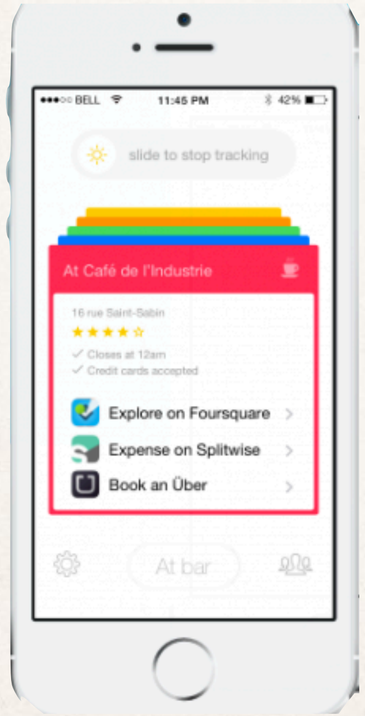
Private Data Aggregation on a Budget

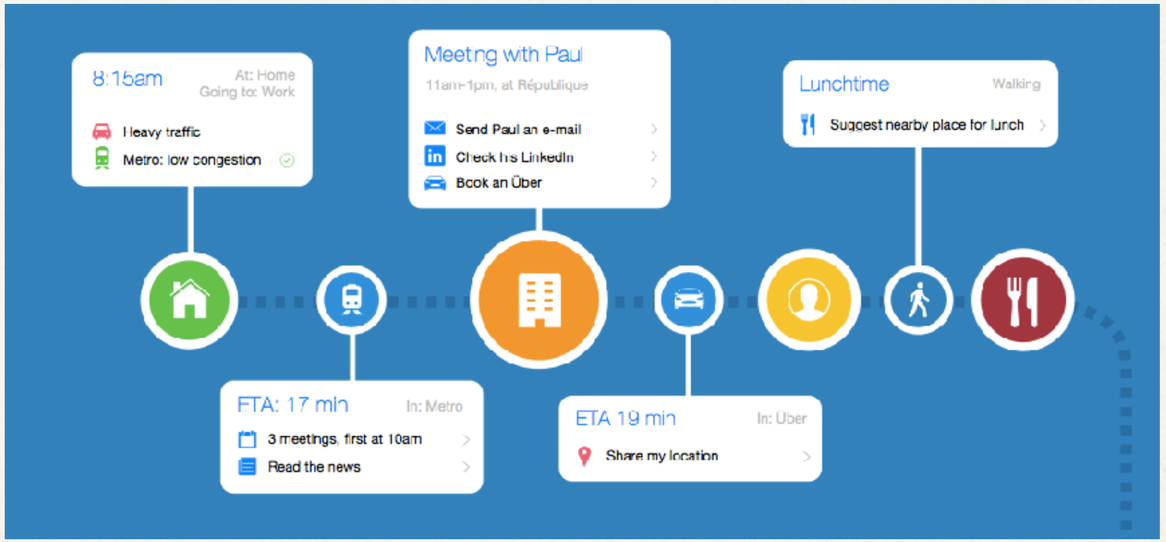
Morten Dahl Valerio Pastro Mathieu Poumeyrol

HEAT Workshop, November '17

Smart Assistant

app giving personalised and context aware recommendations



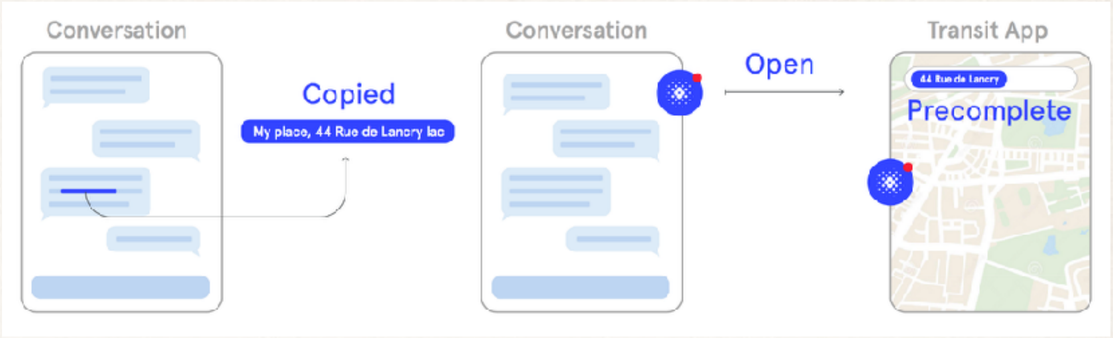


Data

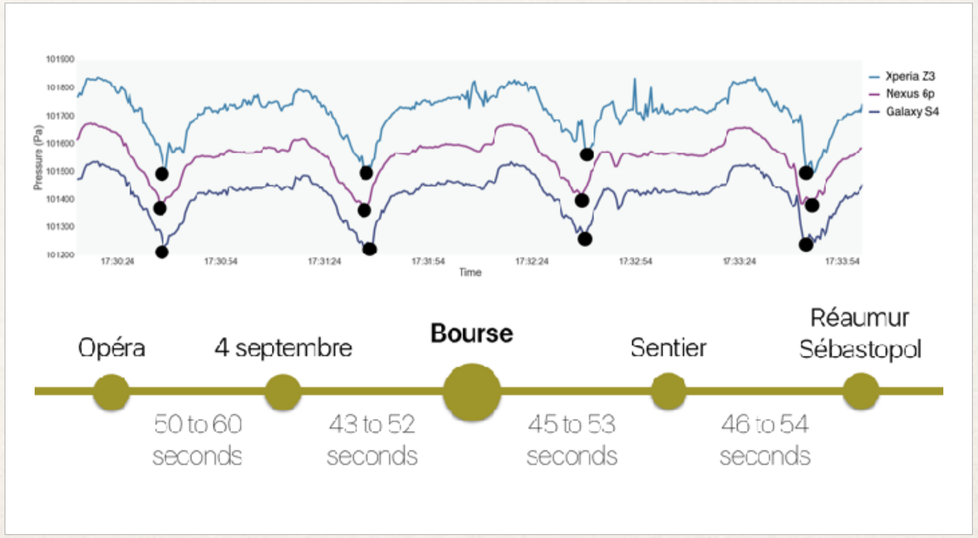
small company

limits of public/generated data

high quality data sources on mobile



https://medium.com/snips-ai/the-contextual-bubble-7576e4bd40c9



https://medium.com/snips-ai/underground-location-tracking-3ea56803dddc

Getting Acceptance

sensitive data

Privacy by Design

run everything locally

Why you should bet big on privacy

Posted May 17, 2016 by Rand Hindi (@randhindi)

Rand Hindi
CONTRIBUTOR

Rand Hindi is the founder of Snips.

Ever felt like you were being watched online? You know, like when you read something about New York, and the next site you visit shows you ads for New York hotels? As it turns out, on my computer, there were more than 130 companies tracking my every move (check yours here, then install this plug-in).

These companies are basically engaging in mass surveillance. Just as governments justify tracking us to prevent terrorist attacks, these companies are tracking us online, without our

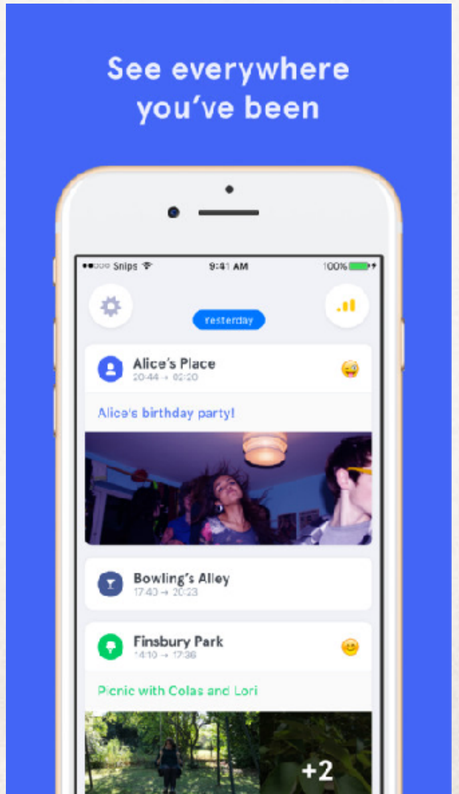
https://techcrunch.com/2016/05/17/why-you-should-bet-big-on-privacy/

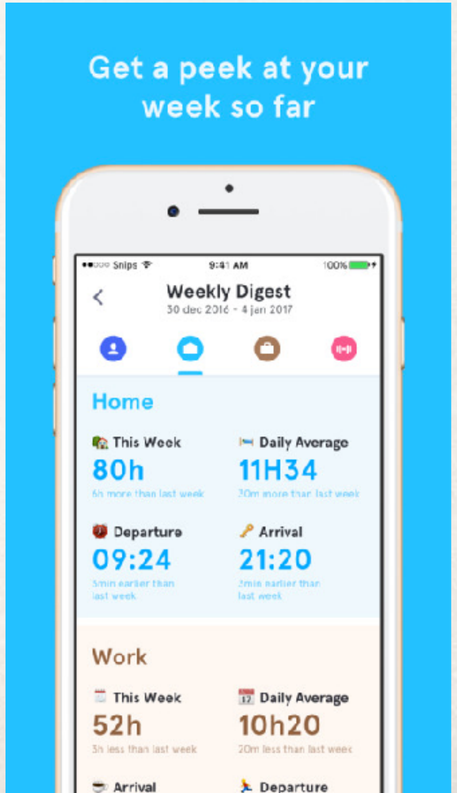
Limits of Running Locally

cold start

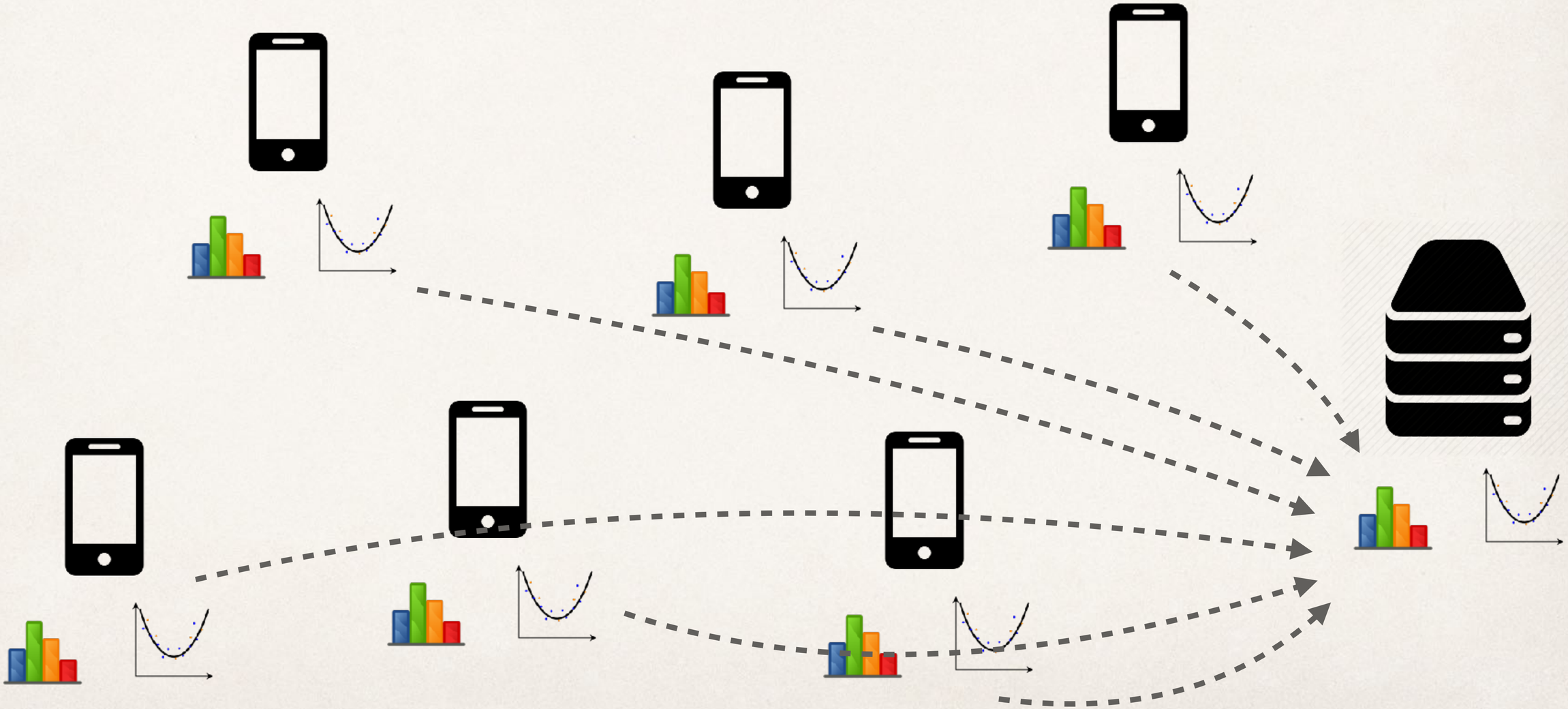
travelling

discovery





Learning from Distributed Data Sets



Constraints

Business

no individual user data

need some kind of crypto

six month cycle

simple solution and implementation

Mobile and Web

limited computation

minimise device processing

limited connectivity

minimise session count + length

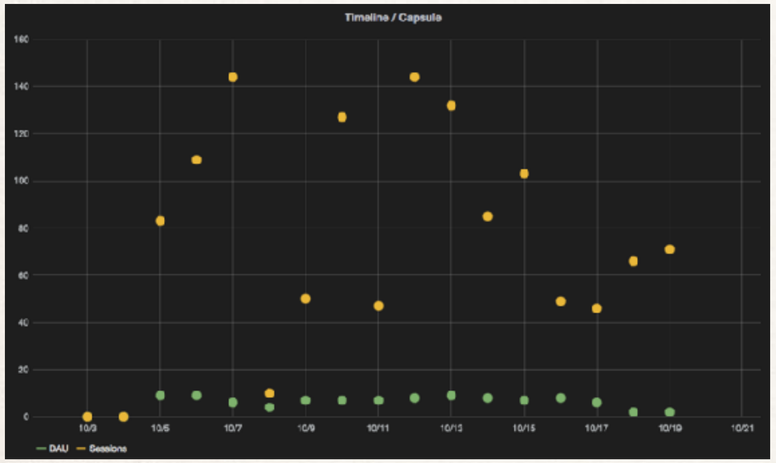
sporadic behaviour

minimise coordination

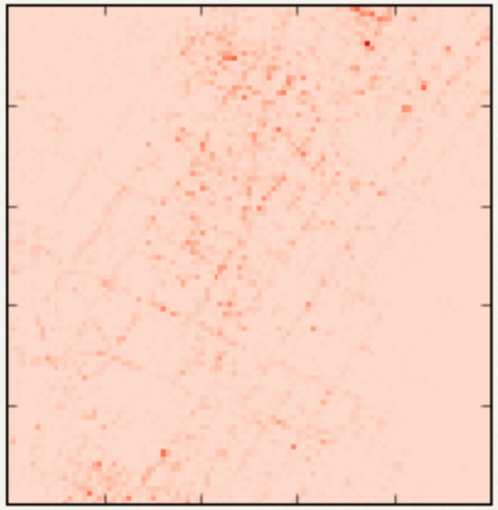
Learning Aggregations

$$x = \sum x_i$$

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



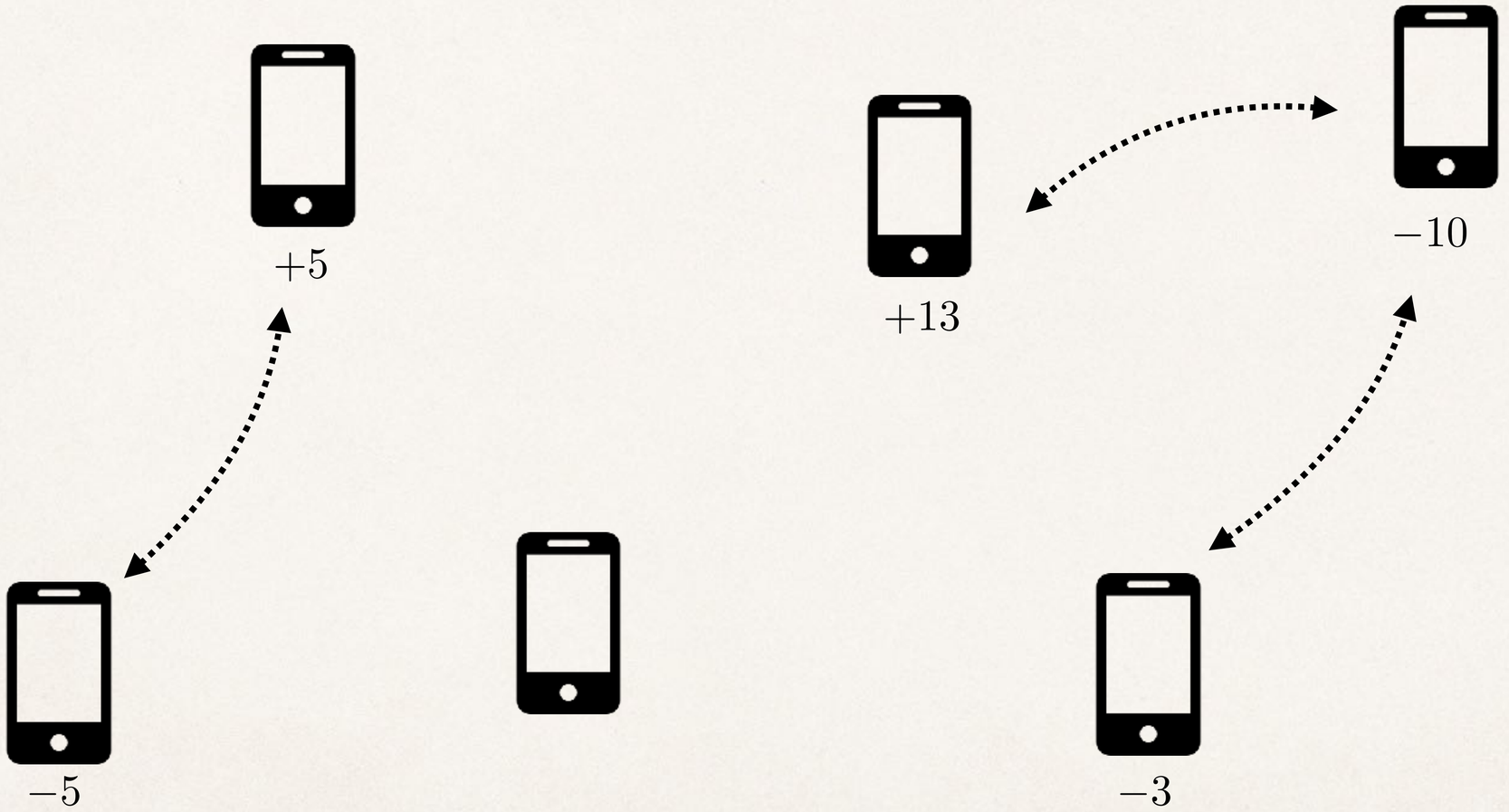
analytics / surveys



discovery

insights

Sensor Networks



$$x = \sum x_i \pm r_i$$

high performance

coordination

(Local) Differential Privacy

$Lap(1/\epsilon)$

$Lap(1/\epsilon)$

$Lap(1/\epsilon)$

$Lap(1/\epsilon)$

$Lap(1/\epsilon)$

$Lap(1/\epsilon)$

$$\hat{x} = \sum x_i + \nu$$

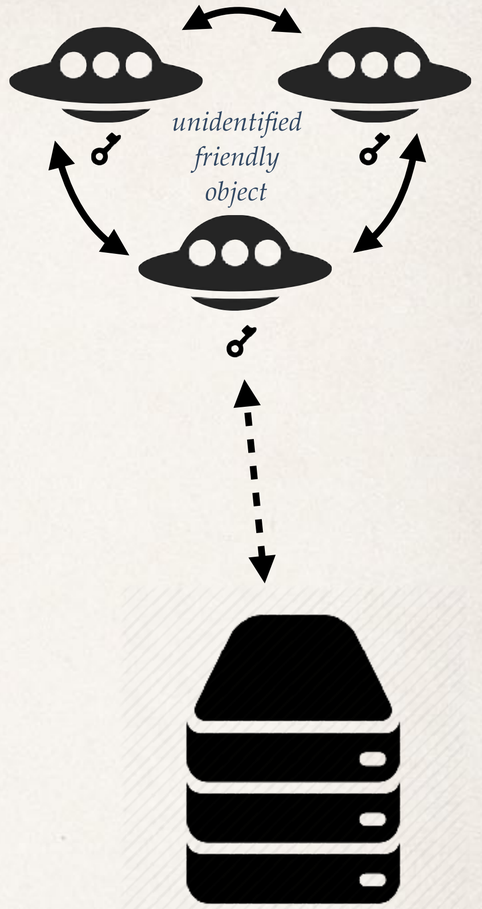
$$x \approx \hat{x}$$

high performance

formal privacy

signal-to-noise

Homomorphic Encryption



$$Enc(x) = \sum Enc(x_i)$$

$$x = Dec(\cdots)$$

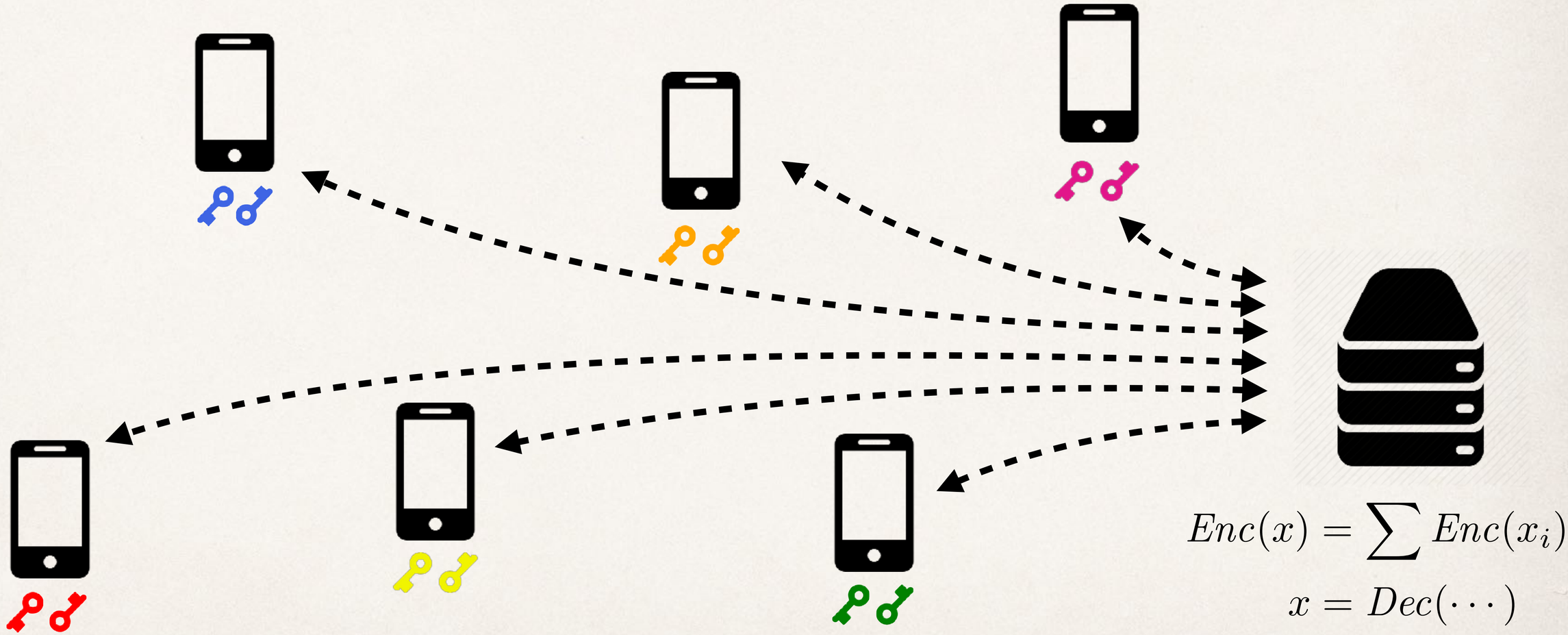
flexible

decent performance

external parties

key setup

Multi-key Homomorphic Encryption



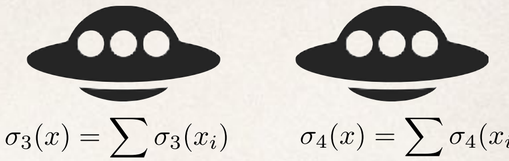
$$Enc(x) = \sum Enc(x_i)$$

$$x = Dec(\cdots)$$

flexible

fragile decryption

Secret Sharing




$\sigma_1(\cdot), \sigma_2(\cdot), \ldots$


$\sigma_1(\cdot), \sigma_2(\cdot), \ldots$


$\sigma_1(\cdot), \sigma_2(\cdot), \ldots$

$\sigma_1(x) = \sum \sigma_1(x_i)$

$x = Rec(\cdots)$


$\sigma_1(\cdot), \sigma_2(\cdot), \ldots$


$\sigma_1(\cdot), \sigma_2(\cdot), \ldots$


$\sigma_1(\cdot), \sigma_2(\cdot), \ldots$

flexible

decent performance

external parties

external computation

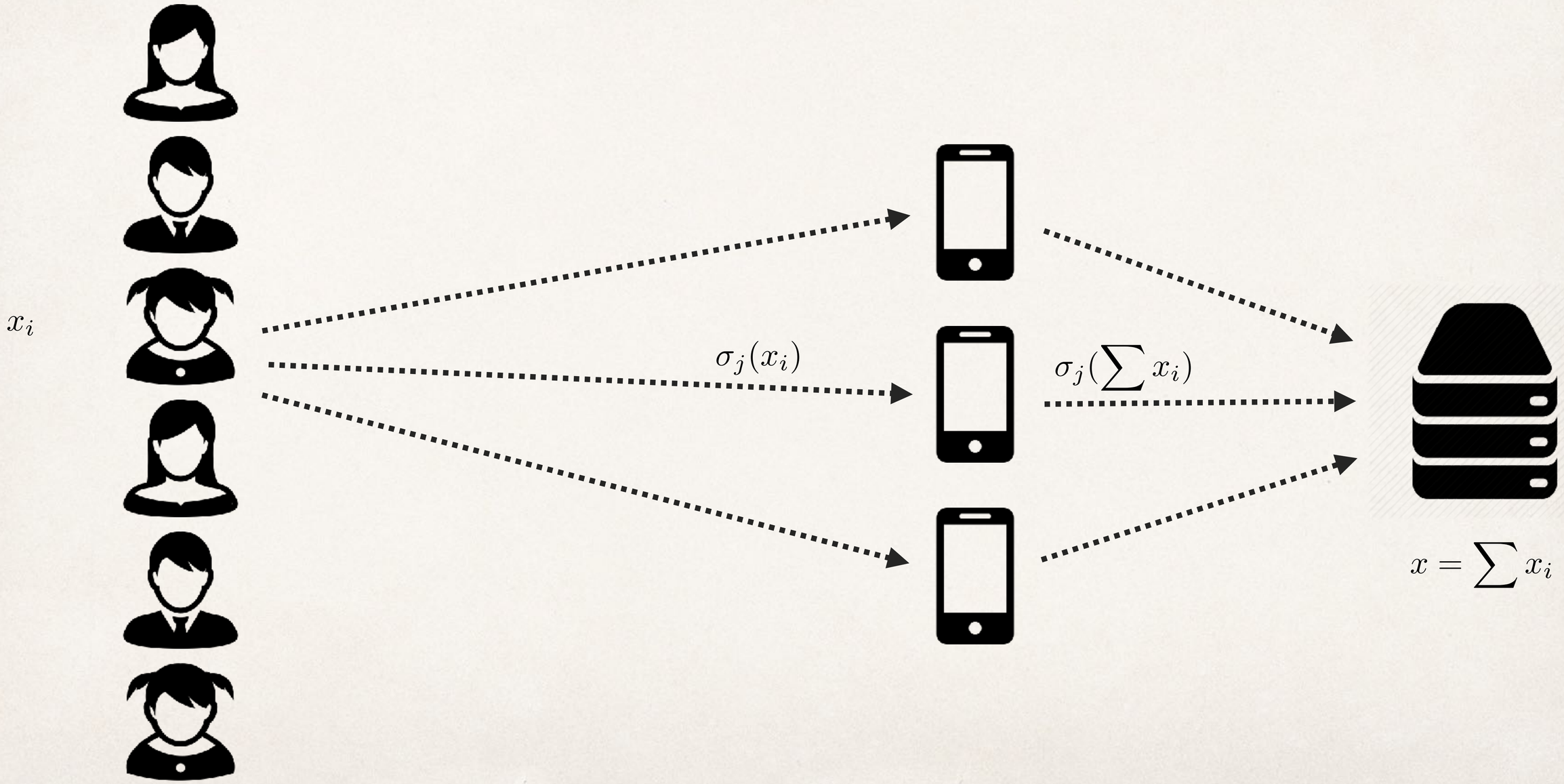
Aggregation Protocol

mix!

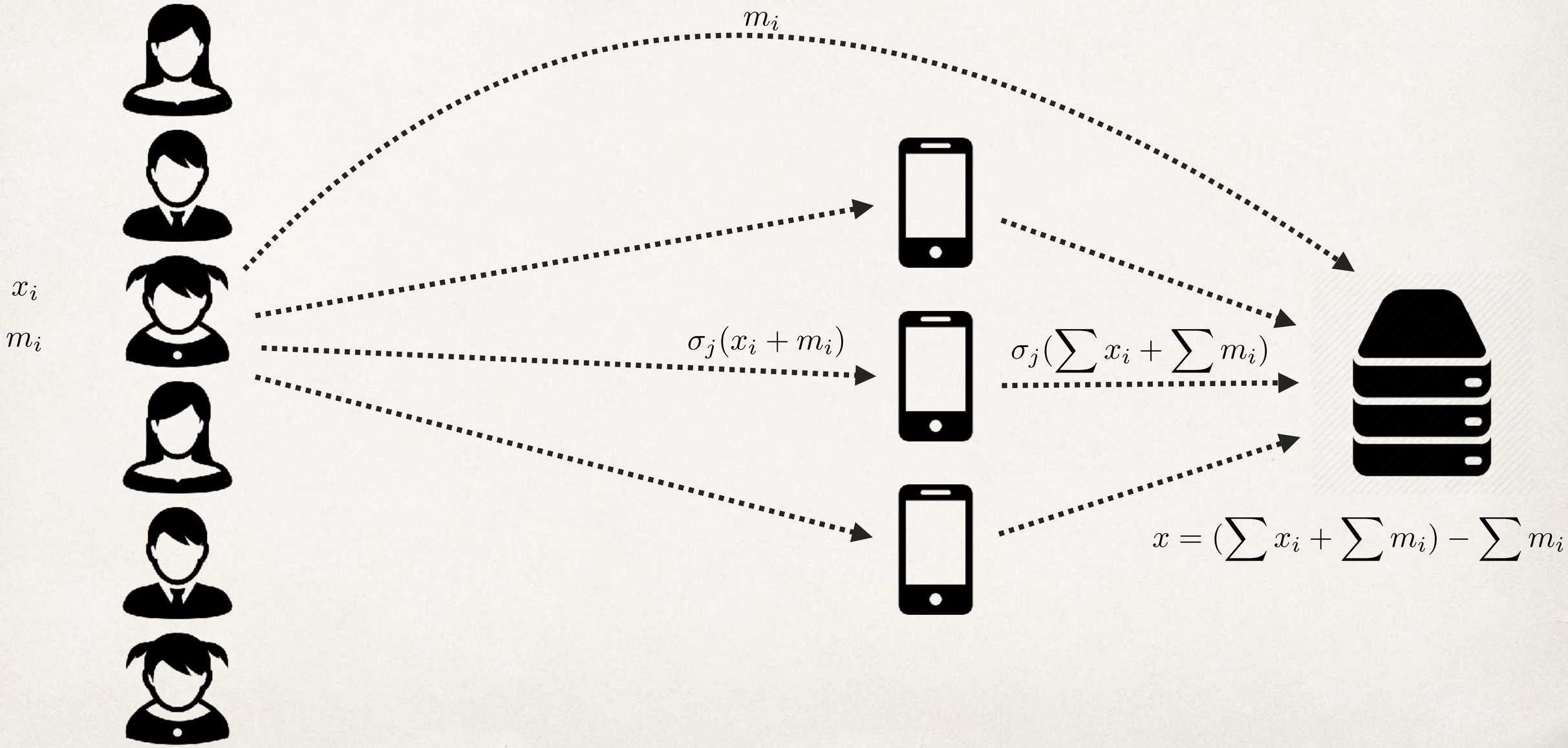
Community Trust

only one (powerful) server

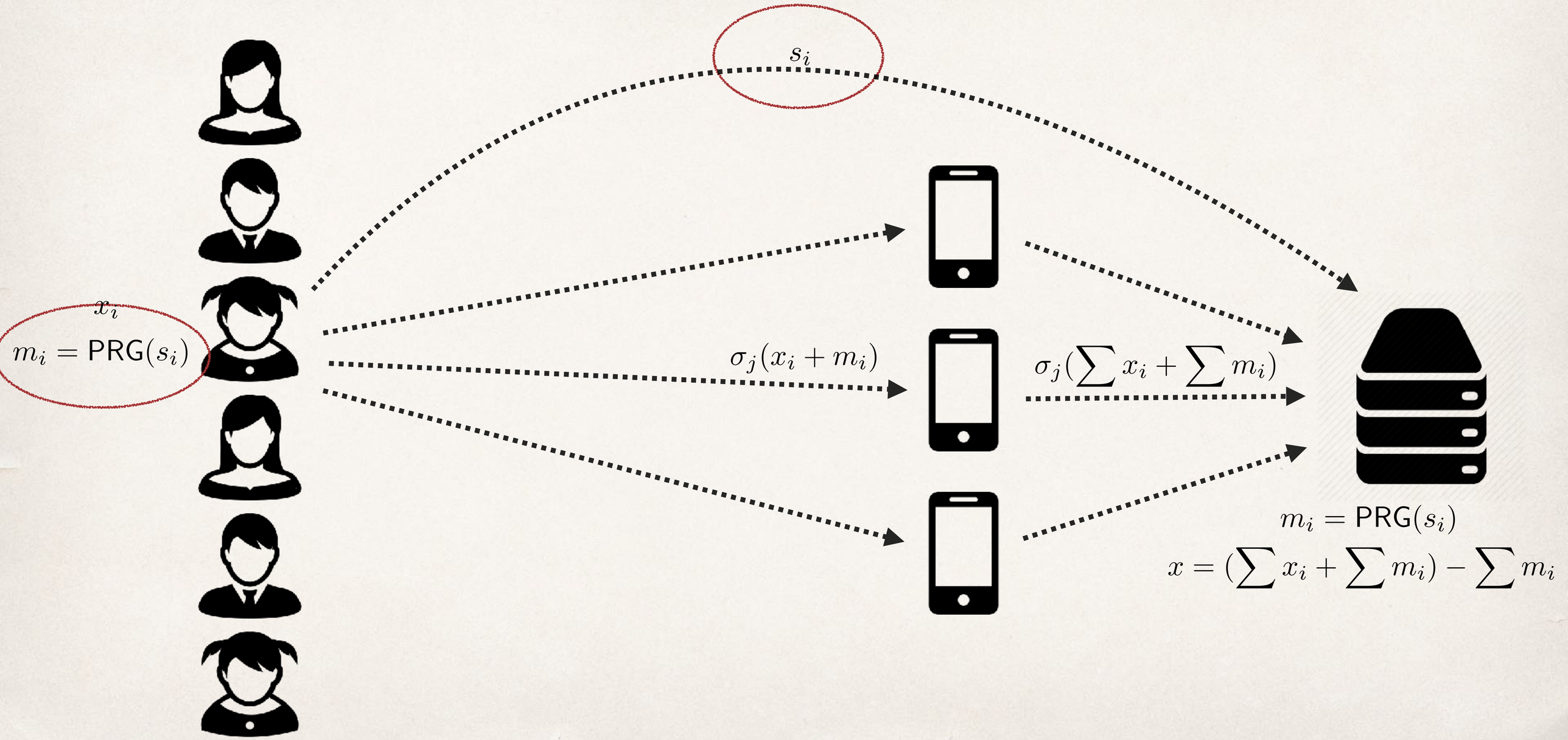
Aggregation Protocol



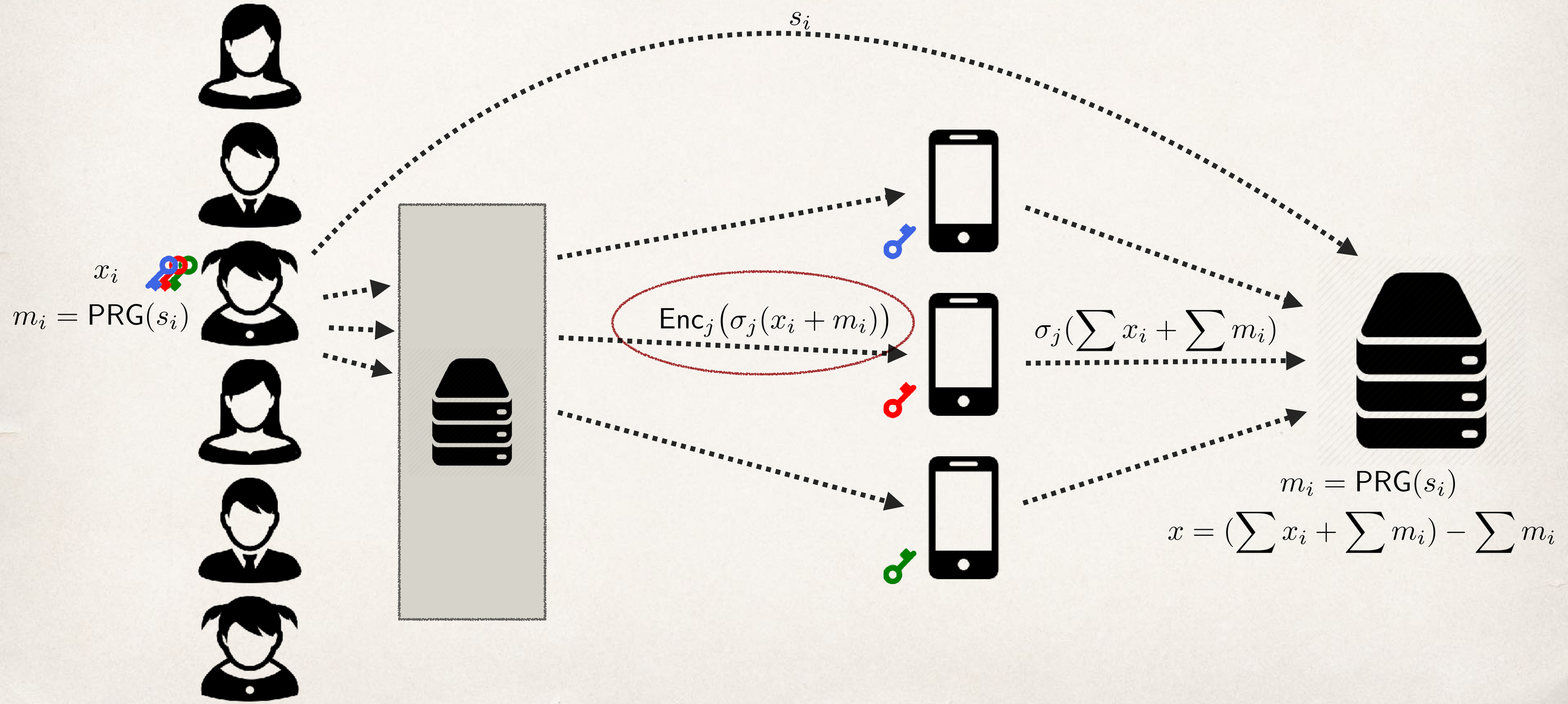
Aggregation Protocol



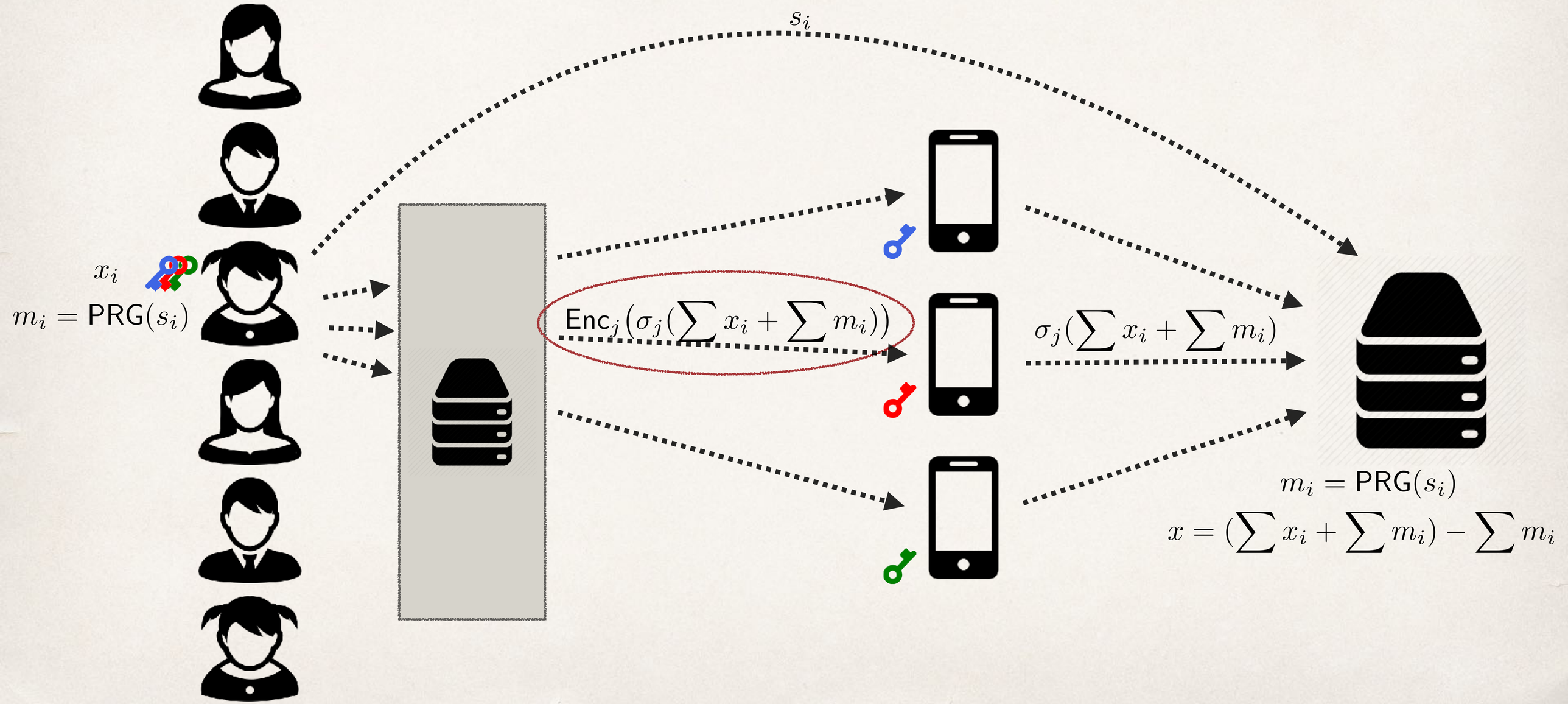
Aggregation Protocol



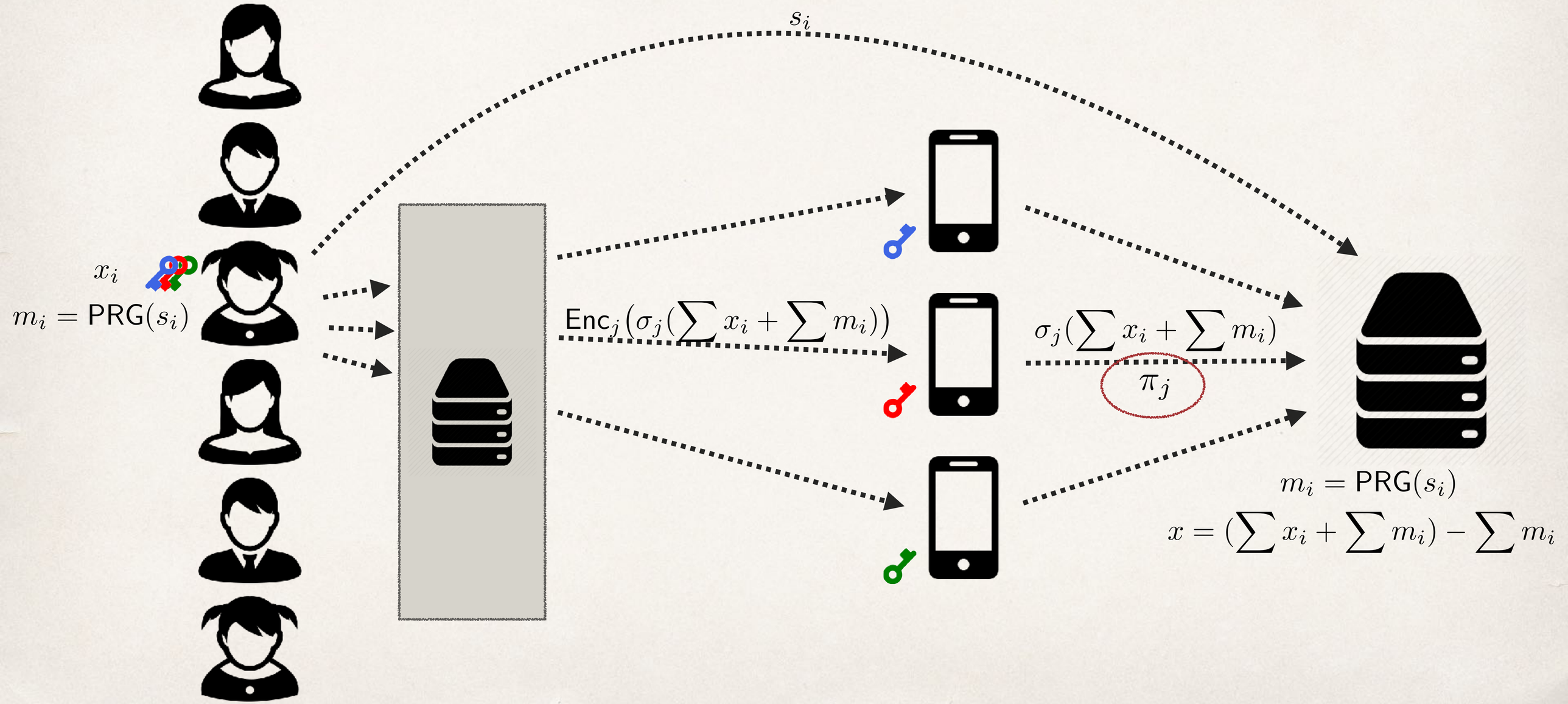
Aggregation Protocol



Aggregation Protocol



Aggregation Protocol



Result

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

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

Implementation

Rust	Secret sharing	Encryption
Laptops	Additive	Sodium
iPhone / Android	Shamir	**Paillier**
Raspberry Pi	**Packed Shamir**	
(web)		

Paillier

$$Enc(m, r) = g^m r^n = c \bmod n^2$$

Packing

$$log_2(n) \sim 2048$$

$$Enc(m_0, \ldots, m_\ell, r) = Enc(m_0\beta^0 + \cdots + m_\ell\beta^\ell, r)$$

$$m_i < \mu \qquad \mu\alpha < \beta$$

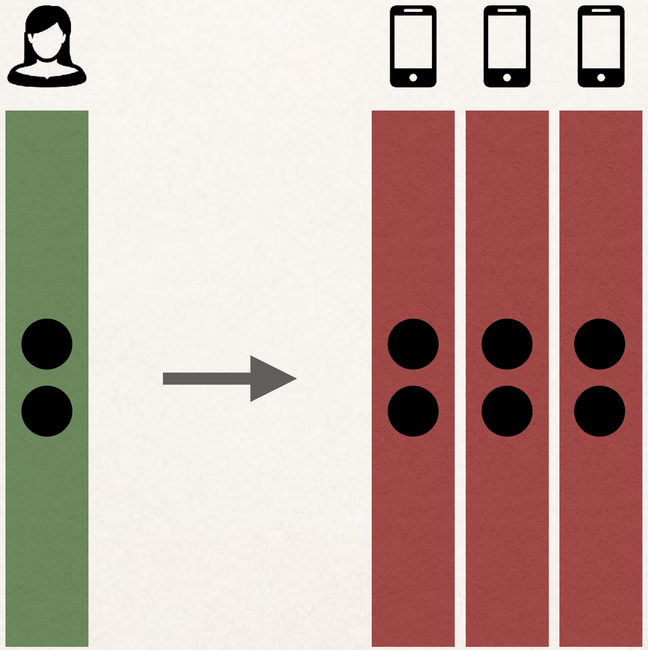
Decryption Proof

$$Dec(c) = (m, r)$$

Secret Sharing & Work Distribution

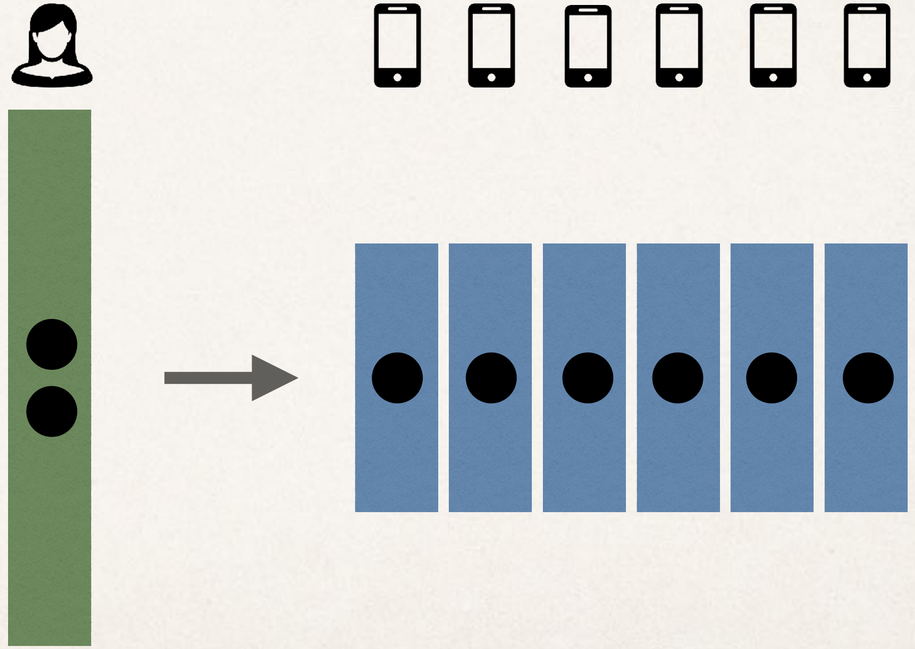
Shamir

$p(0) = x_i$ $deg(p) = t$ $\sigma_j(x_i) = p(j)$



Packed

$p(-k) = x_{i,k}$ $deg(p) = kt$ $\sigma_j(x_i) = p(j)$



Experiments

Packed Shamir

Sodium + Paillier

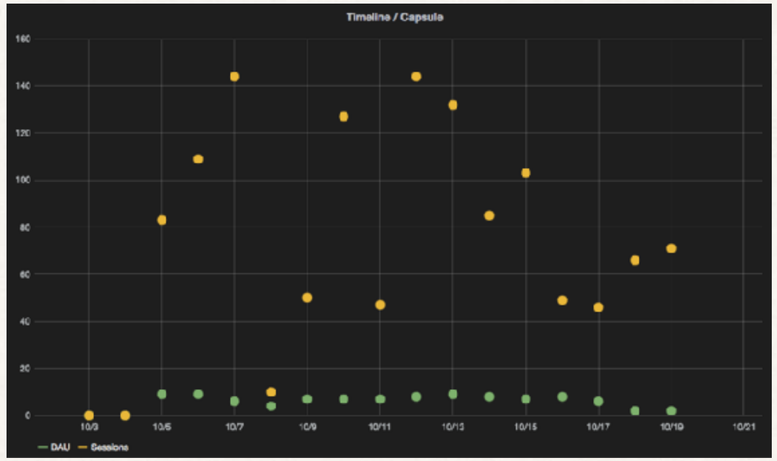
	clerks	threshold	packing
small	26	5	10
medium	80	16	47
large	728	145	366

Analytics

Communication

dimension 100

Sodium



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

Mean

Computation

dimension 35k

Paillier

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

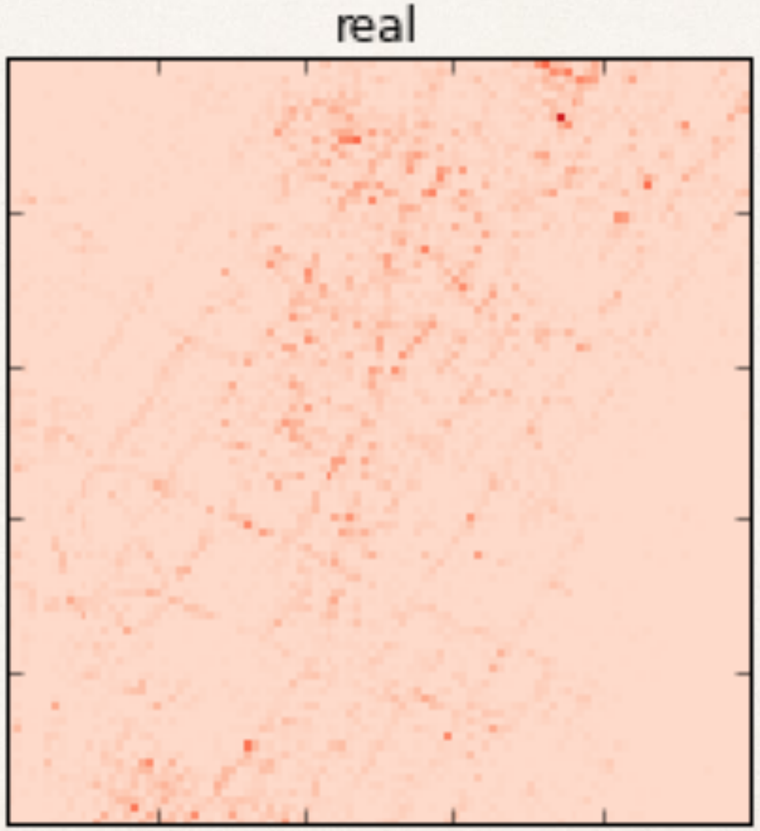
share	small	medium	large
RP	0.3s	0.2s	0.2s
iPhone	0.3s	0.3s	0.3s
MBP	0.06s	0.03s	0.03s

Place Discovery

Approximating

dimension 160k

big data tools



Min-Count Sketch

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

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

Min-Count Sketch

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

+

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

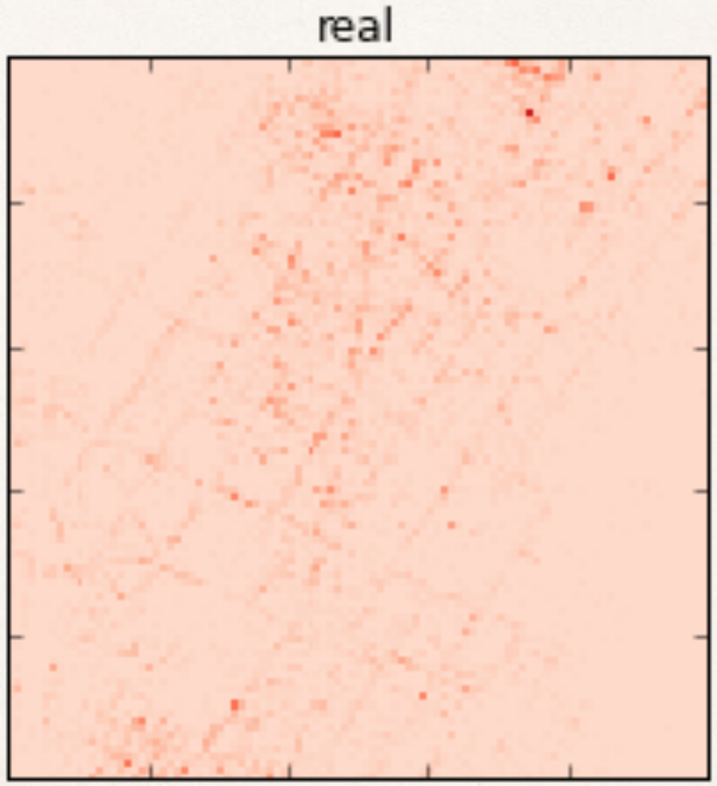
+

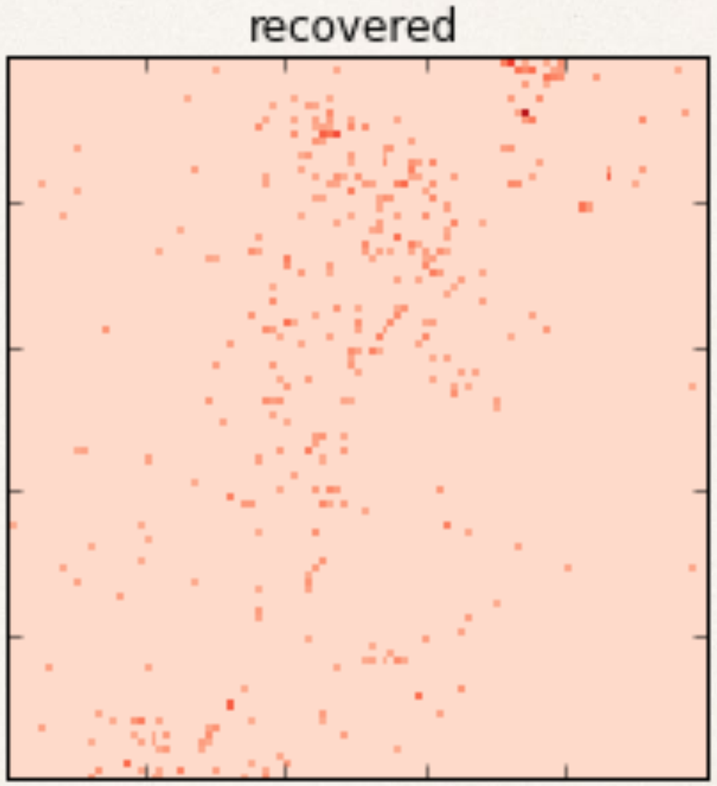
$h_1(x)$			+1						
$h_2(x)$				+1					
$h_3(x)$		+1							
$h_4(x)$							+1		
$h_5(x)$					+1				

$=$

$h_1(x)$			30						
$h_2(x)$				12					
$h_3(x)$		47							
$h_4(x)$								52	
$h_5(x)$					28				

Approximation





8x compression

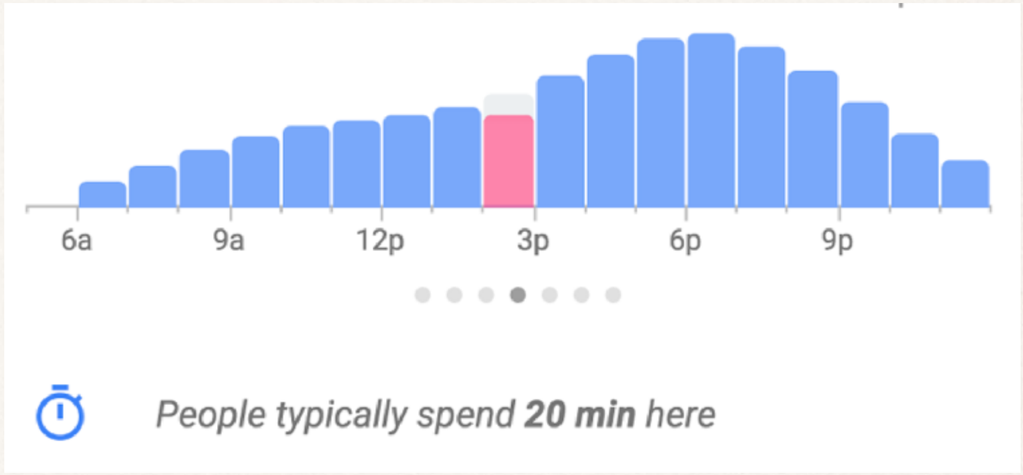
26KB (Paillier)

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)

Secure computation for single company

Community trust

Outsource and distribute work load

Thank you!

@mortendahlcs

https://eprint.iacr.org/2017/643.pdf