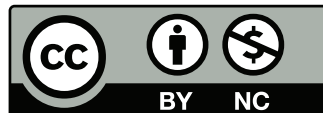


# Preventing Intimate Image Abuse via Privacy-Preserving Credentials

Technical paper: [Jacob Gorman](#), Nikhil Mehta, Marie Nganele, [Janet Zhang](#),  
[Steven M. Bellovin](#)

Law paper: [Janet Zhang](#), [Steven M. Bellovin](#)



# Non-Consensual Pornography (NCP)

- Non-consensual pornography (sometimes called intimate image abuse or revenge porn) has become a serious problem
- The issue: uploading intimate images—often taken or shared with a partner consensually—without consent
- Illegal in almost all states; some also permit civil suits
- But: recourse can be hard
- *Who did the original upload, and how do you prove it?*

# Danielle Citron's Proposal

- Web sites should take certain steps if they wish full §230 protection
- One step: logging relevant information
- But—logging IP addresses doesn't work well
  - Public hotspots (with NATs and no logging)
  - Phones (carrier-grade NAT—do the web sites and carriers log port numbers?)
  - *Doesn't help if other individuals download the pictures and upload them somewhere else*

# Strawman Solution

- Suppose that all images were digitally signed
  - Put the signatures and certificates into the EXIF metadata
- A serious privacy risk
- And: the Supreme Court has repeatedly stated that anonymous speech is constitutionally protected under the First Amendment
- Also: what of news organizations, whistleblower sites, etc.?

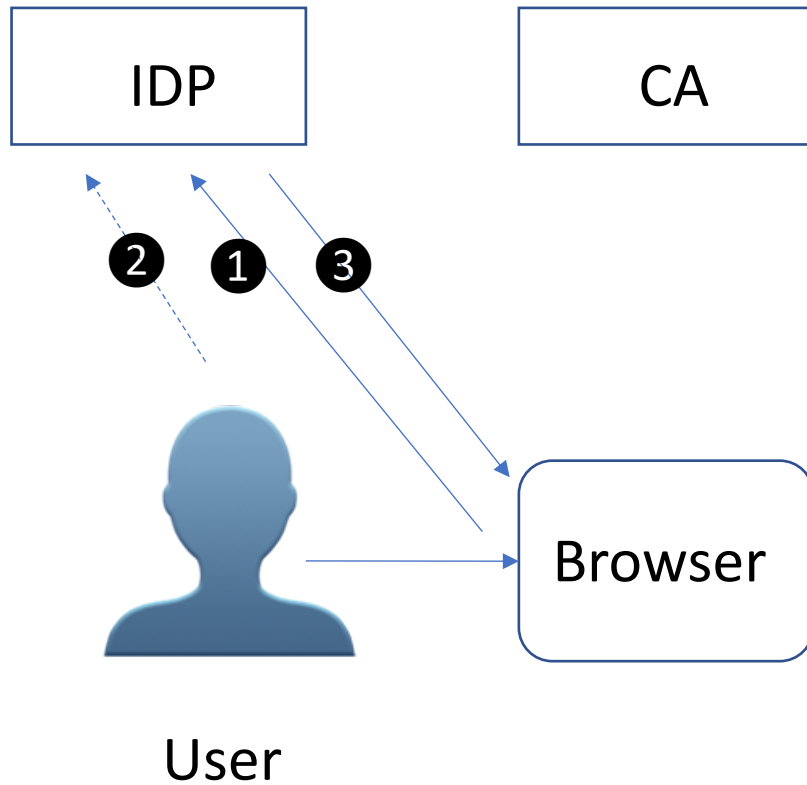
# Our Solution (From 30,000 Feet)

- Use privacy-preserving credentials to sign images
- Web sites don't have to participate (but see Citron re §230 protection)
- Unlinkable between websites
- Require the cooperation of three different parties to deanonymize the signer
- But—how do we do this?
- But—is the requirement constitutional?

# Camenisch-Lysyanskaya Credentials

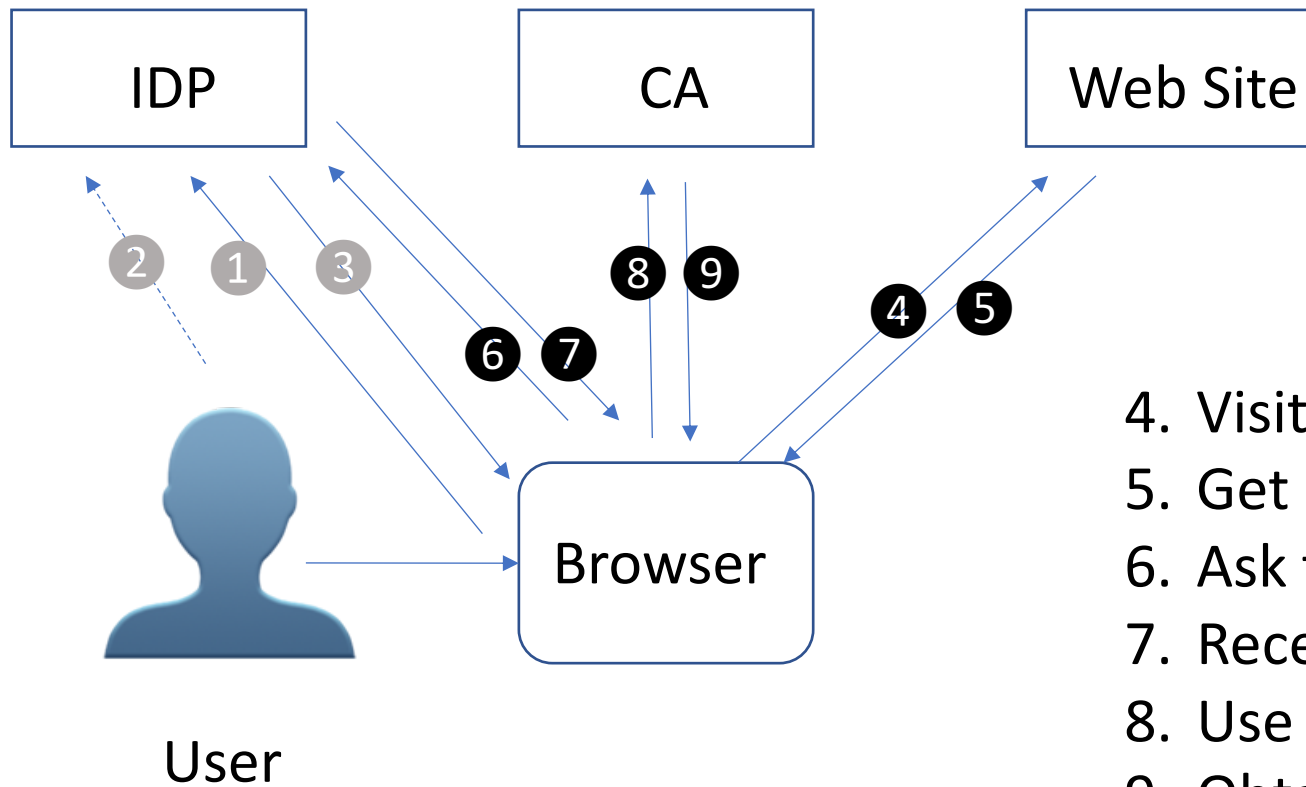
- Obtain a *primary credential*
- Use the primary credential to obtain as many *subcredentials* as you want. The subcredentials are not linkable to each other.
- The subcredentials can contain an encrypted *deanonymization string*
- When presenting the subcredentials to someone, use *zero knowledge proofs* to show that
  - a) they are valid;
  - b) they're derived from a valid primary credential issued by some mutually trusted issuer; and
  - c) the deanonymization string is valid

# Getting a Primary Credential



1. Register with an IDP
2. Visit the IDP with proof of identity, e.g., a photoID
3. Obtain a primary credential

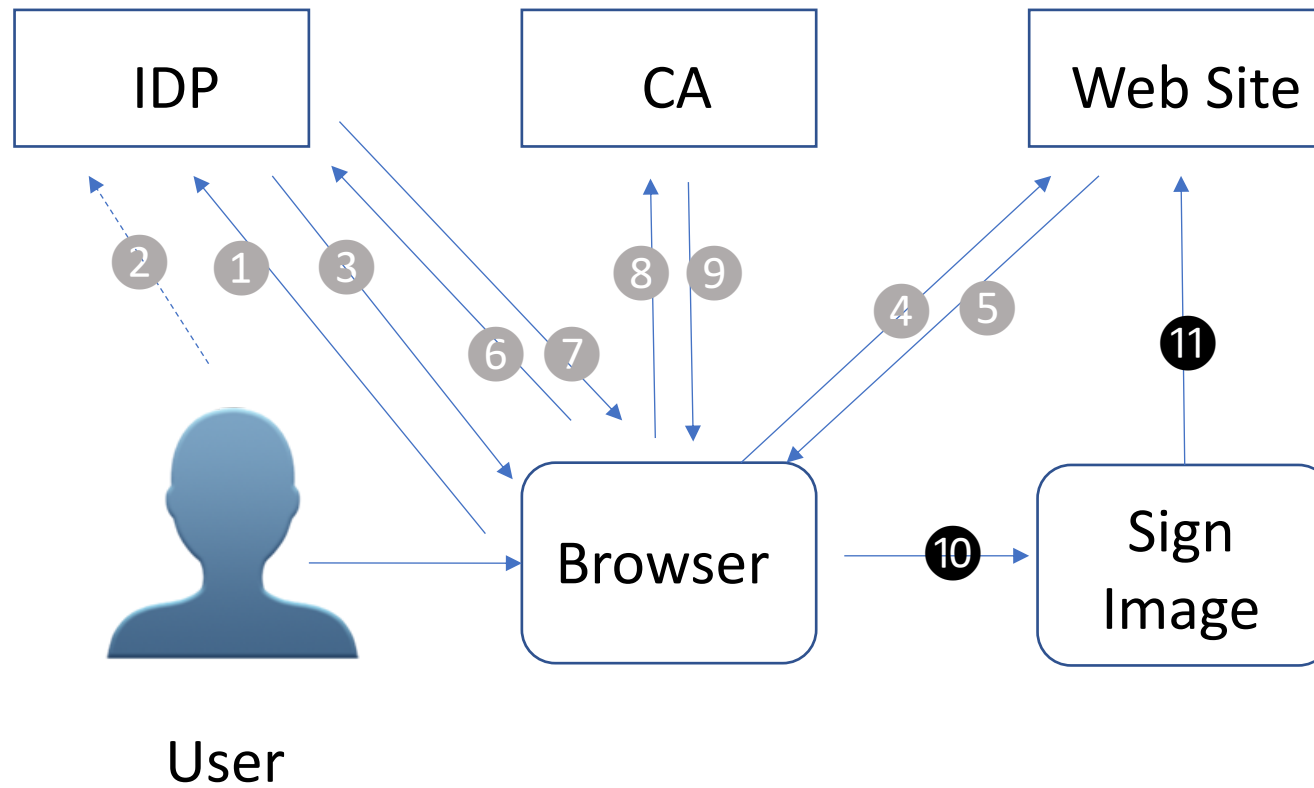
# First Visit to a Web Site



4. Visit a web site
5. Get an image upload page
6. Ask the IDP for a subcredential
7. Receive the subcredential
8. Use the subcredential to log in to a CA
9. Obtain an X.509 certificate

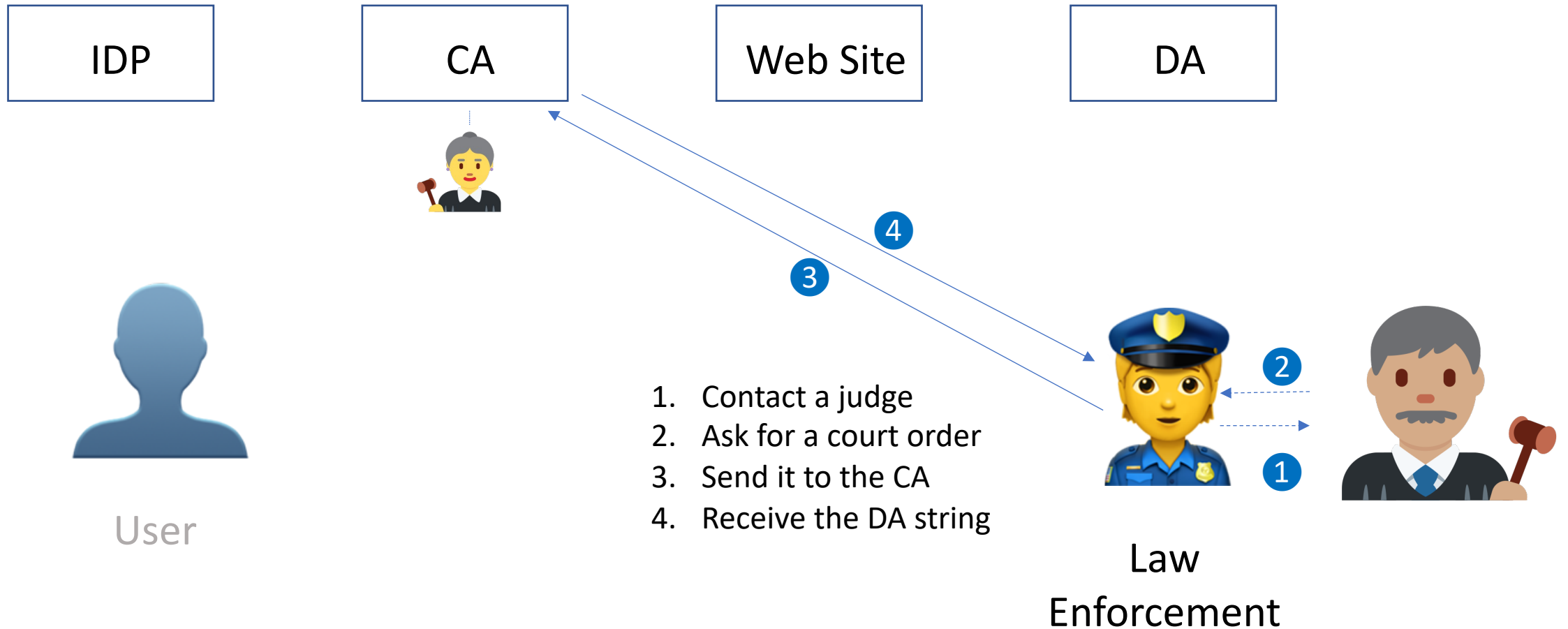


# Upload an Image



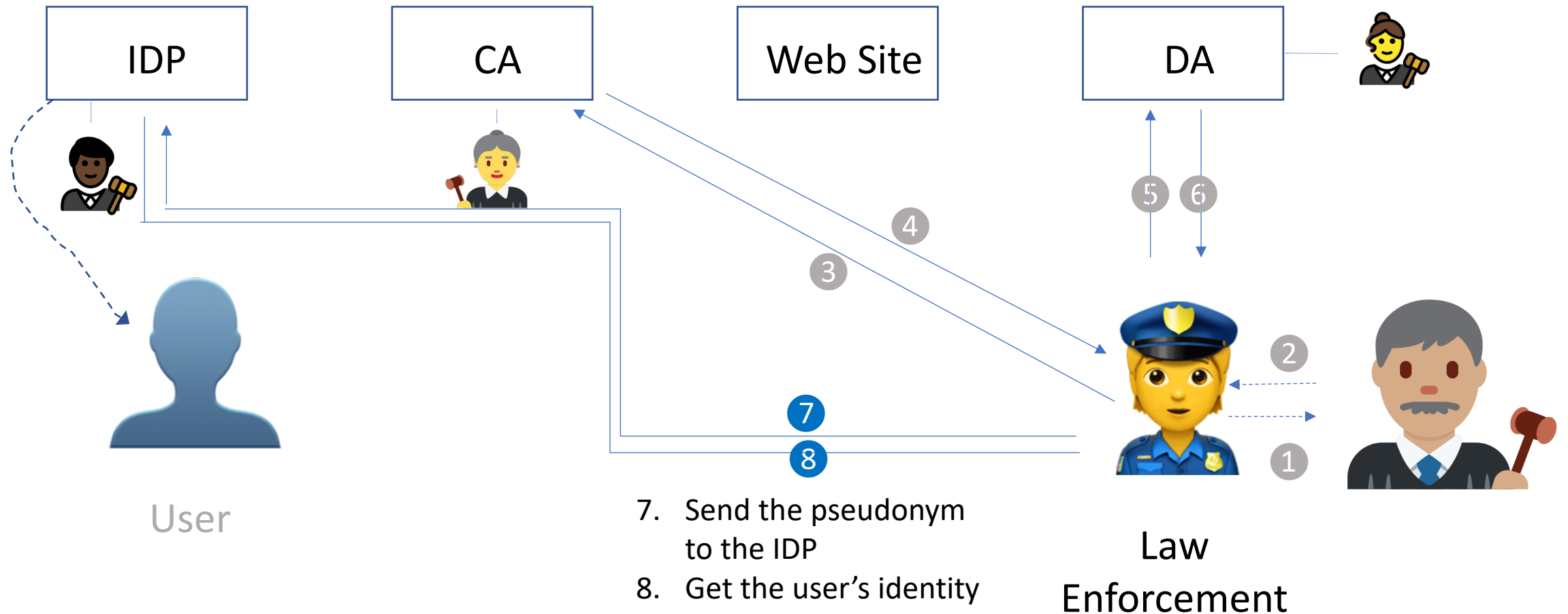
10. Sign the image  
11. Upload the image

# Getting the Deanonimization String

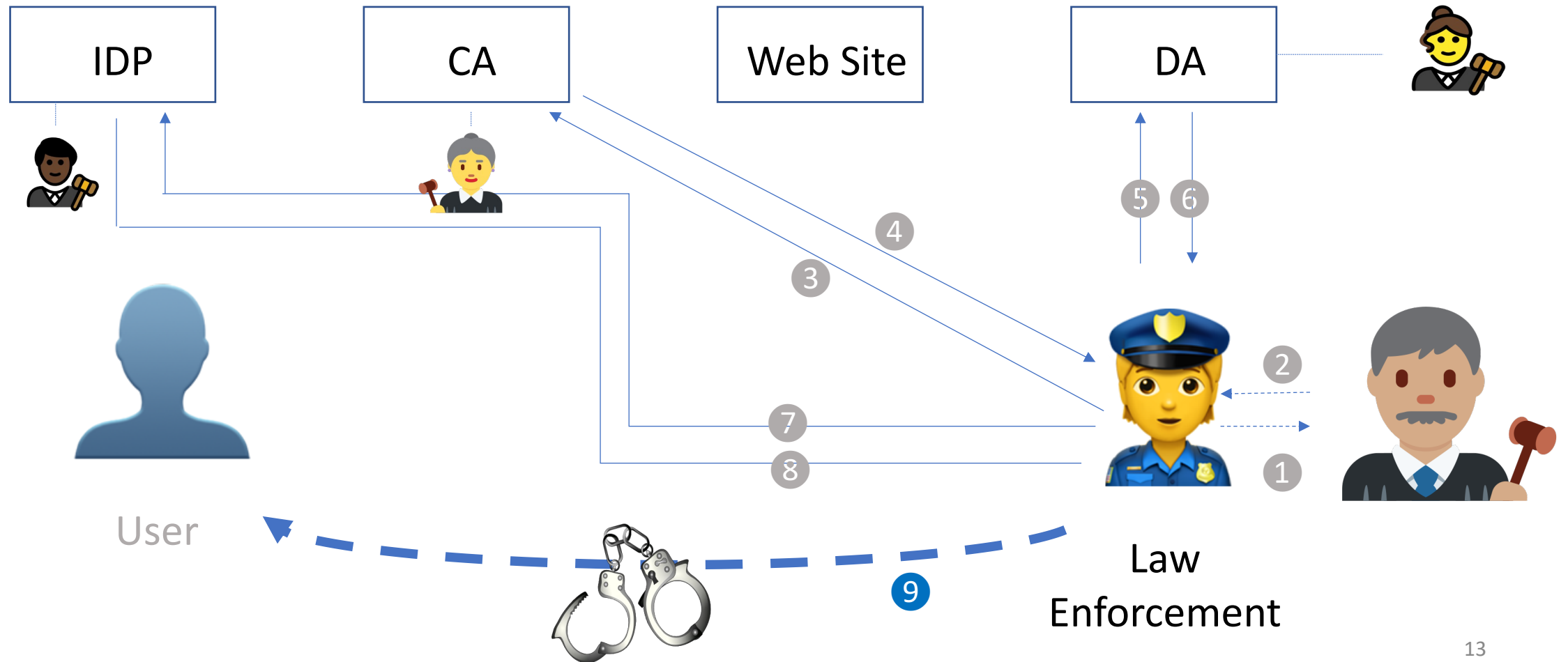




# Getting the User's Identity



# Consequences...



# Legal/Social Questions

- Is this constitutional?
  - (We defer to Citron on the constitutionality of the §230 changes)
  - Does this unduly burden the right to anonymous (free) speech?
- Does this impose undue burdens on minorities, poor people, rural residents, etc.?
- What are the regulatory issues?
- Who pays for all of this?
- Mission creep—how do we restrict deanonymization to non-consensual pornography?

# References

- Janet Zhang and Steven M. Bellovin. “Preventing intimate image abuse via privacy-preserving anonymous credentials”. *SMU Science and Technology Law Review*, 2023 ([unedited draft](#)).
- Jacob Gorman, Nikhil Mehta, Marie Nganele, Janet Zhang, “Privacy-Preserving Accountability for Non-Consensual Pornography”, in preparation.

# Questions?



Barred owl with chipmunk, Central Park, October 11, 2020



# Backup Slides



# A Proof of Concept Implementation

- Use Camenisch-Lysyanskaya credentials
- Only one IDP, CA, DA
- Only one browser supported
- No attempt at optimization
- No attempt at emulating manual functions

# Our Scheme

- The user registers online with an *identity provider (IDP)*, then provides proof of identity to the standards of a notary public (possibly online). The IDP and the user's browser agree on a *pseudonym*
- The first time a participating website is used for image uploads, a browser extension obtains a site-specific subcredential from the identity provider and uses this to log in to a *certificate authority (CA)*
  - The CA stores the deanonymization string, indexed by certificate serial number
  - A standard X.509 certificate is issued for that website
- The browser extension saves this certificate for future use
- It digitally signs all uploaded images for that site, and embeds the signature and certificate in the EXIF metadata
- Only the *deanonymization agent (DA)* can decrypt the deanonymization string

# Anonymous Speech Issues

- There is a right to anonymous speech (*Talley, McIntyre*)
- There is also a right to sexual privacy (*Griswold, Lawrence, Obergefell*)
- How should these be balanced?
- Exacting scrutiny: “which requires a ‘substantial relation’ between the disclosure requirement and a ‘sufficiently important’ governmental interest.” (*Citizens United*)
- Also: web sites do not need to participate; they have to signal willingness in image upload pages

In other words, there is a balancing test—and courts have generally been willing to deanonymize Internet activity in criminal cases. But we have to go further to prevent deanonymization of legitimate photos.

# Undue Burdens

- Many people (especially poor, rural minorities) do not have government-issued photo IDs
  - We know this from litigation over voting (*Crawford*)
- There may not be a nearby notary public, let alone an identity provider
- We cannot differentially impede speech—uploaded photos—by disadvantaged people
- Possible solution: *social authentication*—someone with suitable documents can vouch for the identity of others
  - Note: you can even use affidavits as a form of identification for passports

# Regulatory Issues

- These entities—the IDP, the CA, and the DA—probably need to be regulated
- They have to be independent of each other—they cannot be part of the same company
- They have to be honest
- They have to cooperate with legitimate court orders, which requires effective jurisdiction

# Who Should Pay?

- Users? They can optimize for cost or for the willingness and (expensive!) ability to strongly oppose deanonymization orders
  - Identity Providers are the users' only direct point of contact
  - Note: the Identity Provider chooses the CA and the DA
- Web sites? They benefit from user-created content.
- Law enforcement? They should at least pay for service to the DA.

This requires more study.

# Mission Creep

- How do we prevent more uses of deanonymization orders?
  - The list of eligible crimes under the Wiretap Act has grown considerably since 1968
- There do not appear to be suitable technical mechanisms
- A statutory provision barring use of identifying information from keys issued before amendments could always be repealed
- Best idea thus far: require a new constitutional analysis under exacting scrutiny
- Or: the Federal Rules of Evidence could bar admissibility of evidence obtained this way from credentials issued before the change in the law



# Identifying an Offender

- Law enforcement extracts the certificate from the image
- They obtain appropriate legal process from a judge, based on probable cause
- They send the image and the legal process to the CA to get the deanonymization string
  - The CA *by law* will have standing to challenge that order, e.g., if they don't think it's NCP
- The DA decrypts the deanonymization string and retrieves the pseudonym
  - The DA also has standing to challenge the order
- The IDP can return the user's real identity
  - The IDP also has standing to challenge the order, and will notify the user to permit them to challenge it