Private Data Aggregation on a Budget

Morten Dahl
Snips

Valerio Pastro
Yale University

Mathieu Poumeyrol
Snips

Theory and Practice of Multi-Party Computation '17

snips

Context

Snips

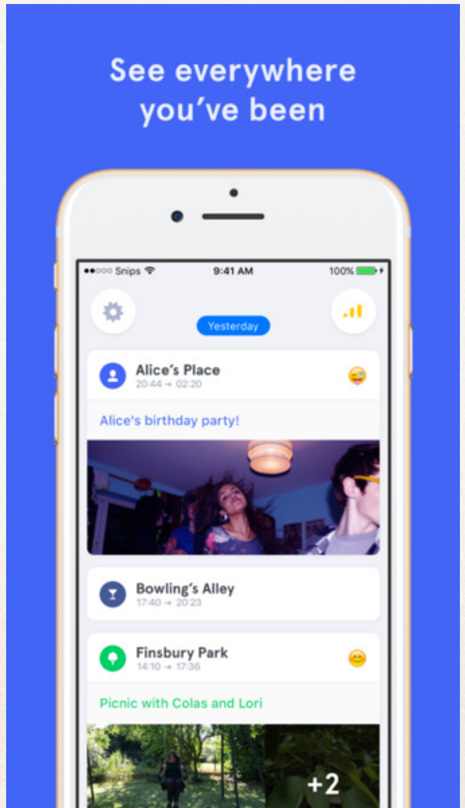
Machine learning SDK
for mobile devices and IoT

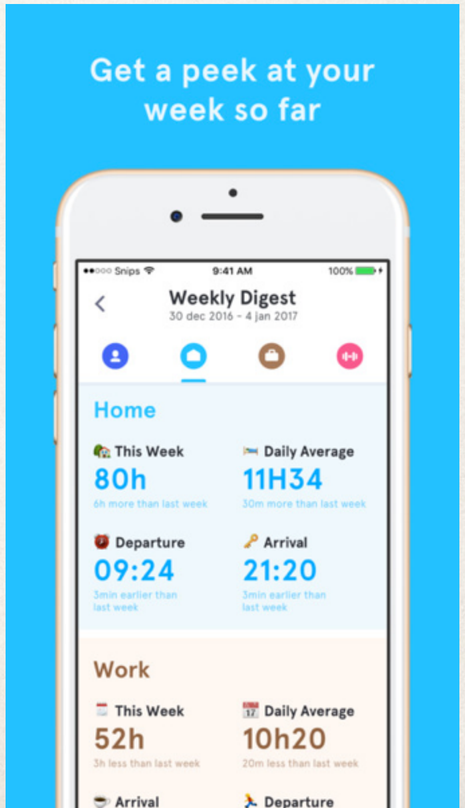
Private by Design

Algorithms run locally

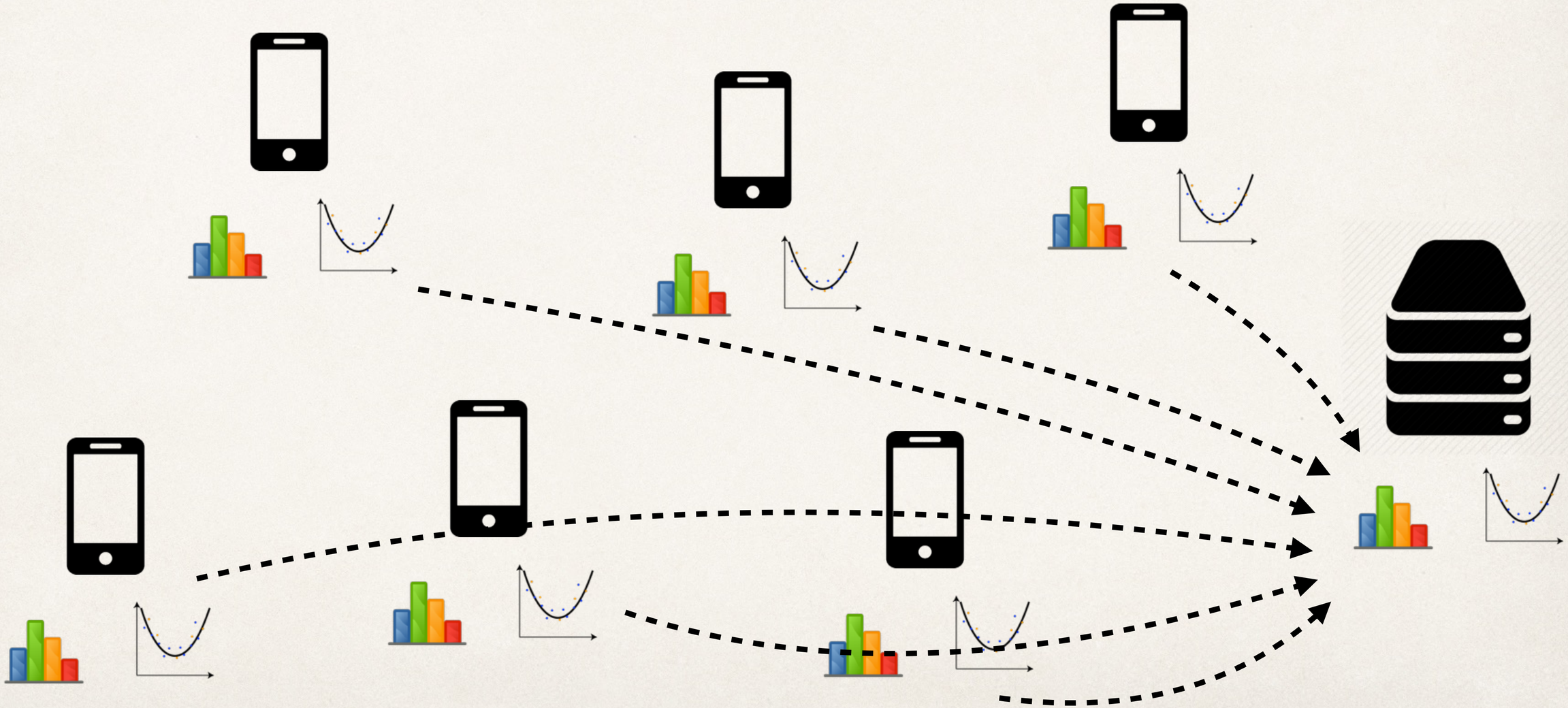
Rich local data sets

Context awareness etc.



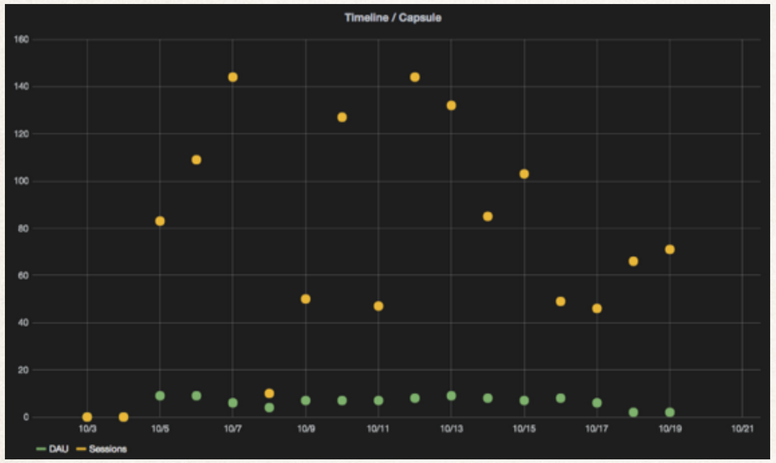


Learning from Distributed Data Sets

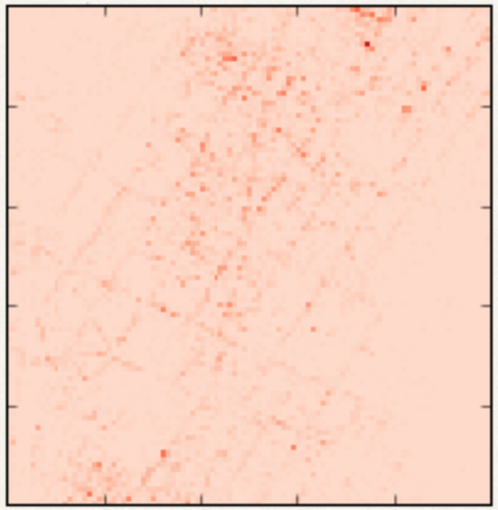


Aggregation of Distributed Data Sets

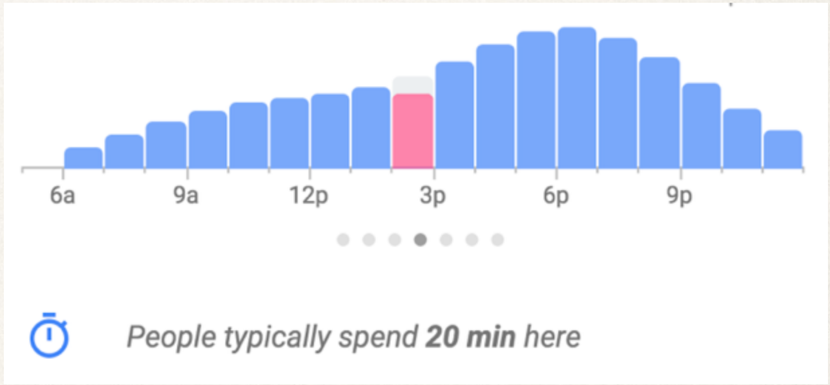
analytics



discovery



recommendations



$$x = \sum x_i$$

$$dim(x_i) \geq 10k$$

Constraints

no individual user data

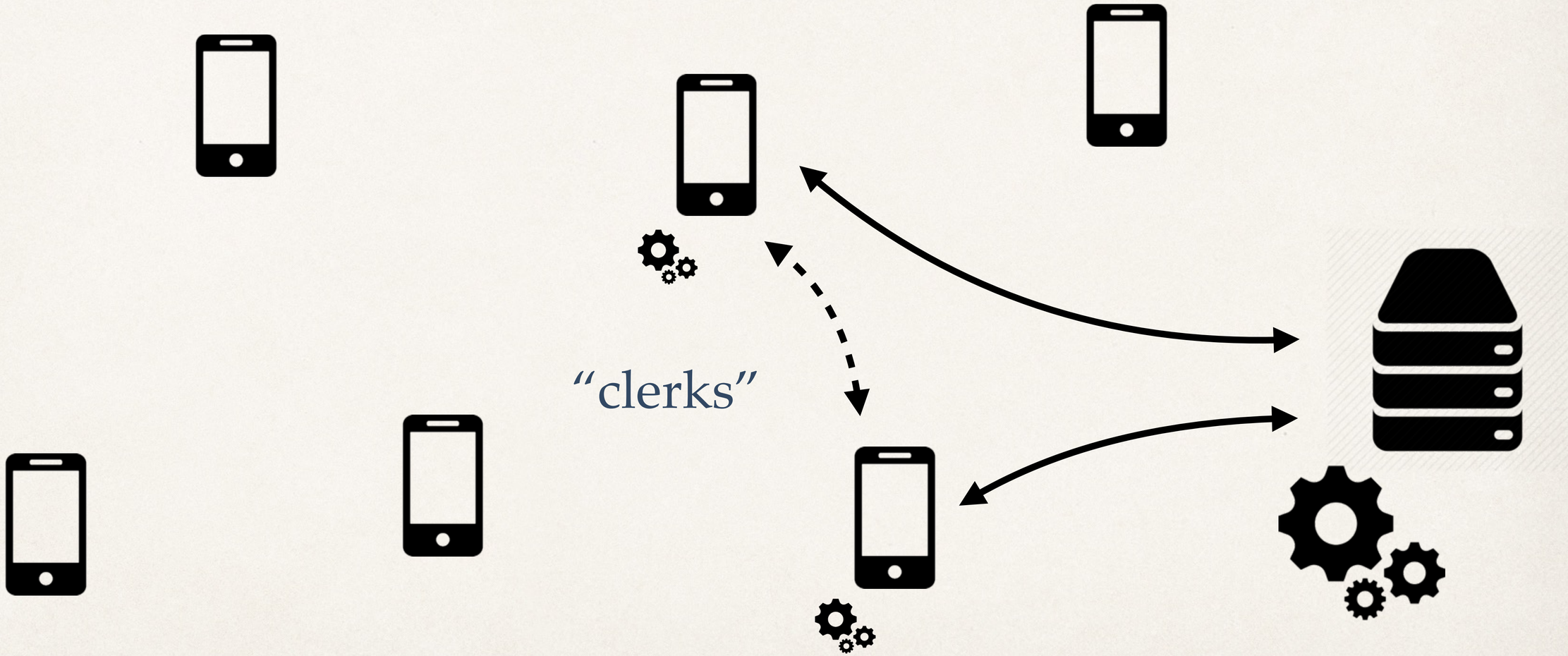
limited computation	limited connectivity	sporadic behaviour
minimise device processing	minimise session count + length	minimise coordination

Solutions

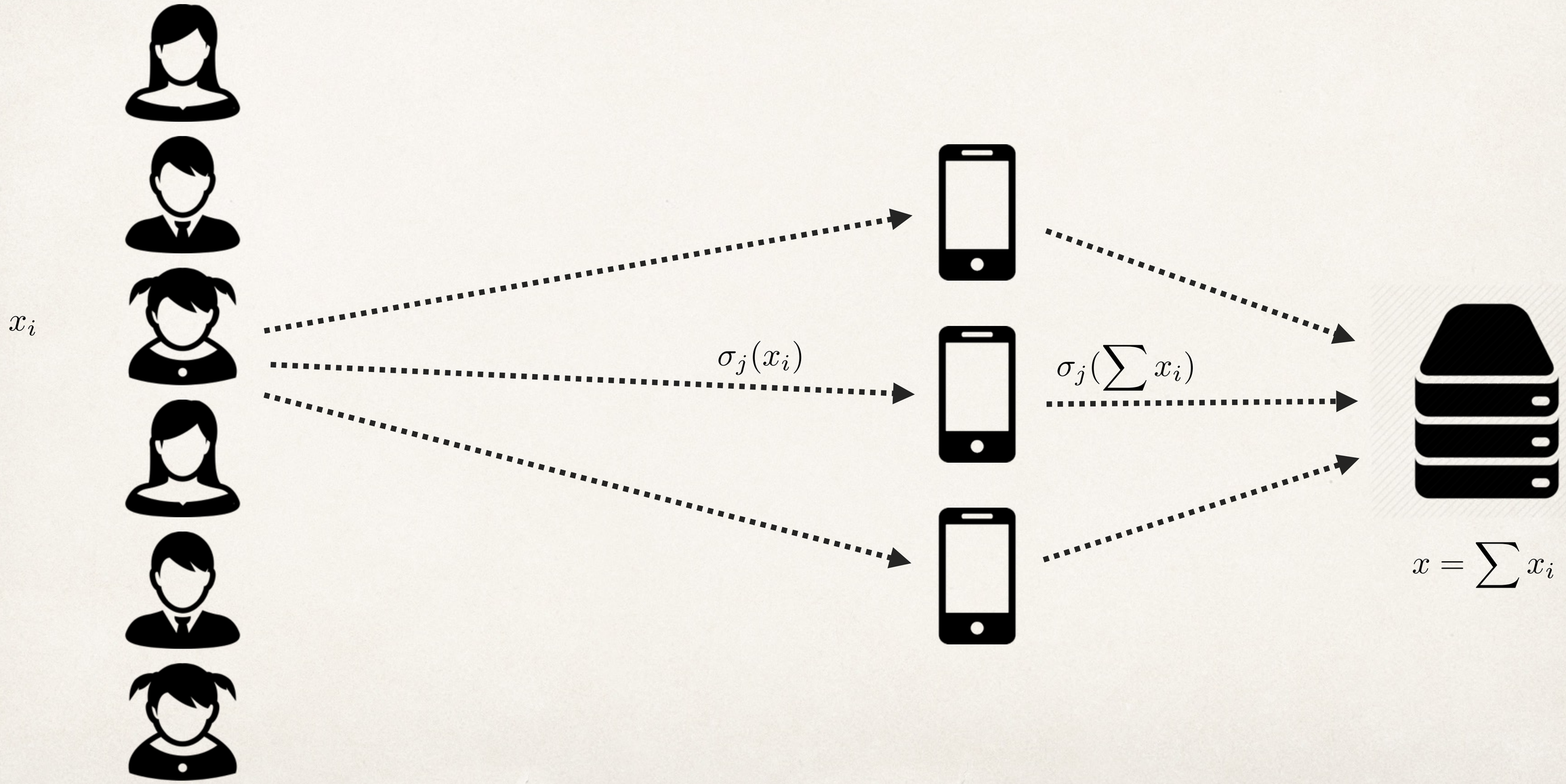
sensor networks	local DP	server-aided MPC
high performance	high performance	flexible
coordination	signal-to-noise	no one to play with

address problem of only one powerful server

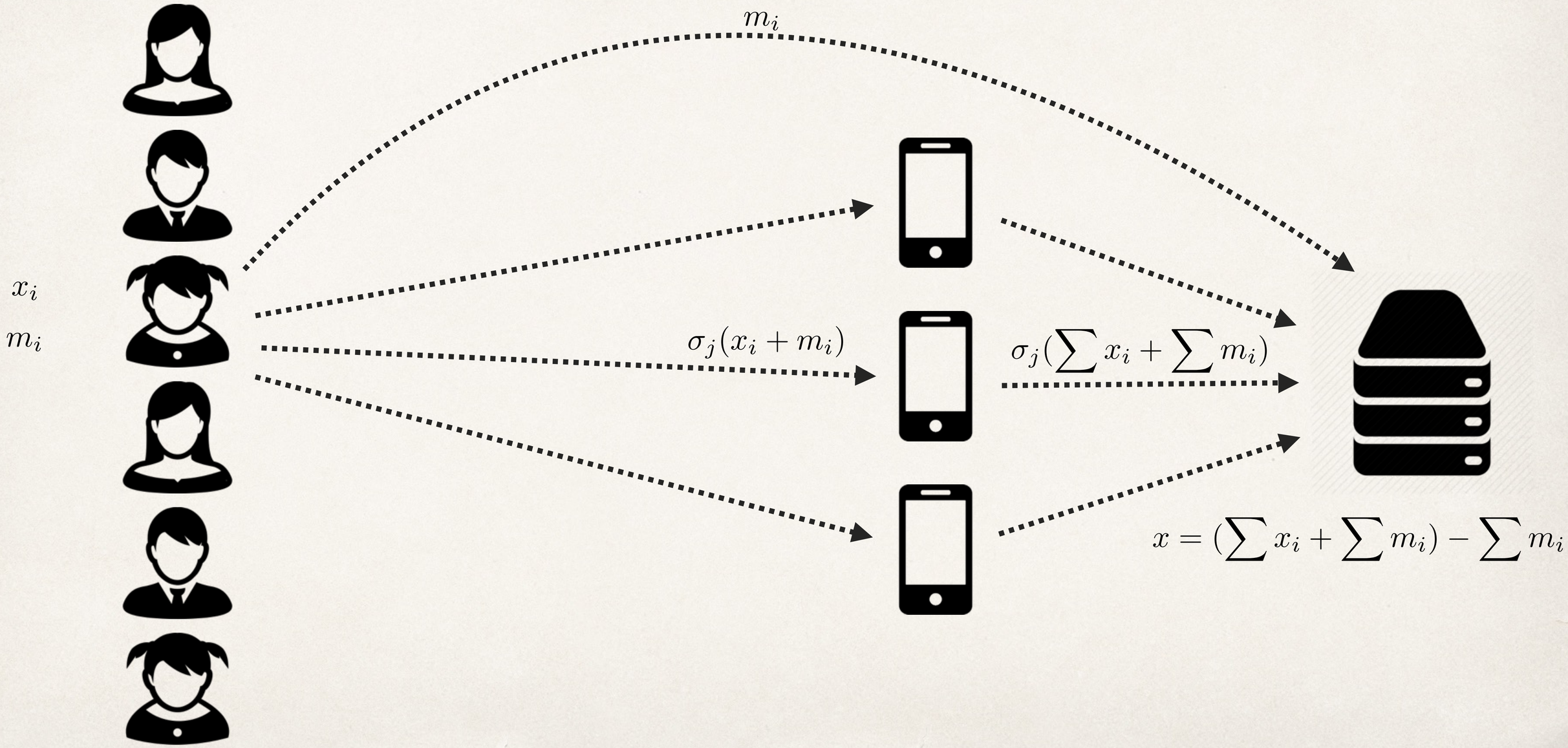
Community



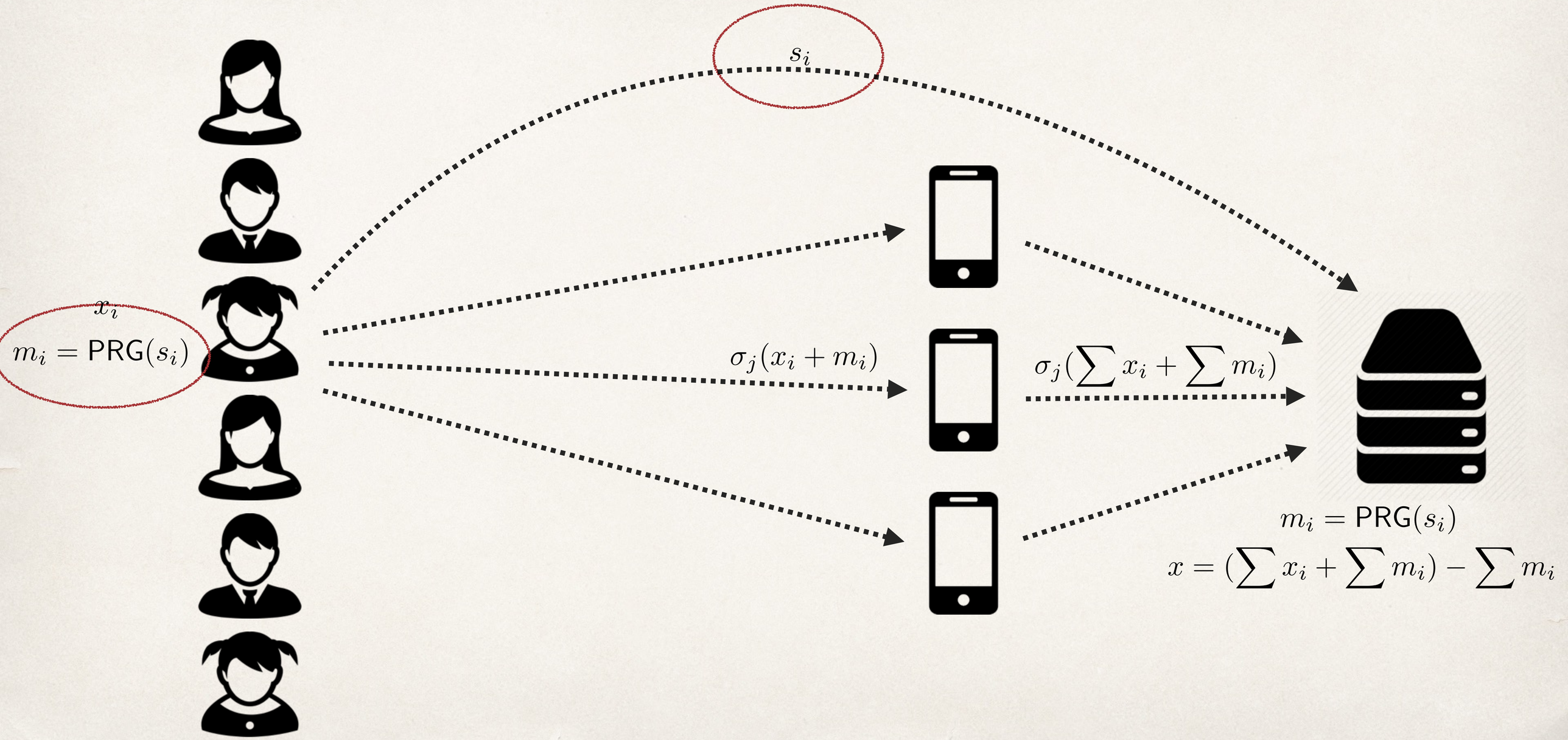
Protocol



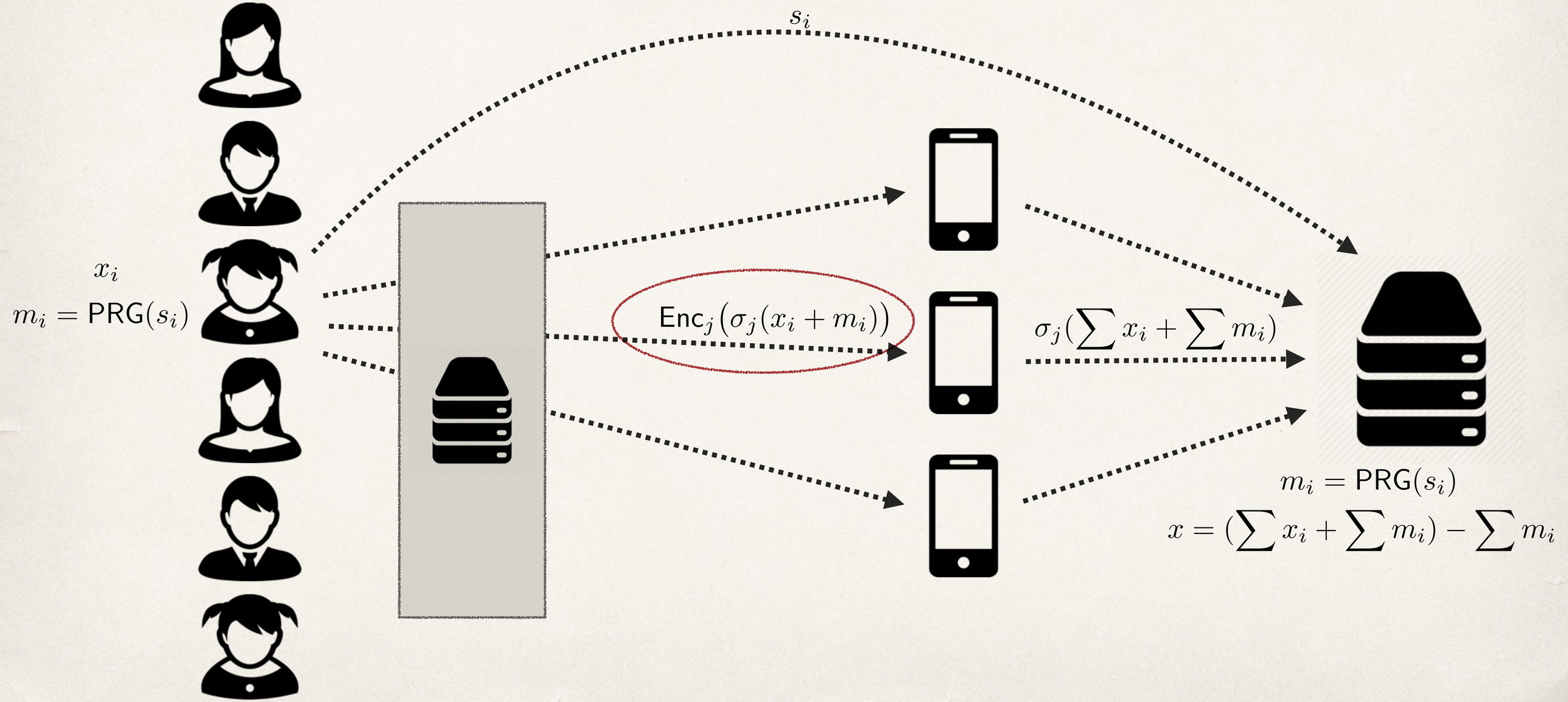
Protocol



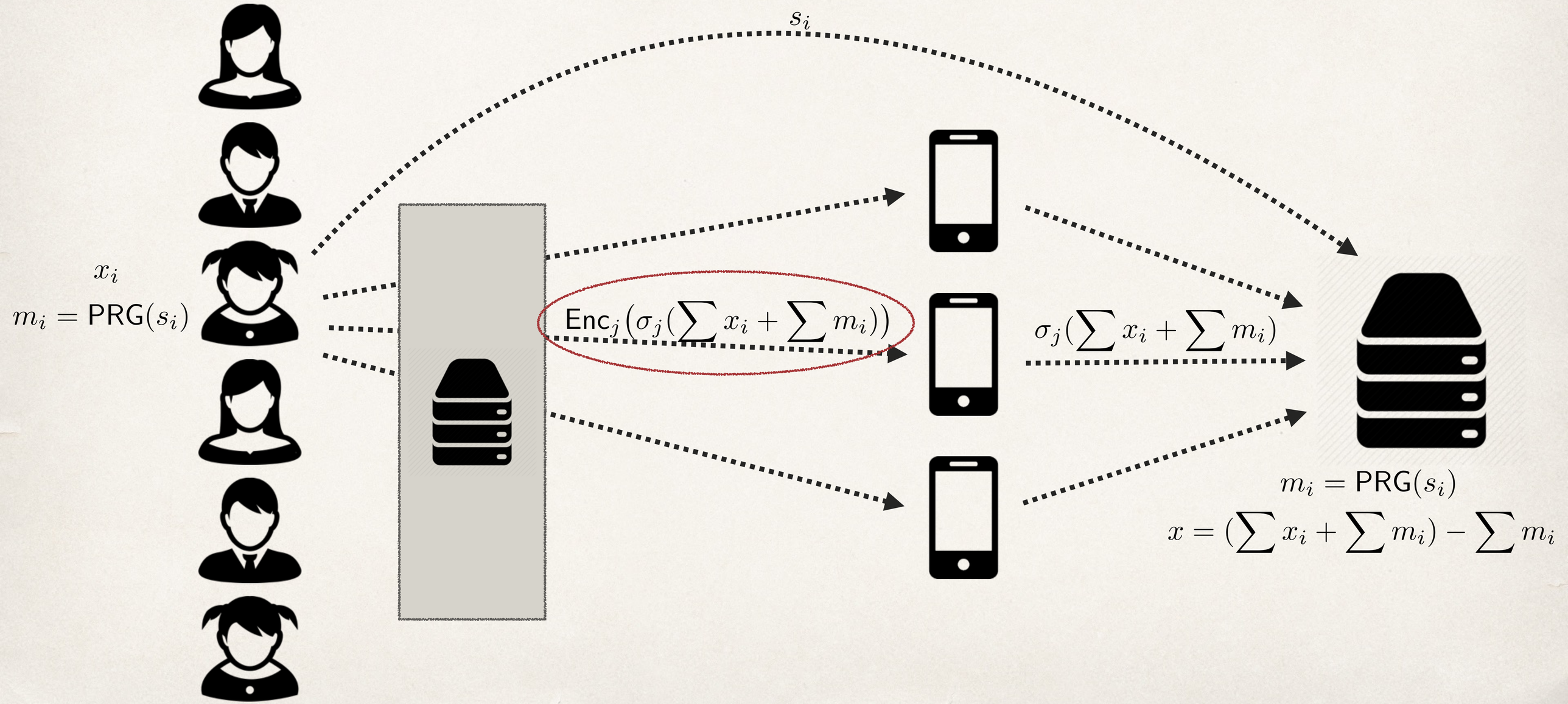
Protocol


s_i
x_i
$m_i = \mathsf{PRG}(s_i)$
$\sigma_j(x_i + m_i)$
$\sigma_j(\sum x_i + \sum m_i)$
$m_i = \mathsf{PRG}(s_i)$
$x = (\sum x_i + \sum m_i) - \sum m_i$

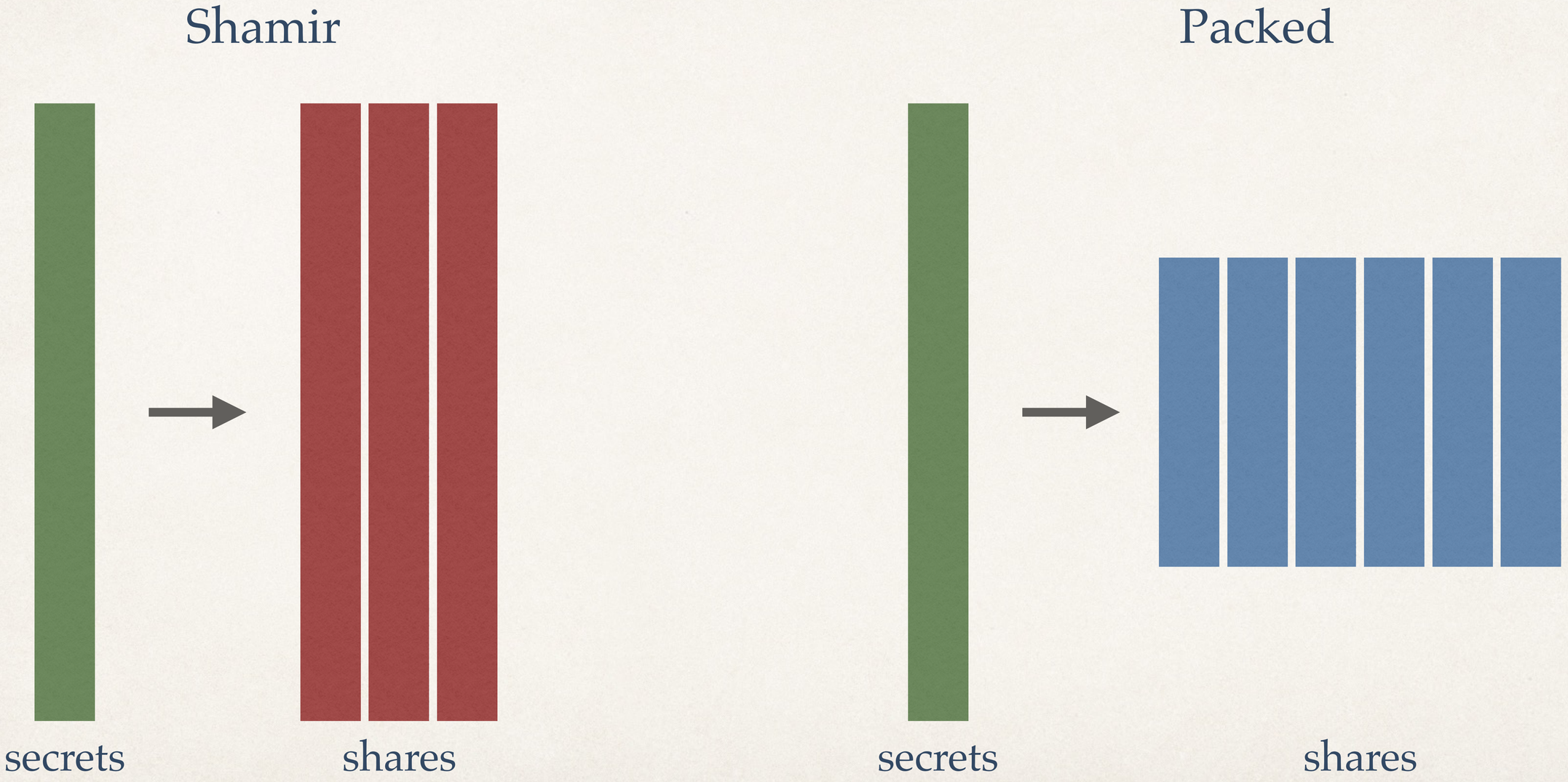
Protocol


s_i
x_i
$m_i = \mathsf{PRG}(s_i)$
$\mathsf{Enc}_j(\sigma_j(x_i + m_i))$
$\sigma_j(\sum x_i + \sum m_i)$
$m_i = \mathsf{PRG}(s_i)$
$x = (\sum x_i + \sum m_i) - \sum m_i$

Protocol


s_i
x_i
$m_i = \mathsf{PRG}(s_i)$
$\mathsf{Enc}_j(\sigma_j(\sum x_i + \sum m_i))$
$\sigma_j(\sum x_i + \sum m_i)$
$m_i = \mathsf{PRG}(s_i)$
$x = (\sum x_i + \sum m_i) - \sum m_i$

Packed Secret Sharing



Result

lightweight MPC protocol for linear functions,
tailored for large-scale high-dimension aggregation

Users	Clerks	Server	DP
single message	easy setup	most of the work	one extra round
passive security	work distribution	output only	passive security
	level of resilience		
	some active security		

Implementation and Experimentation

Rust

Laptops

iPhone/Android

Raspberry Pi

(web)

Secret sharing

Additive

Packed Shamir

Encryption

NaCl/Sodium

Paillier (packed)

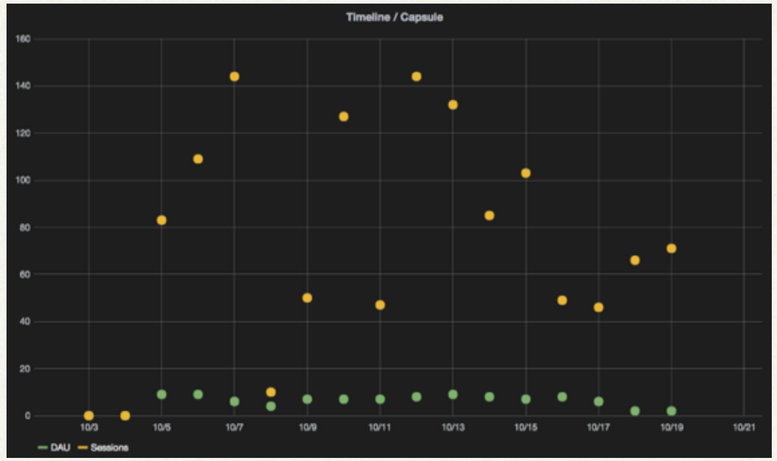
	clerks	threshold	packing
small	26	5	10
medium	80	16	47
large	728	145	366

Analytics

Communication

dimension 100

non-homomorphic



download	small	medium	large
25 000	977KB		
80 000		938KB	
250 000			977KB

Averages

Computation

dimension 35k

homomorphic

unlimited users

decrypt	small	medium	large
RP	21.6s	4.6s	0.6s
iPhone	6.2s	1.3s	0.2s
MBP	0.9s	0.2s	0.03s

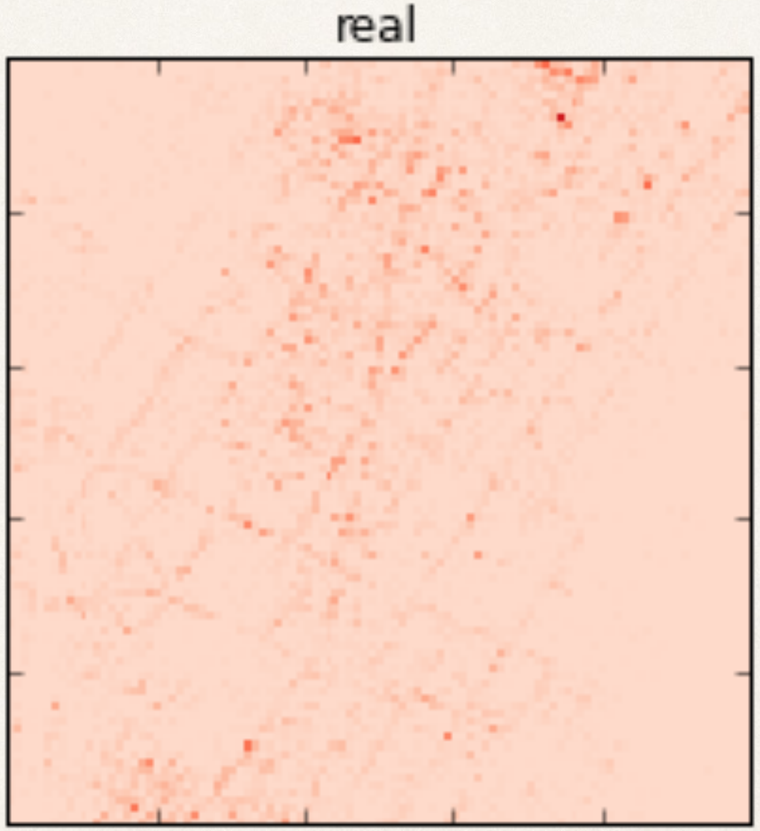
share	small	medium	large
RP	0.3s	0.2s	0.2s
iPhone	0.3s	0.3s	0.3s
MBP	0.06s	0.03s	0.03s

Discover Popular Places

Approximating

dimension 160k (20k)

mix with tools from big data

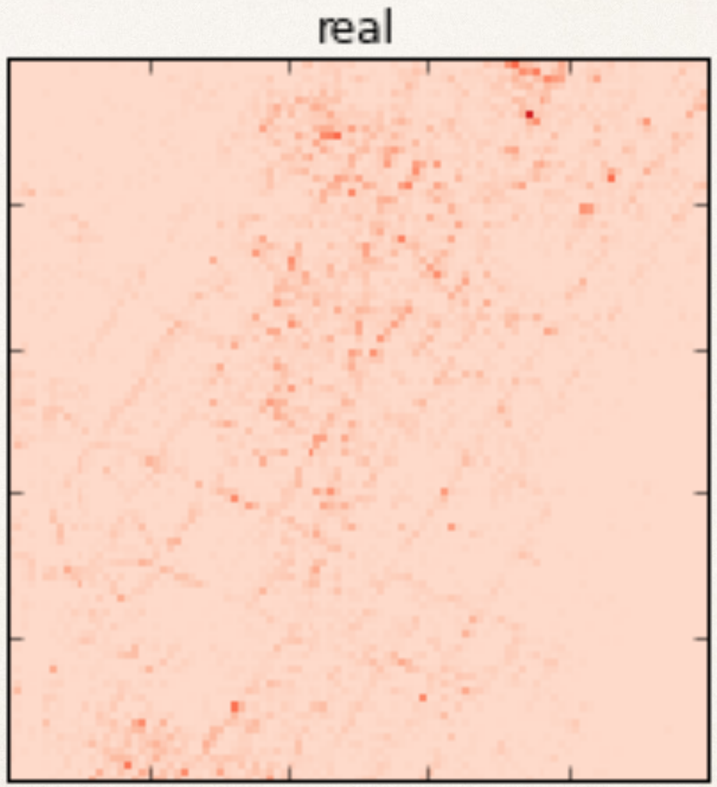


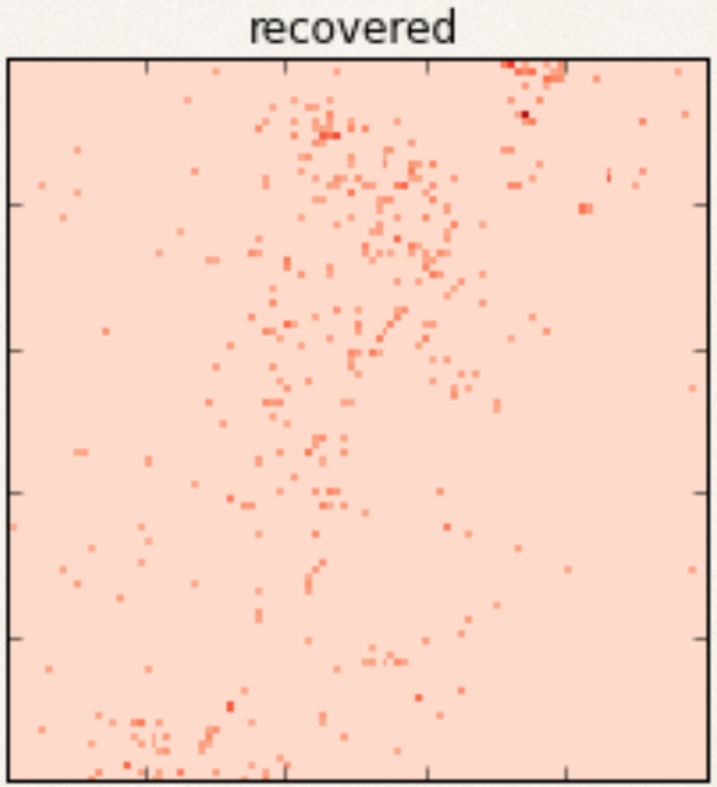
Min-Count Sketch

$$h_i(x) \mapsto [1, w]$$

$h_1(x)$			$+1$							
$h_2(x)$				$+1$						
$h_3(x)$		$+1$								
$h_4(x)$								$+1$		
$h_5(x)$					$+1$					

Approximated Popular Places



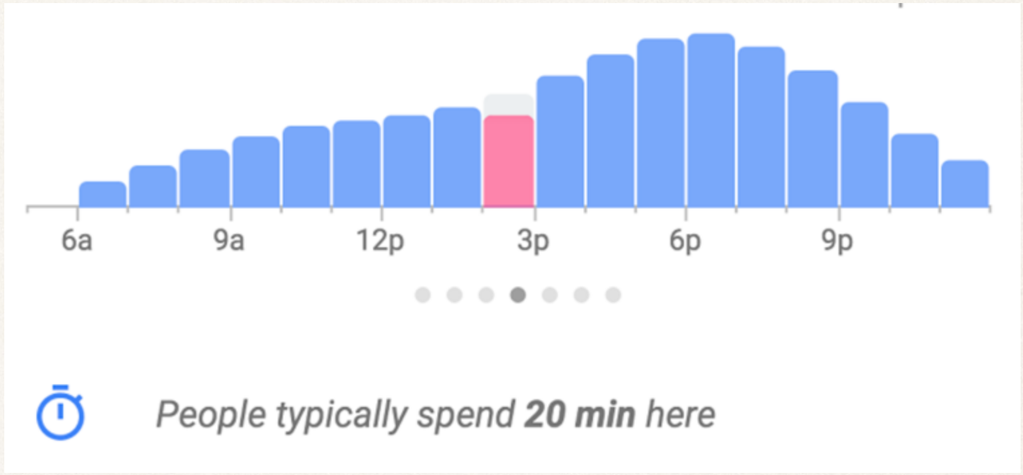


Busy Hours

Several Rounds

approximate top 1,000 places
(dimension 80k/10k)

time and duration for these
(dimension 1,000 * 12 * 4)



Learn from distributed data sets (on mobile phones)

MPC for single company

Community of volunteers

Outsource and distribute work load

Thank you!

https://github.com/snipsco/sda
https://github.com/snipsco/rust-paillier
https://github.com/snipsco/rust-threshold-secret-sharing

https://project.inria.fr/pamela/