COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

Finishing Video Analytics; Database Privacy I

Information Privacy with Applications

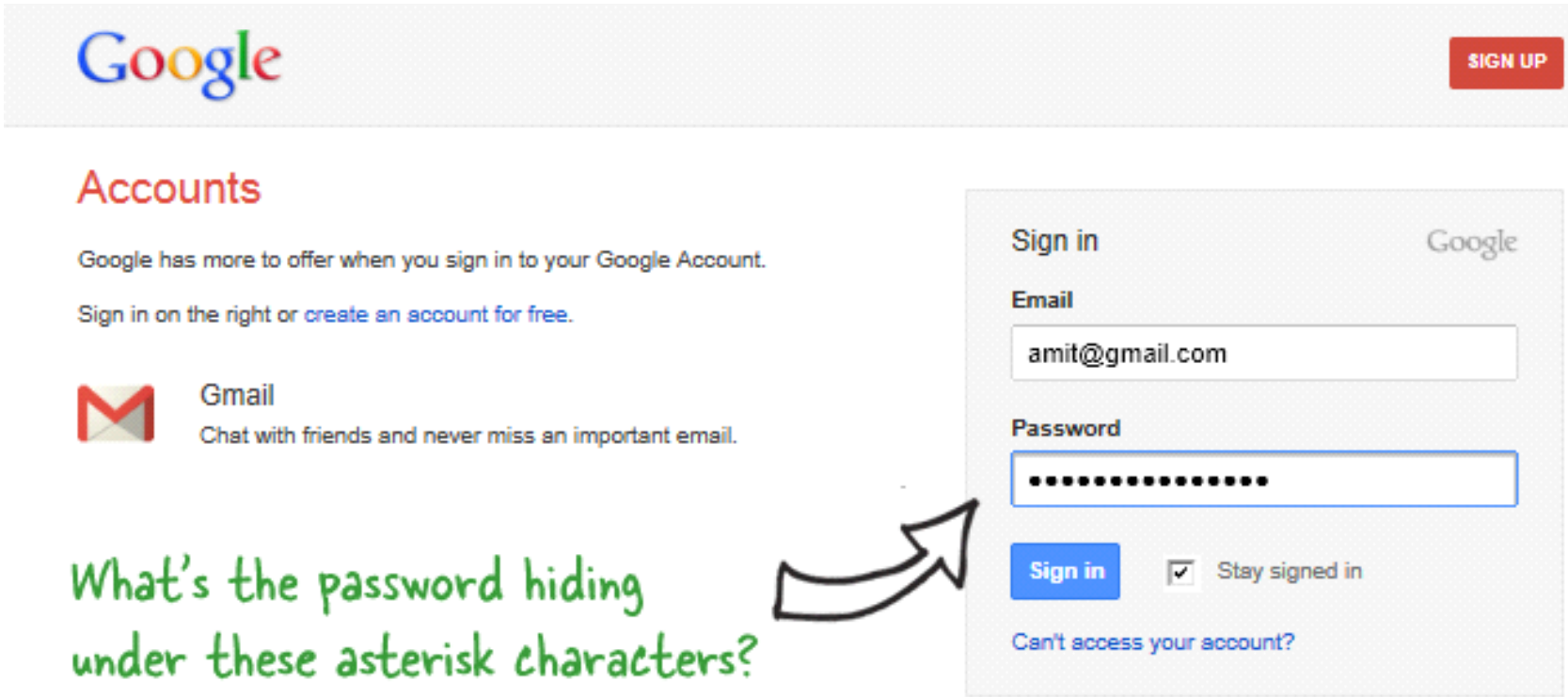
David Sidi (dsidi@email.arizona.edu)

Administrative

- Integrated session: thoughts?
  - Last assignment. Share your cameras!
- Face detection assignment will be posted after class; it is due next Tuesday
  - let’s review it

Finishing up video analytics: Analog Keyhole Problems


KANYE WEST ACCIDENTALLY REVEALED
HIS IPHONE PASSCODE
SAT
12°
SUN
15°
MON
11°

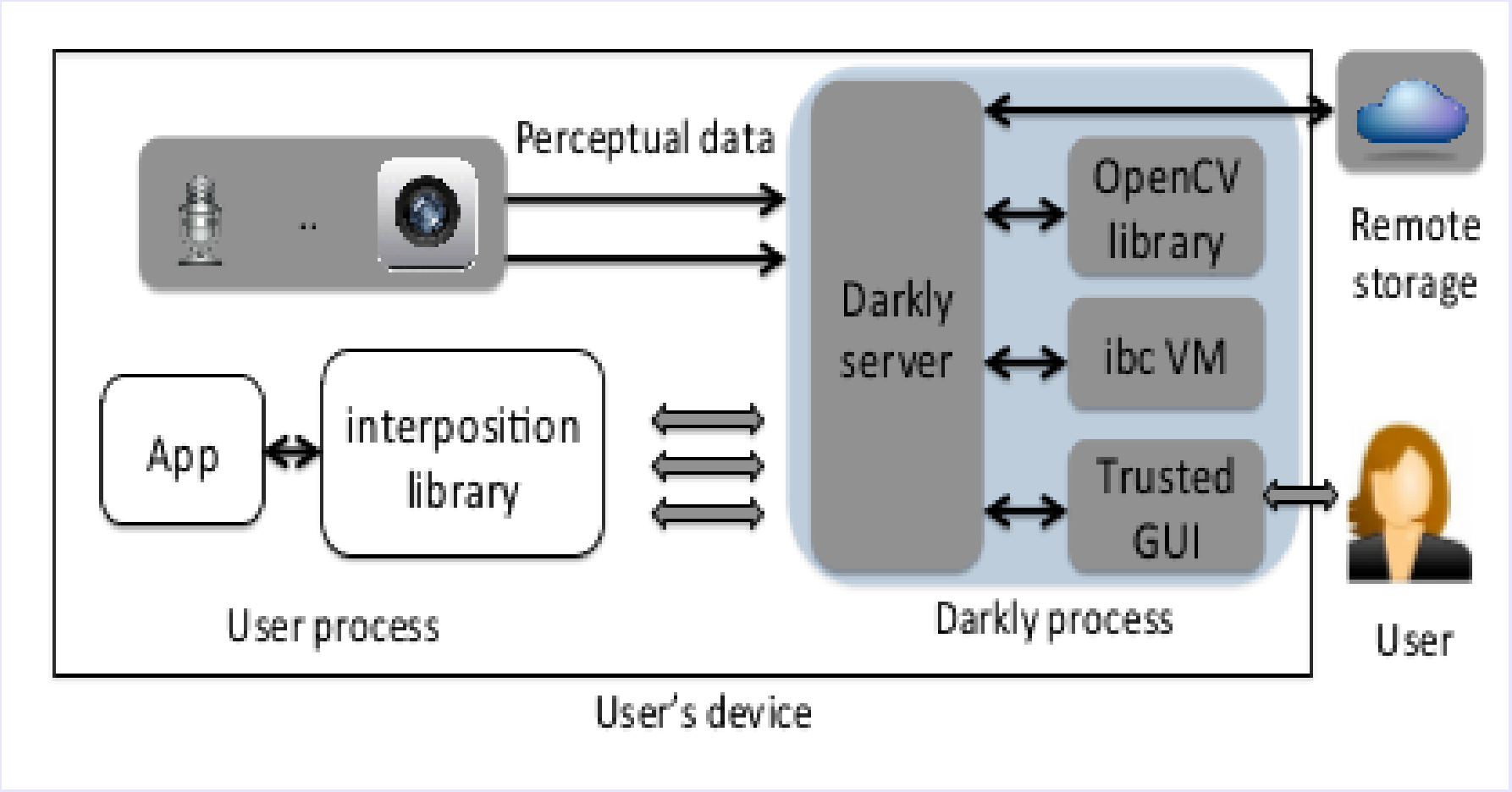

Google
SIGN UP
Accounts
Google has more to offer when you sign in to your Google Account.
Sign in on the right or create an account for free.
Gmail
Chat with friends and never miss an important email.
What's the password hiding
under these asterisk characters?
Sign in
Google
Email
amit@gmail.com
Password
Sign in
Stay signed in
Can't access your account?

DARKLY is for only a limited threat related to the AKP

- Suppose you own devices with perceptual capabilities and want to be sure that the apps that use those capabilities don’t misbehave
- Darkly (@ 35:33 - 43:00)
- Trust includes
  - device operating system
  - the device hardware, including its perceptual sensors
- Trust does not include a third party application running on your device

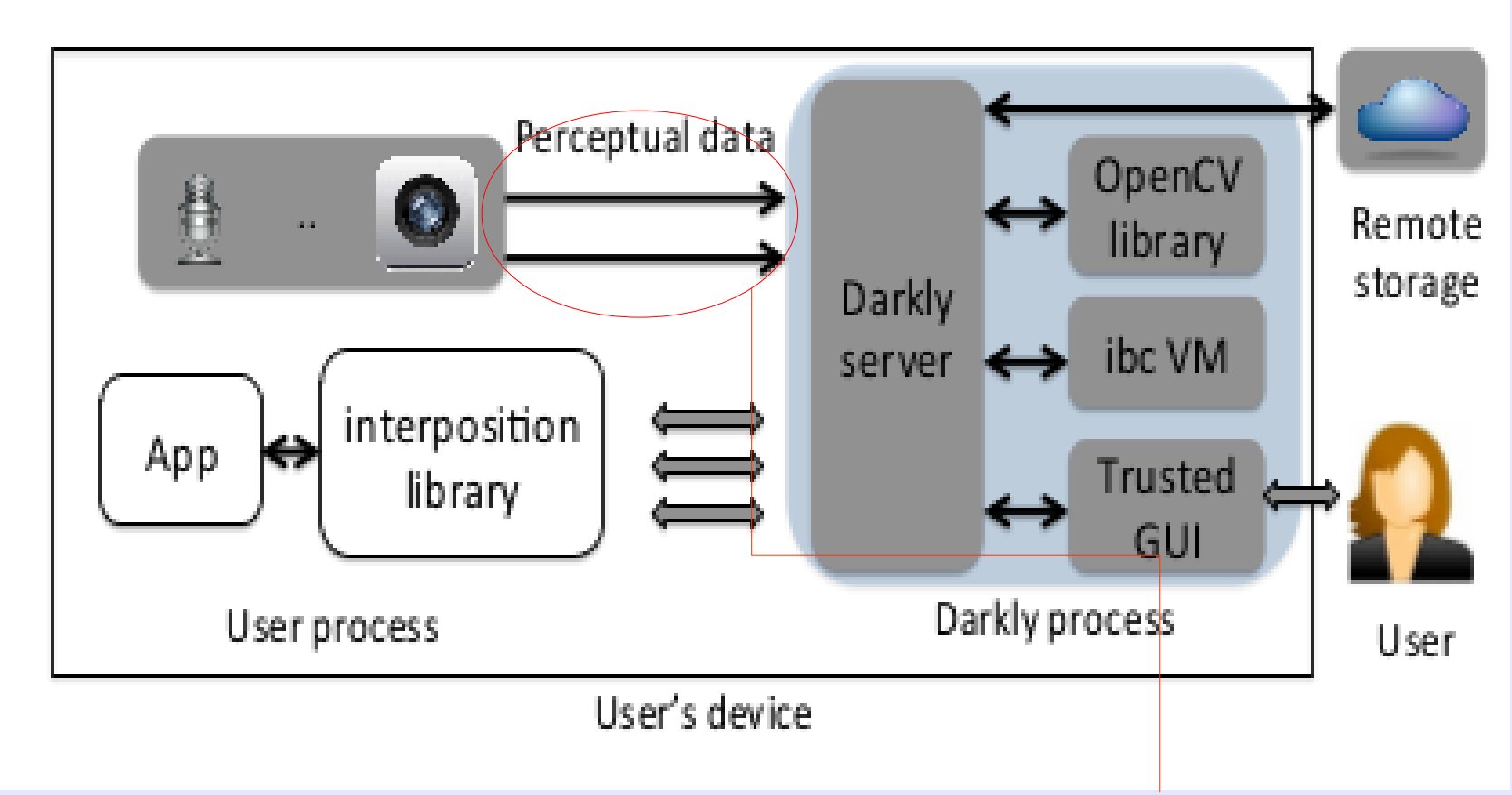
Opaque References

- “the application will never have access to the raw pixels”



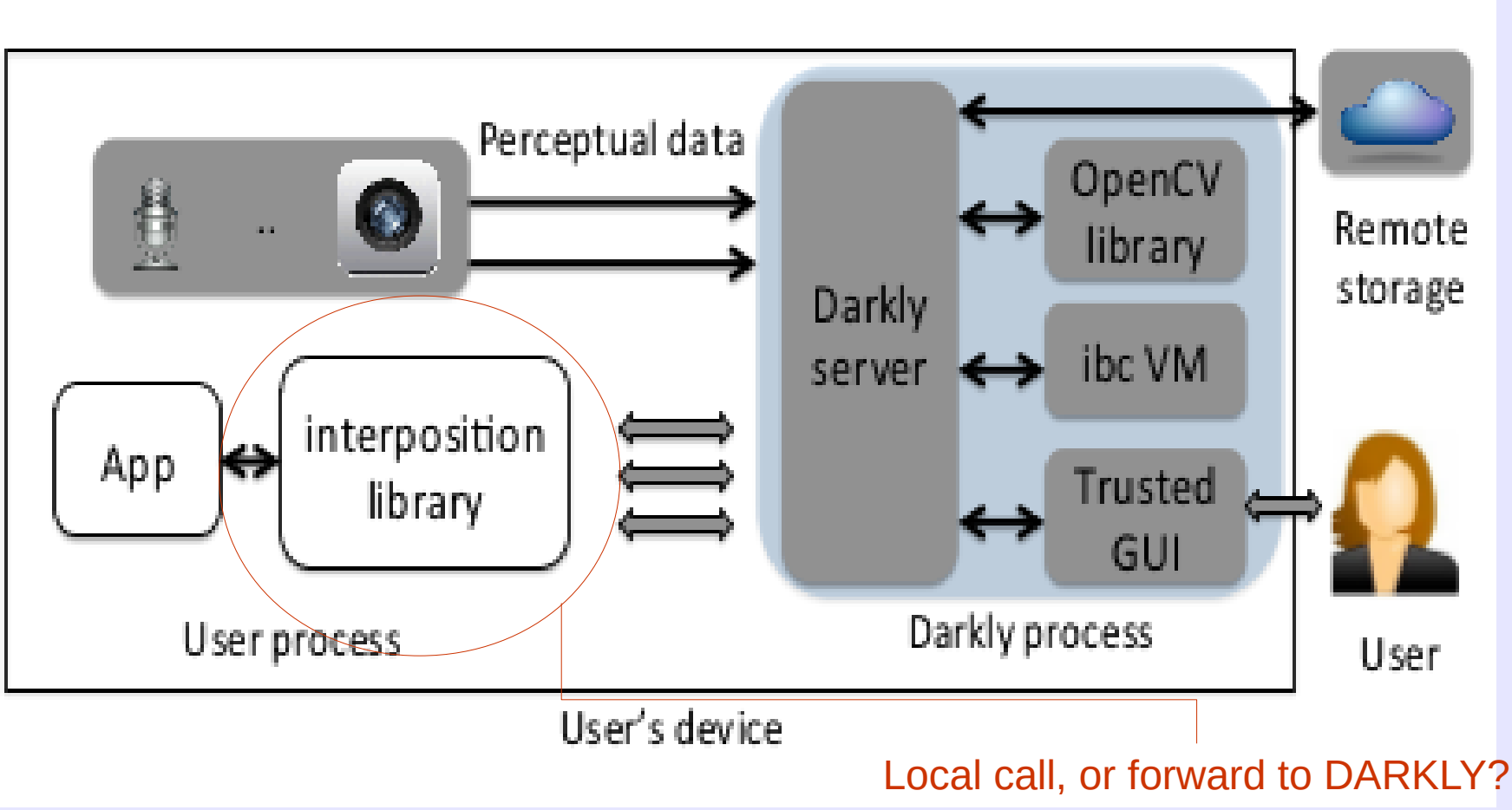
Opaque References

- “the application will never have access to the raw pixels”
  - opaque references



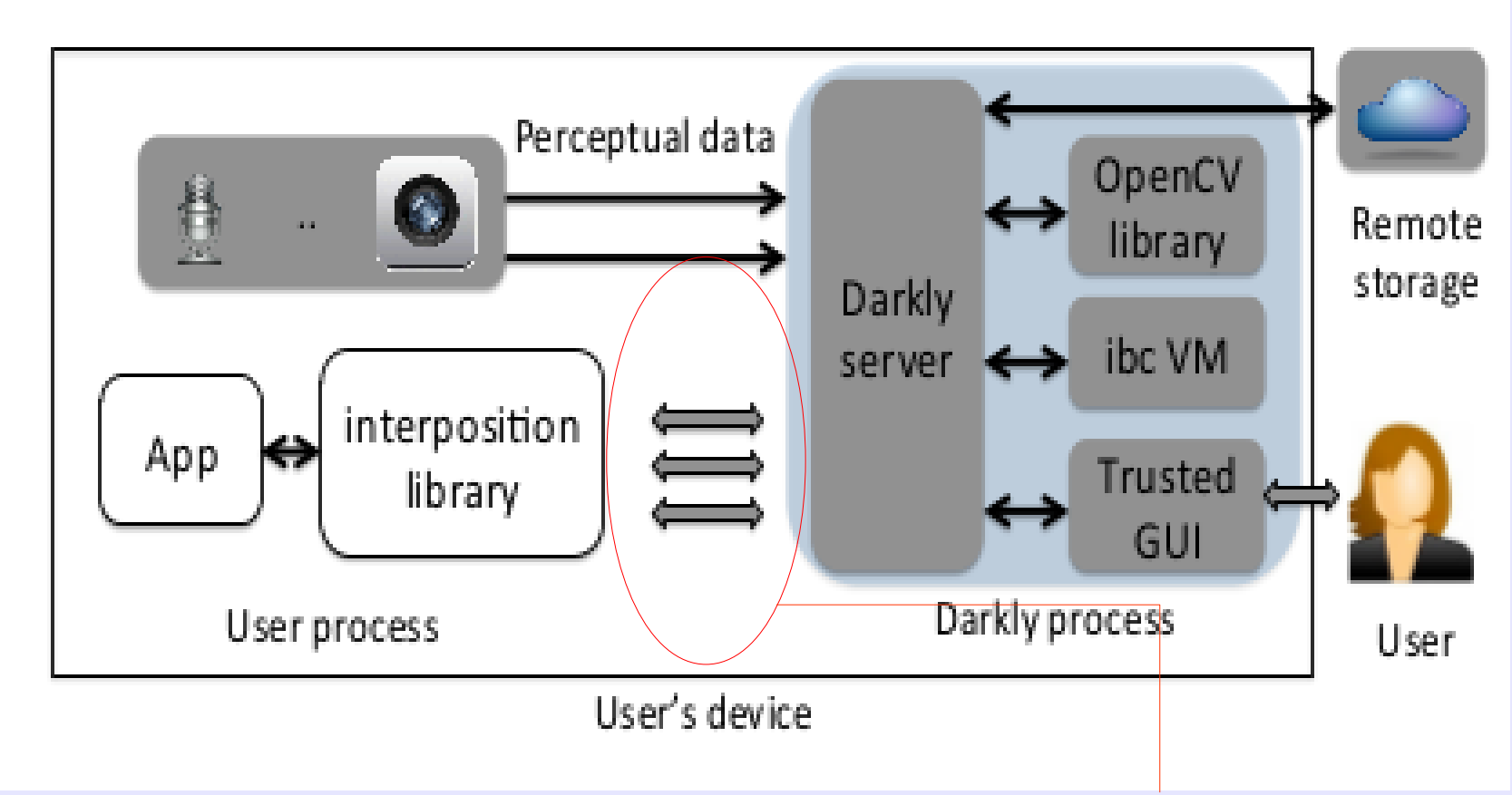
Opaque References

- “the application will never have access to the raw pixels”
  - opaque references



Opaque References

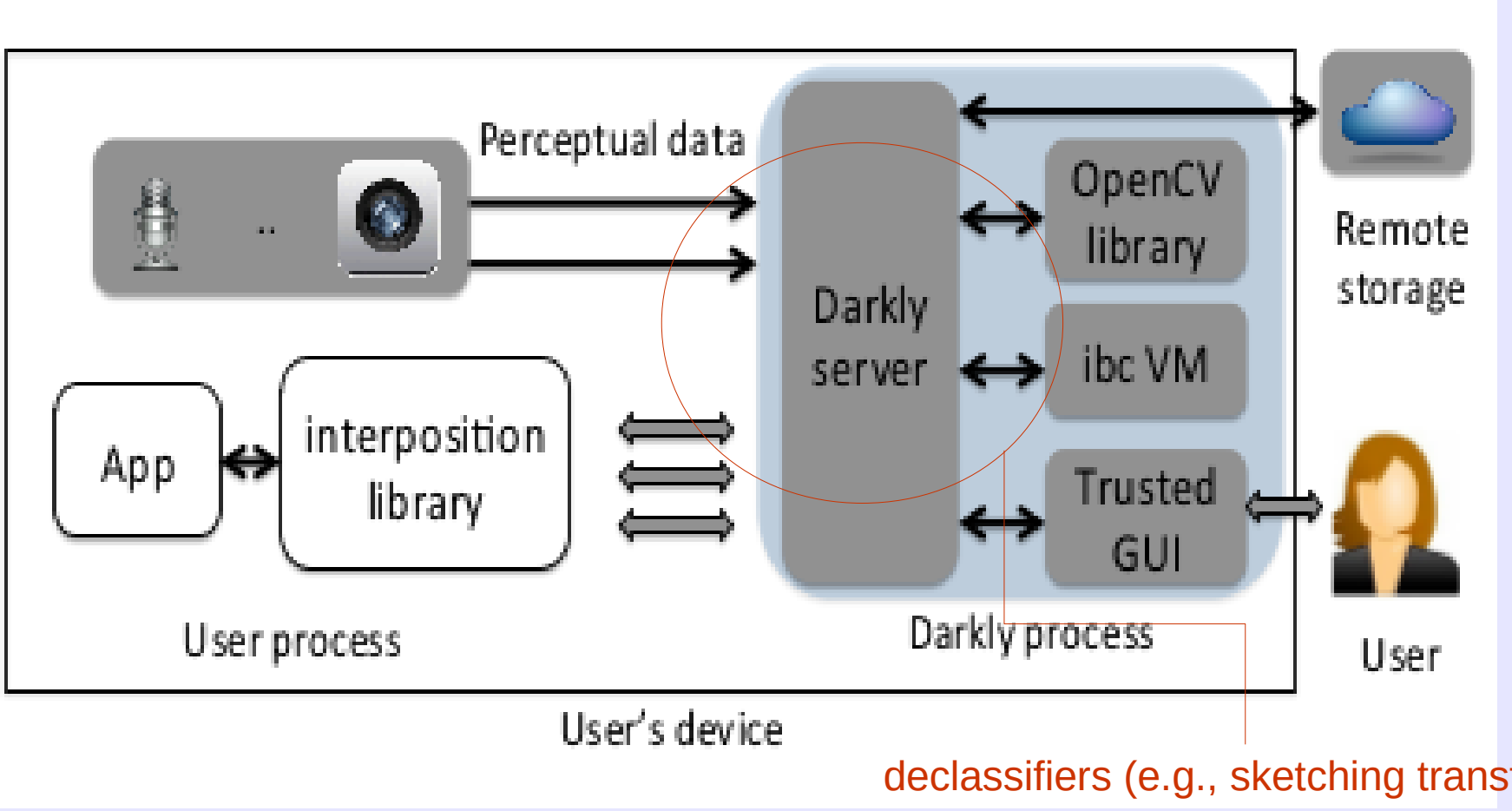
- “the application will never have access to the raw pixels”



any opaque references?

Declassifiers for privacy transformation

- "the application will never have access to the raw pixels"
  - opaque references



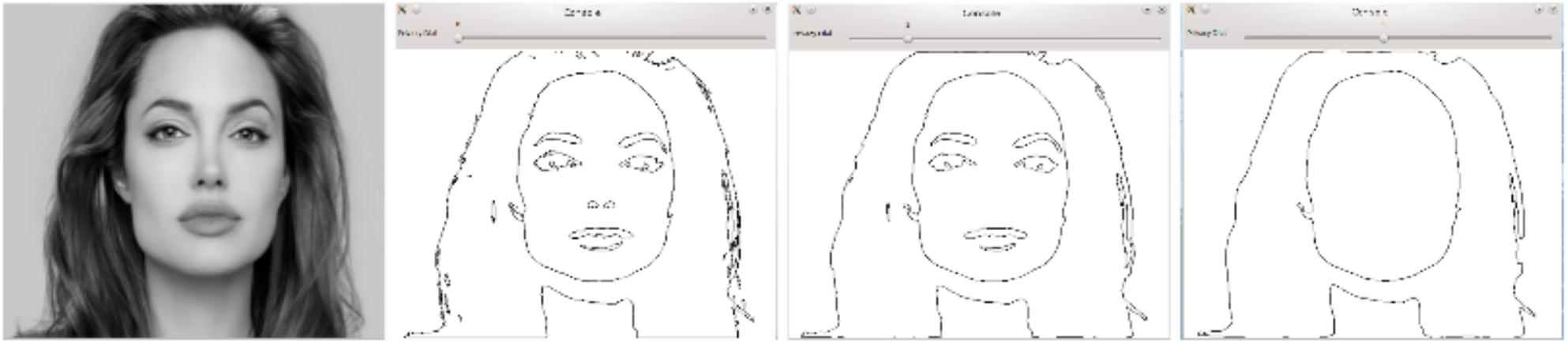


Figure 2. Output of the sketching transform on a female face image at different privacy levels.

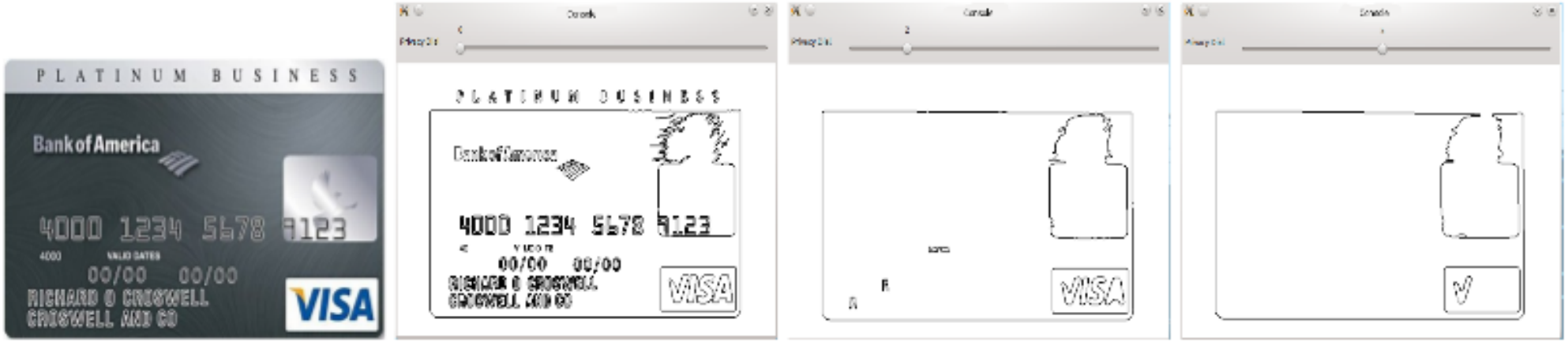
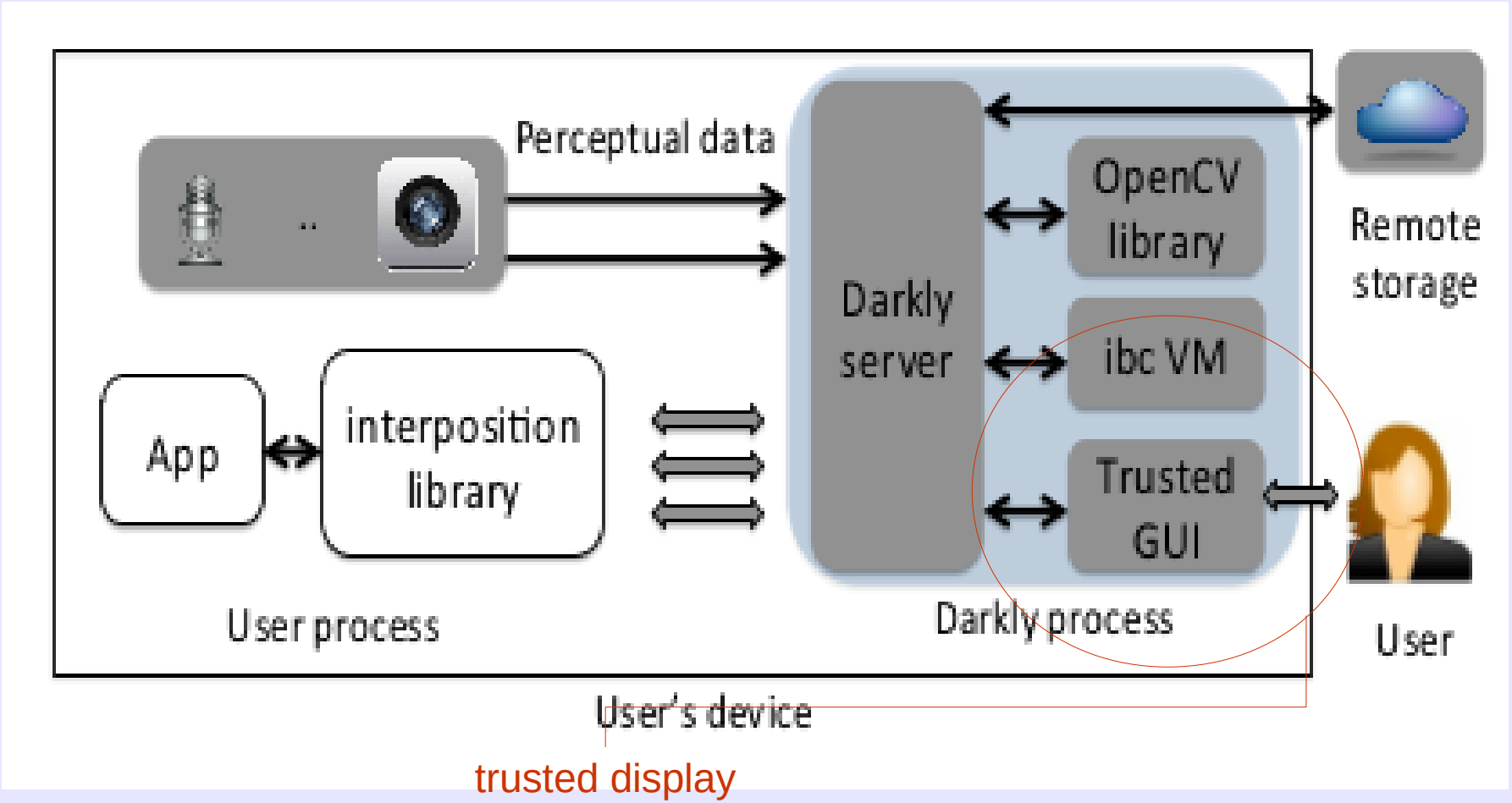


Figure 3. Output of the sketching transform on a credit card image at different privacy levels.

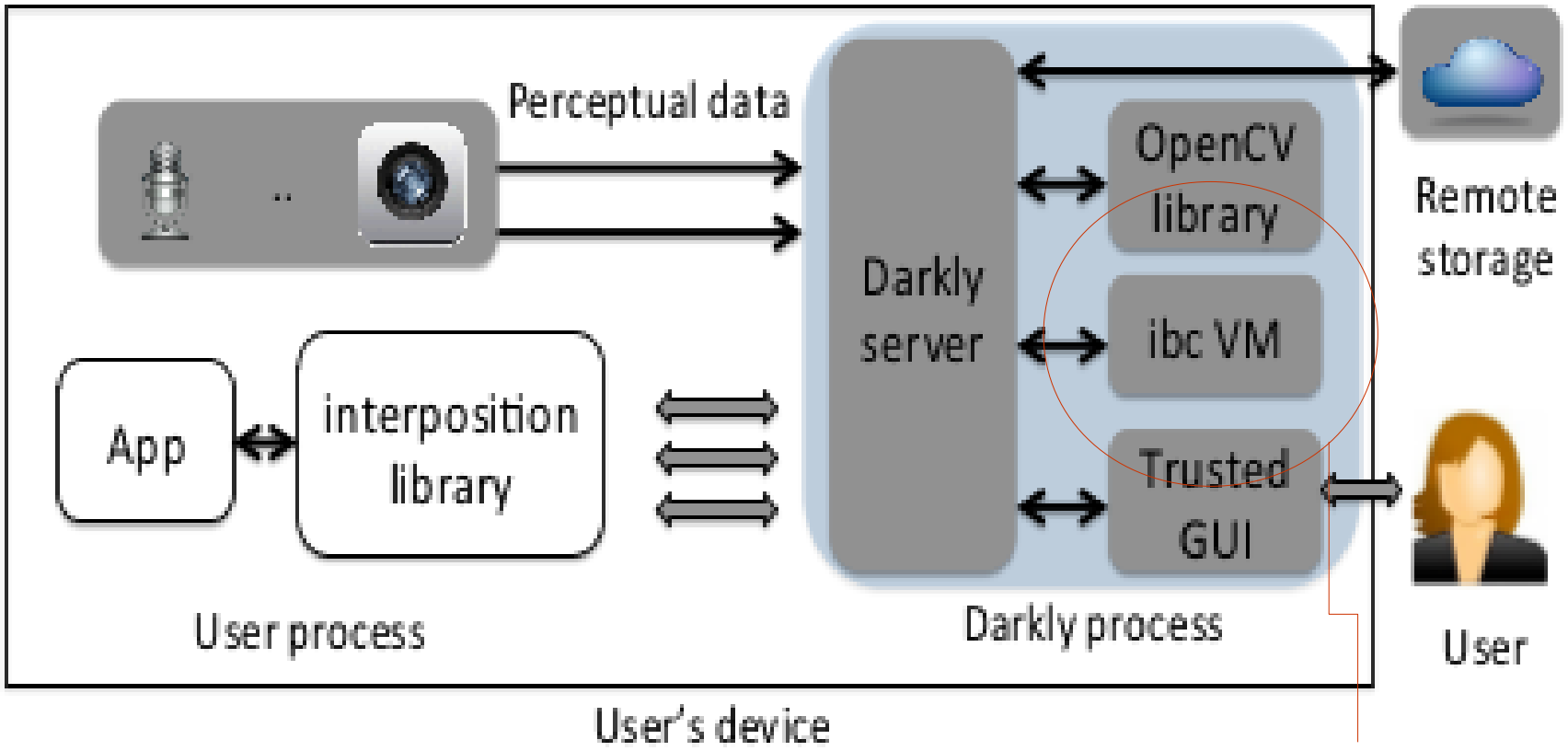
Trusted GUI and Storage

- “the application will never have access to the raw pixels”



ibc for untrusted arbitrary computation

- “the application will never have access to the raw pixels” **(mostly)**



DARKLY is domain-specific

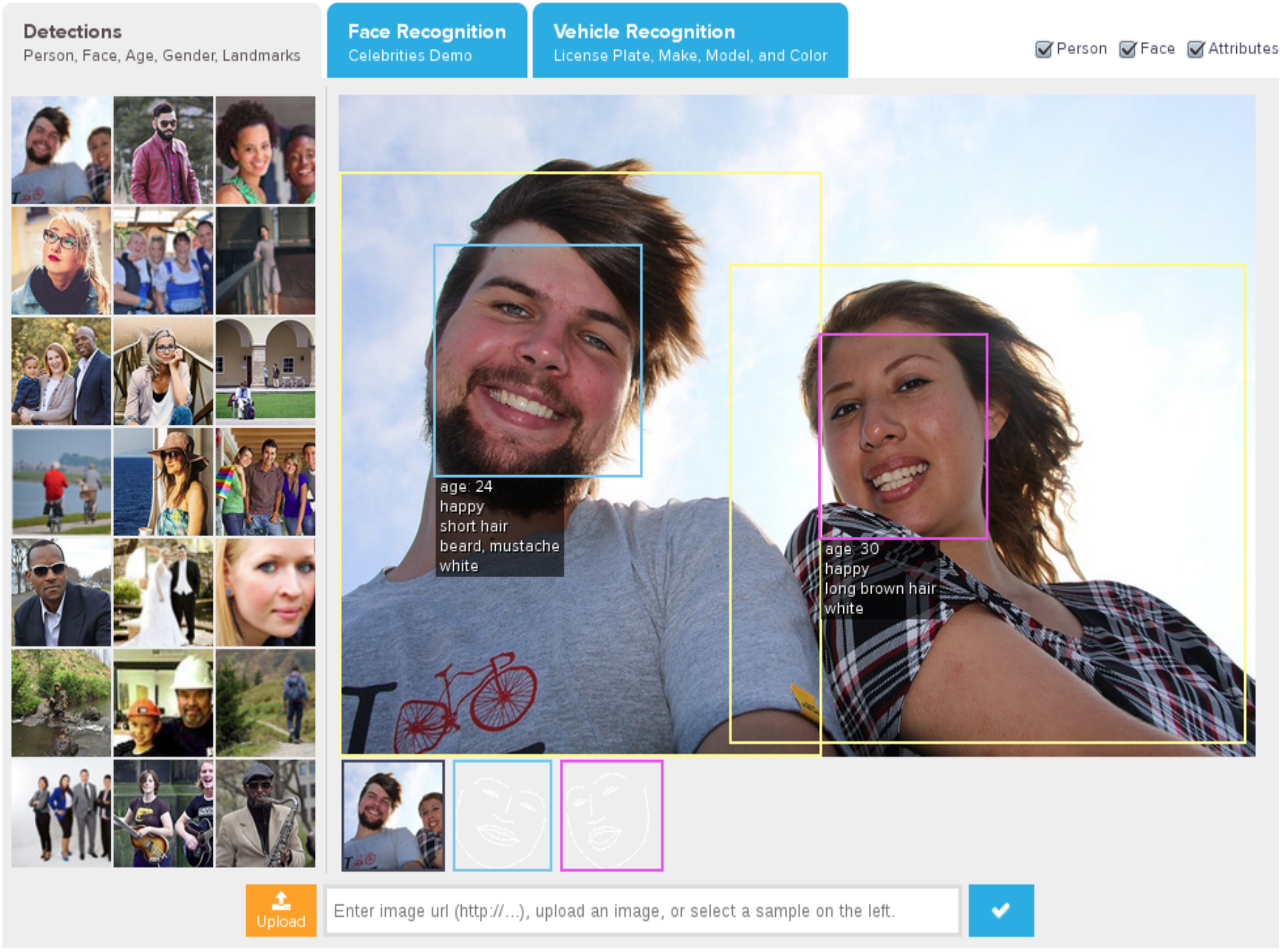
- Architecture is general in principle, but in practice lots of OpenCV specific tinkering required
  - “DARKLY exploits the fact that most OpenCV data structures for images and video include a separate pointer to the actual pixel data. For example, IplImage’s data pointer is stored in the imageData field; CvMat’s data pointer is in the data field. For these objects, DARKLY creates a copy of the data structure, fills the meta-data, but puts the opaque reference in place of the data pointer. Existing applications can thus run without any modifications as long as they do not dereference the pointer to the pixels”

Its hard to make privacy transforms *principled*

- Not always clear what a system needs to perform its work, and manual intervention is problematic
  - “The sketch of an image is intended to convey its high-level features while hiding more specific privacy-sensitive details. A loose analogy is publicly releasing statistical aggregates of a dataset while withholding individual records.”
  - May reduce performance in unexpected ways
  - May reduce privacy in unexpected ways
    - Not always intuitive what privacy protections are guaranteed by different transformations of visual input: sketching transform
    - Example: Gaussian blur

Detection API & Recognition API

Sighthound Cloud offers a Detection API for person, face, gender, age, and facial landmark detections; and a Recognition API that developers can use for face and vehicle recognition applications. Try out the following demo to see the Detection API and Recognition API in action.



New module: Database privacy

COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

Hilda Schrader Whitcher is an employee at Woolworth.

075-05-1120 is an employee at Woolworth.

```
-= EQUIHAX STATEMENT =-

   O
  <\==-   -   - -   -  -  - ---
 ./ \                          _/\_\O <<< EQUIFAX

 ^^ EQUIHAX

FOR PUBLIC RELEASE:
We like the fake hackers idea wanting 600 BTC,
We decide we want crowdfund 600 BTC or 8400 ETH for public release. :)

PUBIC ADDRESSES:
BTC: 1KELNpR9ECN46QaNGxPhoJDL4iqaa7Hgch
ETH: 0x8D992F58f3887cCD72A14FE29aD22Ed0789f70Ef

PRIVATE BUY:
- 4  BTC per 1,000,000 Entries
- 56 ETC per 1,000,000 Entries

INSTRUCTIONS FOR PRIVATE BUY:
STEP 1. SEND EXACTLY 0.2 BTC or 3 ETH TO PUBLIC ADDRESS.
STEP 2. EMAIL TRANSACTION ID TO EQUIHAX AT PROTONMAIL.COM (USE PGP KEY).
STEP 3. EQUIHAX WILL SEND YOU A UNIQUE ADDRESS FOR PRIVATE PURCHASE.

IF YOU DO NOT EMAIL A TRANSACTION ID TO EQUIHAX WITH PGP KEY, WE WILL NOT RESPOND.

SAMPLES FROM OUR TREASURE TROVE:
```

```
Thank you for participating in this year's Equihax! xD
```

```
{
  "requestId" : null,
  "ssn" : [redacted],
  "dob" : "06/14/1946",
  "firstName" : "DONALD",
  "lastName" : "TRUMP",
  "middleName" : "JOHN",
  "city" : "NEW YORK",
  "state" : "NY",
  "streetName" : "5TH",
  "streetType" : "AVE",
  "streetNumber" : "725",
  "postalCode" : "10022",
  "fraudCode" : null,
  "fraudMessage" : null,
  "totalLiability" : null,
  "equifaxStatus" : null,
  "creditScore" : 819,
  "creditReportResourceId" : 2389,
  "creditReportFileSize" : null,
  "creditReportName" : "[redacted]",
  "creditReportMimeType" : "pdf",
  "creditReportSha1CheckSum" : "85b1ab26e5b2d45805a5daaf78d5d6c234300c61",
  "errorCode" : 0,
  "errorMessage" : null,
  "equifaxErrorInfo" : null,
  "createdTime" : "2017-05-20T12:19:12.000+0000",
  "updatedTime" : "2017-05-20T14:39:36.000+0000",
}

{
  "requestId" : null,
  "ssn" : [redacted],
  "dob" : "10/21/1980",
  "firstName" : "KIMBERLY",
  "lastName" : "KARDASHIAN",
  "middleName" : "NOEL",
  "city" : "HIDDEN HILLS",
  "state" : "CA",
  "streetName" : "ELDORADO MEADOW",
  "streetType" : "RD",
  "streetNumber" : "25254",
  "postalCode" : "91302",
  "fraudCode" : null,
  "fraudMessage" : null,
  "totalLiability" : null,
  "equifaxStatus" : null,
  "creditScore" : 643,
  "creditReportResourceId" : 4297,
  "creditReportFileSize" : null,
  "creditReportName" : "[redacted]",
  "creditReportMimeType" : "pdf",
  "creditReportSha1CheckSum" : "7d295821a08c40a5162e7f446aadb2eb9e46710c",
  "errorCode" : 0,
  "errorMessage" : null,
  "equifaxErrorInfo" : null,
  "createdTime" : "2017-05-20T12:19:12.000+0000",
  "updatedTime" : "2017-05-20T14:39:36.000+0000",
}

{
  "requestId" : null,
  "ssn" : [redacted],
  "dob" : "10/28/1955",
  "firstName" : "WILLIAM",
  "lastName" : "GATES"
```


redacted, but present in original

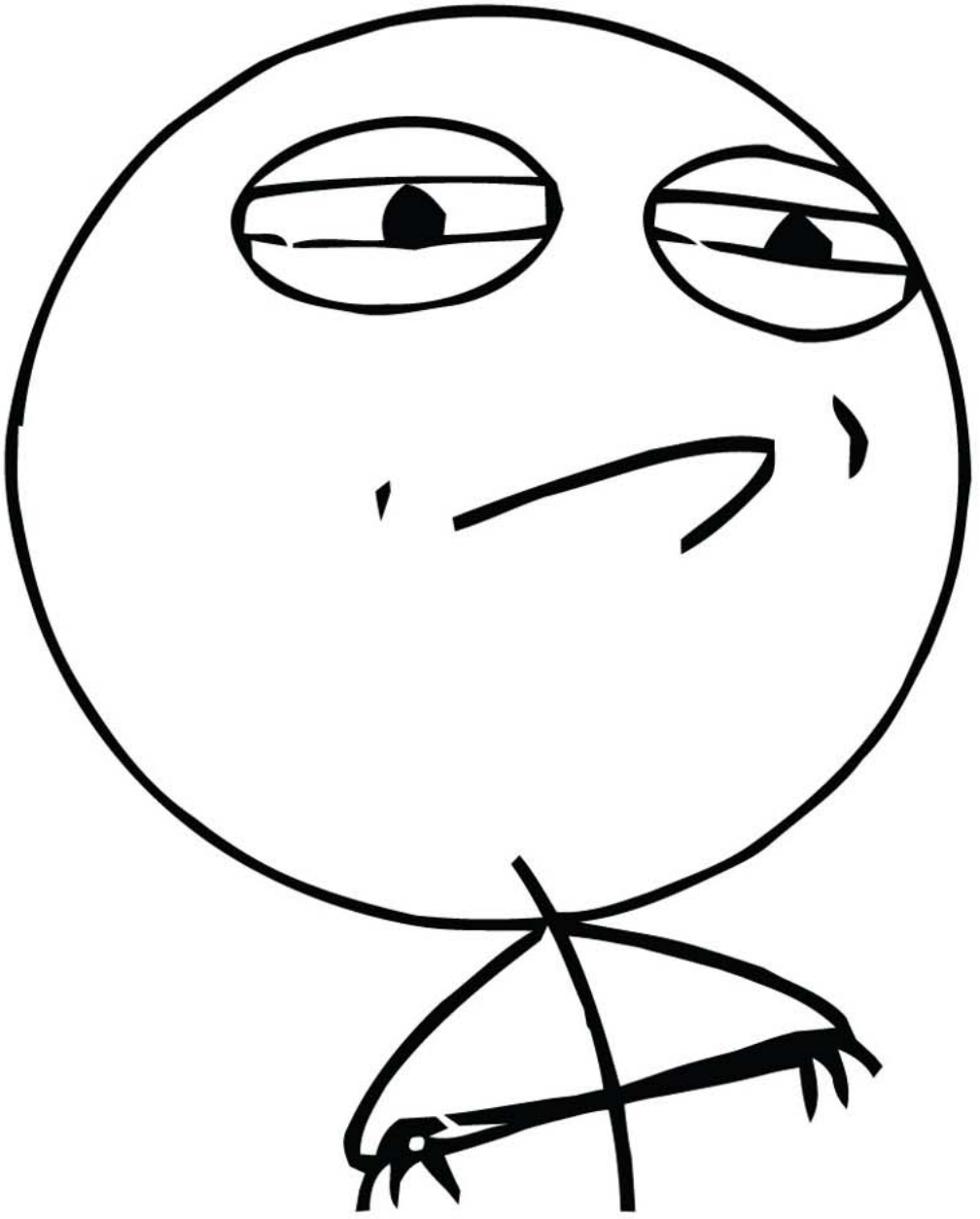
- Occupation: Professional Basketball player (NBA)
- Height:
  - Muggsy Bogues: 5’3”
  - Manute Bol: 7’7”
- Occupation, year active, height is an indirect identifier for these two

Newcombe's version of Muggsy

- In English speaking countries, which of these has more distinguishing power?
  - Zbigniew Zabrinsky
  - John Smith
- Newcombe used odds ratios, with cutoffs used to indicate links
- Fellegi and Sunter: showed optimality under fixed upper bounds on the false link (match) rates and the false non-link (non-match) rates

See Peter Christen, **Data Matching**

- Pre-HIPAA
- “De-identified” hospital records
- Attributes included ZIP, DOB, gender



CHALLENGE ACCEPTED

Governor William Weld, after his data was identified (dramatic reenactment)

- Pseudoidentifiers are built by combining attributes
  - example from earlier in the semester web browsing?

List of fingerprinting sources

1. UserAgent
2. Language
3. Color Depth
4. Screen Resolution
5. Timezone
6. Has session storage or not
7. Has local storage or not
8. Has indexed DB
9. Has IE specific 'AddBehavior'
10. Has open DB
11. CPU class
12. Platform
13. DoNotTrack or not
14. Full list of installed fonts (maintaining their order, which increases the entropy), implemented with Flash.
15. A list of installed fonts, detected with JS/CSS (side-channel technique) - can detect up to 500 installed fonts without flash
16. Canvas fingerprinting
17. WebGL fingerprinting
18. Plugins (IE included)
19. Is AdBlock installed or not
20. Has the user tampered with its languages [1]
21. Has the user tampered with its screen resolution [1]
22. Has the user tampered with its OS [1]
23. Has the user tampered with its browser [1]
24. Touch screen detection and capabilities
25. Pixel Ratio
26. System's total number of logical processors available to the user agent.
27. Device memory

- Pseudoidentifiers are build by combining attributes
- As a matter of US law, only records that are **public** are part of the body of information that can be used to assess “personally-identifiable information”
- But all bets are off if you’re trying to deanonymize using data matching: information gathering can be cumulative, building on itself
- Let’s look at how data integration works in more detail