

Anonymous Communication and Traffic Analysis

Information Privacy with Applications
David Sidi (dsidi@email.arizona.edu)



Warm-up

- Give the definition of `anonymity' from either Torra, or Danezis and Diaz.

Small mention of interesting things

- Assignment 1 deadline has been extended to next Tuesday



COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

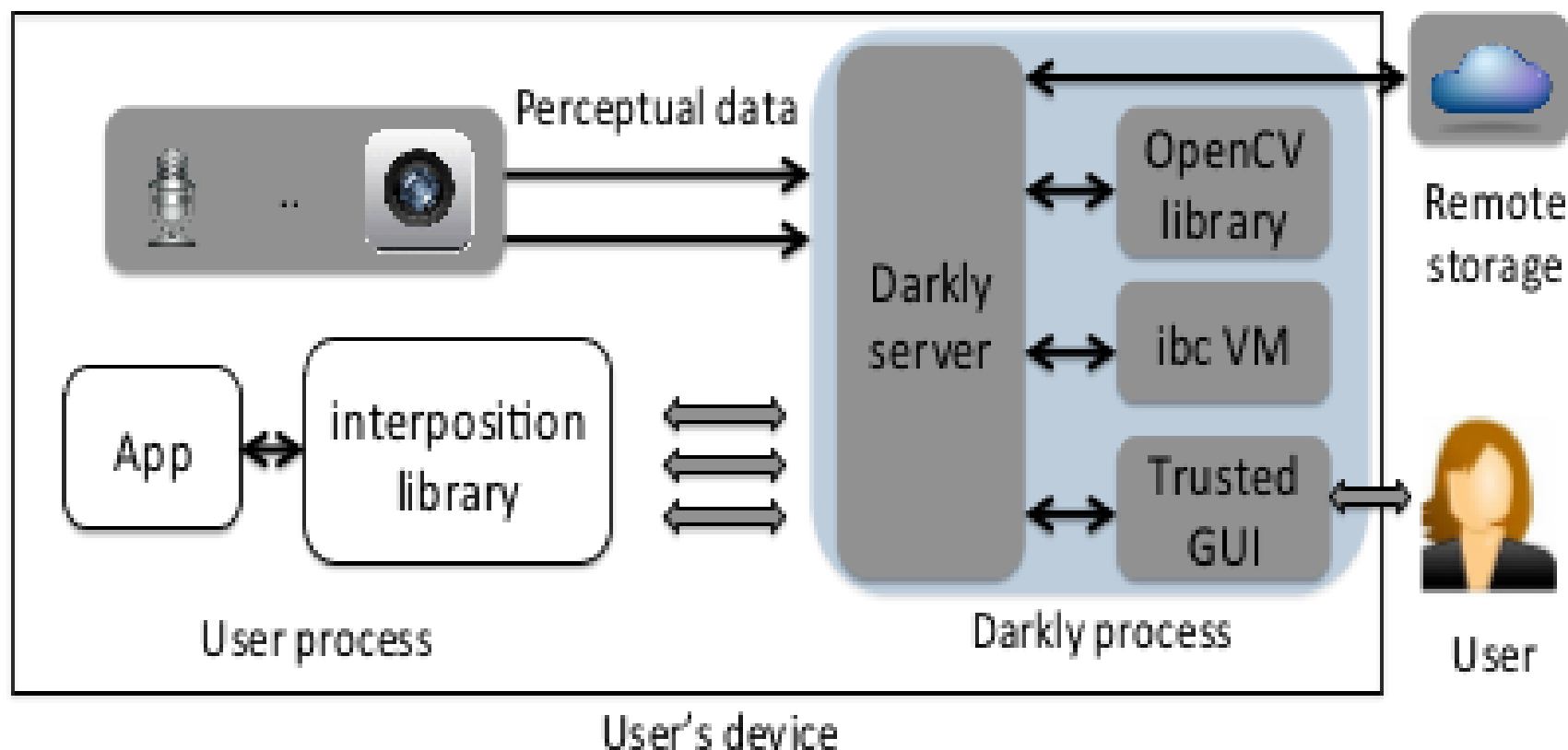
Finishing last time

Infrastructural approaches

- Suppose you own devices with perceptual capabilities (for example, at home), and want to be sure that they don't misbehave
- [Darkly](#) (@ 31:49 - 43:00)
- Trust includes
 - device operating system
 - the hardware of its perceptual sensors
- Trust does not include a third party application running on your device

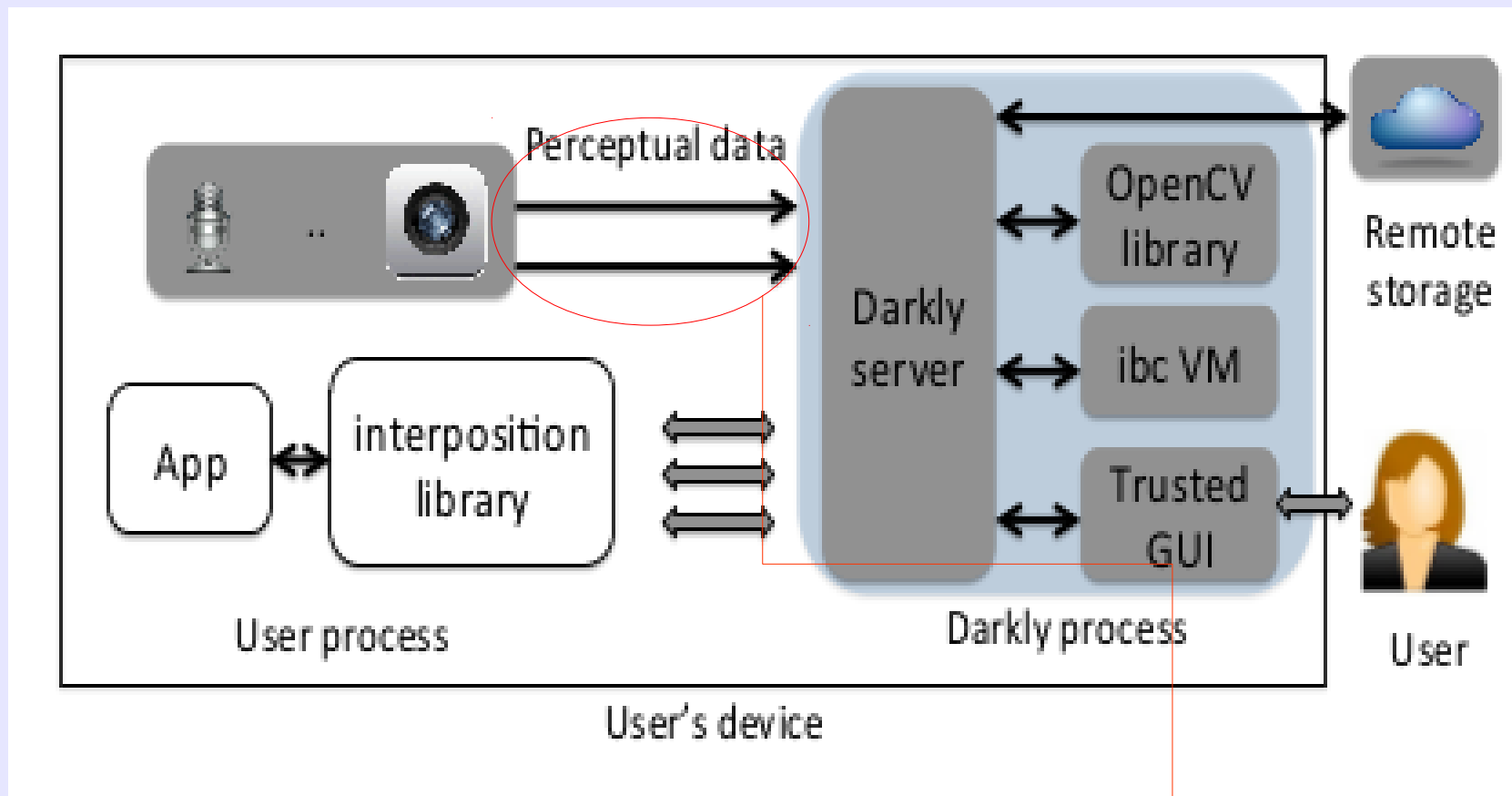
Infrastructural approaches

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



Infrastructural approaches

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



opaque references

Opaque references rely on OS user isolation in *nix

Opaque references are to addresses in the kernel address space.

```
$readelf -l davids_demo_proggy
```

```
Elf file type is EXEC (Executable file)
```

```
Entry point 0x400480
```

```
There are 9 program headers, starting at offset 64
```

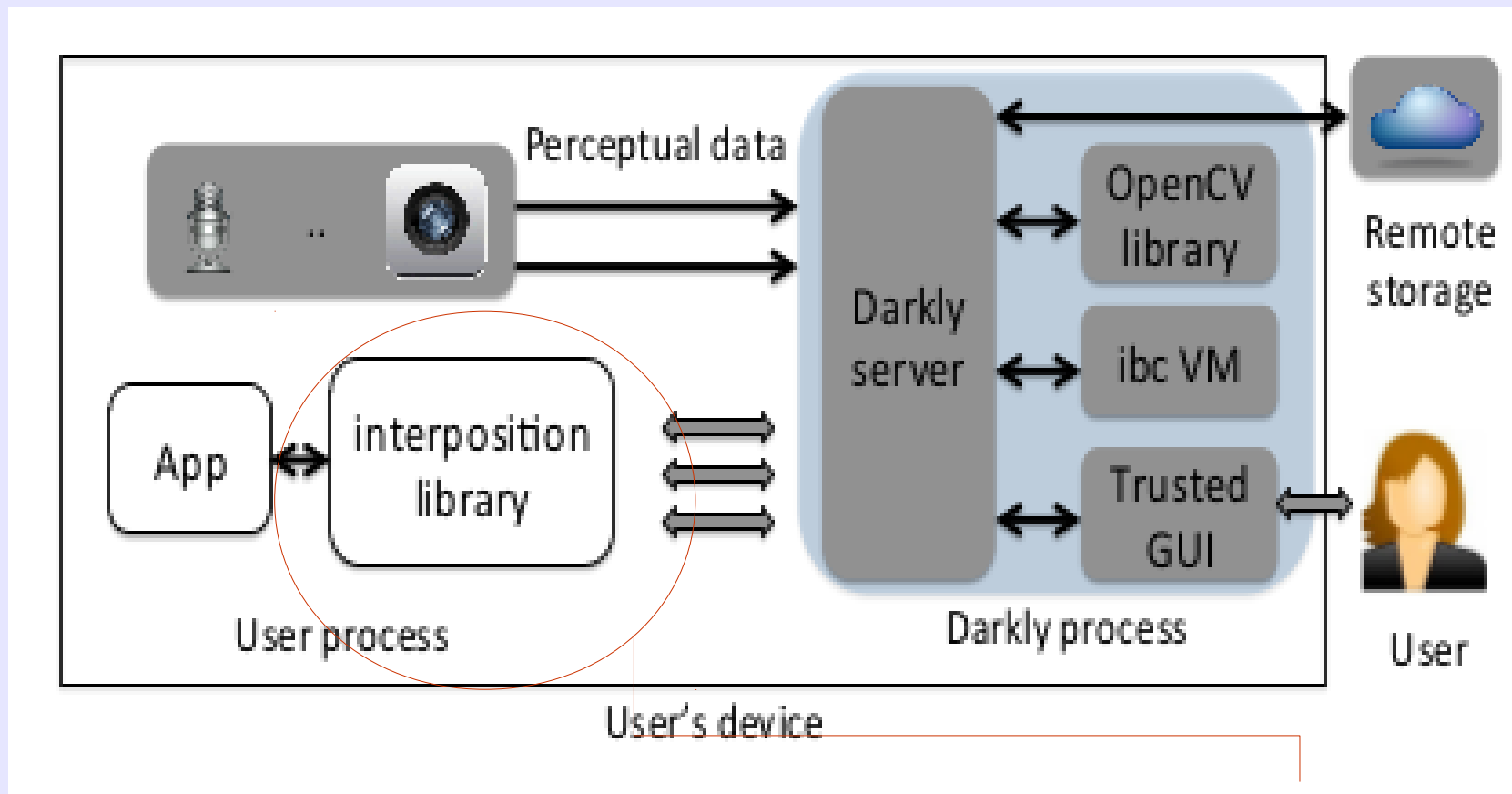
```
Program Headers:
```

Type	Offset FileSiz	VirtAddr MemSiz	PhysAddr Flags Align
PHDR	0x0000000000000040	0x0000000000400040	0x0000000000400040
	0x00000000000001f8	0x00000000000001f8	R E 8
INTERP	0x0000000000000238	0x0000000000400238	0x0000000000400238
	0x00000000000001c	0x00000000000001c	R 1
[Requesting program interpreter: /lib64/ld-linux-x86-64.so.2]			
LOAD	0x0000000000000000	0x0000000000400000	0x0000000000400000
	0x00000000000007fc	0x00000000000007fc	R E 200000

```
--snip--
```

Infrastructural approaches

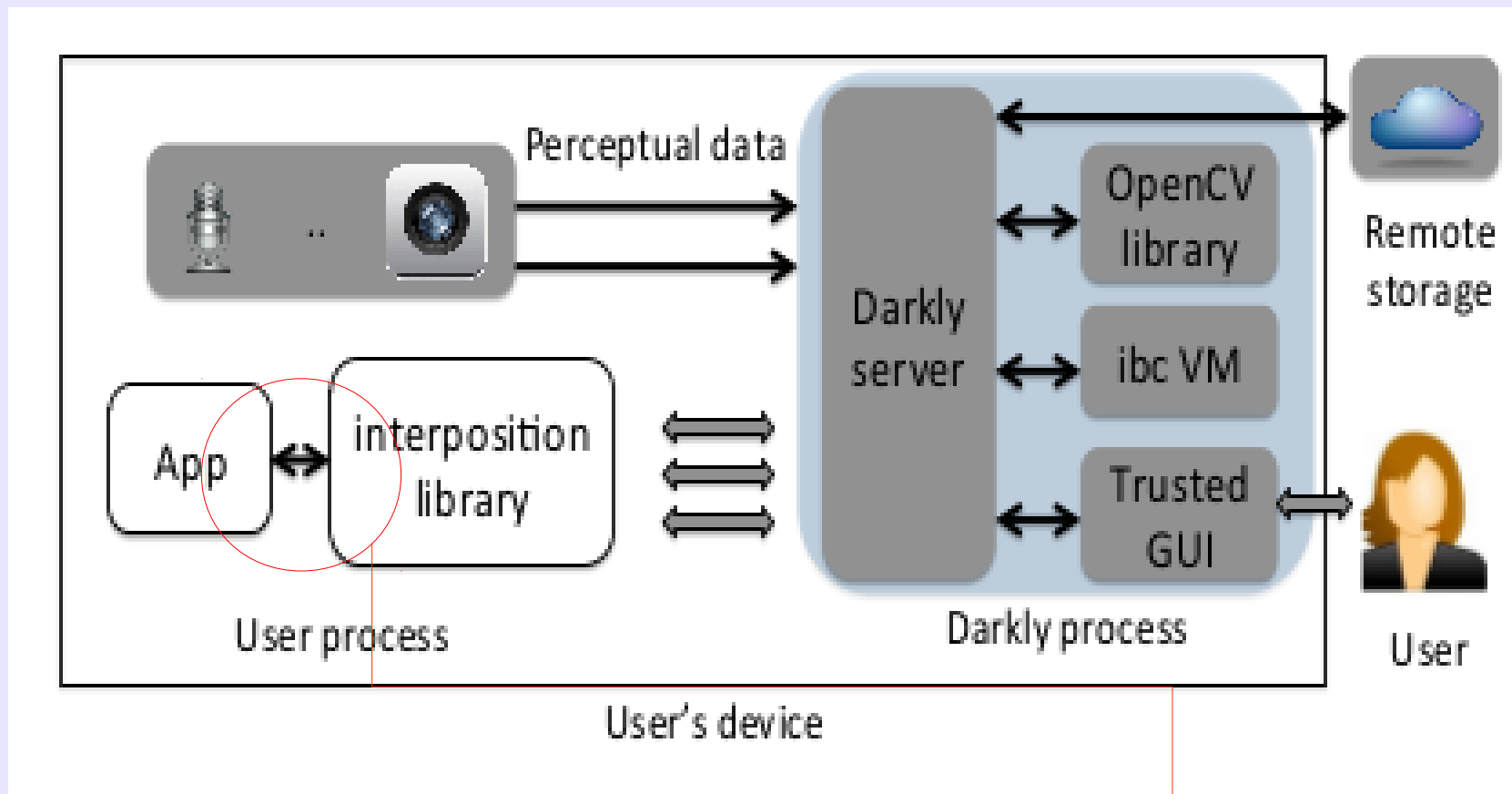
- “the application will never have access to the raw pixels”
 - opaque references



Local call, or forward to DARKLY?

Infrastructural approaches

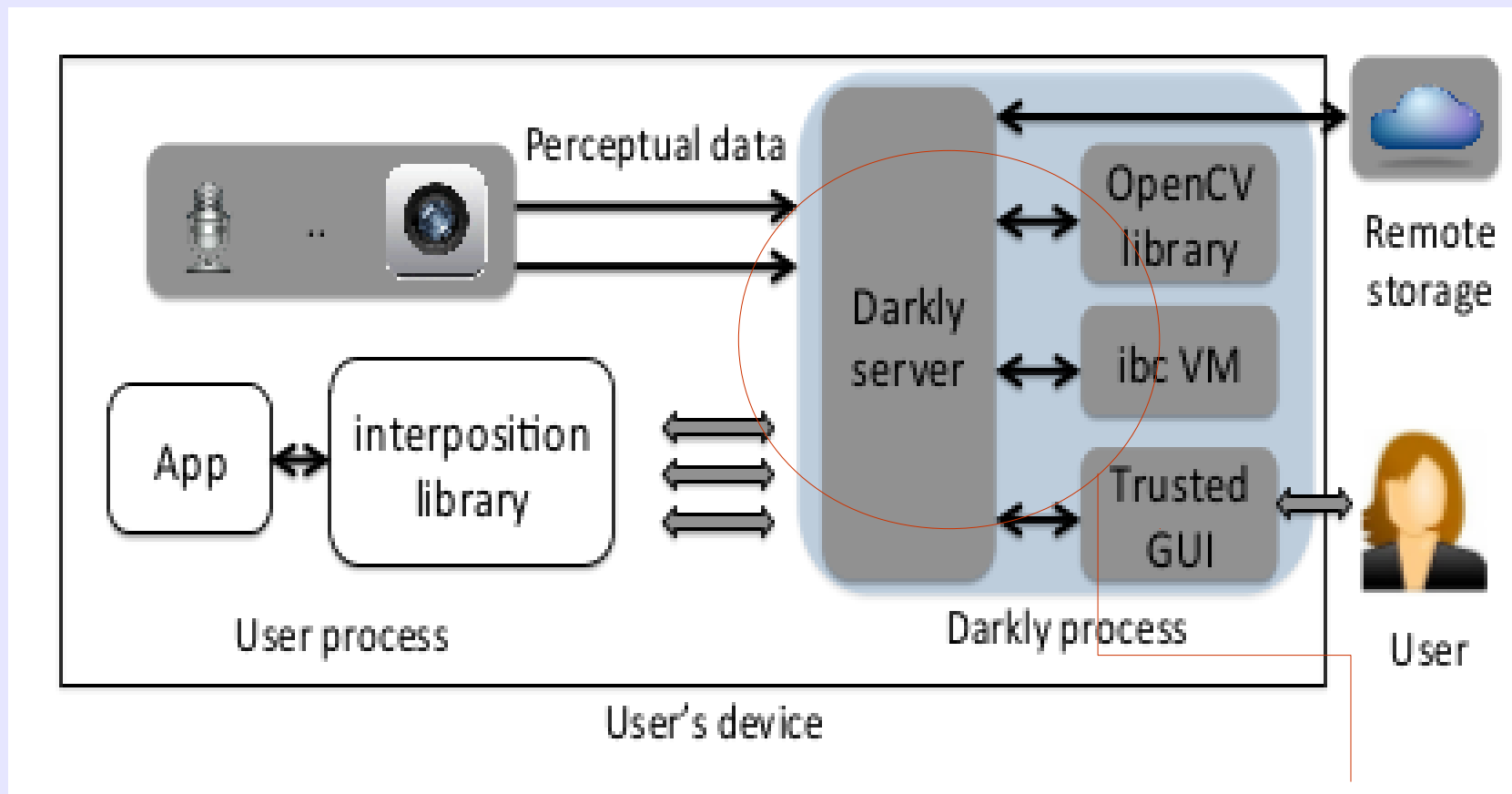
- “the application will never have access to the raw pixels”
 - opaque references



any opaque references?

Infrastructural approaches

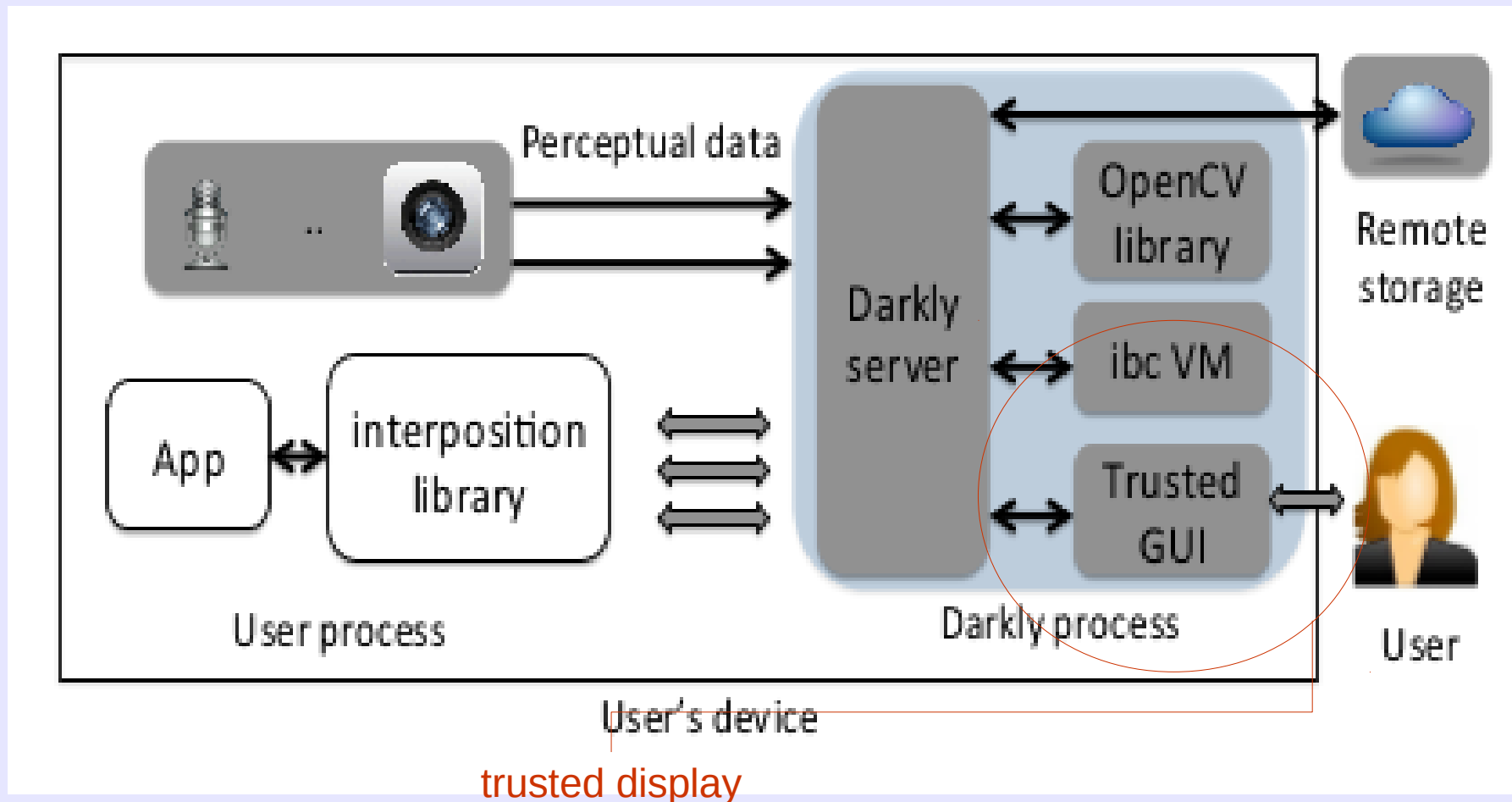
- “the application will never have access to the raw pixels”
 - opaque references



declassifiers (e.g., sketching transform)

Infrastructural approaches

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



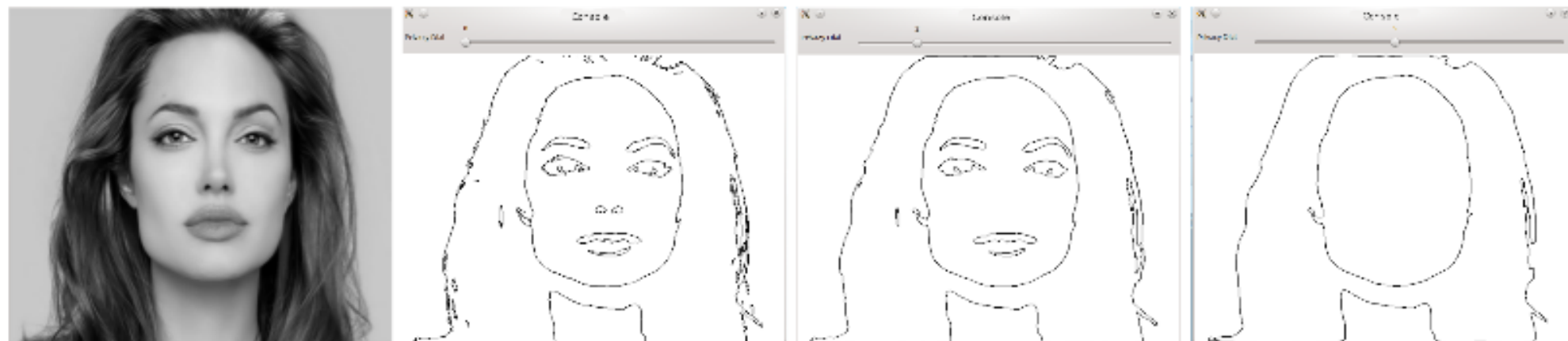


Figure 2. Output of the sketching transform on a female face image at different privacy levels.

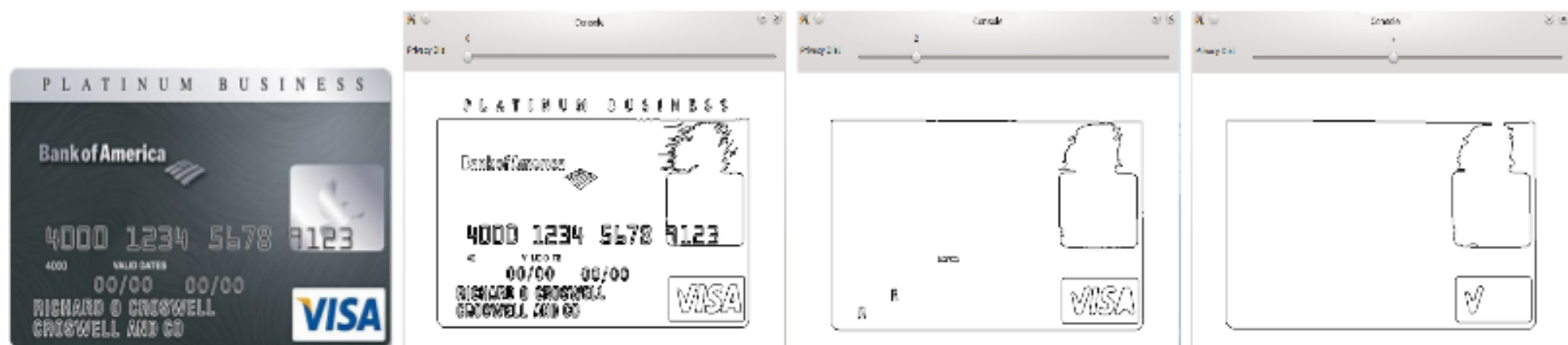


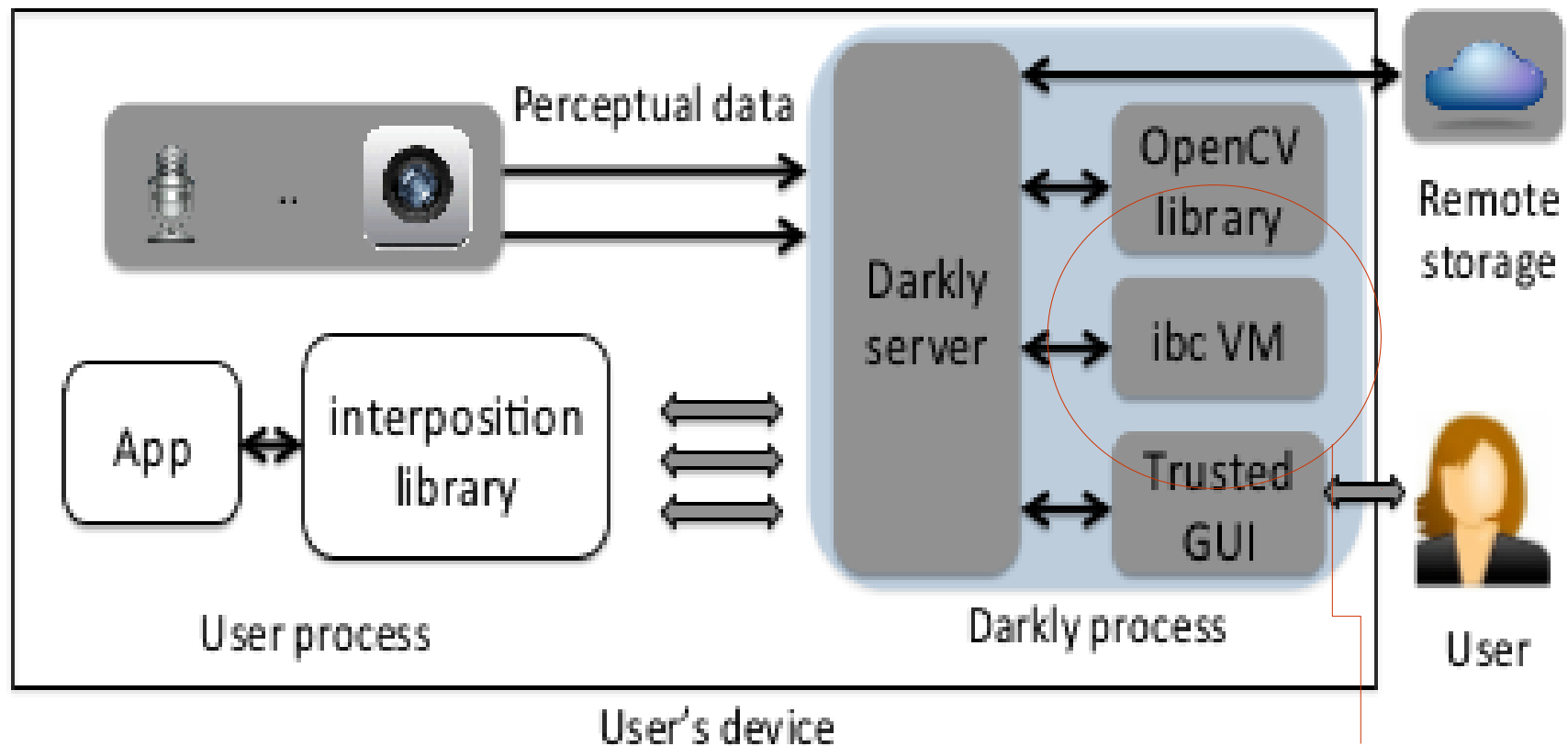
Figure 3. Output of the sketching transform on a credit card image at different privacy levels.

Why is the display sensitive? (2 minutes)

Why is the display sensitive? (2 minutes)
ANALOG HOLE

Infrastructural approaches

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



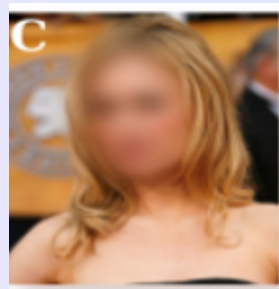
isolate untrusted code running on raw input

Limitations of DARKLY

- DARKLY is awesome and forward-looking, but like any project it has limitations
- 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”

Limitations of DARKLY

- 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



Examples of approaches based on prevention

- Bodyguard FLARE home security camera ([link](#). Also, among the [funniest videos I've seen](#))
- A depolarized monitor matched to polarizing glasses ([link](#))



Anonymity



COLLEGE OF SOCIAL & BEHAVIORAL SCIENCES
School of Information

Terminology Review



Anonymity set

- ‘Anonymity’ is defined with respect to a subset of the possible senders, called the anonymity set.
- Think of it as answering “who might you be?”

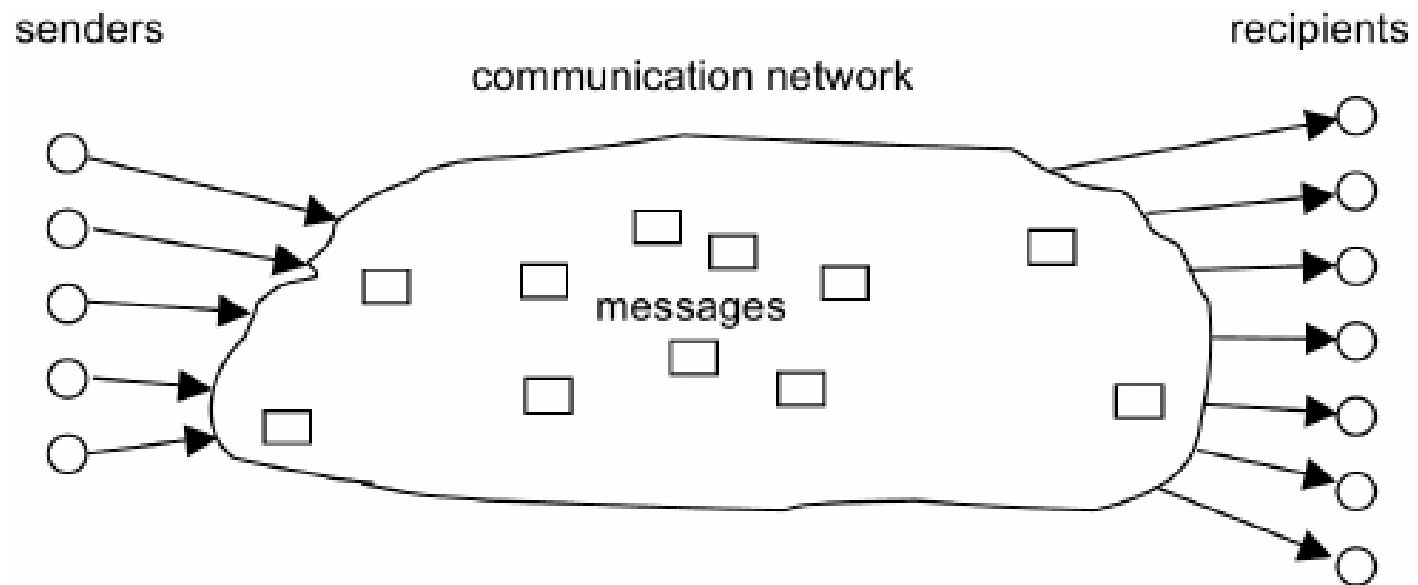
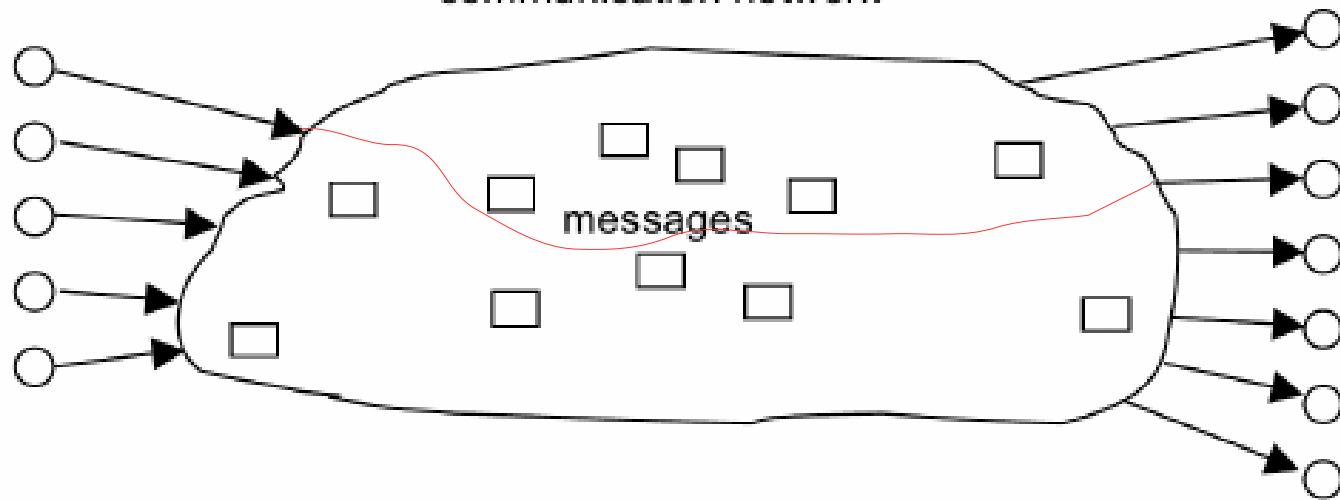


Image credit (before modification):
Christina Pöpper
Ruhr-University Bochum

senders

communication network

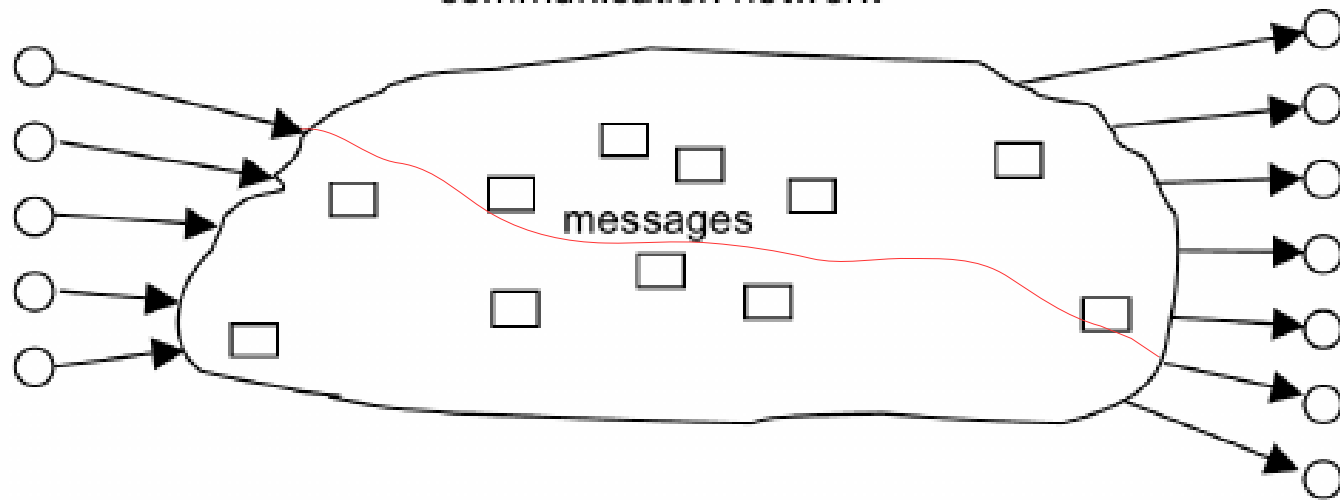
recipients



senders

communication network

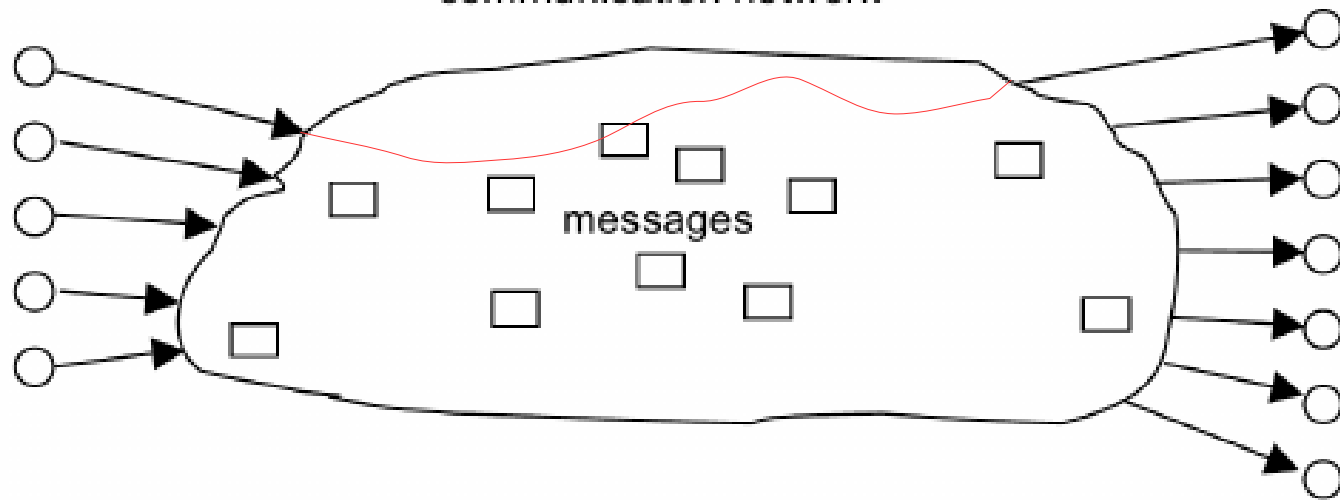
recipients

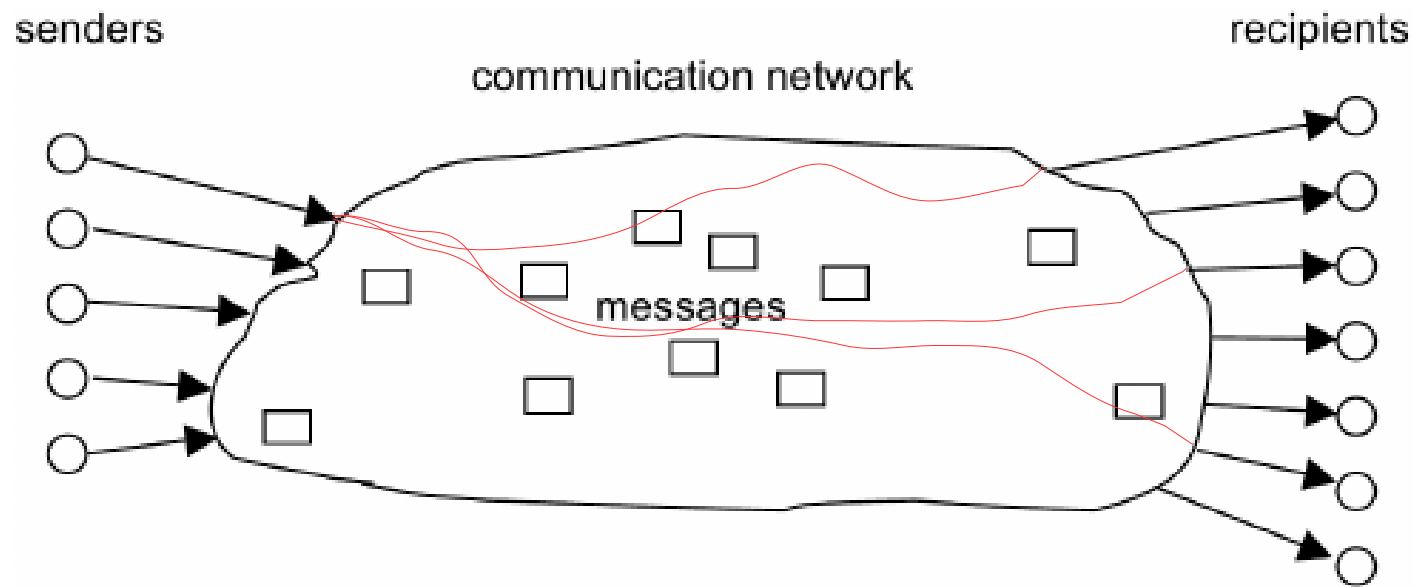


senders

communication network

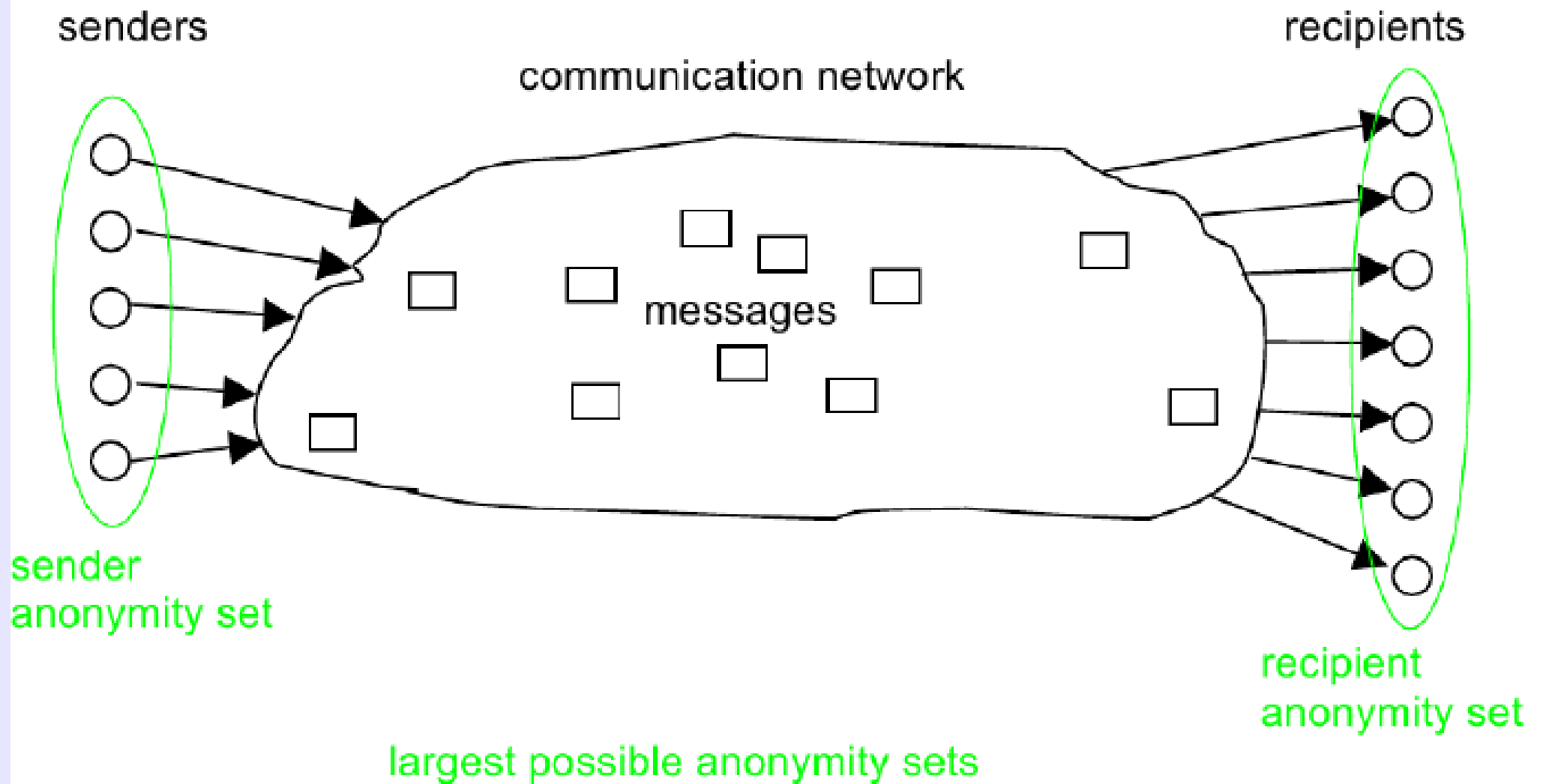
recipients





Anonymity set

- Can you clearly describe the limiting cases for the anonymity set?



Two linked definitional questions

- What is a sender? i.e., how do we get the set of all senders? (Think about the definitions)

Two linked definitional questions

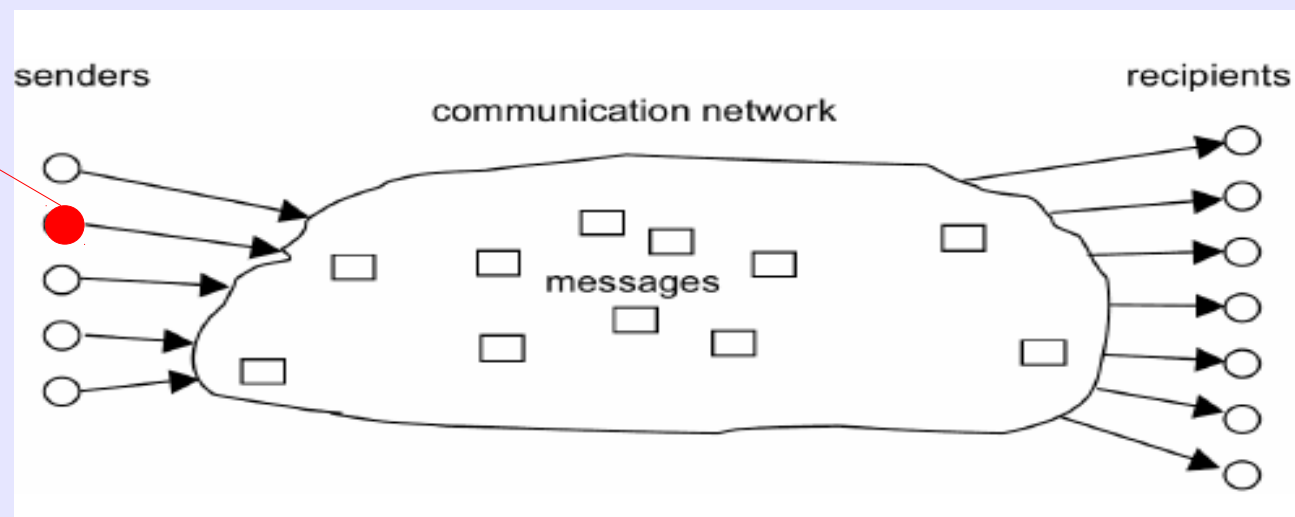
- What is a sender? i.e., how do we get the set of all senders? (Think about the definitions)
 - something that sends messages over the network to recipients (implements protocols, etc.). People, personal computers, cameras, phones, etc.

Two linked definitional questions

- What is a sender? i.e., how do we get the set of all senders? (Think about the definitions)
 - something that sends messages over the network to recipients (implements protocols, etc.). People, personal computers, cameras, phones, etc
- if that were all, all senders would be the same! But the anonymity set is intended to be useful, not trivial, in its separation of senders that cannot be distinguished from those that can be

Senders and recipients are thought of as rows in a database table

	MAC	Browser_fingerprint	IP	Sites_visited
SNDER_1	00:a0:ef:eb:5v:ff	af7f098c39728f8cb67 6e3df82ced01a149ee 3aa92af2b88c20c494 8a5fad5fd		torproject.org, ischool.arizona.edu, maps.google.com...
SNDER_2	00:c0:ff:dd:ff:ef	a5fad5fdd01a149eeaf 7f098c39728f8cb676e 3df82ce3aa92af2b88c 20c4948		nytimes.com, purple.com





Question

- Suppose I include in a record of UA students a person's weight and height as 150 lbs, 5'3". Is the player anonymous? (think: anonymity set)



Question

- Suppose I include in a record a person's weight and height as 150 lbs, 5'3".
- Now suppose further that I do so for a database of male UA basketball players. Is the player anonymous?



Question

- Suppose I include in a record a person's weight and height as 150 lbs, 5'3".
- Now suppose further that I do so for a database of male UA basketball players. Is the player anonymous?
- Where might you find combinations of attributes that are as rare as a short UA basketball player, but in other contexts?
 - Web: DNS, third-party tracking, browser fingerprinting

DNS

- You are required by law to say that DNS is “like a phone book” give it an easy-to-remember name get an IP address for the server hosting the website
- Forwarding DNS server
- Recursive DNS server (or resolver)
- (Root nameserver)
- (Top Level Domain nameserver)
- Authoritative nameserver

DNS

```
$dig arizona.edu
```

```
; <<>> DiG 9.9.5-9+deb8u14-Debian <<>> arizona.edu
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 14058
;; flags: qr rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:
;arizona.edu.                IN      A

;; ANSWER SECTION:
arizona.edu.                6813IN   A      128.196.128.233

;; Query time: 32 msec
;; SERVER: 208.67.222.222#53(208.67.222.222)
;; WHEN: Mon Oct 23 12:43:03 MST 2017
;; MSG SIZE rcvd: 56
```

DNS

- Suppose I use a VPN to tunnel my traffic to a server I control. What can you learn about me from my DNS requests?

DNS

- Suppose I use a VPN to tunnel my traffic to a server I control. What can you learn about me from my DNS requests?
- Many sites to do with local things in Tucson, AZ
- Sites to do with the University of Arizona
- Sites for groups with small memberships (Xerocraft)
- Many hits for a site with a public record attached to one person (sidiprojects.us)

Browser fingerprinting

- UserAgent
- Language
- Color Depth
- Screen Resolution
- Timezone
- Has session storage or not
- Has local storage or not
- Has indexed DB
- Has IE specific 'AddBehavior'
- Has open DB
- CPU class
- Platform
- DoNotTrack or not
- Full list of installed fonts (maintaining their order, which increases the entropy), implemented with Flash.
- A list of installed fonts, detected with JS/CSS (side-channel technique) - can detect up to 500 installed fonts without flash
- Canvas fingerprinting
- WebGL fingerprintingPlugins (IE included)
- Is AdBlock installed or not
- Has the user tampered with its languages 1
- Has the user tampered with its screen resolution 1
- Has the user tampered with its OS 1
- Has the user tampered with its browser 1
- Touch screen detection and capabilities
- Pixel Ratio
- System's total number of logical processors available to the user agent.



Browser fingerprinting

- Multi-monitor detection,
- Internal HashTable implementation detection
- WebRTC fingerprinting
- Math constants
- Accessibility fingerprinting
- Camera information
- DRM support
- Accelerometer support
- Virtual keyboards
- List of supported gestures (for touch-enabled devices)
- Pixel density
- Video and audio codecs availability
- Audio stack fingerprinting

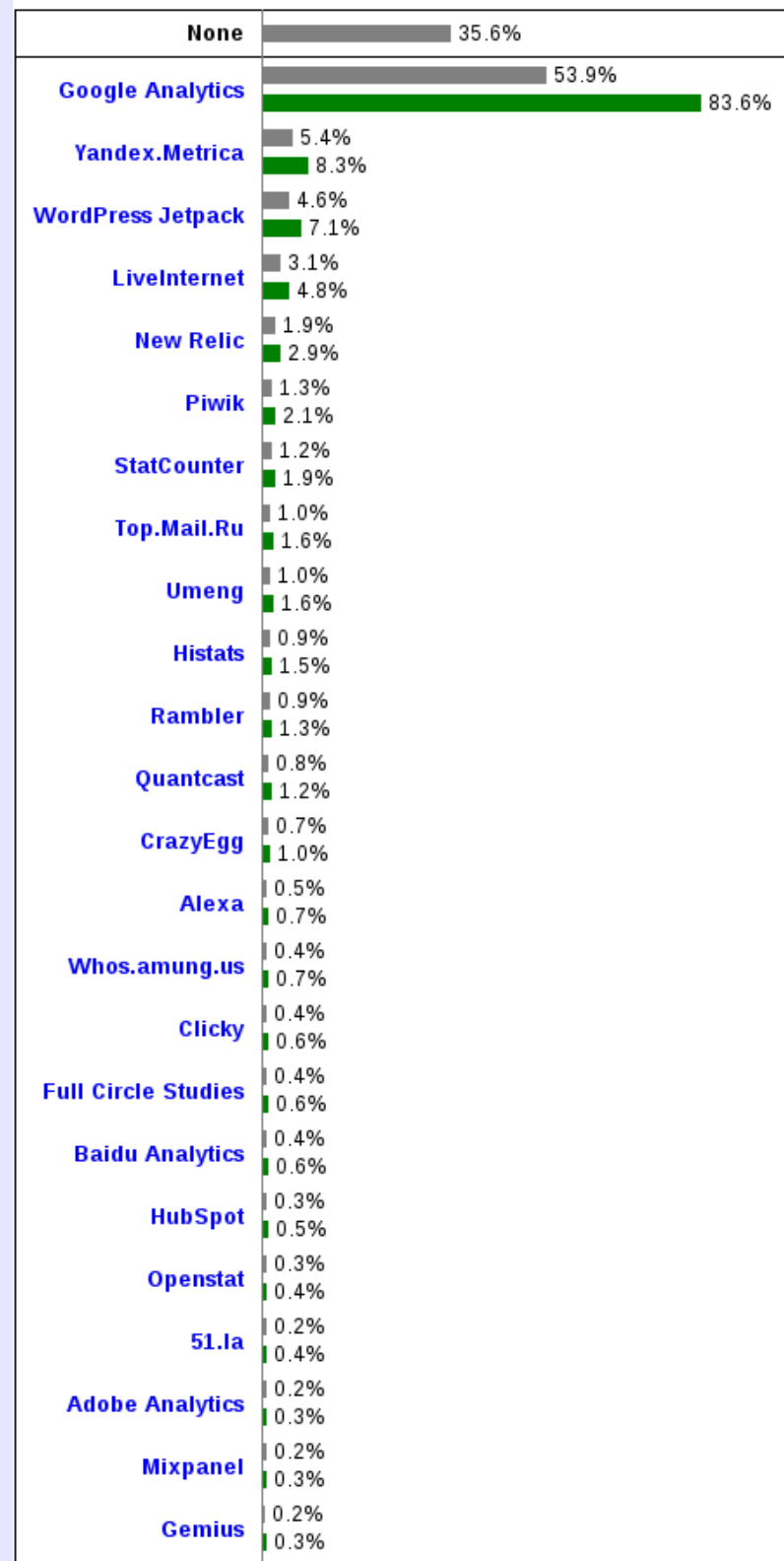


Third-party analytics

- 53.9% of all sites use Google Analytics

https://w3techs.com/technologies/overview/traffic_analysis/all

CC-SA License by David Sidi



Third-party analytics

- 53.9% of all sites use Google Analytics

Creating a complete picture

Begin by centralizing your data. Analytics 360 pulls in data across:



Online Sources



Offline Sources



Google Sources

Data from your site, app, internet-connected devices, and even offline sources will be connected in one place. If you're using Google and DoubleClick advertising products, seamless, out-of-the-box integrations mean you can pull in that information to create a single, complete data source across all customer touchpoints.

<https://www.google.com/analytics/analytics/features/>

Third-party analytics

- 53.9% of all sites use Google Analytics

If you want to learn more about how your customers behave away from your site, you can share your Analytics 360 customer segments with Audience Center 360 to get additional insights like demographics, interests, and in-market information.

<https://www.google.com/analytics/analytics/features/>