Systems for Differential Privacy

Research systems

- RAPPOR
- PINQ / wPINQ
- GUPT
- Airavat
- Chorus
- Fuzz / DFuzz / ADAFuzz
- Ektelo

Local DP

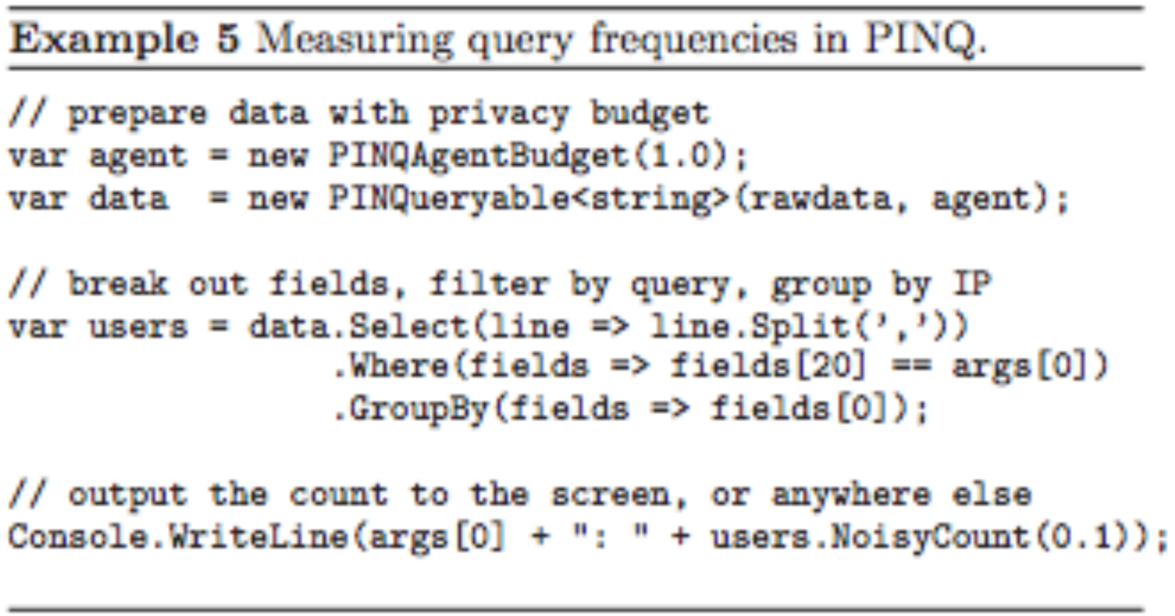
Deployed systems

- RAPPOR
- Apple
- Uber
- U.S. Census

Local DP

Privacy Integrated Queries (PINQ)

- Early system for enforcing differential privacy
- Microsoft LINQ queries
- Open-source implementation



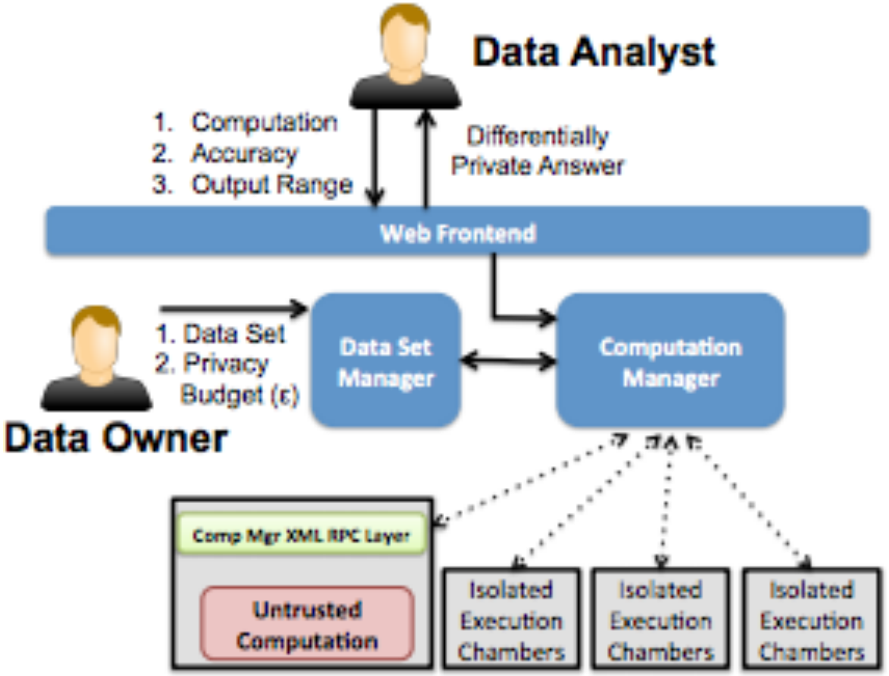
Example 5 Measuring query frequencies in PINQ.

```
// prepare data with privacy budget
var agent = new PINQAgentBudget(1.0);
var data  = new PINQueryable<string>(rawdata, agent);

// break out fields, filter by query, group by IP
var users = data.Select(line => line.Split(','))
                .Where(fields => fields[20] == args[0])
                .GroupBy(fields => fields[0]);

// output the count to the screen, or anywhere else
Console.WriteLine(args[0] + ": " + users.NoisyCount(0.1));
```

https://www.microsoft.com/en-us/research/project/privacy-integrated-queries-pinq/

GUPT

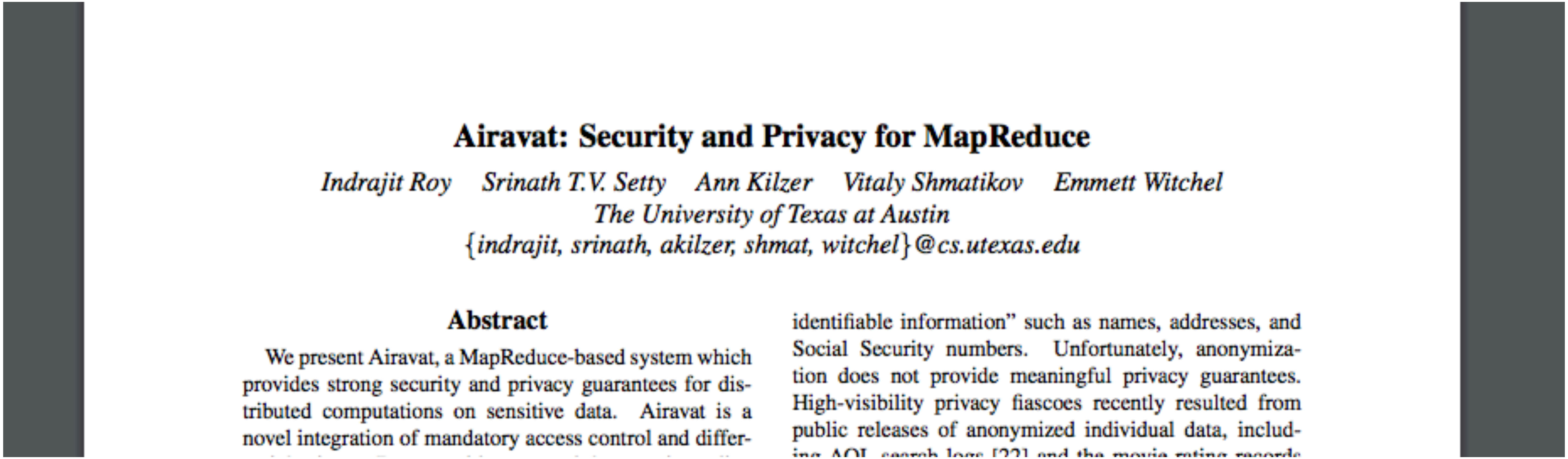
- Implementation of sample & aggregate
- Works for any query
  - Query written as a Python program
  - Treated as "untrusted computation"



https://github.com/prashmohan/GUPT

Airavat

- Implementation of sample & aggregate for MapReduce



Airavat: Security and Privacy for MapReduce

Indrajit Roy   Srinath T.V. Setty   Ann Kilzer   Vitaly Shmatikov   Emmett Witchel
The University of Texas at Austin
{indrajit, srinath, akilzer, shmat, witchel}@cs.utexas.edu

Abstract

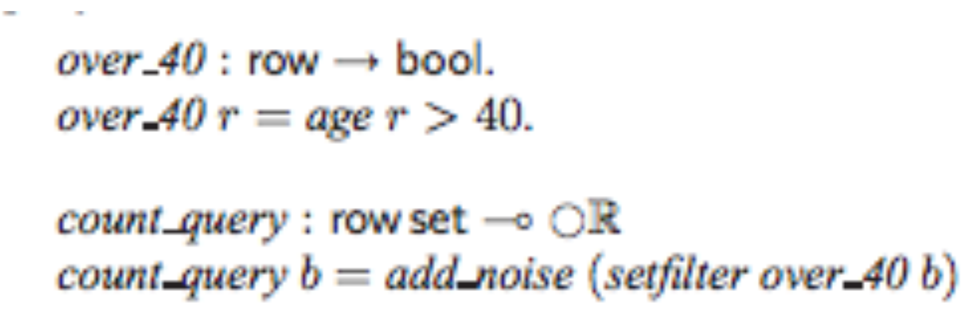
We present Airavat, a MapReduce-based system which provides strong security and privacy guarantees for distributed computations on sensitive data. Airavat is a novel integration of mandatory access control and differ-

identifiable information" such as names, addresses, and Social Security numbers. Unfortunately, anonymization does not provide meaningful privacy guarantees. High-visibility privacy fiascoes recently resulted from public releases of anonymized individual data, including AOL search logs [22] and the movie rating records

https://z.cs.utexas.edu/users/osa/airavat/

Fuzz, DFuzz, ADAFuzz

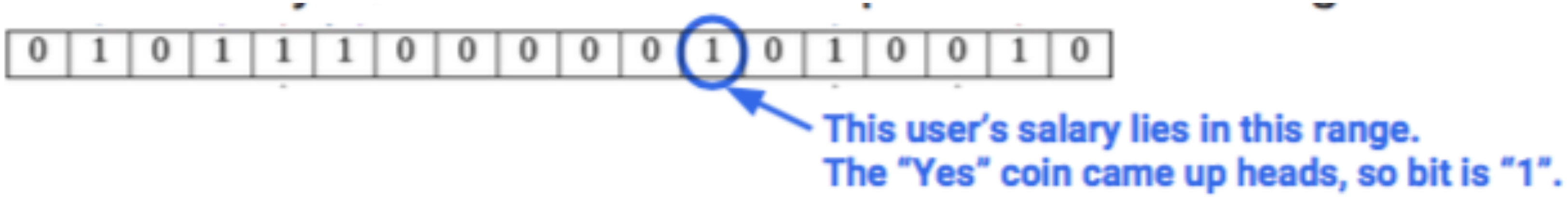
- Type system-based approach
- Enforce differential privacy for functional programs
- Fuzz: fixed parameters
  - http://www.cis.upenn.edu/~bcpierce/papers/dp.pdf
- DFuzz: dependent types for variable parameters
  - http://www.cis.upenn.edu/~ahae/papers/dfuzz-popl2013.pdf
- ADAFuzz: run-time enforcement for (ε, δ)-differential privacy
  - http://www.cis.upenn.edu/~ahae/papers/adafuzz-long.pdf



```
over_40 : row → bool.
over_40 r = age r > 40.

count_query : row set ⊸ ○ℝ
count_query b = add_noise (setfilter over_40 b)
```

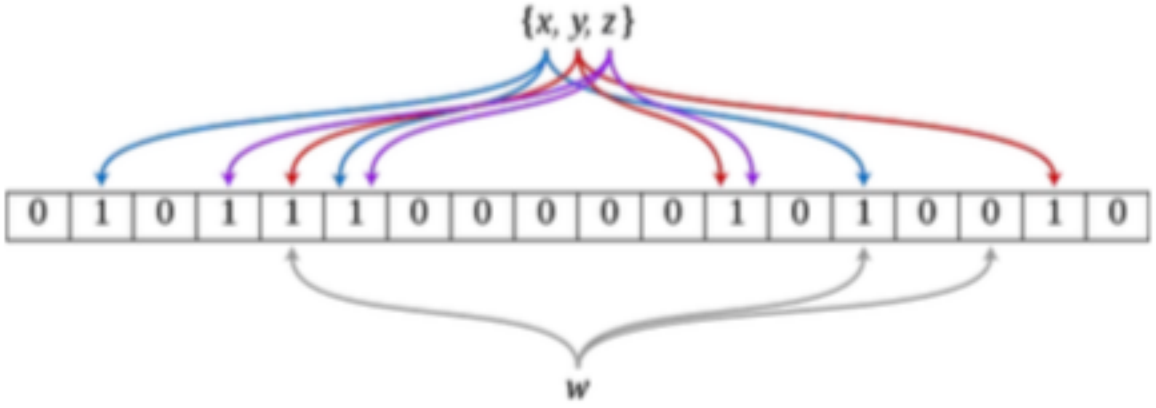
RAPPOR

- Deployed in Chrome & Chromium to record home pages
- Based on randomized response for bit-vectors



This user's salary lies in this range.
The "Yes" coin came up heads, so bit is "1".

- Bloom filters used to reduce dimensionality



RAPPOR Algorithm

1. Hash a value *v* into Bloom filter *B* using *h* hash functions
2. Memoize a **Permanent Randomized Response** *B′*

$$B'_i = \begin{cases} 1, & \text{with probability } \frac{1}{2}f \\ 0, & \text{with probability } \frac{1}{2}f \\ B_i, & \text{with probability } 1 - f \end{cases}$$

$f = ½$
for example

3. Report an **Instantaneous Randomized Response** *S*

$$P(S_i = 1) = \begin{cases} q, & \text{if } B'_i = 1. \\ p, & \text{if } B'_i = 0. \end{cases}$$

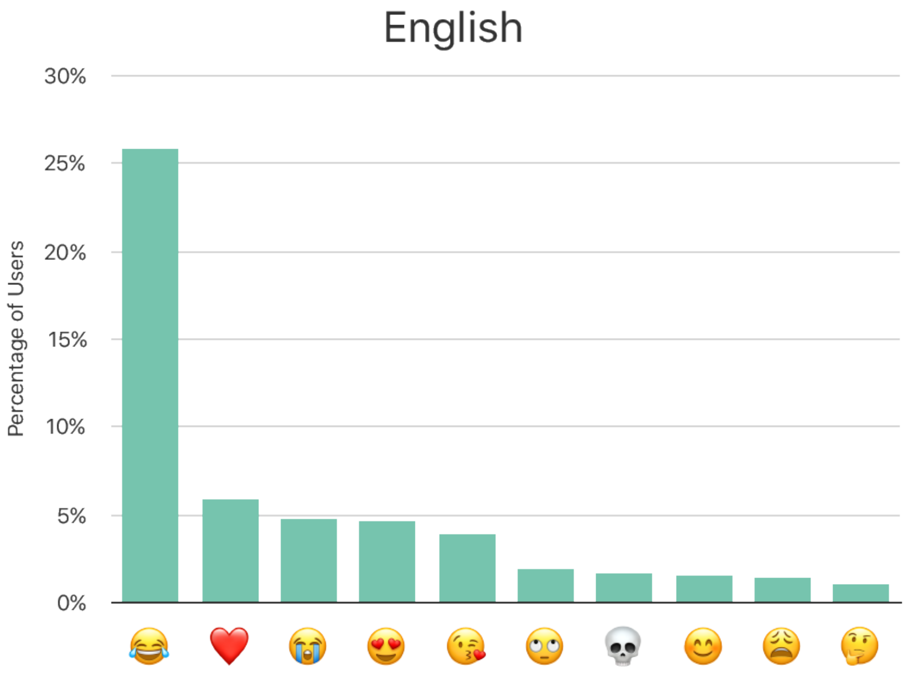
$q = ¾$ and $p = ½$
for example

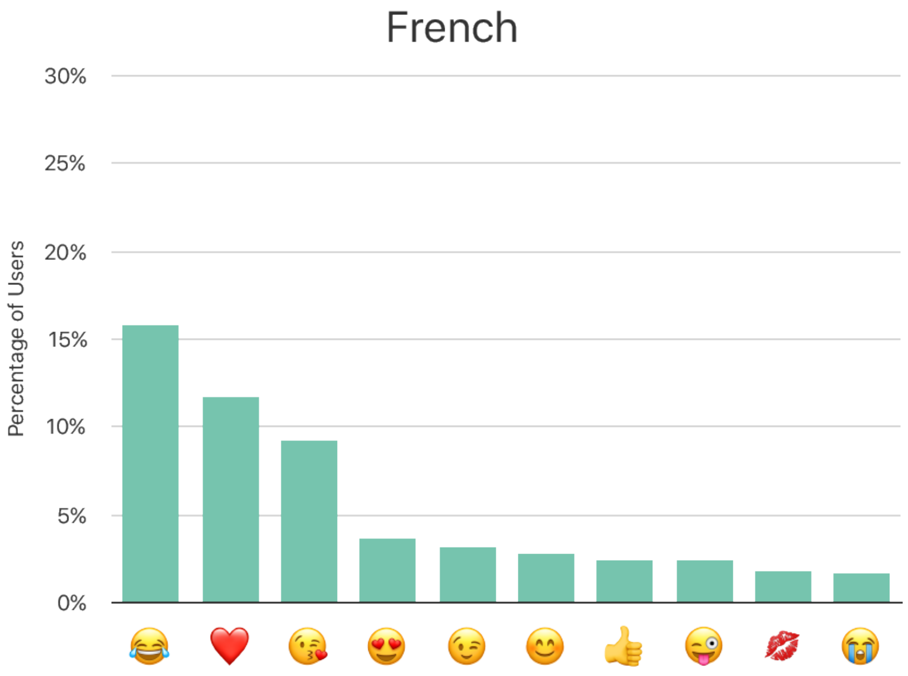
RAPPOR References

- Research implementation: https://github.com/google/rappor
- Implementation in Chrome: http://www.chromium.org/developers/design-documents/rappor
- Paper: https://arxiv.org/abs/1407.6981

Apple’s Differential Privacy System

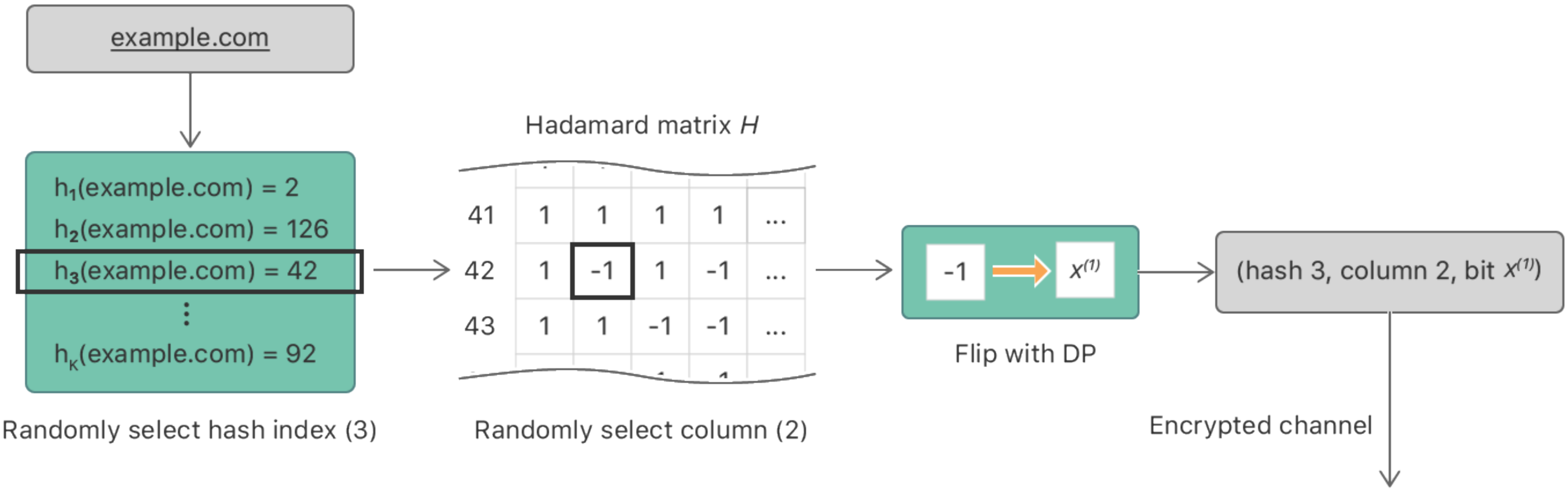
- Basic idea: same as RAPPOR
- In the beginning: recorded frequency of emoji use





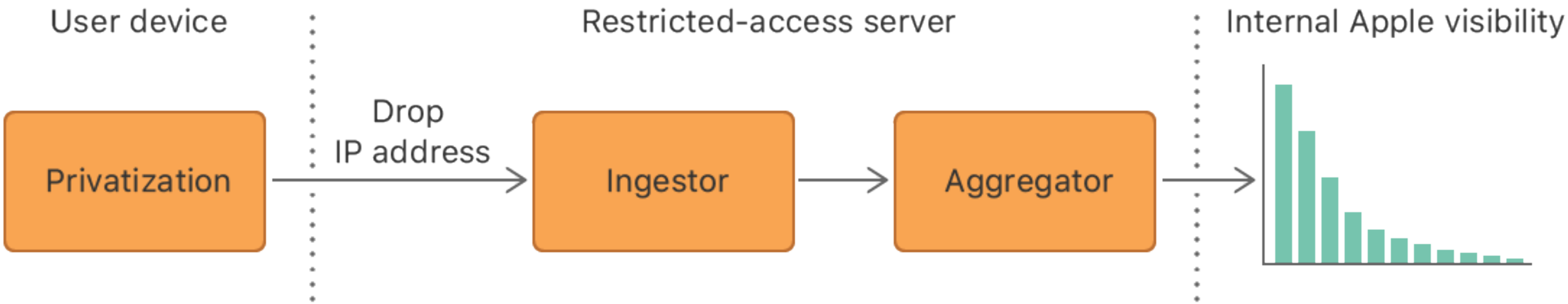
Private Hadamard Count-Mean Sketch

Client-side algorithm



41	1	1	1	1	...
42	1	-1	1	-1	...
43	1	1	-1	-1	...

Emphasis on Privacy for the User



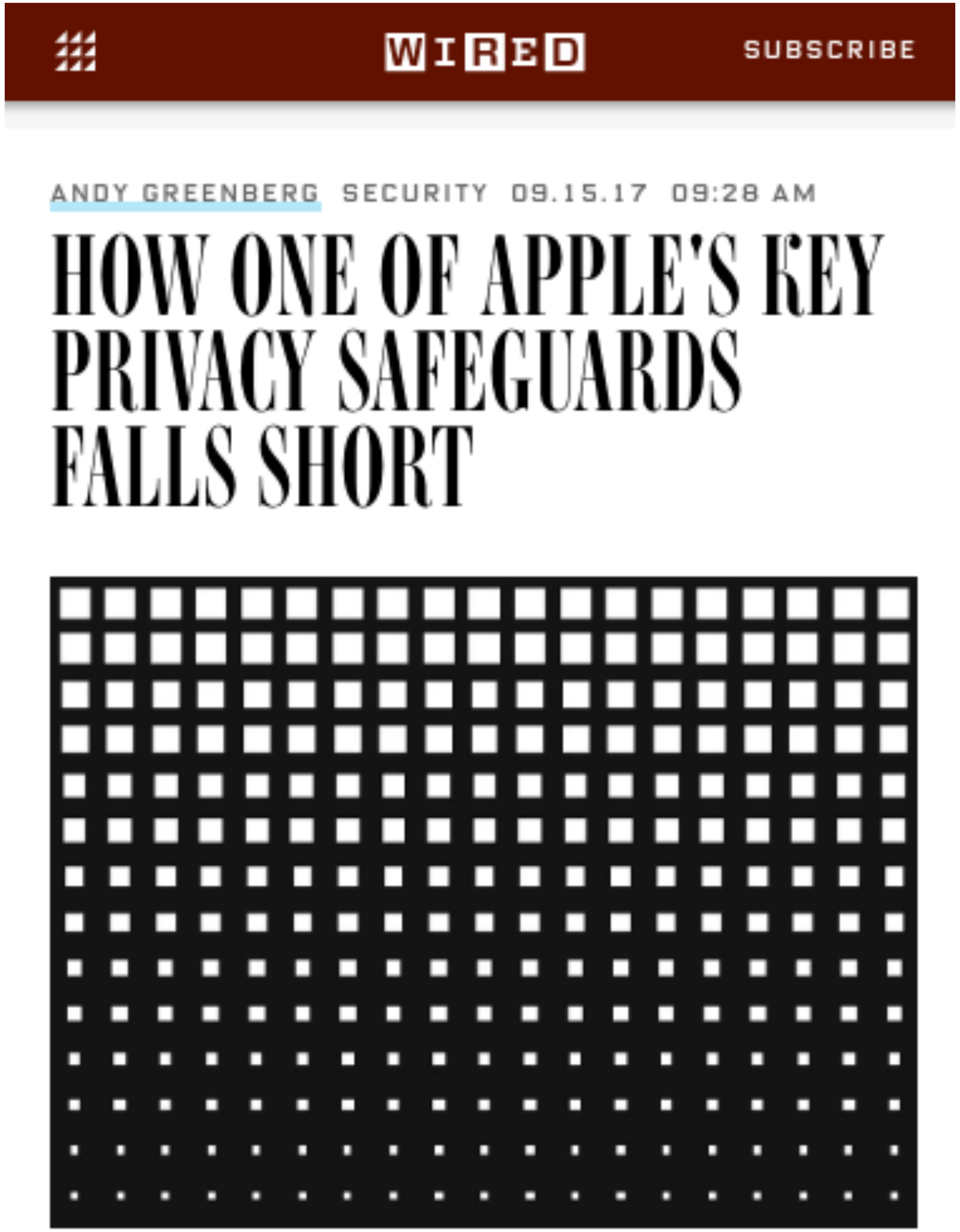
What about Epsilon?

"Apple has put some kind of handcuffs on in how they interact with your data…it just turns out those handcuffs are made out of tissue paper"

- *Frank McSherry*

- Epsilon = 14 for some data sources
- Epsilon (maybe) resets every day?
- But: no attack proposed!
- Open question: are epsilons comparable?

https://www.wired.com/story/apple-differential-privacy-shortcomings/



ANDY GREENBERG SECURITY 09.15.17 09:28 AM

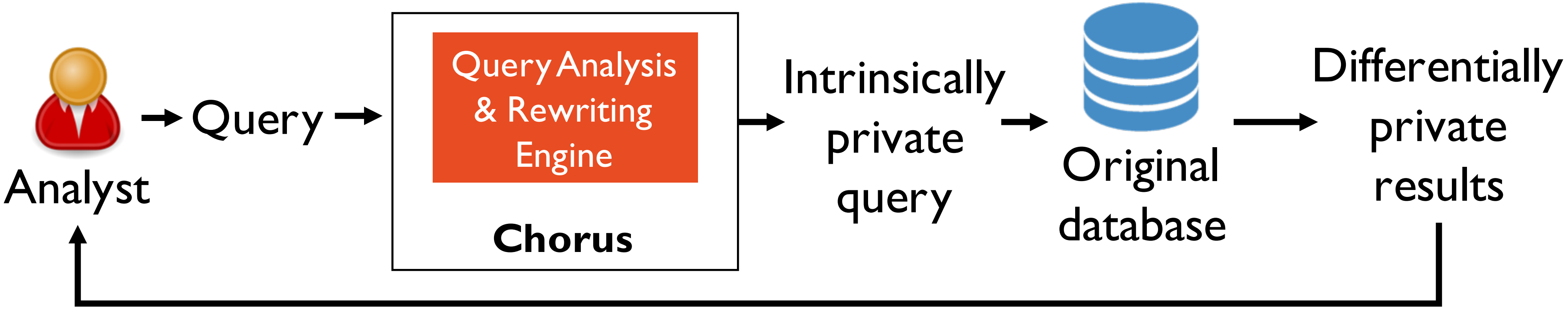
HOW ONE OF APPLE'S KEY PRIVACY SAFEGUARDS FALLS SHORT

References

- Apple announcement: https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf
- More detailed description: https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html
- WIRED Article: https://www.wired.com/story/apple-differential-privacy-shortcomings/
- Budgeting paper: https://arxiv.org/pdf/1709.02753.pdf

Chorus

- Differential privacy for "real SQL queries"
- Local sensitivity-based approach for queries with joins
- Works with any SQL database by rewriting the query


Analyst
Query
Query Analysis & Rewriting Engine
Chorus
Intrinsically private query
Original database
Differentially private results

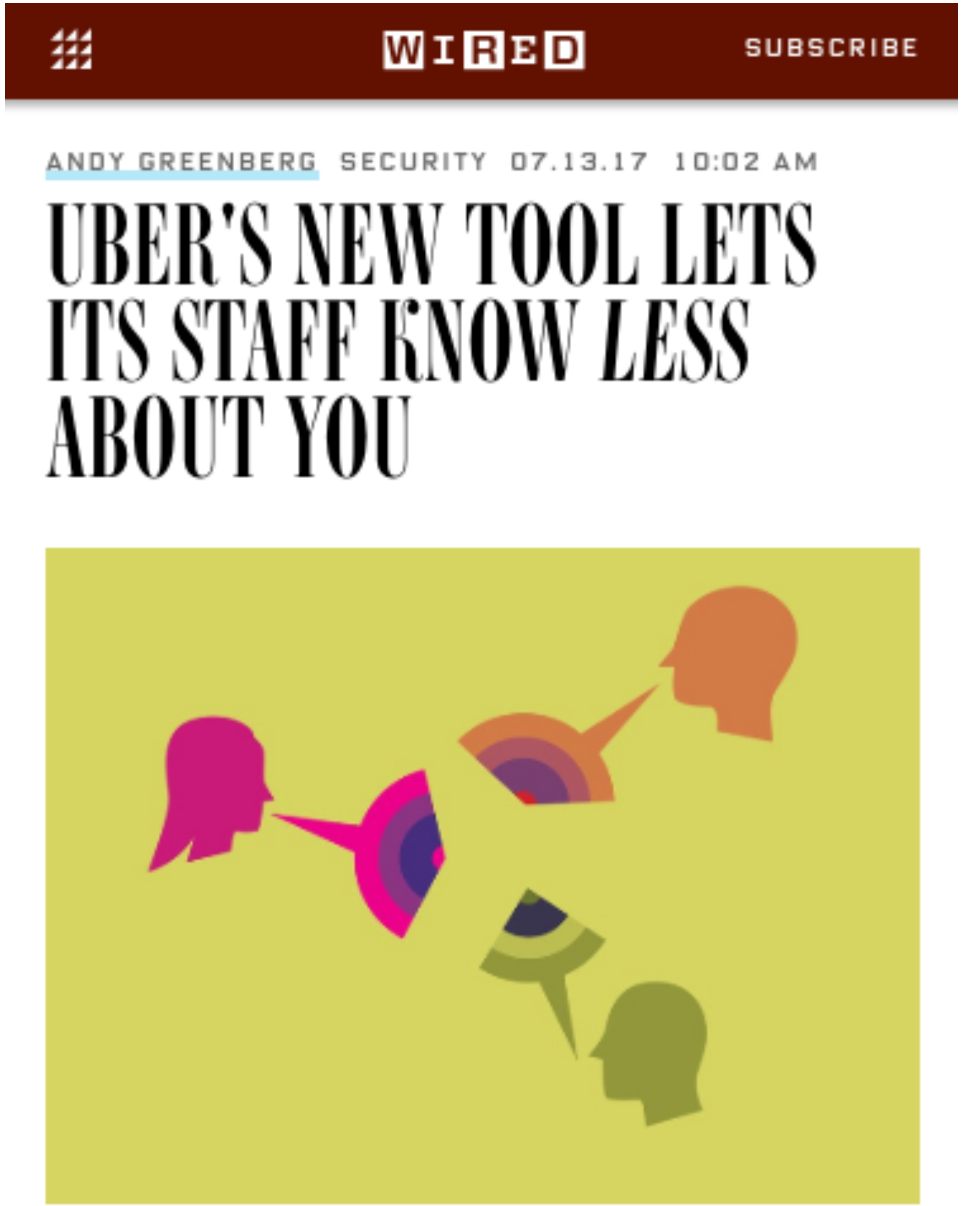
https://github.com/uber/sql-differential-privacy

https://arxiv.org/abs/1706.09479 https://arxiv.org/abs/1809.07750

Deployment of Chorus

- Enforces differential privacy
- Satisfies GDPR requirements
- Processes majority of internal queries
- But: DP not yet exposed to analysts (i.e. deployment is *not finished*)

https://www.wired.com/story/uber-privacy-elastic-sensitivity/

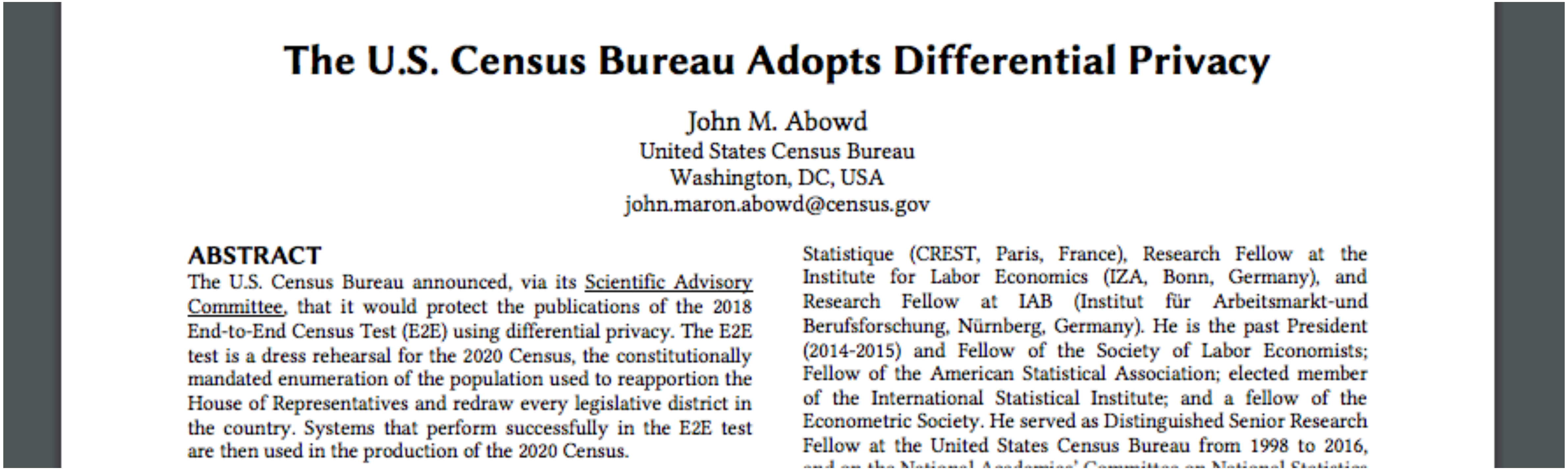


WIRED SUBSCRIBE

ANDY GREENBERG SECURITY 07.13.17 10:02 AM

UBER'S NEW TOOL LETS ITS STAFF KNOW *LESS* ABOUT YOU

Differential Privacy at U.S. Census



The U.S. Census Bureau Adopts Differential Privacy

John M. Abowd
United States Census Bureau
Washington, DC, USA
john.maron.abowd@census.gov

ABSTRACT

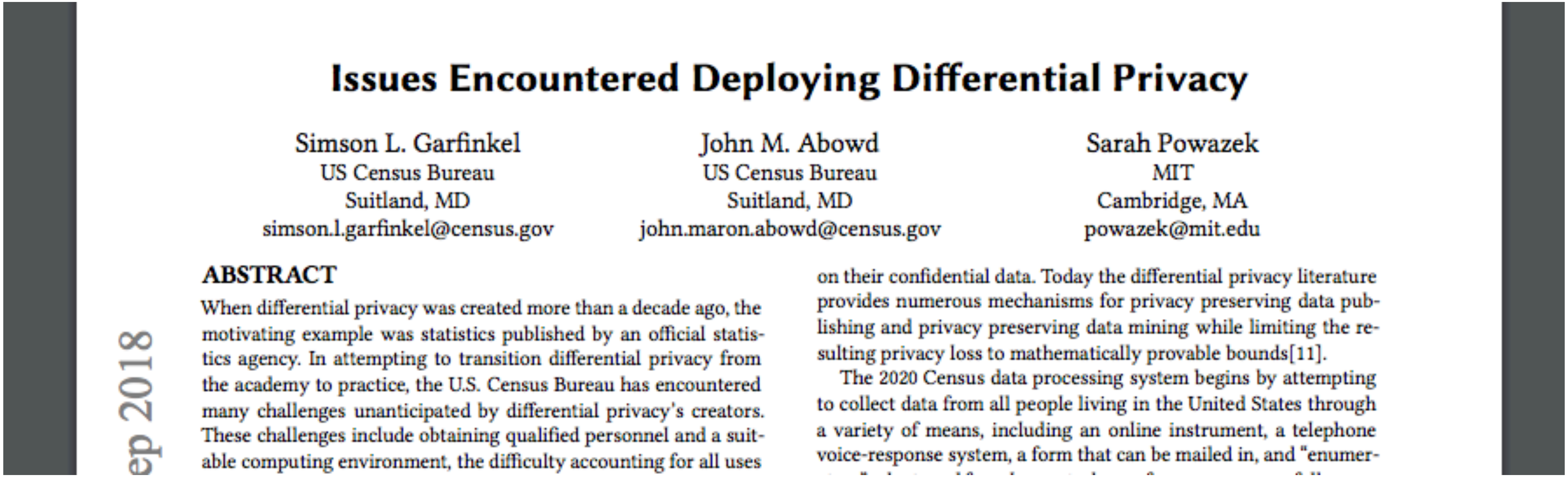
The U.S. Census Bureau announced, via its Scientific Advisory Committee, that it would protect the publications of the 2018 End-to-End Census Test (E2E) using differential privacy. The E2E test is a dress rehearsal for the 2020 Census, the constitutionally mandated enumeration of the population used to reapportion the House of Representatives and redraw every legislative district in the country. Systems that perform successfully in the E2E test are then used in the production of the 2020 Census.

Statistique (CREST, Paris, France), Research Fellow at the Institute for Labor Economics (IZA, Bonn, Germany), and Research Fellow at IAB (Institut für Arbeitsmarkt-und Berufsforschung, Nürnberg, Germany). He is the past President (2014-2015) and Fellow of the Society of Labor Economists; Fellow of the American Statistical Association; elected member of the International Statistical Institute; and a fellow of the Econometric Society. He served as Distinguished Senior Research Fellow at the United States Census Bureau from 1998 to 2016, and on the National Academies' Committee on National Statistics

https://dl.acm.org/citation.cfm?id=3226070

https://www.census.gov/newsroom/blogs/research-matters/2018/08/protecting_the_confi.html

Census: Huge Test for DP!



Issues Encountered Deploying Differential Privacy

Simson L. Garfinkel
US Census Bureau
Suitland, MD
simson.l.garfinkel@census.gov

John M. Abowd
US Census Bureau
Suitland, MD
john.maron.abowd@census.gov

Sarah Powazek
MIT
Cambridge, MA
powazek@mit.edu

ABSTRACT

When differential privacy was created more than a decade ago, the motivating example was statistics published by an official statistics agency. In attempting to transition differential privacy from the academy to practice, the U.S. Census Bureau has encountered many challenges unanticipated by differential privacy's creators. These challenges include obtaining qualified personnel and a suitable computing environment, the difficulty accounting for all uses

on their confidential data. Today the differential privacy literature provides numerous mechanisms for privacy preserving data publishing and privacy preserving data mining while limiting the resulting privacy loss to mathematically provable bounds[11].

The 2020 Census data processing system begins by attempting to collect data from all people living in the United States through a variety of means, including an online instrument, a telephone voice-response system, a form that can be mailed in, and "enumer-

Obtaining Qualified Personnel and Tools

Difficulties Arising from Increased Transparency

Misunderstandings about Randomness and Noise Infusion

https://arxiv.org/pdf/1809.02201.pdf

Ektelo: DP Programming Framework

- Focus on algorithms for query workloads
- Focus on pre- and post-processing
- Aided development of HDMM
  - New mechanism proposed for Census

Ektelo

Ektelo is a novel programming framework and system for implementing both existing and new privacy algorithms.

View the Project on GitHub
https://github.com/ektelo/ektelo

https://ektelo.github.io/

https://people.cs.umass.edu/~miklau/assets/pubs/dp/mckenna18hdmm.pdf