Preventing Intimate Image Abuse via Privacy-Preserving Credentials

Or: Why I Do Law as Well as Computer Science

Technical paper: Jacob Gorman, Nikhil Mehta, Marie Nganele, Janet Zhang, Steven M. Bellovin

Law paper: Janet Zhang, Steven M. Bellovin

CS @CU

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?*

Section 230

- Under a provision of Federal law commonly known as *Section 230* (more formally, 47 U.S.C. §230), sites are not liable for content uploaded by their users
- In other words: if someone uploads NCP to YouTube or Instagram, Google and Meta are not liable
- The uploader is liable—if you can find them and prove that they did it

Danielle Citron's Proposal

- Web sites should take certain steps if they wish full §230 protection
- One step: logging relevant information, e.g., IP address of uploader
- 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.?

EXIF Metadata

Aperture Value	**6.919**
Body Serial Number	**3028903**
Color Space	**sRGB**
Components Configurati...	**1, 2, 3, 0**
Contrast	**Normal**
Custom Rendered	**Custom process**
Date Time Digitized	**Apr 8, 2024 at 3:28:4...**
Date Time Original	**Apr 8, 2024 at 3:28:4...**
Exif Version	**2.3.1**
Exposure Bias Value	**-1**
Exposure Mode	**Manual exposure**
Exposure Program	**Manual**
Exposure Time	**1/100**
File Source	**DSC**
Flash	**No Flash**
FlashPix Version	**1.0**
FNumber	**11**
Focal Length	**800**
Focal Length In 35mm Fi...	**800**

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?

Our Scheme, in More Detail

- 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 a *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

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

What's a Zero-Knowledge Proof?

- Prove that you know something without disclosing some secret
- Bad example…
  - It's easy to square numbers; it's much harder to calculate a square root (and for most of us, impossible by hand)
  - I claim that I can do it
  - Repeat until you're convinced
    - You give me a number
    - I give you the square root
    - You square that and see if the answer matches

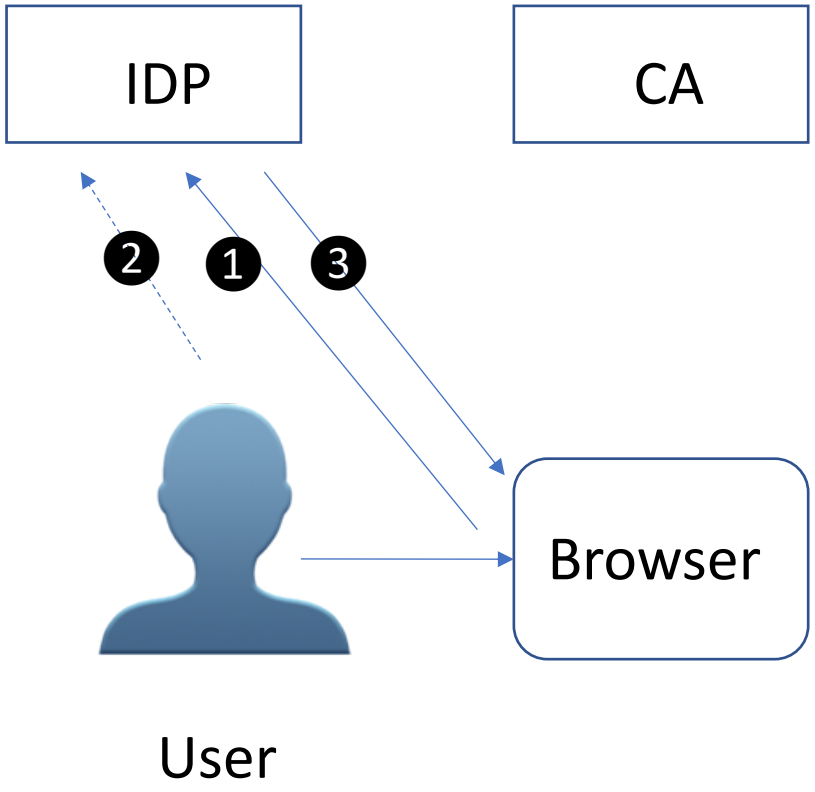
Identifying an Offender

- 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

Standing

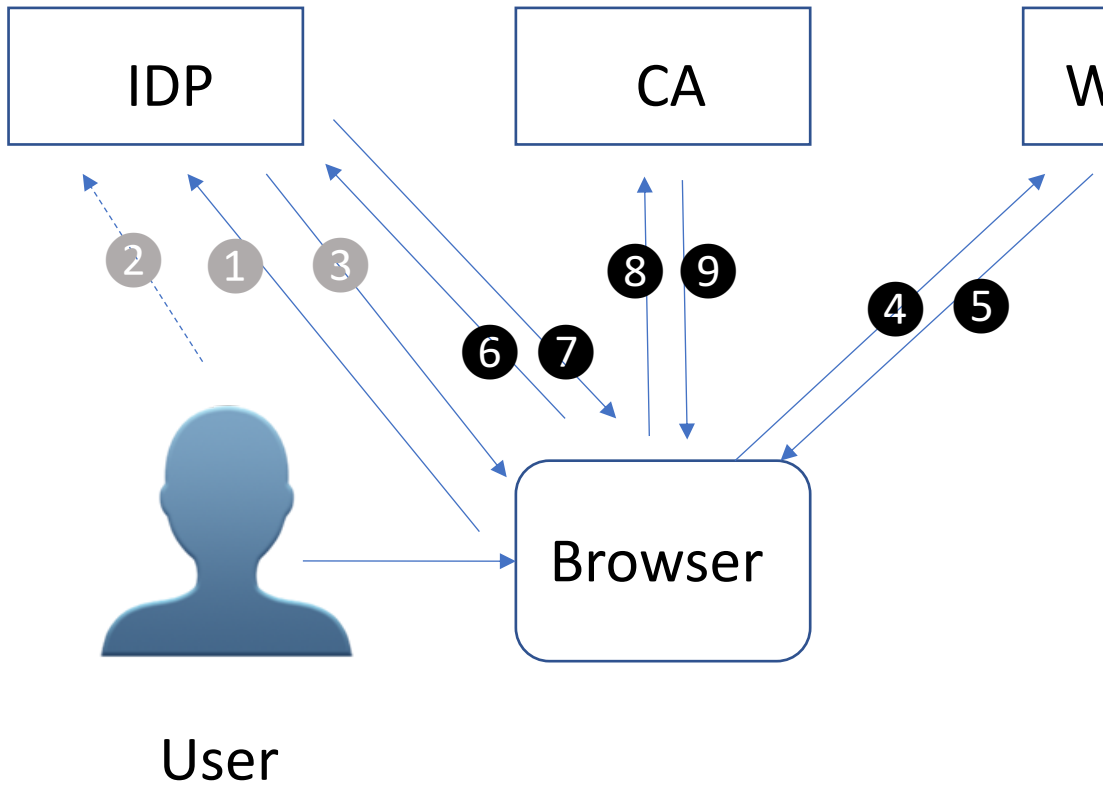
- What is *standing*?
- (Very) briefly: it's the right to be able to file a lawsuit
- Complex legal topic; many facets
- Real world example: you're offended by news reports of NSA's activities. You don't have standing to sue unless you can show that *your* traffic was collected.

Getting a Primary Credential



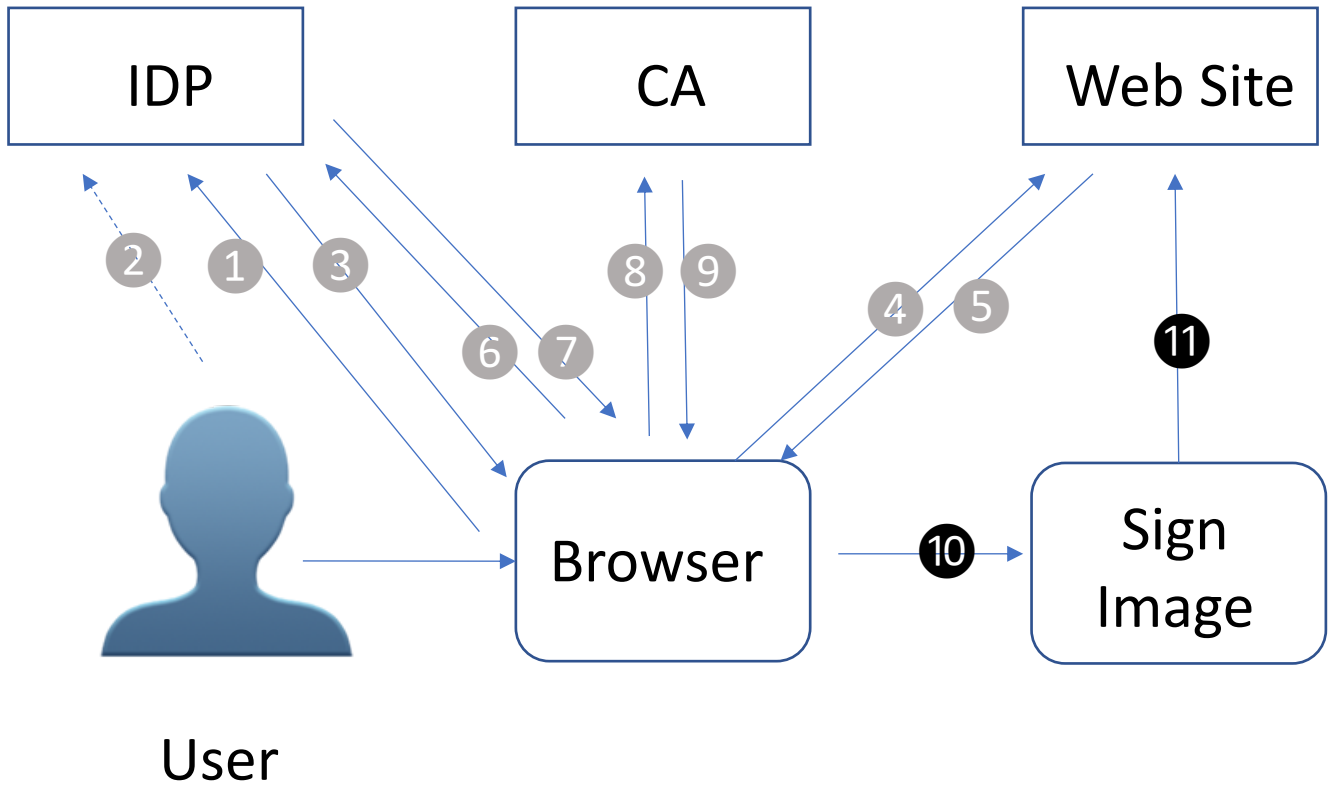
1. Register with an IDP
2. Visit the IDP with proof of identity, e.g., a photoID
3. Obtain a primary credential

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; it validates the subcredential and stores the deanonymization string
9. Obtain an X.509 certificate

Upload an Image

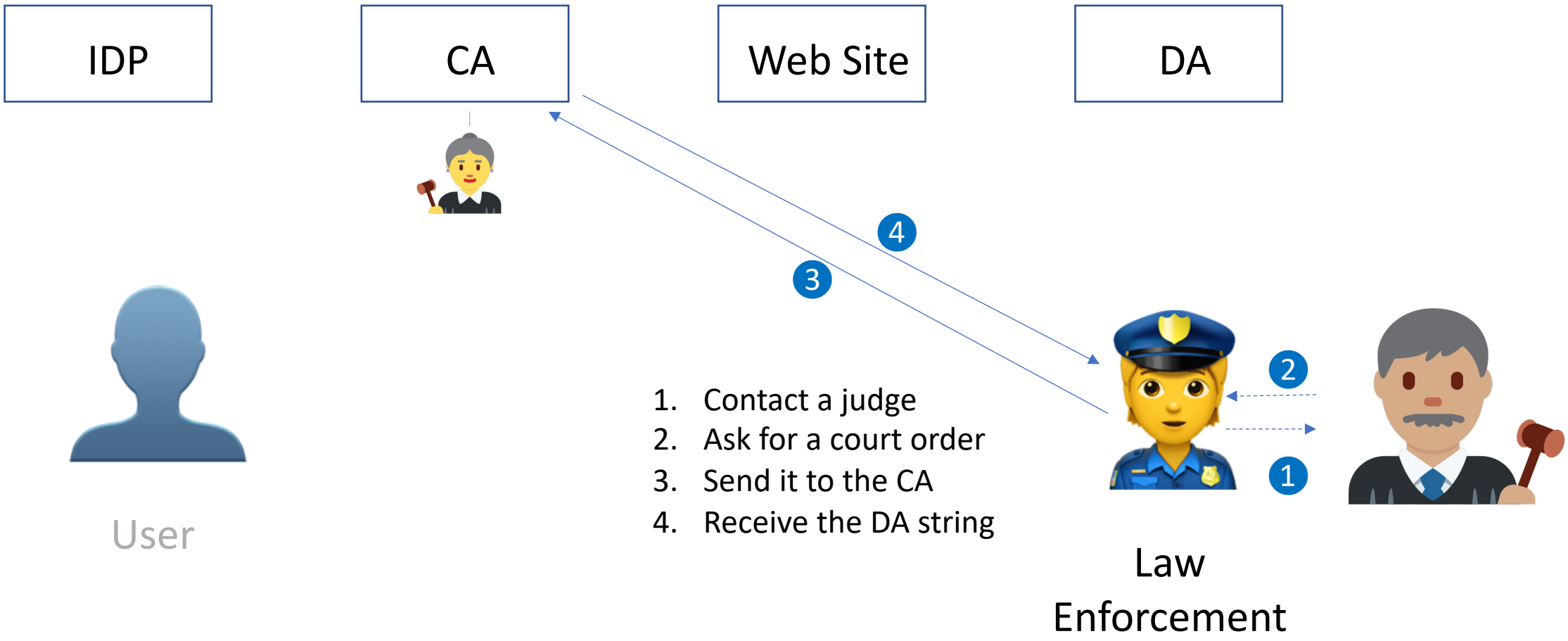


10.Sign the image
11.Upload the image

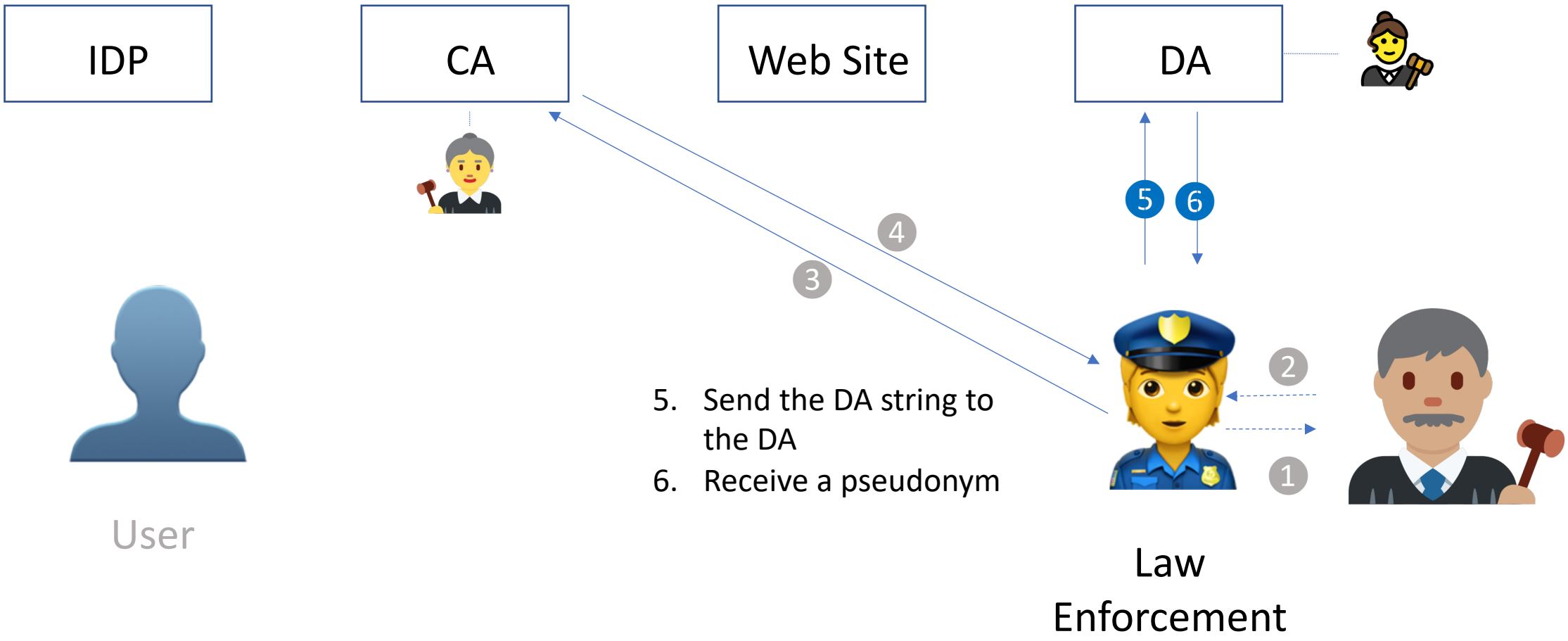
And if NCP is Uploaded?

- The complainant (probably the victim) reports it to law enforcement
- They find out who uploaded it

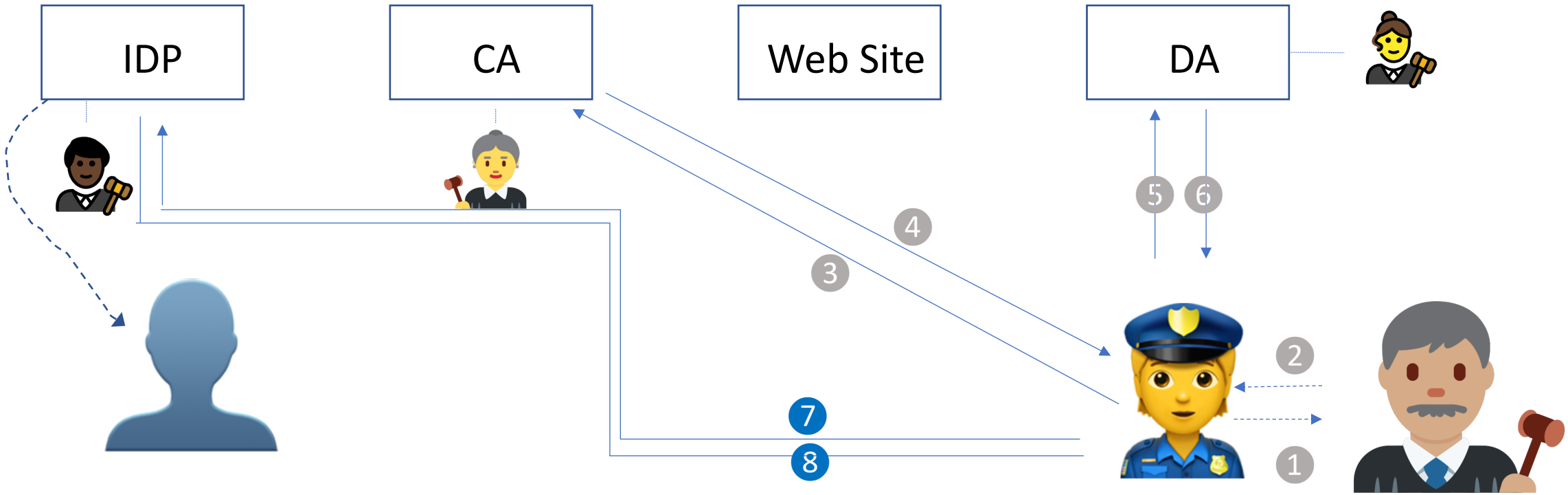
Getting the Deanonymization String



Getting the Pseudonym

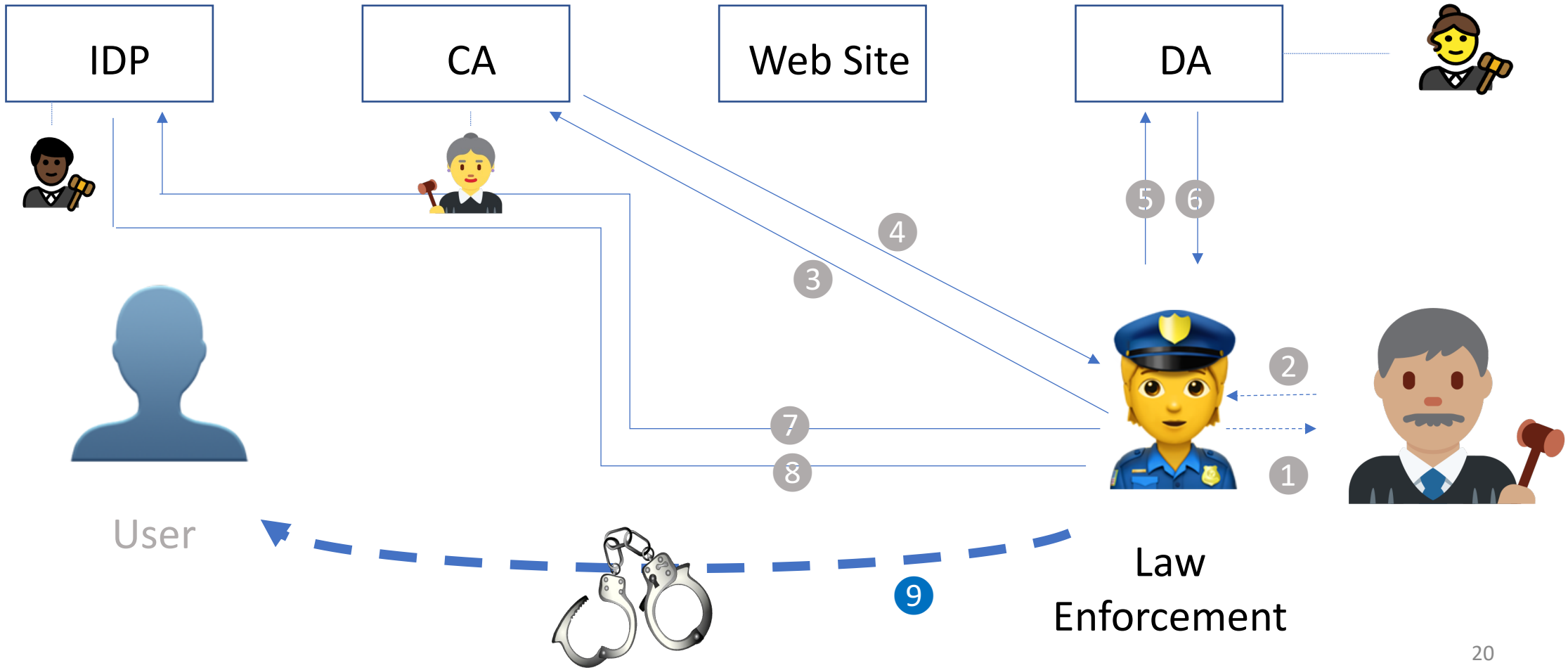


Getting the User's Identity



7. Send the pseudonym to the IDP
8. Get 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?

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.

Scrutiny

- The Supreme Court sometimes applies different levels of scrutiny when assessing the constitutionality of a law
  - Rational basis
  - Intermediate scrutiny
  - Strict scrutiny
  - Exacting scrutiny

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

Mission Creep

- How do we prevent more uses of deanonymization orders?
  - The list of eligible crimes under the Wiretap Act has grown considerably since it was originally enacted in 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

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 choses 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.

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

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

Skills and Knowledge Needed

- Knowledge of cryptography
- Coding, for the proof-of-concept implementation
- Knowledge of law (free and anonymous speech issues)
- Social issues

References

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

Legal References

- Talley *v. California,* 362 U.S. 60 (1960)
- *McIntyre v. Ohio Elections Commission*, 514 U.S. 334 (1995)
- *Griswold v. Connecticut*, 381 U.S. 479 (1965)
- *Lawrence v. Texas*, 539 U.S. 588 (2003)
- *Obergefell v. Hodges*, 576 U.S. 644 (2015)
- *Crawford v. Marion County Election Board*, 553 U.S. 181 (2008)
- *Citizens United v. FEC*, 558 U.S. 310 (2010)
- Wiretap Act: 18 U.S.C. §2510 *et seq.*

Questions?

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