technologies — decentralized privacy overlays (DePOs) — that allows users to remain on cOSNs but share select secret content through decentralized content distribution networks. Our goal was to investigate future technologies where people can resist and reverse the centralization of online social networking without needing to de-platform themselves in the process. To succeed in gaining widespread adoption and lessening the unilateral control cOSNs have over users and their data, DePOs will require not only tool development but human-centered design processes that foreground and center the voices of the historically marginalized populations. To evaluate how end-users might view and use DePOs, we implemented and evaluated a proof-of-concept design probe, Image DePO, that allows people to share secret images on Facebook through the IPFS peer-to-peer protocol. Our qualitative evaluation of Image DePO with queer and BIPOC participants revealed that while participants were indeed concerned about institutional actors as threats to data agency and privacy, these threats were not the most salient when participants made specific sharing decisions. Instead, participants were generally more concerned about social threats when making individual posting decisions and Image DePO only peripherally addresses these concerns. Finally, we identified design considerations for future DePOs that should better address people's immediate sharing concerns while providing protection from and resistance against cOSNs and the other institutional threats that are emboldened or enabled by cOSNs. With Image DePO we illustrated the feasibility of DePOs as a class of privacy-enhancing technology and through our qualitative assessment with marginalized populations we identified how DePOs may find success outside of the security and privacy community.

POSITIONALITY STATEMENT

We designed our analysis to highlight the perspectives of people adversely affected by normative assumptions embedded into centralized western institutions, such as heterosexism and white supremacy. The first author is an American, cisgender, white queer man from a middle-upper class background. All the authors are deeply concerned with fortifying the safety and autonomy of individuals and groups who face systemic discrimination. We are committed to ensuring people in marginalized groups are the focus of work to mitigate the externalities and negative effects of institutions and technologies.

REFERENCES

[1] Nathaniel D Amsden and Lei Chen. 2015. Analysis of Facebook steganographic capabilities. In *2015 International Conference on Computing, Networking and Communications (ICNC)*. 67–71. https://doi.org/10.1109/ICCNC.2015.7069317

[2] Mark Andrejevic. 2011. Exploitation in the Data Mine. In *Internet and Surveillance*. Routledge. Num Pages: 18.

[3] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, and Daniel Starin. 2009. Persona: an online social network with user-defined privacy. In *Proceedings of the ACM SIGCOMM 2009 conference on Data communication (SIGCOMM '09)*. Association for Computing Machinery, New York, NY, USA, 135–146. https://doi.org/10.1145/1592568.1592585

[4] Lujo Bauer, Lorrie Faith Cranor, Saranga Komanduri, Michelle L. Mazurek, Michael K. Reiter, Manya Sleeper, and Blase Ur. 2013. The post anachronism: the temporal dimension of facebook privacy. In *Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society (WPES '13)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/2517840.2517859

[5] E. Baumer. 2007. Questioning the Technological Panacea: Three Reflective Questions for Designers. *undefined* (2007). https://www.semanticscholar.org/paper/Questioning-the-Technological-Panacea%3A-Three-for-Baumer/344139846c86e96c4854477d35e5d0ee519af9f3

[6] Eric P.S. Baumer, Phil Adams, Vera D. Khovanskaya, Tony C. Liao, Madeline E. Smith, Victoria Schwanda Sosik, and Kaiton Williams. 2013. Limiting, leaving, and (re)lapsing: an exploration of facebook non-use practices and experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3257–3266. https://doi.org/10.1145/2470654.2466446

[7] Eric P.S. Baumer and M. Six Silberman. 2011. When the implication is not to design (technology). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2271–2274. https://doi.org/10.1145/1978942.1979275

[8] Eric P. S. Baumer, Shion Guha, Patrick Skeba, and Geraldine Gay. 2019. All Users are (Not) Created Equal: Predictors Vary for Different Forms of Facebook Non/use. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 80:1–80:28. https://doi.org/10.1145/3359182

[9] Nancy Baym, Danah Boyd, Kate Crawford, Tarleton Gillespie, and Mary Gray. 2011. "If you don't like it, don't use it. It's that simple." ORLY? https://socialmediacollective.org/2011/08/11/if-you-dont-like-it-dont-use-it-its-that-simple-orly/

[10] Filipe Beato, Iulia Ion, Srdjan Čapkun, Bart Preneel, and Marc Langheinrich. 2013. For some eyes only: protecting online information sharing. In *Proceedings of the third ACM conference on Data and application security and privacy (CODASPY '13)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/2435349.2435351

[11] Juan Benet. 2014. IPFS - Content Addressed, Versioned, P2P File System. *arXiv:1407.3561 [cs]* (July 2014). http://arxiv.org/abs/1407.3561 arXiv: 1407.3561.

[12] Ruha Benjamin. 2020. Race After Technology: Abolitionist Tools for the New Jim Code. *Social Forces* 98, 4 (June 2020), 1–3. https://doi.org/10.1093/sf/soz162

[13] Lauren Berlant and Michael Warner. 1998. Sex in public. *Critical inquiry* 24, 2 (1998), 547–566. Publisher: University of Chicago Press.

[14] Sam Biddle. 2020. Police Surveilled George Floyd Protests With Help From Twitter-Affiliated Startup Dataminr. https://theintercept.com/2020/07/09/twitter-dataminr-police-spy-surveillance-black-lives-matter-protests/

[15] Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. 2012. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *2012 IEEE Symposium on Security and Privacy*. IEEE, San Francisco, CA, USA, 553–567. https://doi.org/10.1109/SP.2012.44

[16] Russell Brandom. 2016. Facebook, Twitter, and Instagram surveillance tool was used to arrest Baltimore protestors. https://www.theverge.com/2016/10/11/13243890/facebook-twitter-instagram-police-surveillance-geofeedia-api

[17] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[18] Cristian Bravo-Lillo, Saranga Komanduri, Lorrie Faith Cranor, Robert W. Reeder, Manya Sleeper, Julie Downs, and Stuart Schechter. 2013. Your attention please: designing security-decision UIs to make genuine risks harder to ignore. In *Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUPS '13)*. Association for Computing Machinery, Newcastle, United Kingdom, 1–12. https://doi.org/10.1145/2501604.2501610

[19] André Brock. 2012. From the Blackhand Side: Twitter as a Cultural Conversation. *Journal of Broadcasting & Electronic Media* 56, 4 (Oct. 2012), 529–549. https://doi.org/10.1080/08838151.2012.732147 Publisher: Routledge _eprint: https://doi.org/10.1080/08838151.2012.732147.

[20] Simone Browne. 2015. *Dark Matters: On the Surveillance of Blackness*. Duke University Press. Google-Books-ID: snmJCgAAQBAJ.

[21] Axel Bruns, Tim Highfield, and Jean Burgess. 2013. The Arab Spring and Social Media Audiences: English and Arabic Twitter Users and Their Networks. *American Behavioral Scientist* 57, 7 (July 2013), 871–898. https://doi.org/10.1177/0002764213479374 Publisher: SAGE Publications Inc.

[22] Finn Brunton and Helen Fay Nissenbaum. 2015. *Obfuscation: a user's guide for privacy and protest*. MIT Press, Cambridge, Massachusetts.

[23] Jan Butora and Jessica Fridrich. 2019. Effect of JPEG Quality on Steganographic Security. In *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security - IH&MMSec'19*. ACM Press, Paris, France, 47–56. https://doi.org/10.1145/3335203.3335714

[24] Nathalie Casemajor, Stéphane Couture, Mauricio Delfin, Matthew Goerzen, and Alessandro Delfanti. 2015. Non-participation in digital media: toward a framework of mediated political action. *Media, Culture & Society* 37, 6 (Sept. 2015), 850–866. https://doi.org/10.1177/0163443715584098 Publisher: SAGE Publications Ltd.

[25] A. Castiglione, G. Cattaneo, and A. De Santis. 2011. A Forensic Analysis of Images on Online Social Networks. In *2011 Third International Conference on Intelligent Networking and Collaborative Systems*. 679–684. https://doi.org/10.1109/INCoS.2011.17

[26] Rajarathnam Chandramouli, Mehdi Kharrazi, and Nasir Memon. 2003. Image steganography and steganalysis: Concepts and practice. In *International Workshop on Digital Watermarking*. Springer, 35–49.

[27] Chun-Chih Chang and Thung-Hong Lin. 2020. Autocracy login: internet censorship and civil society in the digital age. *Democratization* 27, 5 (July 2020), 874–895. https://doi.org/10.1080/13510347.2020.1747051 Publisher: Routledge _eprint: https://doi.org/10.1080/13510347.2020.1747051.

[28] Marc Chaumont. 2020. 14 - Deep learning in steganography and steganalysis. In *Digital Media Steganography*, Mahmoud Hassaballah (Ed.). Academic Press, 321–349. https://doi.org/10.1016/B978-0-12-819438-6.00022-0

[29] Jeffrey T. Child, Angela R. Duck, Laura A. Andrews, Maria Butauski, and Sandra Petronio. 2015. Young Adults' Management of Privacy on Facebook with Multiple Generations of Family Members. *Journal of Family Communication* 15, 4 (Oct. 2015), 349–367. https://doi.org/10.1080/15267431.2015.1076425 Publisher: Routledge _eprint: https://doi.org/10.1080/15267431.2015.1076425.

[30] Alexander Cho. 2018. Default publicness: Queer youth of color, social media, and being outed by the machine. *New Media & Society* 20, 9 (Sept. 2018), 3183–3200. https://doi.org/10.1177/1461444817744784 Publisher: SAGE Publications.

[31] Isobel Cockerell. 2019. Inside China's Massive Surveillance Operation. *Wired* (May 2019). https://www.wired.com/story/inside-chinas-massive-surveillance-operation/

[32] E. D. Cristofaro, C. Soriente, G. Tsudik, and A. Williams. 2012. Hummingbird: Privacy at the Time of Twitter. In *2012 IEEE Symposium on Security and Privacy*. 285–299. https://doi.org/10.1109/SP.2012.26 ISSN: 2375-1207.

[33] L. A. Cutillo, R. Molva, and T. Strufe. 2009. Safebook: Feasibility of transitive cooperation for privacy on a decentralized social network. In *2009 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks Workshops*. 1–6. https://doi.org/10.1109/WOWMOM.2009.5282446

[34] Sauvik Das and Adam Kramer. 2013. Self-Censorship on Facebook. *Proceedings of the International AAAI Conference on Web and Social Media* 7, 1 (June 2013). https://ojs.aaai.org/index.php/ICWSM/article/view/14412 Number: 1.

[35] Jenny L. Davis and Nathan Jurgenson. 2014. Context collapse: theorizing context collusions and collisions. *Information, Communication & Society* 17, 4 (April 2014), 476–485. https://doi.org/10.1080/1369118X.2014.888458 Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2014.888458.

[36] Thiago Dias Oliva, Dennys Marcelo Antonialli, and Alessandra Gomes. 2020. Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online. *Sexuality & Culture* (Nov. 2020). https://doi.org/10.1007/s12119-020-09790-w

[37] Angel Diaz. 2019. *New York City Police Department Surveillance Technology*. Technical Report. Brennan Center for Justice. https://www.brennancenter.org/our-work/research-reports/new-york-city-police-department-surveillance-technology

[38] Paul Dourish, Rebecca E. Grinter, Jessica Delgado de la Flor, and Melissa Joseph. 2004. Security in the wild: user strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing* 8, 6 (Nov. 2004), 391–401. https://doi.org/10.1007/s00779-004-0308-5

[39] Stefanie Duguay. 2016. “He has a way gayer Facebook than I do”: Investigating sexual identity disclosure and context collapse on a social networking site. *New Media & Society* 18, 6 (June 2016), 891–907. https://doi.org/10.1177/1461444814549930 Publisher: SAGE Publications.

[40] Serge Egelman and Eyal Peer. 2015. Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS). In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2873–2882. https://doi.org/10.1145/2702123.2702249

[41] Lisa Jo Elliott and Vera Polyakova. 2014. Beyond Facebook: The generalization of social networking site measures. *Computers in Human Behavior* 33 (April 2014), 163–170. https://doi.org/10.1016/j.chb.2014.01.023

[42] Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. 2007. The Benefits of Facebook “Friends:” Social Capital and College Students’ Use of Online Social Network Sites. *Journal of Computer-Mediated Communication* 12, 4 (July 2007), 1143–1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x

[43] S. M. Furnell, A. Jusoh, and D. Katsabas. 2006. The challenges of understanding and using security: A survey of end-users. *Computers & Security* 25, 1 (Feb. 2006), 27–35. https://doi.org/10.1016/j.cose.2005.12.004

[44] Urs Gasser, Jack Goldsmith, Susan Landau, Joseph Nye, David O’Brien, Matt Olsen, Bruce Schneier, and Jonathan Zittrain. 2016. *Don’t Panic: Making Progress on the "Going Dark" Debate | Berkman Klein Center*. Technical Report. Harvard University, Berkman Klein Center for Internet & Society. https://cyber.harvard.edu/publications/2016/Cybersecurity/Dont_Panic

[45] Ben Gilbert. 2018. How Facebook makes money from your data, in Mark Zuckerberg’s words. *Business Insider* (April 2018). https://www.businessinsider.com/how-facebook-makes-money-according-to-mark-zuckerberg-2018-4

[46] Anita R Gohdes. 2015. Pulling the plug: Network disruptions and violence in civil conflict. *Journal of Peace Research* 52, 3 (May 2015), 352–367. https://doi.org/10.1177/0022343314551398 Publisher: SAGE Publications Ltd.

[47] Saikat Guha, Kevin Tang, and Paul Francis. 2008. NOYB: privacy in online social networks. In *Proceedings of the first workshop on Online social networks (WOSN ’08)*. Association for Computing Machinery, New York, NY, USA, 49–54. https://doi.org/10.1145/1397735.1397747

[48] Jean Hardy and Stefani Vargas. 2019. Participatory Design and the Future of Rural LGBTQ Communities. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion (DIS ’19 Companion)*. Association for Computing Machinery, New York, NY, USA, 195–199. https://doi.org/10.1145/3301019.3323894

[49] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer Experience of Obscuring Scene Elements in Photos to Enhance Privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI ’18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173621

[50] Hassan, Rakibul Hasan, Patrick Shaffer, David Crandall, and Eman T. Apu Kapadia. 2017. Cartooning for Enhanced Privacy in Lifelogging and Streaming Videos. 29–38. https://openaccess.thecvf.com/content_cvpr_2017_workshops/w16/html/Kapadia_Cartooning_for_Enhanced_CVPR_2017_paper.html

[51] J. He, B. Liu, D. Kong, X. Bao, N. Wang, H. Jin, and G. Kesidis. 2016. PUPPIES: Transformation-Supported Personalized Privacy Preserving Partial Image Sharing. In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. 359–370. https://doi.org/10.1109/DSN.2016.40 ISSN: 2158-3927.

[52] Sebastian Hellmeier. 2016. The Dictator’s Digital Toolkit: Explaining Variation in Internet Filtering in Authoritarian Regimes. *Politics & Policy* 44, 6 (2016), 1158–1191. https://doi.org/10.1111/polp.12189 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/polp.12189.

[53] Kashmir Hill. 2019. I Cut Google Out Of My Life. It Screwed Up Everything. https://gizmodo.com/i-cut-google-out-of-my-life-it-screwed-up-everything-1830565500

[54] Kashmir Hill. 2020. Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match. *The New York Times* (Dec. 2020). https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html

[55] Jason Hiney, Tejas Dakve, Krzysztof Szczypiorski, and Kris Gaj. 2015. Using Facebook for Image Steganography. In *2015 10th International Conference on Availability, Reliability and Security*. IEEE, Toulouse, France, 442–447. https://doi.org/10.1109/ARES.2015.20

[56] Micah Hodosh, Peter Young, and Julia Hockenmaier. [n.d.]. Flickr8k Dataset. ([n. d.]).

[57] S. Jahid, S. Nilizadeh, P. Mittal, N. Borisov, and A. Kapadia. 2012. DECENT: A decentralized architecture for enforcing privacy in online social networks. In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*. 326–332. https://doi.org/10.1109/PerComW.2012.6197504

[58] Yury Kabanov and Mikhail Karyagin. 2018. Data-Driven Authoritarianism: Non-democracies and Big Data. In *Digital Transformation and Global Society (Communications in Computer and Information Science)*, Daniel A. Alexandrov, Alexander V. Boukhanovsky, Andrei V. Chugunov, Yury Kabanov, and Olessia Koltsova (Eds.). Springer International Publishing, Cham, 144–155. https://doi.org/10.1007/978-3-030-02843-5_12

[59] Jan Kodovský, Tomáš Pevný, and Jessica Fridrich. 2010. Modern steganalysis can detect YASS, Nasir D. Memon, Jana Dittmann, Adnan M. Alattar, and Edward J. Delp III (Eds.). San Jose, California, 754102. https://doi.org/10.1117/12.838768

[60] Ada Lerner, Helen Yuxun He, Anna Kawakami, Silvia Catherine Zeamer, and Roberto Hoyle. 2020. Privacy and Activism in the Transgender Community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376339

[61] Sam Levin. 2016. ACLU finds social media sites gave data to company tracking black protesters. http://www.theguardian.com/technology/2016/oct/11/aclu-geofeedia-facebook-twitter-instagram-black-lives-matter Section: Technology.

[62] Bin Li, Junhui He, Jiwu Huang, and Yun Qing Shi. 2011. A Survey on Image Steganography and Steganalysis. *Journal of Information Hiding and Multimedia Signal Processing* 2, 2 (April 2011), 31.

[63] B. Light. 2014. *Disconnecting with Social Networking Sites*. Springer. Google-Books-ID: RlSoBAAAQBAJ.

[64] Eden Litt and Eszter Hargittai. 2016. The Imagined Audience on Social Network Sites. *Social Media + Society* 2, 1 (Jan. 2016), 2056305116633482. https://doi.org/10.1177/2056305116633482 Publisher: SAGE Publications Ltd.

[65] Matthew M. Lucas and Nikita Borisov. 2008. FlyByNight: mitigating the privacy risks of social networking. In *Proceedings of the 7th ACM workshop on Privacy in the electronic society (WPES '08)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/1456403.1456405

[66] W. Luo, Q. Xie, and U. Hengartner. 2009. FaceCloak: An Architecture for User Privacy on Social Networking Sites. In *2009 International Conference on Computational Science and Engineering*, Vol. 3. 26–33. https://doi.org/10.1109/CSE.2009.387

[67] Aida E Manduley, Andrea Mertens, Iradele Plante, and Anjum Sultana. 2018. The role of social media in sex education: Dispatches from queer, trans, and racialized communities. *Feminism & Psychology* 28, 1 (Feb. 2018), 152–170. https://doi.org/10.1177/0959353517717751 Publisher: SAGE Publications Ltd.

[68] Ryan A. Miller. 2017. "My Voice Is Definitely Strongest in Online Communities": Students Using Social Media for Queer and Disability Identity-Making. *Journal of College Student Development* 58, 4 (2017), 509–525. https://doi.org/10.1353/csd.2017.0040 Publisher: Johns Hopkins University Press.

[69] Sacha Molitorisz. 2020. *Net Privacy: How We Can Be Free in an Age of Surveillance*. McGill-Queen's Press - MQUP. Google-Books-ID: CWzeDwAAQBAJ.

[70] Marcia Mundt, Karen Ross, and Charla M Burnett. 2018. Scaling Social Movements Through Social Media: The Case of Black Lives Matter. *Social Media + Society* 4, 4 (Oct. 2018), 2056305118807911. https://doi.org/10.1177/2056305118807911 Publisher: SAGE Publications Ltd.

[71] Moses Namara, Daricia Wilkinson, Kelly Caine, and Bart P. Knijnenburg. 2020. Emotional and Practical Considerations Towards the Adoption and Abandonment of VPNs as a Privacy-Enhancing Technology. *Proceedings on Privacy Enhancing Technologies* 2020, 1 (Jan. 2020), 83–102. https://doi.org/10.2478/popets-2020-0006

[72] Arvind Narayanan, Vincent Toubiana, Solon Barocas, Helen Nissenbaum, and Dan Boneh. 2012. A Critical Look at Decentralized Personal Data Architectures. (Feb. 2012). https://arxiv.org/abs/1202.4503v1

[73] Shirin Nilizadeh, Sonia Jahid, Prateek Mittal, Nikita Borisov, and Apu Kapadia. 2012. Cachet: A Decentralized Architecture for Privacy Preserving Social Networking with Caching. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies (CoNEXT '12)*. ACM, New York, NY, USA, 337–348. https://doi.org/10.1145/2413176.2413215 event-place: Nice, France.

[74] Safiya Umoja Noble. 2018. *Algorithms of oppression*. New York University Press.

[75] Theresa Senft and Safi ya Umoja Noble. 2013. Race and Social Media. In *The Social Media Handbook*. Routledge. Num Pages: 19.

[76] Ihudiya Finda Ogbonnaya-Ogburu, Angela D.R. Smith, Alexandra To, and Kentaro Toyama. 2020. Critical Race Theory for HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376392

[77] Leysia Palen and Paul Dourish. 2003. Unpacking "privacy" for a networked world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. Association for Computing Machinery, Ft. Lauderdale, Florida, USA, 129–136. https://doi.org/10.1145/642611.642635

[78] Thomas Paul, Antonino Famulari, and Thorsten Strufe. 2014. A survey on decentralized Online Social Networks. *Computer Networks* 75 (Dec. 2014), 437–452. https://doi.org/10.1016/j.comnet.2014.10.005

[79] Laura Portwood-Stacer. 2013. Media refusal and conspicuous non-consumption: The performative and political dimensions of Facebook abstention. *New Media & Society* 15, 7 (Nov. 2013), 1041–1057. https://doi.org/10.1177/1461444812465139 Publisher: SAGE Publications.

[80] Paulina Pospieszna and Aleksandra Galus. 2019. 'Liberation Technology' or 'Net Delusion'? Civic Activists' Perceptions of Social Media as a Platform for Civic Activism in Belarus and Ukraine. *Europe-Asia Studies* 71, 10 (Nov. 2019), 1664–1684. https://doi.org/10.1080/09668136.2019.1623176 Publisher: Routledge _eprint: https://doi.org/10.1080/09668136.2019.1623176.

[81] John Postill. 2014. Democracy in an age of viral reality: A media epidemiography of Spain's indignados movement. *Ethnography* 15, 1 (March 2014), 51–69. https://doi.org/10.1177/1466138113502513 Publisher: SAGE Publications.

[82] Jeffrey G. Proudfoot, David Wilson, Joseph S. Valacich, and Michael D. Byrd. 2018. Saving face on Facebook: privacy concerns, social benefits, and impression management. *Behaviour & Information Technology* 37, 1 (Jan. 2018), 16–37. https://doi.org/10.1080/0144929X.2017.1389988 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/0144929X.2017.1389988.

[83] Moo-Ryong Ra, Ramesh Govindan, and Antonio Ortega. 2013. P3: Toward Privacy-Preserving Photo Sharing. 515–528. https://www.usenix.org/conference/nsdi13/technical-sessions/presentation/ra

[84] Jeff John Roberts. 2011. Facebook Search Warrants, A New Tool For U.S. Law Enforcement. https://www.huffpost.com/entry/facebook-search-warrant_n_896328

[85] Espen Geelmuyden Rød and Nils B Weidmann. 2015. Empowering activists or autocrats? The Internet in authoritarian regimes. *Journal of Peace Research* 52, 3 (May 2015), 338–351. https://doi.org/10.1177/0022343314555782 Publisher: SAGE Publications Ltd.

[86] M Angela Sasse. 2003. Computer Security: Anatomy of a Usability Disaster, and a Plan for Recovery. (April 2003), 5.

[87] Morgan Klaus Scheuerman, Stacy M. Branham, and Foad Hamidi. 2018. Safe Spaces and Safe Places: Unpacking Technology-Mediated Experiences of Safety and Harm with Transgender People. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 155:1–155:27. https://doi.org/10.1145/3274424

[88] Morgan Klaus Scheuerman, Kandrea Wade, Caitlin Lustig, and Jed R. Brubaker. 2020. How We've Taught Algorithms to See Identity: Constructing Race and Gender in Image Databases for Facial Analysis. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (May 2020), 058:1–058:35. https://doi.org/10.1145/3392866

[89] Amre Shakimov, Harold Lim, Ramón Cáceres, Landon P. Cox, Kevin Li, Dongtao Liu, and Alexander Varshavsky. 2011. Vis-à-Vis: Privacy-preserving online social networking via Virtual Individual Servers. In *2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011)*. 1–10. https://doi.org/10.1109/COMSNETS.2011.5716497

[90] Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. 2013. The post that wasn't: exploring self-censorship on facebook. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 793–802. https://doi.org/10.1145/2441776.2441865

[91] Manya Sleeper, William Melicher, Hana Habib, Lujo Bauer, Lorrie Faith Cranor, and Michelle L. Mazurek. 2016. Sharing Personal Content Online: Exploring Channel Choice and Multi-Channel Behaviors. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 101–112. https://doi.org/10.1145/2858036.2858170

[92] Kaushal Solanki, Anindya Sarkar, and B. S. Manjunath. 2007. YASS: Yet Another Steganographic Scheme That Resists Blind Steganalysis. In *Information Hiding*, Teddy Furon, François Cayre, Gwenaël Doërr, and Patrick Bas (Eds.). Vol. 4567. Springer Berlin Heidelberg, Berlin, Heidelberg, 16–31. https://doi.org/10.1007/978-3-540-77370-2_2

[93] Olivia Solon. 2017. 'It's digital colonialism': how Facebook's free internet service has failed its users. *the Guardian* (July 2017). http://www.theguardian.com/technology/2017/jul/27/facebook-free-basics-developing-markets Section: Technology.

[94] Katta Spiel, Os Keyes, Ashley Marie Walker, Michael A. DeVito, Jeremy Birnholtz, Emeline Brulé, Ann Light, Pınar Barlas, Jean Hardy, Alex Ahmed, Jennifer A. Rode, Jed R. Brubaker, and Gopinaath Kannabiran. 2019. Queer(ing) HCI: Moving Forward in Theory and Practice. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–4.

https://doi.org/10.1145/3290607.3311750

[95] Charles Steinfield, Nicole B. Ellison, and Cliff Lampe. 2008. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology* 29, 6 (Nov. 2008), 434–445. https://doi.org/10.1016/j.appdev.2008.07.002

[96] Elizabeth Stoycheff, G. Scott Burgess, and Maria Clara Martucci. 2020. Online censorship and digital surveillance: the relationship between suppression technologies and democratization across countries. *Information, Communication & Society* 23, 4 (March 2020), 474–490. https://doi.org/10.1080/1369118X.2018.1518472 Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2018.1518472.

[97] Frederic Stutzman and Woodrow Hartzog. 2012. Boundary regulation in social media. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 769–778. https://doi.org/10.1145/2145204.2145320

[98] Peter Swire and Kenesa Ahmad. 2011. 'Going Dark' Versus a 'Golden Age for Surveillance.'. *Center for Democracy and Technology* (2011).

[99] Alexandra Tereshonkova. 2016. *Protection Vs. Empowerment: the State of the LGBT Child in Russia*. Ph.D. Dissertation. Columbia University. https://doi.org/10.7916/D8FJ2H84

[100] The ImageMagick Development Team. 2021. ImageMagick. https://imagemagick.org

[101] Yannis Theocharis, Will Lowe, Jan W. van Deth, and Gema García-Albacete. 2015. Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society* 18, 2 (Feb. 2015), 202–220. https://doi.org/10.1080/1369118X.2014.948035 Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2014.948035.

[102] Maria Vasilyeva. 2020. Russian LGBT activist fined for 'gay propaganda' family drawings. *Reuters* (July 2020). https://www.reuters.com/article/us-russia-activist-court-idUSKBN24B2IY

[103] Andreas Westfeld. 2001. F5—A Steganographic Algorithm: High Capacity Despite Better Steganalysis. In *Information Hiding*, Gerhard Goos, Juris Hartmanis, Jan van Leeuwen, and Ira S. Moskowitz (Eds.). Vol. 2137. Springer Berlin Heidelberg, Berlin, Heidelberg, 289–302. https://doi.org/10.1007/3-540-45496-9_21 Series Title: Lecture Notes in Computer Science.

[104] Alma Whitten and J Doug Tygar. 1999. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0.. In *USENIX Security Symposium*, Vol. 348. 169–184.

[105] Tim Wu. 2010. *The master switch: The rise and fall of information empires*. Vintage.

[106] Ching-man Au Yeung, Ilaria Liccardi, Kanghao Lu, Oshani Seneviratne, and Tim Berners-lee. 2009. Decentralization: The future of online social networking. In *In W3C Workshop on the Future of Social Networking Position Papers*.

[107] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 493–502. https://doi.org/10.1145/1240624.1240704

[108] Shoshana Zuboff. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs. Google-Books-ID: lRqrDQAAQBAJ.

Received July 2021; revised November 2021; accepted November 2021