UC RIVERSIDE

CS 260 – Privacy Seminar

https://spalab.cs.ucr.edu/teaching/cs260

Emiliano De Cristofaro
https://emilianodc.com

Course Objectives

1. Learn the “basics” of privacy (and privacy technologies)

   Its connection to security
   Its societal, ethical, and legal aspects
   Its relevance to engineering

2. Expose you to advanced research in CS and privacy in particular

   How to find, read, understand, and explain research papers
   Hands-on work on research projects

Enrolling

- Need explicit approval from me
- Pre-Requisites: Undergraduate or Graduate Security Class at UCR – not negotiable
- Not accepted past the 2nd week

Communication

- Piazza (https://piazza.com/ucr/spring2024/cs260) as the main communication channel
  - Announcements, slides, projects, polls, etc.
  - Discussion and Q&A

    privacydabest

Welcome!

- **Timetable**
  - 10 lectures, ~~Mon 3:30-4:50pm WCH 142~~ → pre-recorded lectures
  - 10 classes, Wed 3:30-4:50pm WCH 142 → mandatory attendance
- **Grading**
  - 50% Project
  - 25% Class Discussions
  - 25% Quizzes/Class Attendance/Class Participation
- **Office Hours (TBC)**
  - Mon 3:30-4:30 pm, in-person or on Zoom
    Please book a slot: https://calendly.com/emilianodc/cs260

Tentative Schedule

		Wednesday
Week 1	Intro to Privacy	Overview of the Projects
Week 2	Anonymity	Surveillance
Week 3	Privacy-oriented Crypto	Crypto Case Studies
Week 4	Differential Privacy (DP)	DP Case Studies
Week 5	Privacy in Machine Learning	Privacy and LLMs
Week 6	Tracking and Profiling	Tracking Case Studies
Week 7	Human Factors	Human Factor Case Studies
Week 8	Privacy and Cybersafety	Privacy and Law
Week 9	[Memorial Day]	Privacy by Design
Week 10	Project Presentations	

Lectures **Discussions**

Project

- You can work in groups of 2-3 students (non-negotiable)
  - The amount of expected *individual* work is an invariant
  - Each student will have to submit an *individual* project report

Timeline

- Project ranking due April 10
- Project proposal due April 24
- Weekly progress report due every Wednesday, May 1-29
- Project presentations June 3rd and 5th
- Project submission (report + codebase) due June 7th

Project List

1. Browser fingerprinting evolution through Internet archive
2. Auditing FP-Fed
3. Improving FP-Fed
4. Federated Learning for Hate Speech
5. Looking at r/Privacy for privacy advice
6. Tor bridge on Raspberry PI
7. Python “Private Set Intersection” Toolkit
8. Does Alexa listen to me?
9. How to set up a privacy clinic
10. Update the petlib library

What is Browser Fingerprinting (BFP)

- An invasive tracking technique
  - Stateless: no information is stored on the browser (e.g., cookies)

- Collecting a set of uniquely identifiable information related to device
  - Hardware (# CPU cores, screen size, etc.)
  - Software (Fonts installed, keyboard layout, etc.)

- Typically deployed via JS scripts in browser (e.g. fingerprintjs)

- Widely recognized as a threat to privacy
  - Can track users without consent, stable for long periods of time

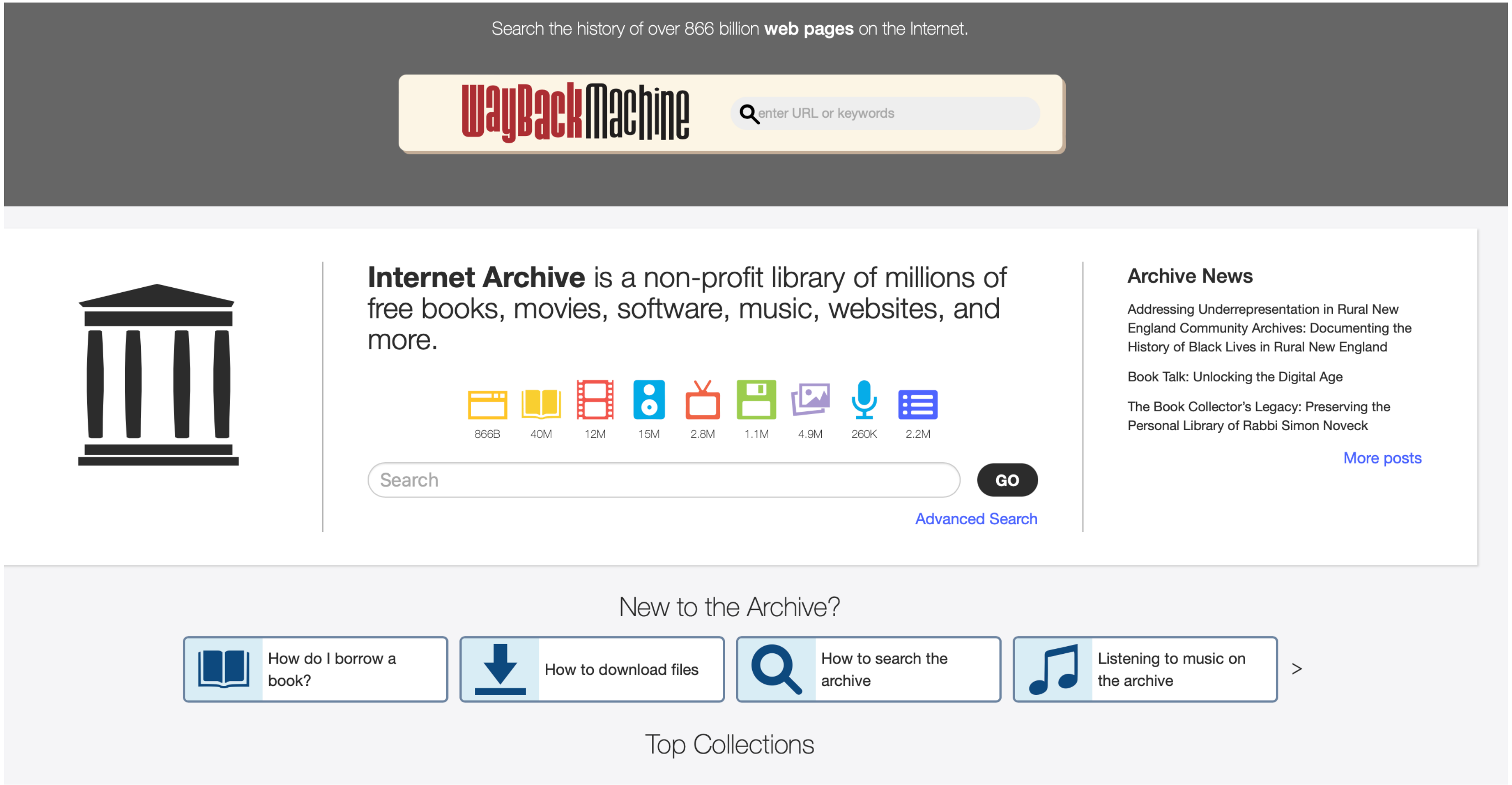
BFP Examples

- Canvas
  - Differences in ways images are rendered on different devices
- Canvas Font
  - Differences in ways text is rendered if it is installed vs not installed
- WebRTC
  - Uniqueness of peers present in WebRTC protocol
- Audio Context
  - Differences in how audio signals are processed by different hardware
- And more…
  - OS info (navigator.platform), screen size, keyboard layout, Java/Flash version, etc.

Project List

1. **Browser fingerprinting evolution through Internet archive**
2. Auditing FP-Fed
3. Improving FP-Fed
4. Federated Learning for Hate Speech
5. Looking at r/Privacy for privacy advice
6. Tor bridge on Raspberry PI
7. Python “Private Set Intersection” Toolkit
8. Does Alexa listen to me?
9. How to set up a privacy clinic
10. Update the petlib library

Internet Archive



Search the history of over 866 billion **web pages** on the Internet.

Internet Archive is a non-profit library of millions of free books, movies, software, music, websites, and more.

Archive News

Addressing Underrepresentation in Rural New England Community Archives: Documenting the History of Black Lives in Rural New England

Book Talk: Unlocking the Digital Age

The Book Collector's Legacy: Preserving the Personal Library of Rabbi Simon Noveck

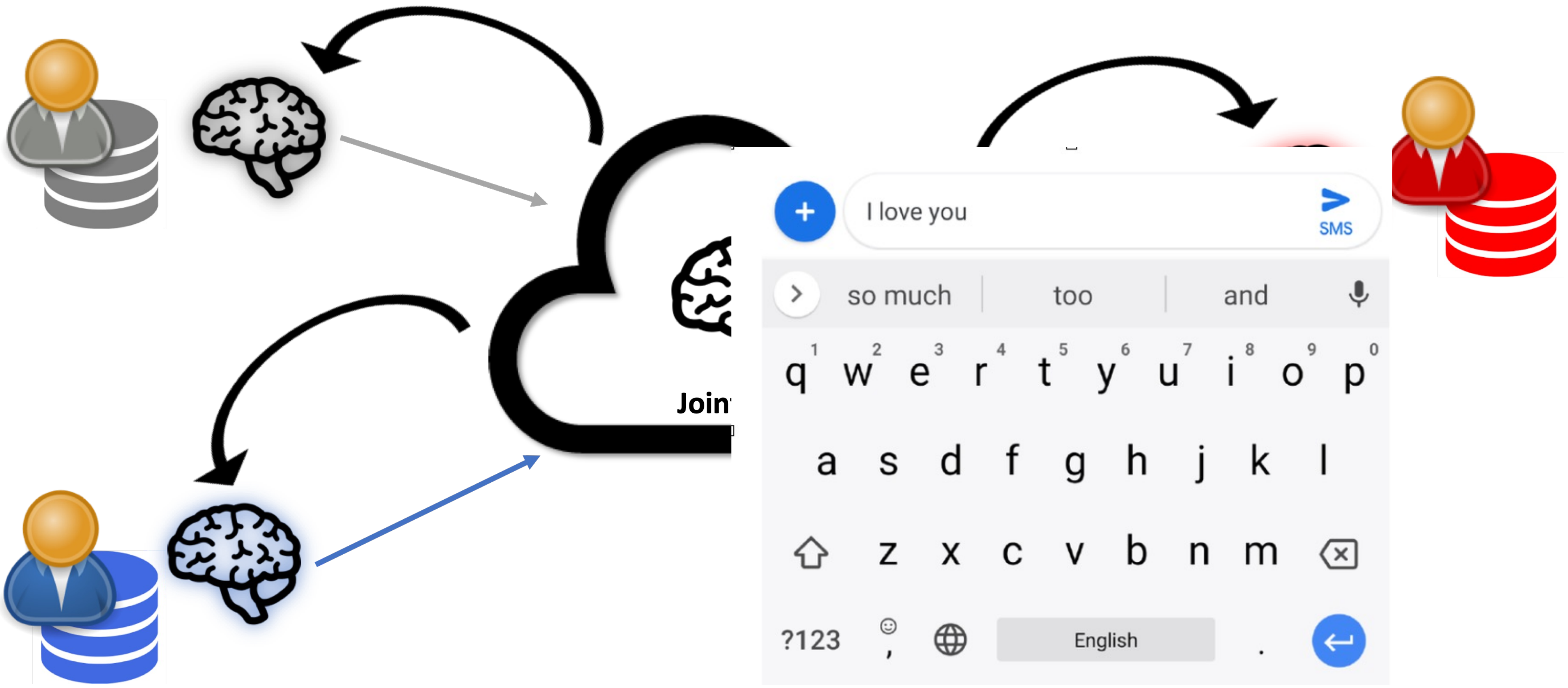
More posts

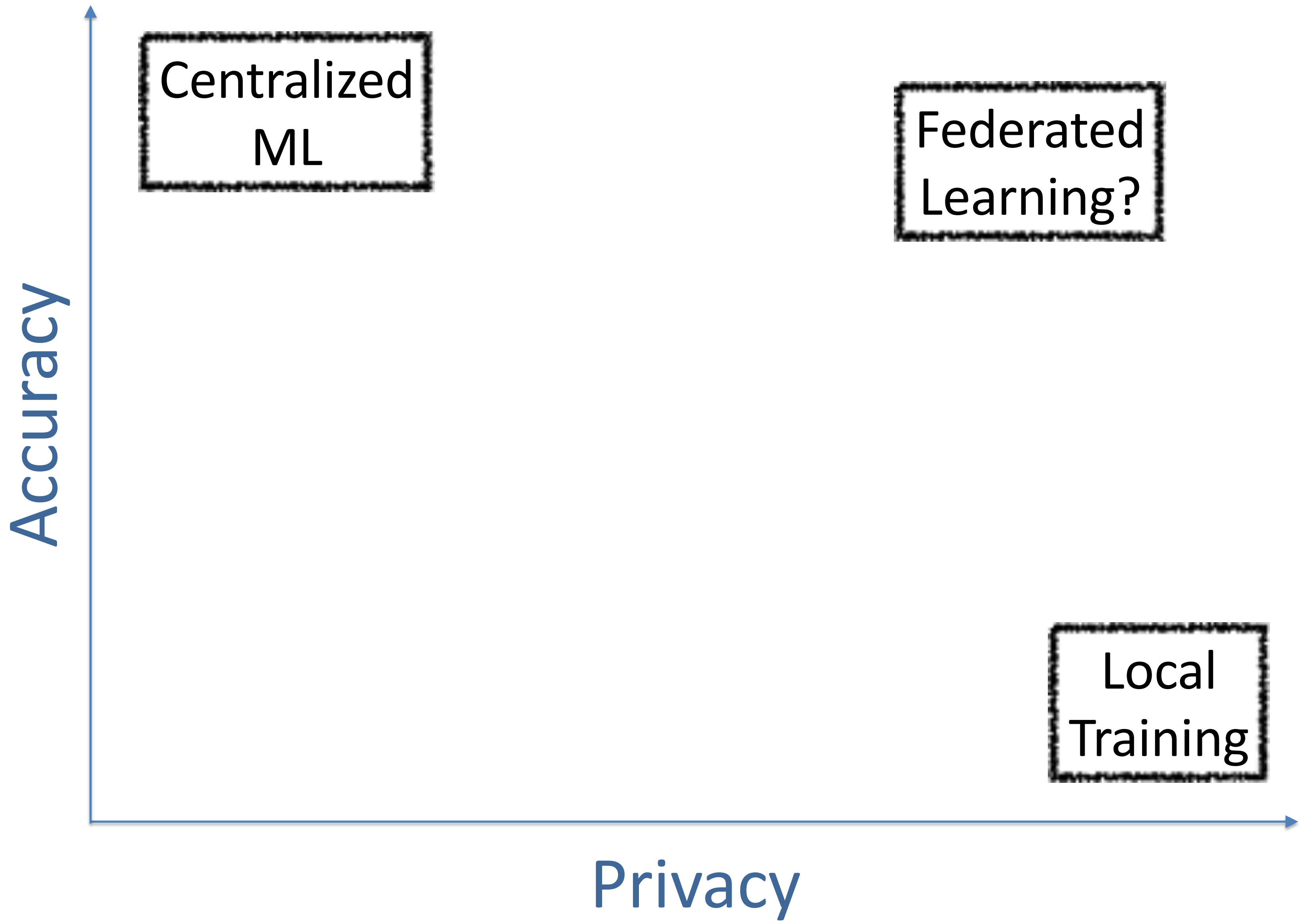
New to the Archive?

Top Collections

https://archive.org

Federated Learning




Centralized ML
Federated Learning?
Local Training
Accuracy
Privacy

Privacy in FL

- Sharing gradients better for privacy than sharing raw data
  - But prior work still shows (aggregate) gradients can be used to violate individuals' privacy

- Solution: share noisy gradients
  - Using the formal framework of Differential Privacy

Differential Privacy

Let X be the “data universe”

Let $D \subset X$ be the “dataset”

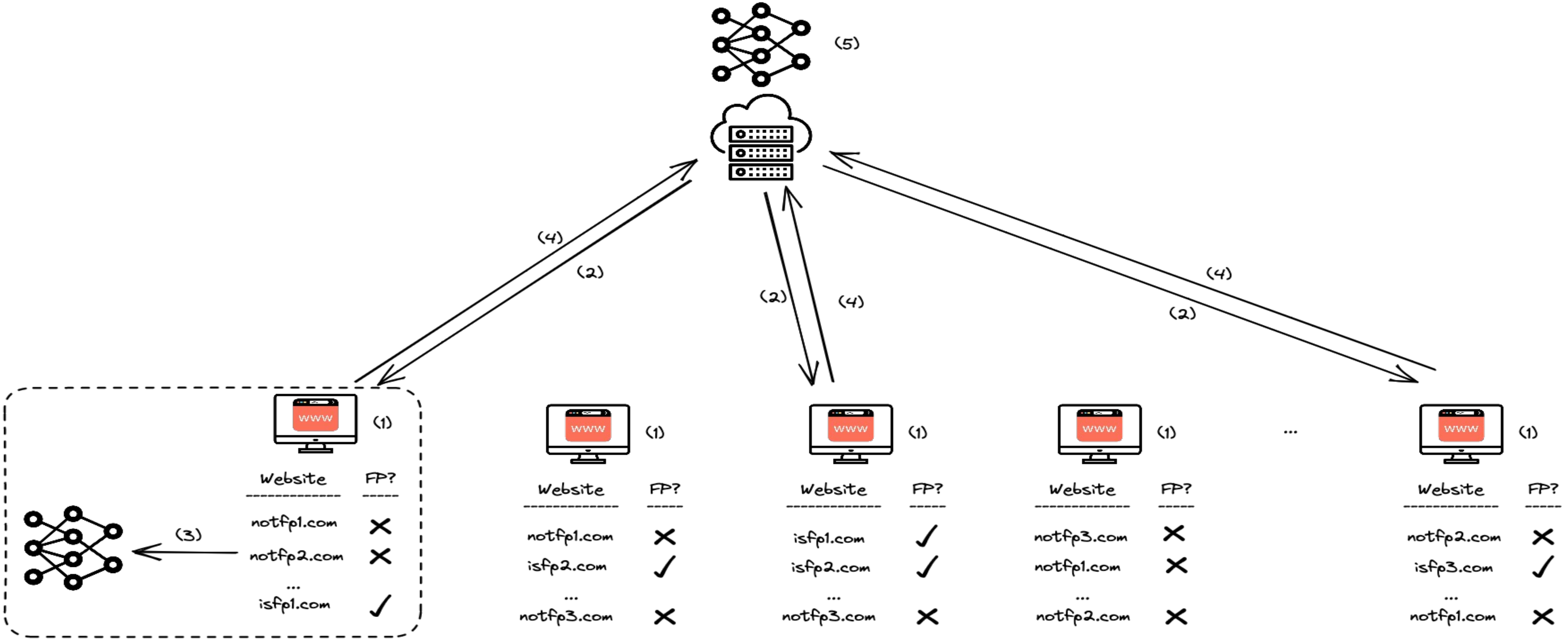
Definition: An Algorithm M is (ε,δ)-differentially private if for all pairs of neighboring datasets (D,D’), and for all outputs x:

$$\Pr[M(D)=x] \leq \exp(\varepsilon) * \Pr[M(D')=x] + \delta$$

quantifies information leakage

allows for a small probability of failure

FP-Fed



Website	FP?
notfp1.com	✗
notfp2.com	✗
...	
isfp1.com	✓

Website	FP?
notfp1.com	✗
isfp2.com	✓
...	
notfp3.com	✗

Website	FP?
isfp1.com	✓
isfp2.com	✓
...	
notfp3.com	✗

Website	FP?
notfp3.com	✗
notfp1.com	✗
...	
notfp2.com	✗

Website	FP?
notfp2.com	✗
isfp3.com	✓
...	
notfp1.com	✗

Project List

1. Browser fingerprinting evolution through Internet archive
2. **Auditing FP-Fed**
3. **Improving FP-Fed**
4. Federated Learning for Hate Speech
5. Looking at r/Privacy for privacy advice
6. Tor bridge on Raspberry PI
7. Python “Private Set Intersection” Toolkit
8. Does Alexa listen to me?
9. How to set up a privacy clinic
10. Update the petlib library

Auditing FP-Fed

- Differential Privacy provides a theoretical privacy guarantee
- Real-world attacks provide empirical privacy metrics
- How close are they to each other?

Improving FP-Fed

- Use different classifiers
- Use different features

Project List

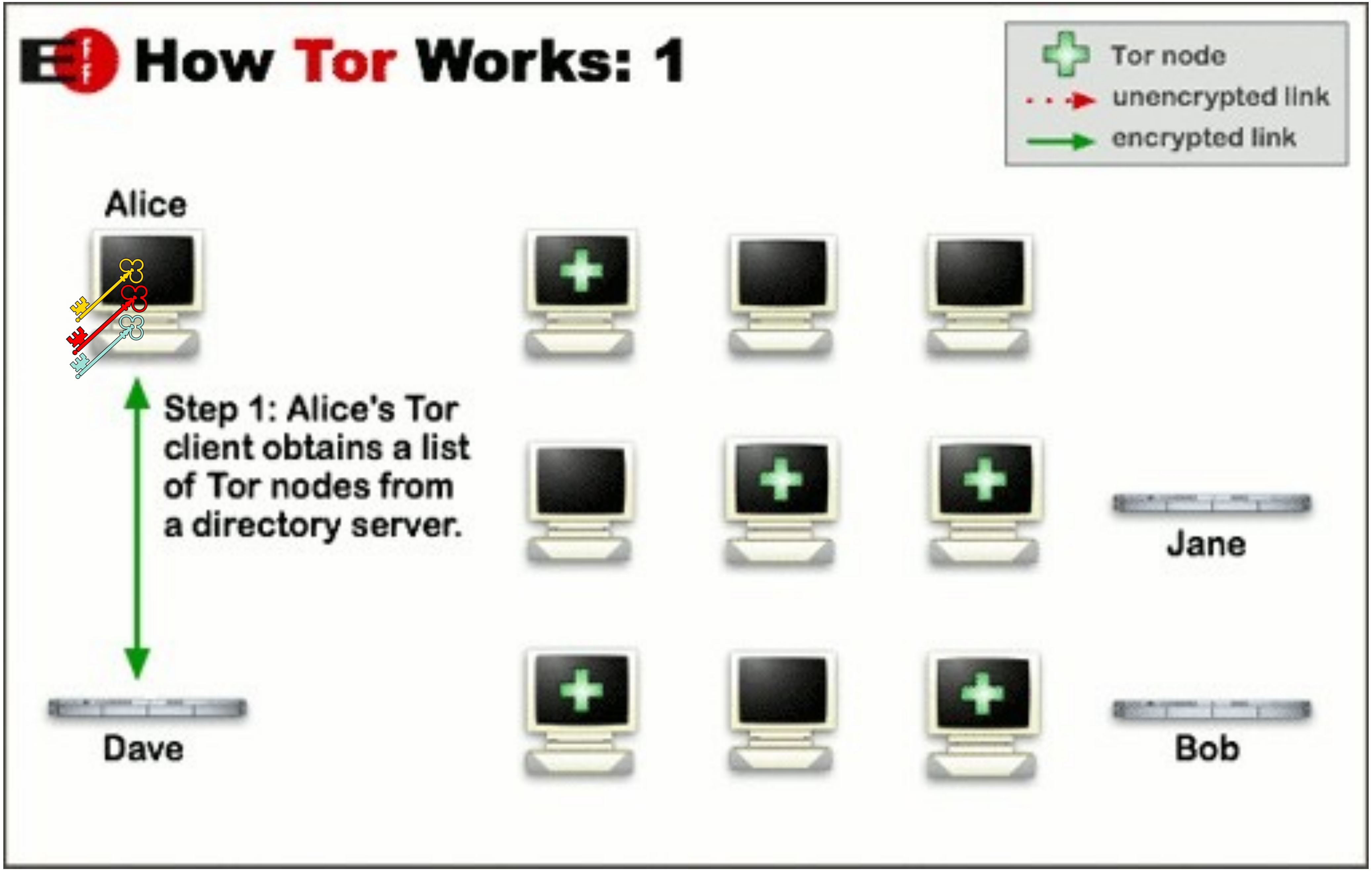
1. Browser fingerprinting evolution through Internet archive
2. Auditing FP-Fed
3. Improving FP-Fed
4. **Federated Learning for Hate Speech**
5. Looking at r/Privacy for privacy advice
6. Tor bridge on Raspberry PI
7. Python “Private Set Intersection” Toolkit
8. Does Alexa listen to me?
9. How to set up a privacy clinic
10. Update the petlib library

Project List

1. Browser fingerprinting evolution through Internet archive
2. Auditing FP-Fed
3. Improving FP-Fed
4. Federated Learning for Hate Speech
5. **Looking at r/Privacy for privacy advice**
6. Tor bridge on Raspberry PI
7. Python “Private Set Intersection” Toolkit
8. Does Alexa listen to me?
9. How to set up a privacy clinic
10. Update the petlib library

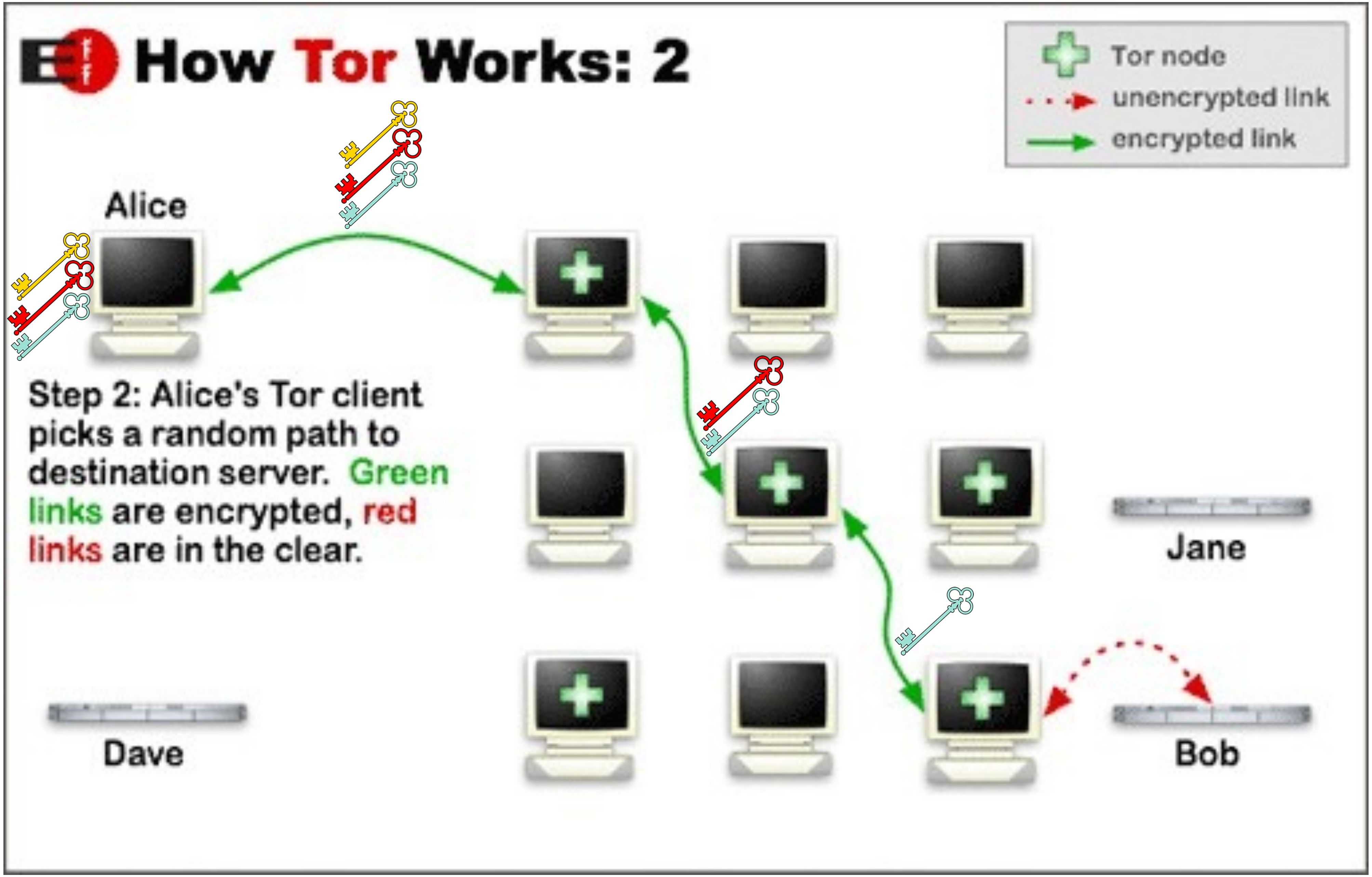
Project List

1. Browser fingerprinting evolution through Internet archive
2. Auditing FP-Fed
3. Improving FP-Fed
4. Federated Learning for Hate Speech
5. Looking at r/Privacy for privacy advice
6. **Tor bridge on Raspberry PI**
7. Python “Private Set Intersection” Toolkit
8. Does Alexa listen to me?
9. How to set up a privacy clinic
10. Update the petlib library



How Tor Works: 1

Step 1: Alice's Tor client obtains a list of Tor nodes from a directory server.



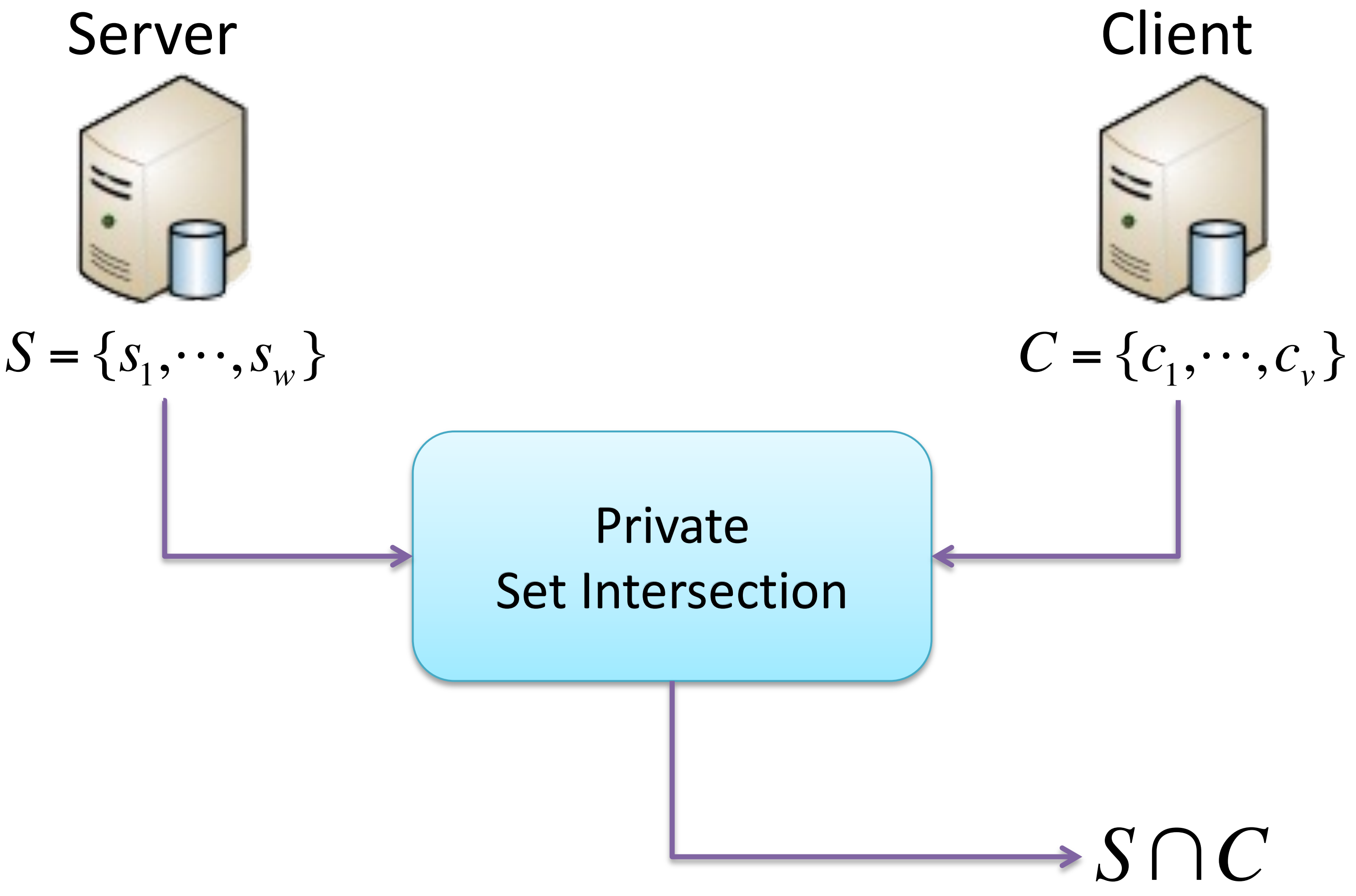
How Tor Works: 2

Step 2: Alice's Tor client picks a random path to destination server. Green links are encrypted, red links are in the clear.

Project List

1. Browser fingerprinting evolution through Internet archive
2. Auditing FP-Fed
3. Improving FP-Fed
4. Federated Learning for Hate Speech
5. Looking at r/Privacy for privacy advice
6. Tor bridge on Raspberry PI
7. **Python “Private Set Intersection” Toolkit**
8. Does Alexa listen to me?
9. How to set up a privacy clinic
10. Update the petlib library

Private Set Intersection (PSI)



Server

Client

$S = \{s_1, \cdots, s_w\}$

$C = \{c_1, \cdots, c_v\}$

Private
Set Intersection

$S \cap C$

Private Set Intersection (2)

- **Alice** (Facebook Friend List) and **Bob** (Facebook Friend List)
  - Find out the list of common friends

- **DHS** (Terrorist Watch List) and **Airline** (Passenger List)
  - Find out whether any suspect is on a given flight

- **IRS** (Tax Evaders) and **Swiss Bank** (Customers)
  - Discover if tax evaders have accounts at foreign banks

- **Hoag Hospital** (Patients) and **SSA** (Social Security DB)
  - Patients with fake Social Security Number

Project List

1. Browser fingerprinting evolution through Internet archive
2. Auditing FP-Fed
3. Improving FP-Fed
4. Federated Learning for Hate Speech
5. Looking at r/Privacy for privacy advice
6. Tor bridge on Raspberry PI
7. Python “Private Set Intersection” Toolkit
8. **Does Alexa listen to me?**
9. How to set up a privacy clinic
10. Update the petlib library

Project List

1. Browser fingerprinting evolution through Internet archive
2. Auditing FP-Fed
3. Improving FP-Fed
4. Federated Learning for Hate Speech
5. Looking at r/Privacy for privacy advice
6. Tor bridge on Raspberry PI
7. Python “Private Set Intersection” Toolkit
8. Does Alexa listen to me?
9. **How to set up a privacy clinic**
10. Update the petlib library

Project List

1. Browser fingerprinting evolution through Internet archive
2. Auditing FP-Fed
3. Improving FP-Fed
4. Federated Learning for Hate Speech
5. Looking at r/Privacy for privacy advice
6. Tor bridge on Raspberry PI
7. Python “Private Set Intersection” Toolkit
8. Does Alexa listen to me?
9. How to set up a privacy clinic
10. **Update the petlib library**

petlib

https://github.com/gdanezis/petlib