

# Privacy-preserving Firefox telemetry with Prio

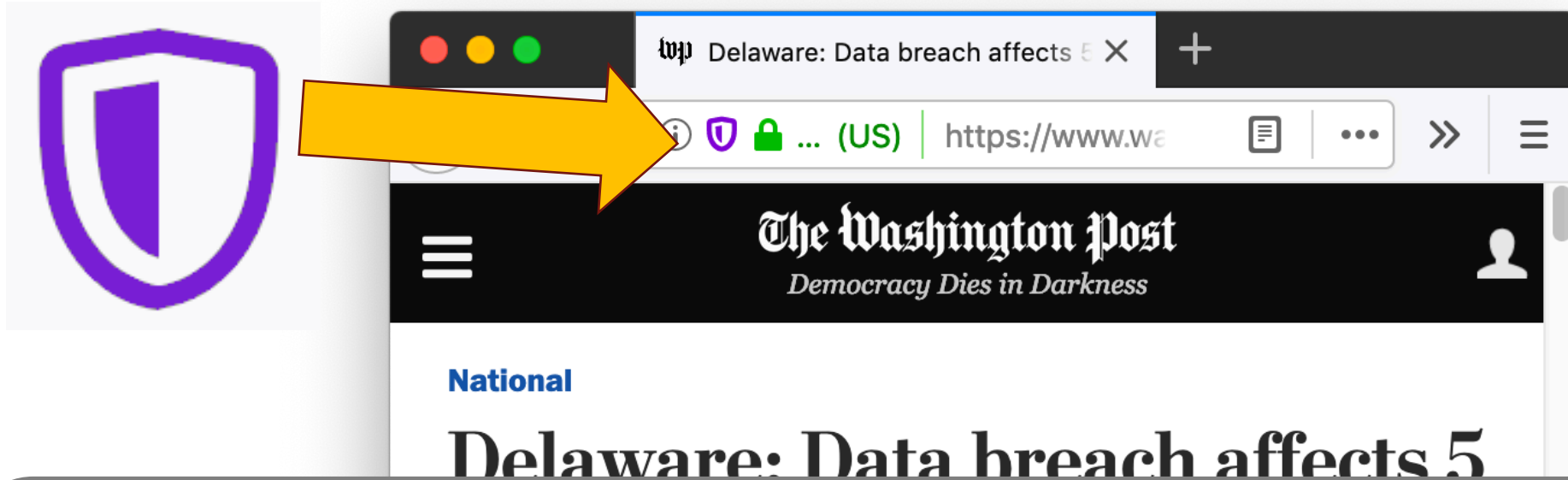
Henry Corrigan-Gibbs  
(EPFL → MIT CSAIL)

**In collaboration with:** Dan Boneh (Stanford),  
Gary Chen, Steven Englehardt, Robert Helmer, Chris Hutten-Czapski,  
Anthony Miyaguchi, Eric Rescorla, and Peter Saint-Andre (Mozilla)

# Running example: Measuring effectiveness of tracking protection



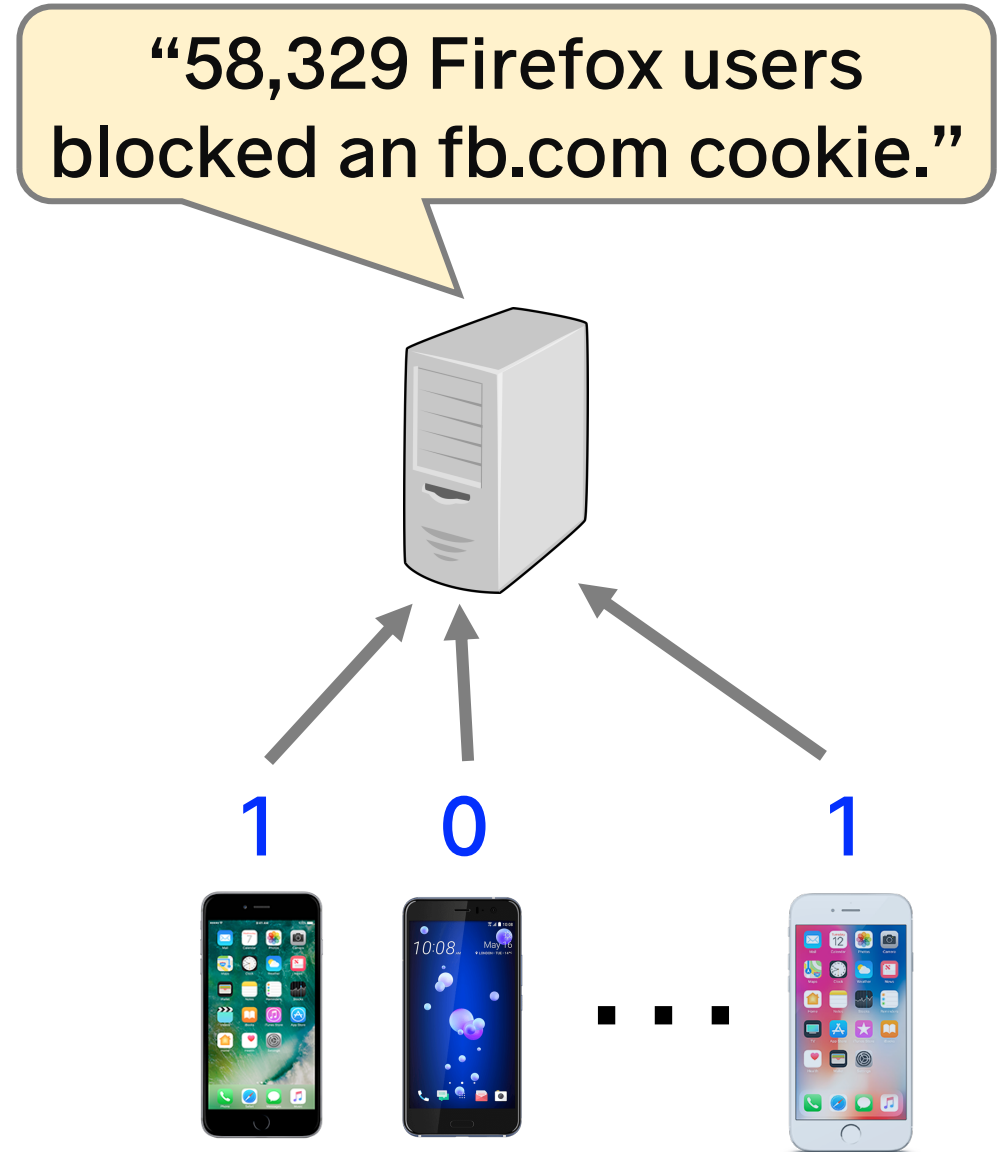
# Running example: Measuring effectiveness of tracking protection



Mozilla wants to know:  
“How many Firefox users blocked  
a tracking cookie from fb.com?”

Software vendors often answer these questions by collecting **sensitive** usage data directly.

→ **Single point of failure.**

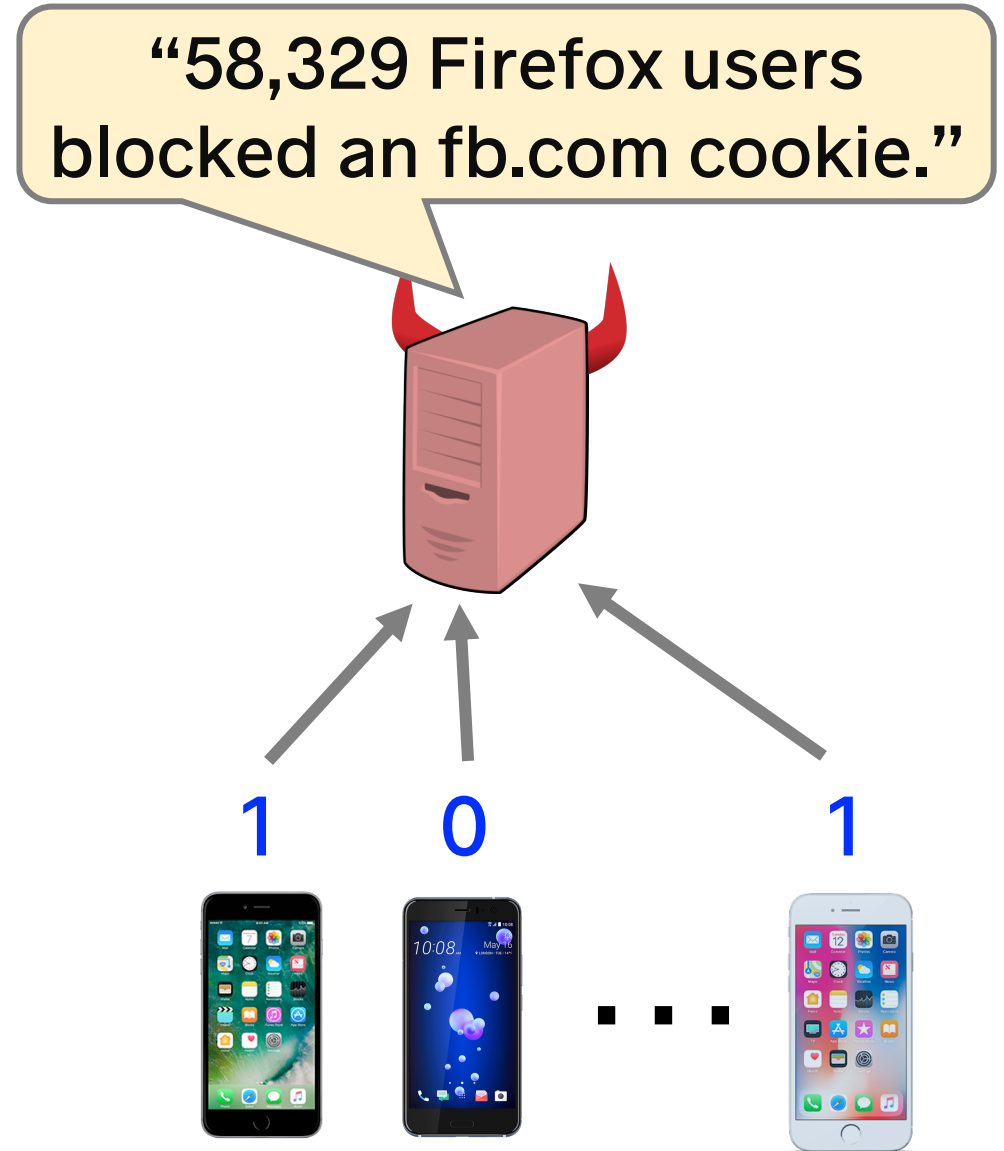




Software vendors often answer these questions by collecting **sensitive** usage data directly.

→ **Single point of failure.**

- Theft by attackers
- Abuse by malicious insiders
- Snooping by governments



# Prio: Aggregate data without the privacy risks

C-G and Boneh (NSDI 2017)

- Collect aggregate usage data **without seeing any single user's data.**
- New cryptography makes this system practical
  - Proofs on secret-shared data
- Basis for Mozilla's new privacy-preserving telemetry system
  - In pilot phase: Enabled by default in Firefox's "Nightly" build
  - Largest deployment of technology based on PCPs (probabilistically checkable proofs)

# Running example: Measuring effectiveness of tracking protection

- There are  $n \approx 2,500$  domains on the tracking-protection blacklist
- For each blocked domain, each user  $i$  has a bit
  - Bit is “1” iff user  $i$ ’s browser ever blocked cookies from domain.com
  - These bits are **sensitive** – reveal user’s browsing history

	fb.com	orkut.com	ru4.com	onad.eu	nugg.ad	xa.net	po.st	sas.com	cams.com	tapit.com	ucoz.ae	gmail.com	ibm.com	...	Domain $n$
User 1	<1	0	1	0	1	0	1	0	0	0	0	0	0	...	1>
User 2	<1	1	1	0	1	0	1	0	0	1	0	0	1	...	0>
...														...	
User $U$	<0	0	0	0	1	0	1	0	0	0	0	1	0	...	0>

# Running example: Measuring effectiveness of tracking protection

- Mozilla wants the sum of these vectors over all users  $i$

	fb.com	orkut.com	ru4.com	onad.eu	nugg.ad	xa.net	po.st	sas.com	cams.com	tapit.com	ucoz.ae	gmail.com	ibm.com	...	Domain $n$
User 1	<1	0	1	0	1	0	1	0	0	0	0	0	0	...	1>
User 2	<1	1	1	0	1	0	1	0	0	1	0	0	1	...	0>
...															
User $U$	<0	0	0	0	1	0	1	0	0	0	0	1	0	...	0>

# Running example: Measuring effectiveness of tracking protection

- Mozilla wants the sum of these vectors over all users  $i$

	<i>fb.com</i>	<i>orkut.com</i>	<i>ru4.com</i>	<i>onad.eu</i>	<i>nugg.ad</i>	<i>xa.net</i>	<i>po.st</i>	<i>sas.com</i>	<i>cams.com</i>	<i>tapit.com</i>	<i>ucoz.ae</i>	<i>gmail.com</i>	<i>ibm.com</i>	...	Domain $n$
<b>User 1</b>	<1	0	1	0	1	0	1	0	0	0	0	0	0	...	1>
<b>User 2</b>	<1	1	1	0	1	0	1	0	0	1	0	0	1	...	0>
...															
<b>User <math>U</math></b>	<0	0	0	0	1	0	1	0	0	0	0	1	0	...	0>
<b>SUM</b>	31,	91,	6,	0,	8,	29,	81,	0,	0,	88,	10,	5,	59,	...,	50

# Running example: Measuring effectiveness of tracking protection

- Mozilla wants the sum of these vectors over all users  $i$

	<i>fb.com</i>	<i>orkut.com</i>	<i>ru4.com</i>	<i>onad.eu</i>	<i>nugg.ad</i>	<i>xa.net</i>	<i>po.st</i>	<i>sas.com</i>	<i>cams.com</i>	<i>tapit.com</i>	<i>ucoz.ae</i>	<i>gmail.com</i>	<i>ibm.com</i>	...	Domain $n$
<b>User 1</b>	<1	0	1	0	1	0	1	0	0	0	0	0	0	...	1>
<b>User 2</b>	<1	1	1	0	1	0	1	0	0	1	0	0	1	...	0>
...															
<b>User <math>U</math></b>	<0	0	0	0	1	0	1	0	0	0	0	1	0	...	0>
<b>SUM</b>	<b>31,</b>	91,	6,	0,	8,	29,	81,	0,	0,	88,	10,	5,	59,	...	50

How many users blocked fb.com cookies via tracking protection

# Running example: Measuring effectiveness of tracking protection

- Mozilla wants the sum of these vectors over all users  $i$

	<i>fb.com</i>	<i>orkut.com</i>	<i>ru4.com</i>	<i>onad.eu</i>	<i>nugg.ad</i>	<i>xa.net</i>	<i>po.st</i>	<i>sas.com</i>	<i>cams.com</i>	<i>tapit.com</i>	<i>ucoz.ae</i>	<i>gmail.com</i>	<i>ibm.com</i>	...	<i>Domain n</i>
<b>User 1</b>	<1	0	1	0	1	0	1	0	0	0	0	0	0	...	1>
<b>User 2</b>	<1	1	1	0	1	0	1	0	0	1	0	0	1	...	0>
...															
<b>User <math>U</math></b>	<0	0	0	0	1	0	1	0	0	0	0	1	0	...	0>
<b>SUM</b>	31,	91,	6,	0,	8,	29,	81,	0,	0,	88,	10,	5,	59,	...	50

# Running example: Measuring effectiveness of tracking protection

- Mozilla wants the sum of these vectors over all users  $i$

		fb.com	orkut.com	ru4.com	onad.eu	nugg.ad	xa.net	po.st	sas.com	cams.com	tapit.com	ucoz.ae	gmail.com	ibm.com	Domain $n$	
User 1	$x_1$	0	1	0	1	0	1	0	0	0	0	0	0	0	...	1
User 2	$x_2$	1	1	0	1	0	1	0	0	1	0	0	1	...	0	
...	...															
User $U$	$x_U$	0	0	0	1	0	1	0	0	0	0	1	0	...	0	
<b>SUM</b>	$\sum_{i=1}^U x_i$	6	0	8	29	81	0	0	88	10	5	59	...	50		



# Running example: Measuring effectiveness of tracking protection

- Mozilla wants the sum of these vectors over all users  $i$

		fb.com	orkut.com	ru4.com	onad.eu	nugg.ad	xa.net	po.st	sas.com	cams.com	tapit.com	ucoz.ae	gmail.com	ibm.com	Domain $n$
User 1	$x_1$	0	1	0	1	0	1	0	0	0	0	0	0	...	1
User 2	$x_2$	1	1	0	1	0	1	0	0	1	0	0	1	...	0
...	...													...	
User $U$	$x_U$	0	0	0	1	0	1	0	0	0	0	1	0	...	0
<b>SUM</b>	$\sum_{i=1}^U x_i$	6	2	0	2	0	2	0	0	1	0	1	1	...	1

We run the system many times in parallel to compute the statistics for all domains

# Prio: System goals

1. **Correctness.** If clients and servers are honest, servers learn  $\sum_i x_i$

Extension: Maintain correctness in spite of server faults

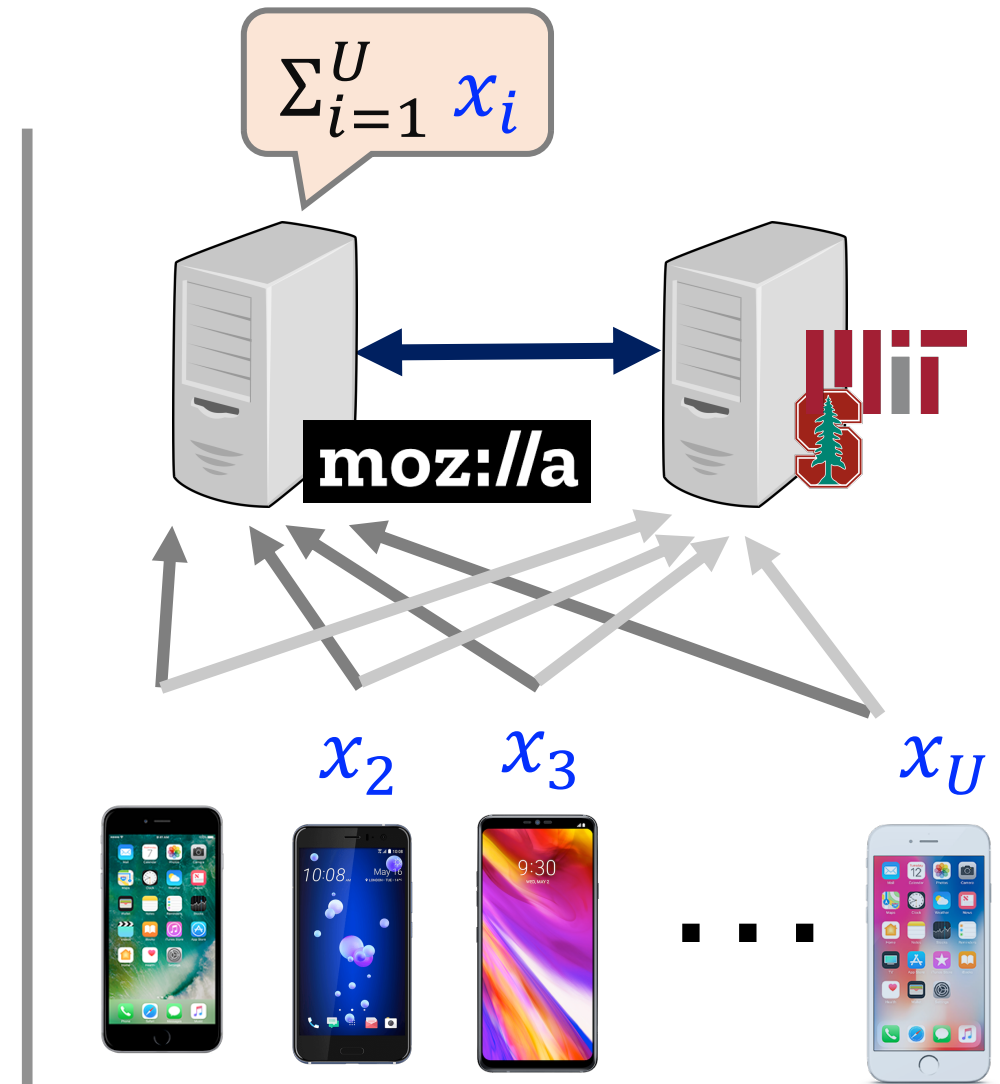
2.  **$f$ -Privacy.** Attacker must compromise all servers to learn more than  $\sum_i x_i$

Extension: Differential privacy [DMNS06]

3. **Disruption resistance.**

The worst that a malicious client can do is lie about her input.

4. **Efficiency.** Handle millions of submissions per server per hour



# Prio: System goals

1. **Correctness.** If clients and servers are honest, servers learn  $\sum_i x_i$

Extension: Maintain correctness in spite of server faults

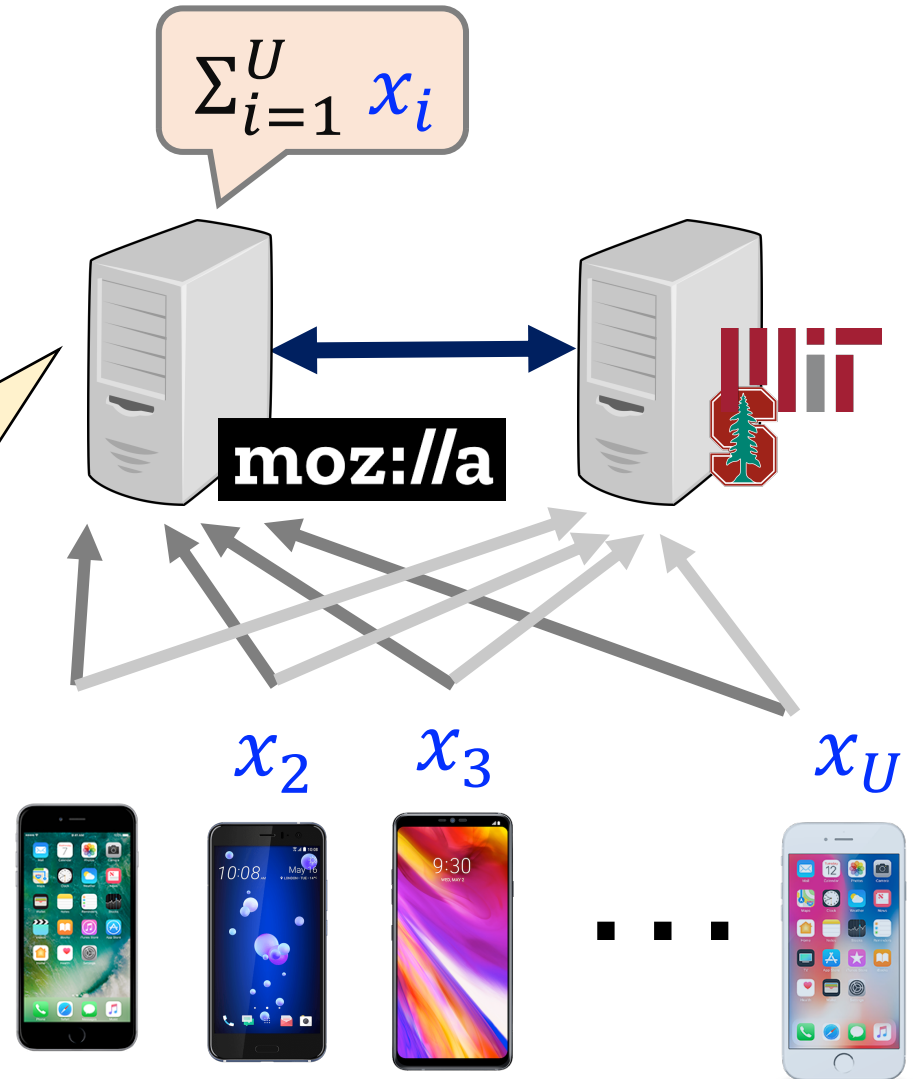
2.  **$f$ -Privacy.** An attacker must compromise all more than  $\sum_i x_i$

Extension: Different

3. **Disruption** The worst that a malicious client can do is lie about her input.

4. **Efficiency.** Handle millions of submissions per server per hour

**Attacker must compromise all servers to learn private data.**



# Prio: System goals

1. **Correctness.** If clients and servers are honest, servers learn  $\sum_i x_i$

Extension: Maintain correctness in spite of server faults

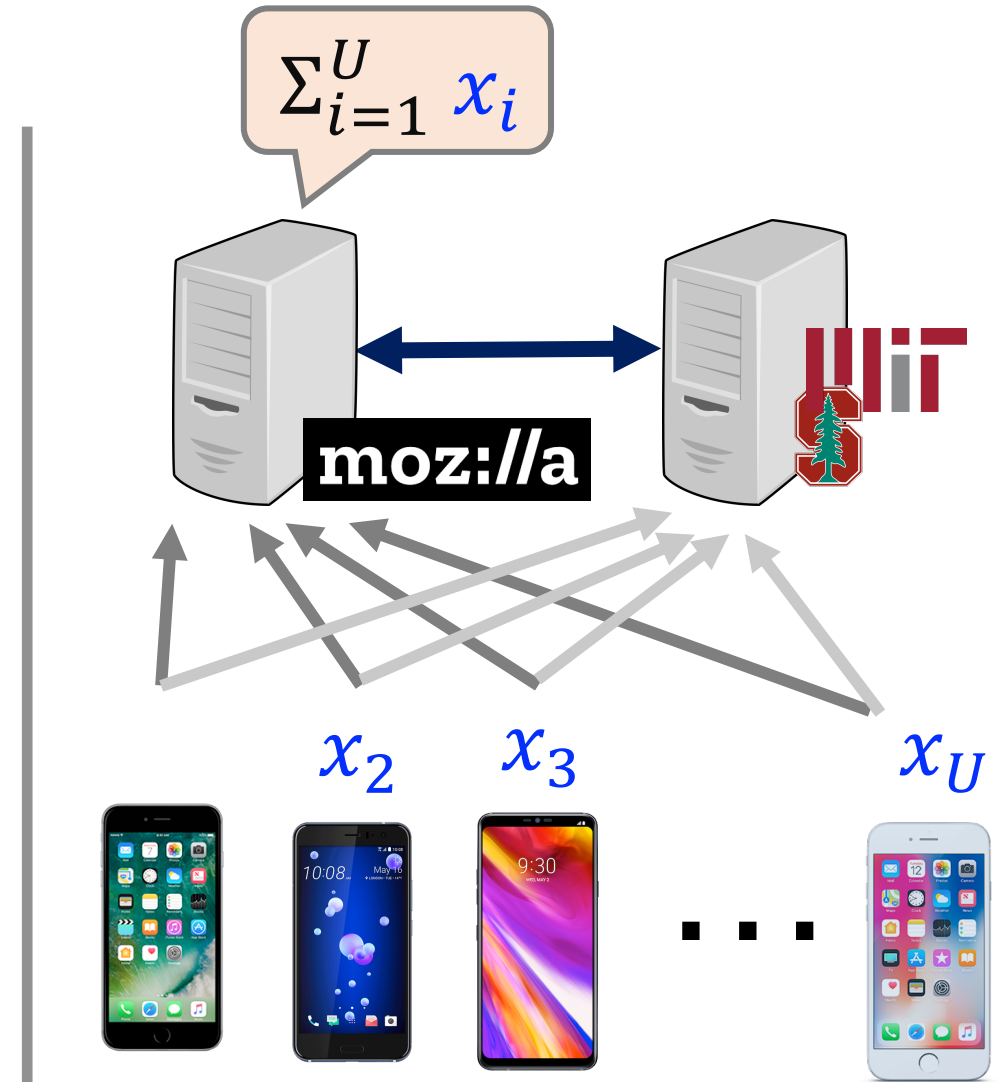
2.  **$f$ -Privacy.** Attacker must compromise all servers to learn more than  $\sum_i x_i$

Extension: Differential privacy [DMNS06]

3. **Disruption resistance.**

The worst that a malicious client can do is lie about her input.

4. **Efficiency.** Handle millions of submissions per server per hour



# Prio: System goals

**1. Correctness.** If clients and servers are honest, servers learn  $\sum_i x_i$

Extension: Maintain correctness in spite of server faults

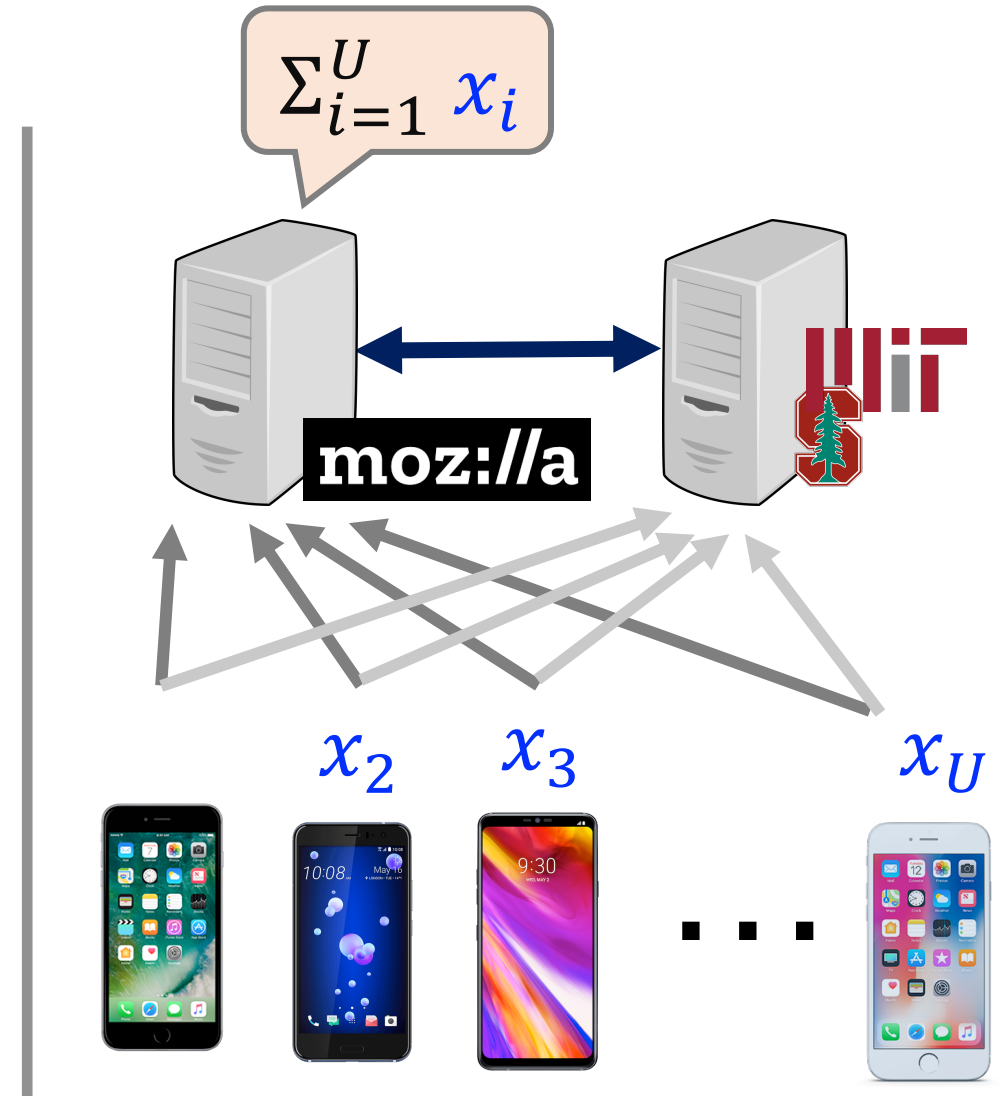
**2.  $f$ -Privacy.** Attacker must compromise all servers to learn more than  $\sum_i x_i$

Extension: Differential privacy [DMNS06]

**3. Disruption resistance.**

The worst that a malicious client can do is lie about her input.

**4. Efficiency.** Handle millions of submissions per server per hour



# Prio: System goals

**1. Correctness.** If clients and servers are honest, servers learn  $\sum_i x_i$

Extension: Maintain correctness in spite of server faults

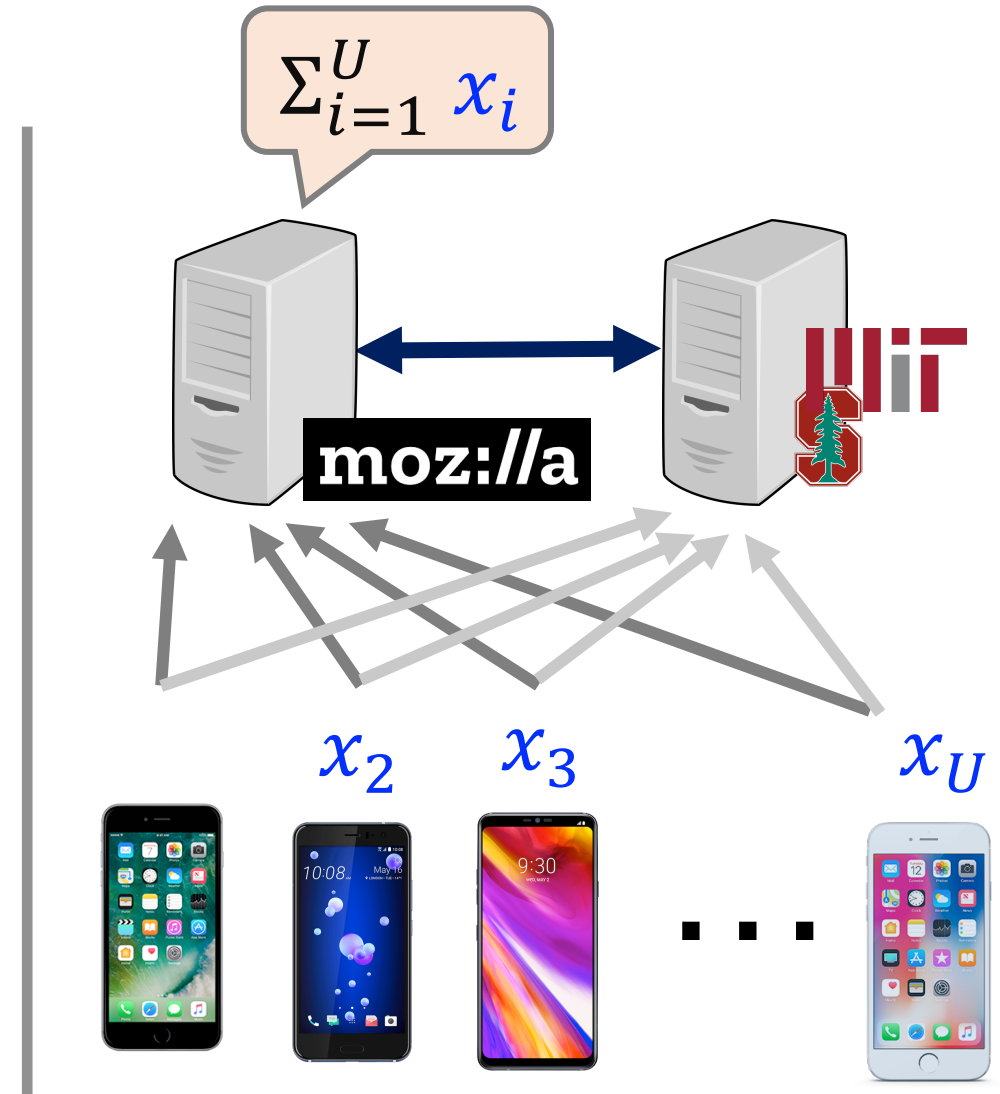
**2.  $f$ -Privacy.** Attacker must compromise all servers to learn more than  $\sum_i x_i$

Extension: Differential privacy [DMNS06]

**3. Disruption resistance.**

The worst that a malicious client can do is lie about her input.

**4. Efficiency.** Handle millions of submissions per server per hour



# Prio: System goals

**1. Correctness.** If clients and servers are honest, servers learn  $\sum_i x_i$

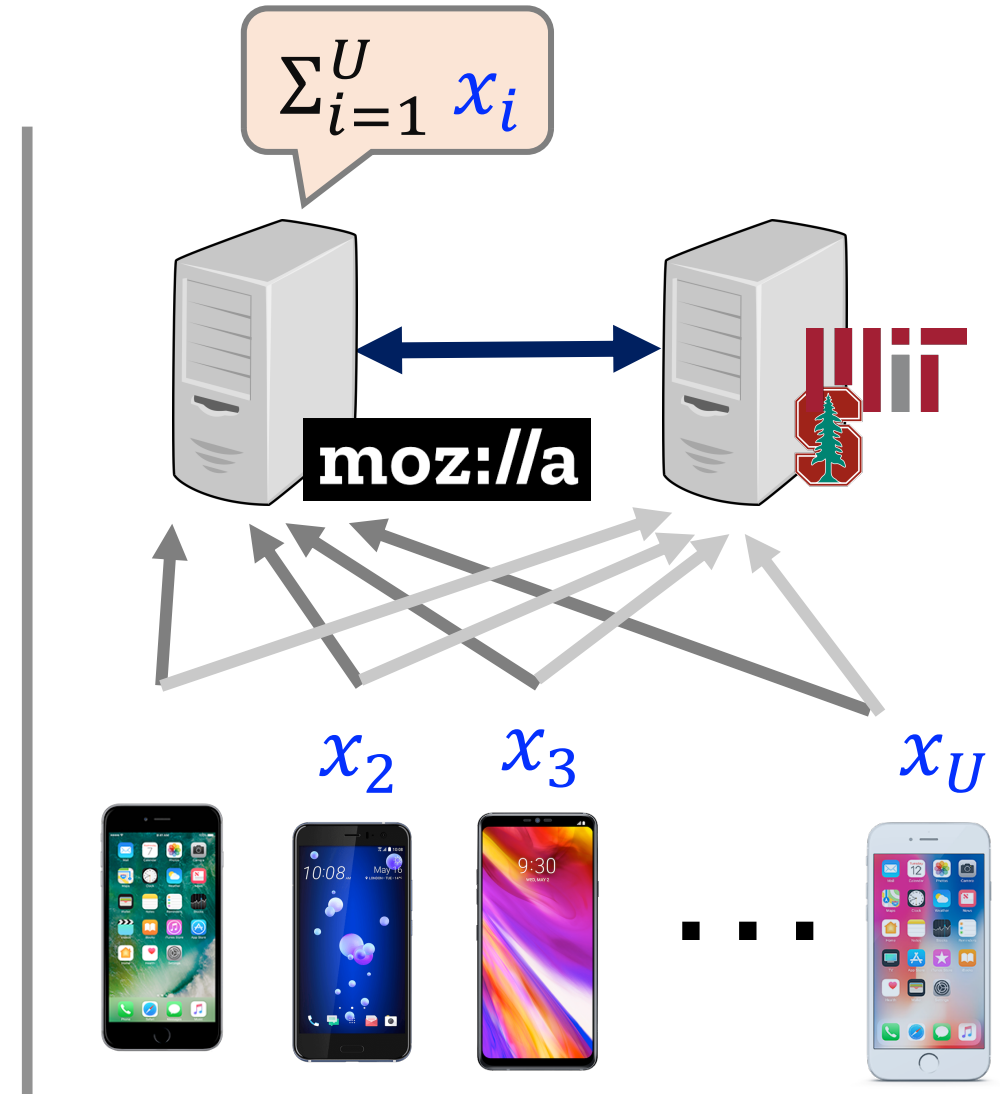
Extension: Maintain correctness in spite of server faults

**2.  $f$ -Privacy.** Attacker must compromise all servers to learn more than  $\sum_i x_i$

Extension: Differential privacy [DMNS06]

**3. Disruption resistance.** The worst that a malicious client can do is lie about her input.

**4. Efficiency.** Handle millions of submissions per server per hour



# Prio: System goals

**1. Correctness.** If clients and servers are honest, servers learn  $\sum_i x_i$

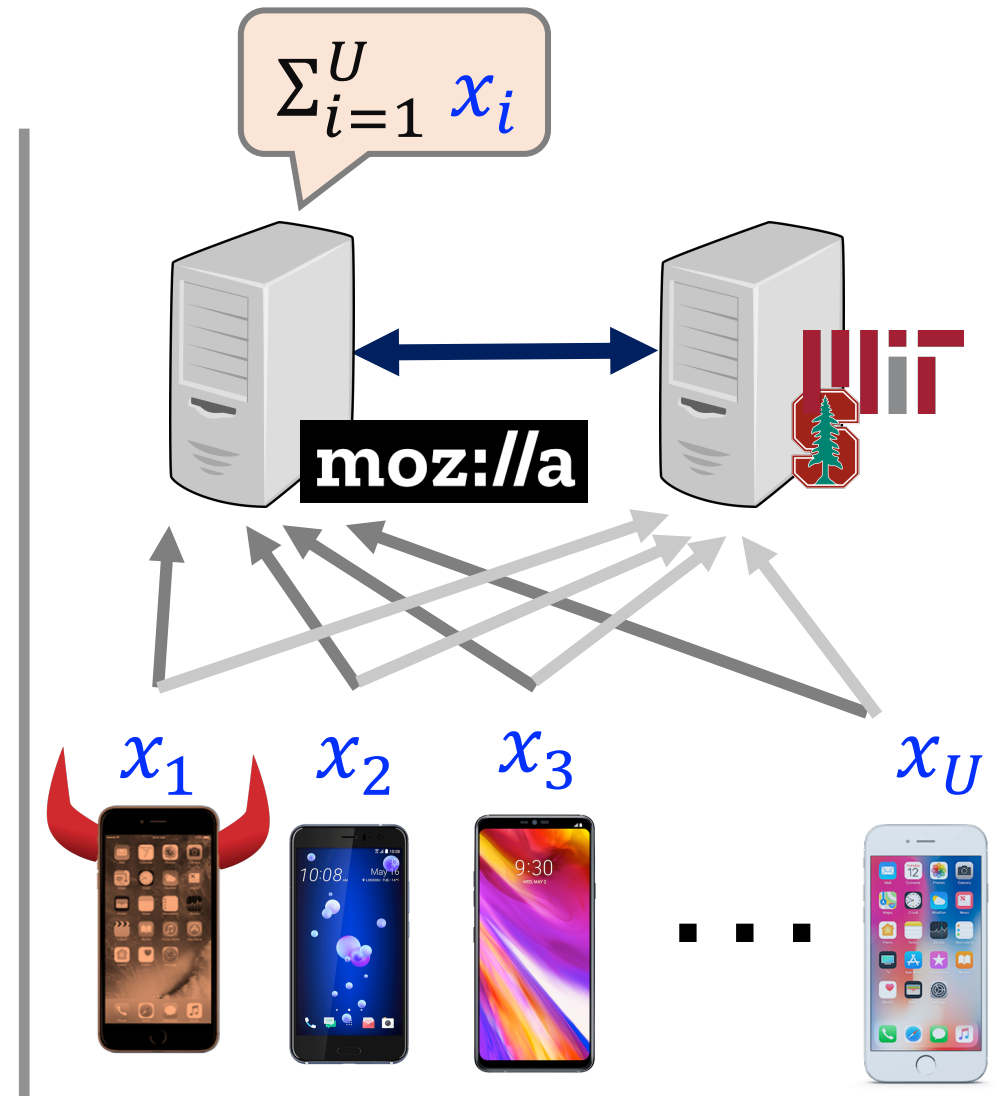
Extension: Maintain correctness in spite of server faults

**2.  $f$ -Privacy.** Attacker must compromise all servers to learn more than  $\sum_i x_i$

Extension: Differential privacy [DMNS06]

**3. Disruption resistance.** The worst that a malicious client can do is lie about her input.

**4. Efficiency.** Handle millions of submissions per server per hour





# Prio: System goals

**1. Correctness.** If clients and servers are honest, servers learn  $\sum_i x_i$

Extension: Maintain correctness in spite of server faults

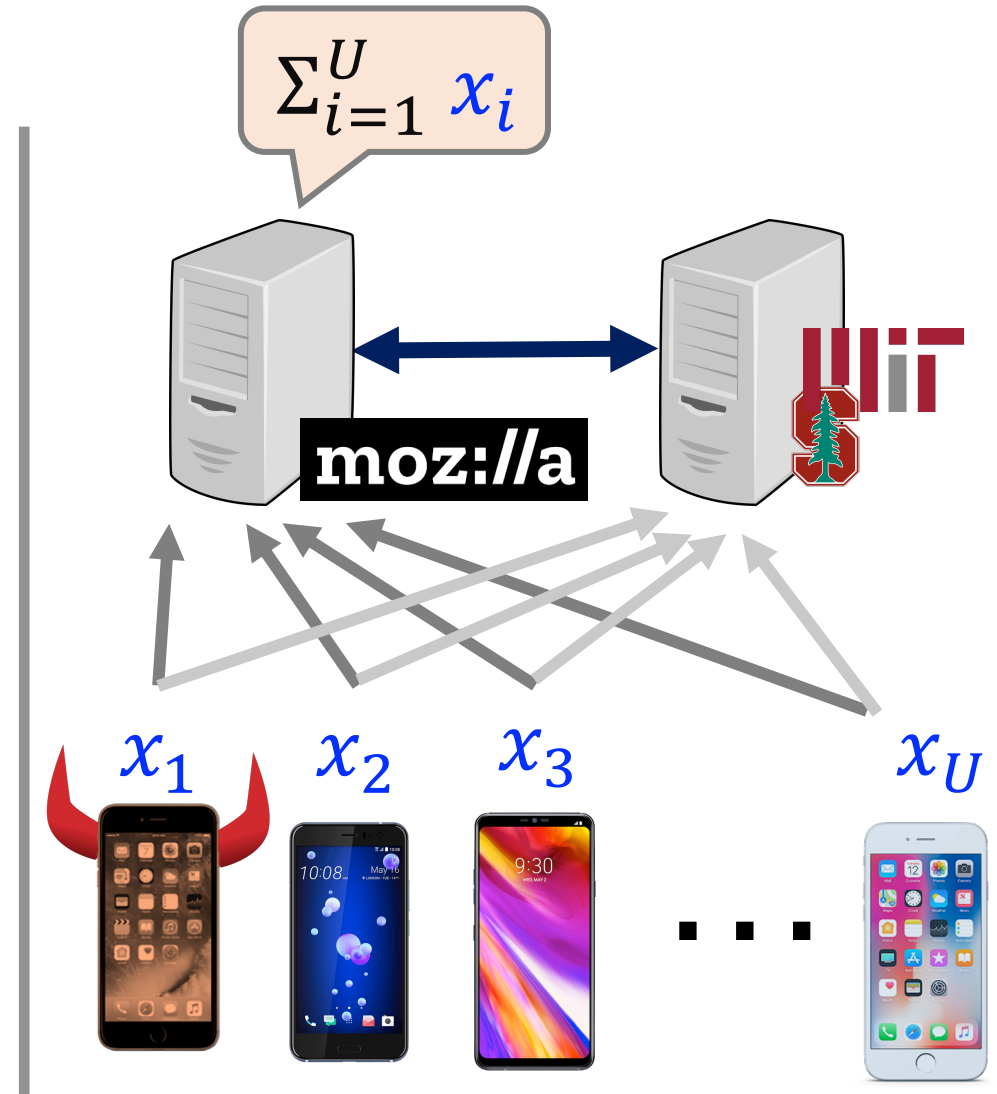
**2.  $f$ -Privacy.** Attacker must compromise all servers to learn more than  $\sum_i x_i$

Extension: Differential privacy [DMNS06]

**3. Disruption resistance.**

The worst that a malicious client can do is lie about her input.

**4. Efficiency.** Handle millions of submissions per server per hour



## Relax correctness

Randomized response: [W65], [DMNS06], [DJW13], [BS15]  
RAPPOR (2014, 2016), Wang et al. (2017),  
Ding et al. (2017)...

## Relax privacy model

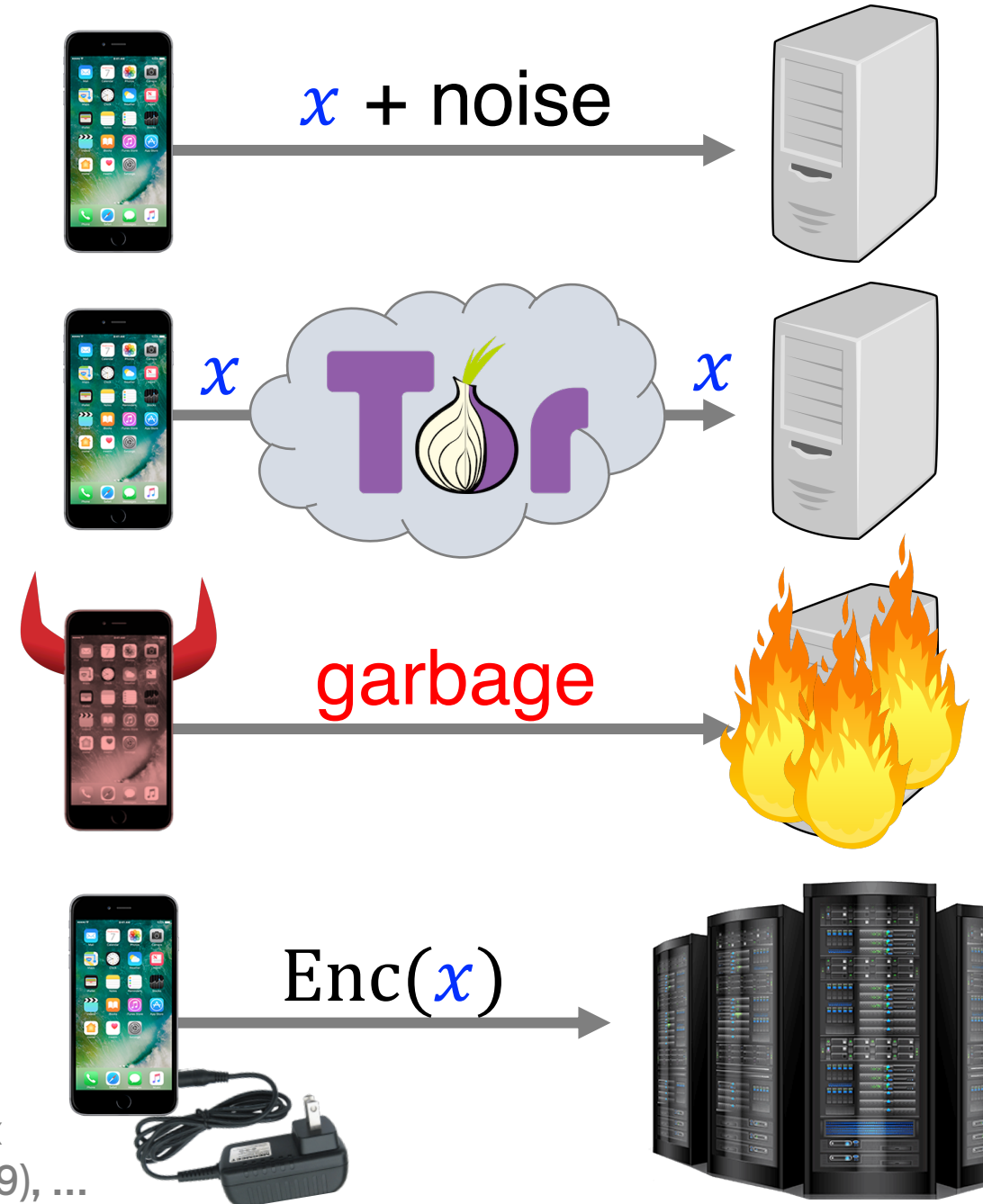
Tor: PrivStats (2011), ANONIZE (2014), ...  
SGX: Prochlo (2017), SGX-BigMatrix (2017), ...  
Honest but curious: PDDP (2012), SplitX (2013), ...

## Relax disruption resistance

Private metering (2011), PrivEx-S2 (2014),  
PrivCount (2016), Federated ML (2016, 2017), ...

## Relax efficiency

P4P (2010), Grid aggregation (2011), Haze (2013),  
PrivEx-D2 (2014), Succinct sketches (2016), HisTor $\epsilon$  (2017), ...  
General MPC [GMW87], [BGW88]: FairPlay (2004), FairplayMP  
(2008), SEPIA (2010), Private matrix factorization (2013), UnLynx  
(2017), Private ridge regression (2018), Shuffle model (2017, 2019), ...

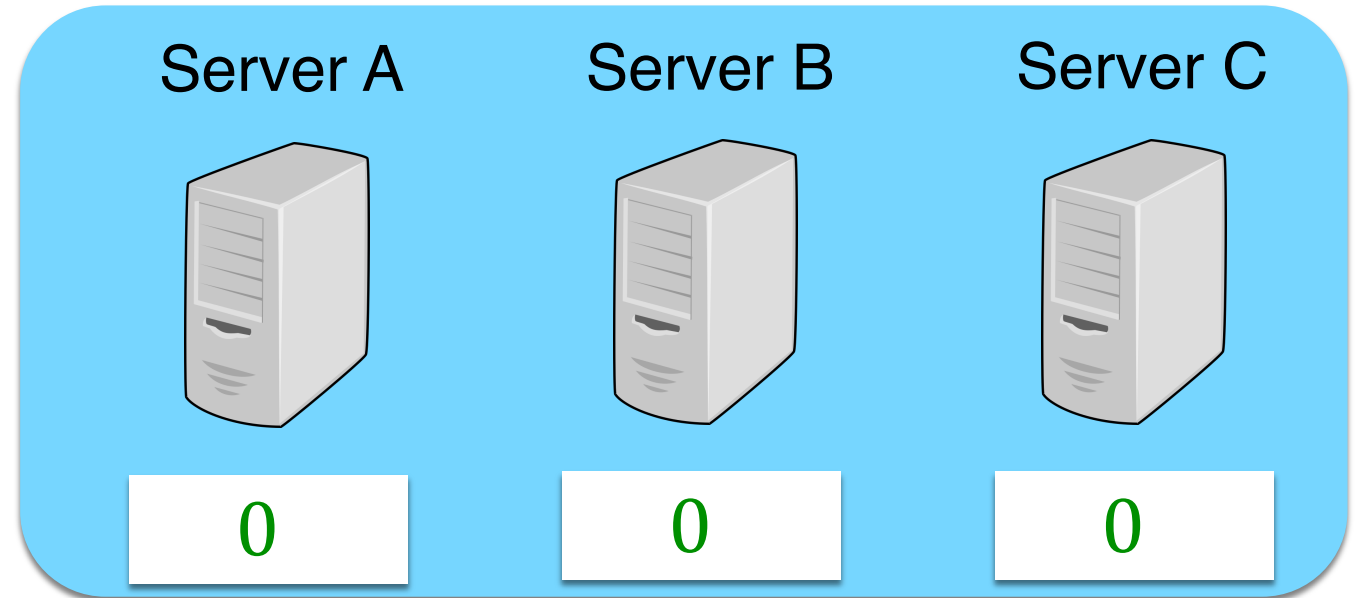


# Straw-man scheme

## Private sums without disruption resistance

[C88], [BGW88], ...

[KDK11] [DFKZ13] [PrivEx14] ...



$$x_1 = 1$$

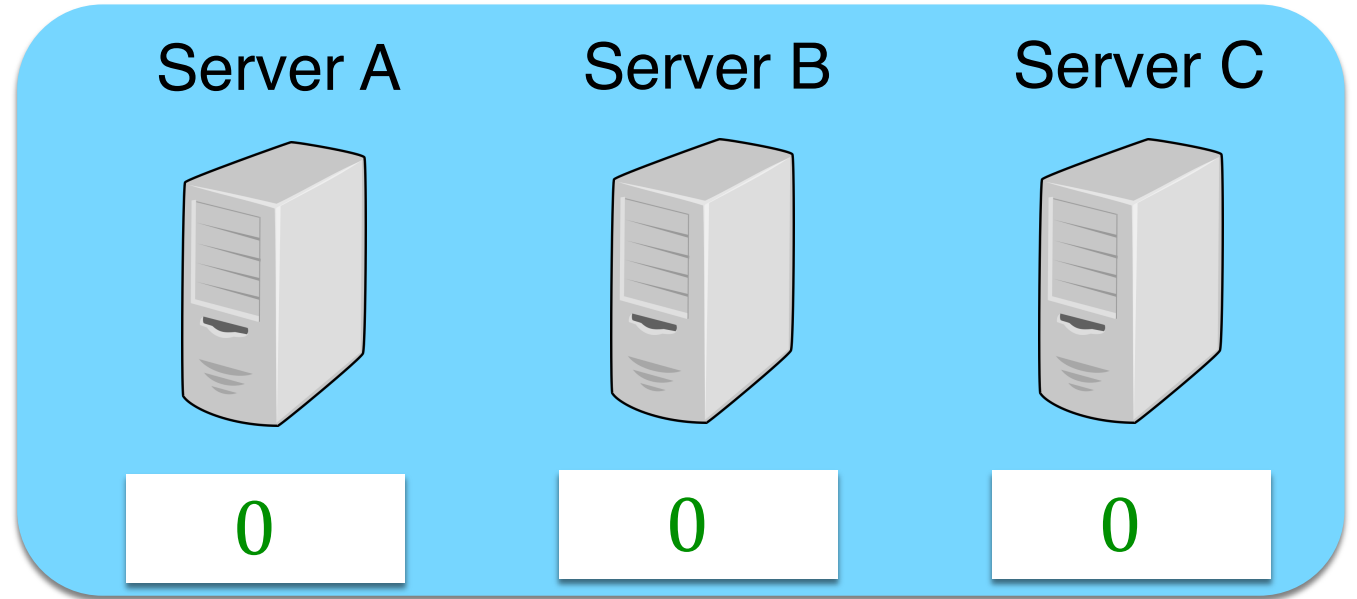


# Straw-man scheme

Private sums without  
disruption resistance

[C88], [BGW88], ...

[KDK11] [DFKZ13] [PrivEx14] ...



$$x_1 = 1$$



Pick three random “shares” that sum to  $x_1 = 1$ .

$$1 = 15 + (-12) + (-2) \quad (\text{mod } p)$$

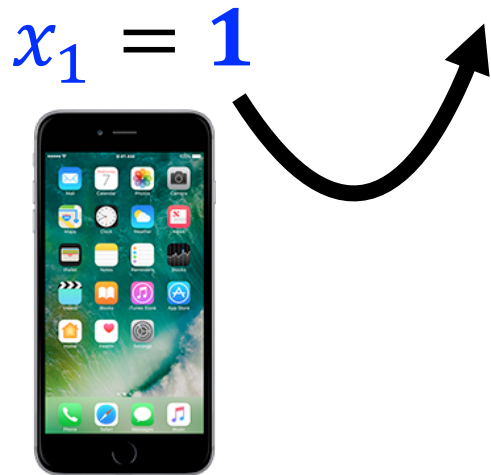
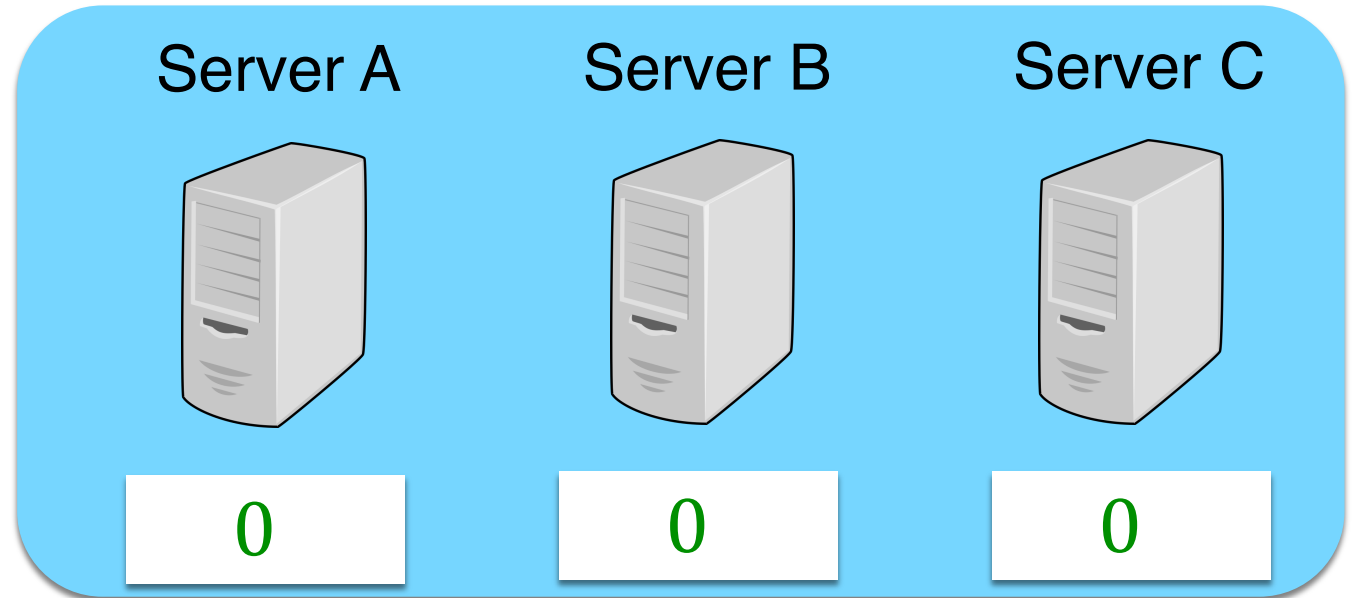
Send one share to each server.

# Straw-man scheme

Private sums without  
disruption resistance

[C88], [BGW88], ...

[KDK11] [DFKZ13] [PrivEx14] ...



Pick three random “shares” that sum to  $x_1 = 1$ .

$$1 = 15 + (-12) + (-2) \quad (\text{mod } p)$$

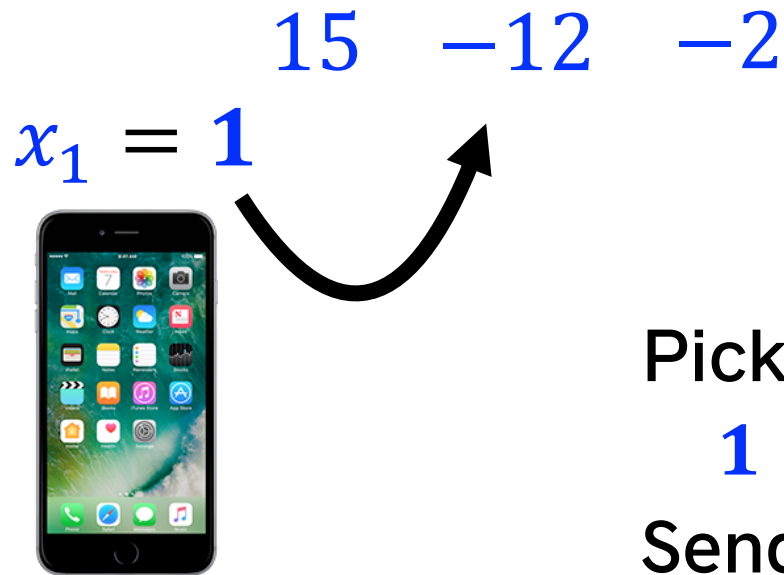
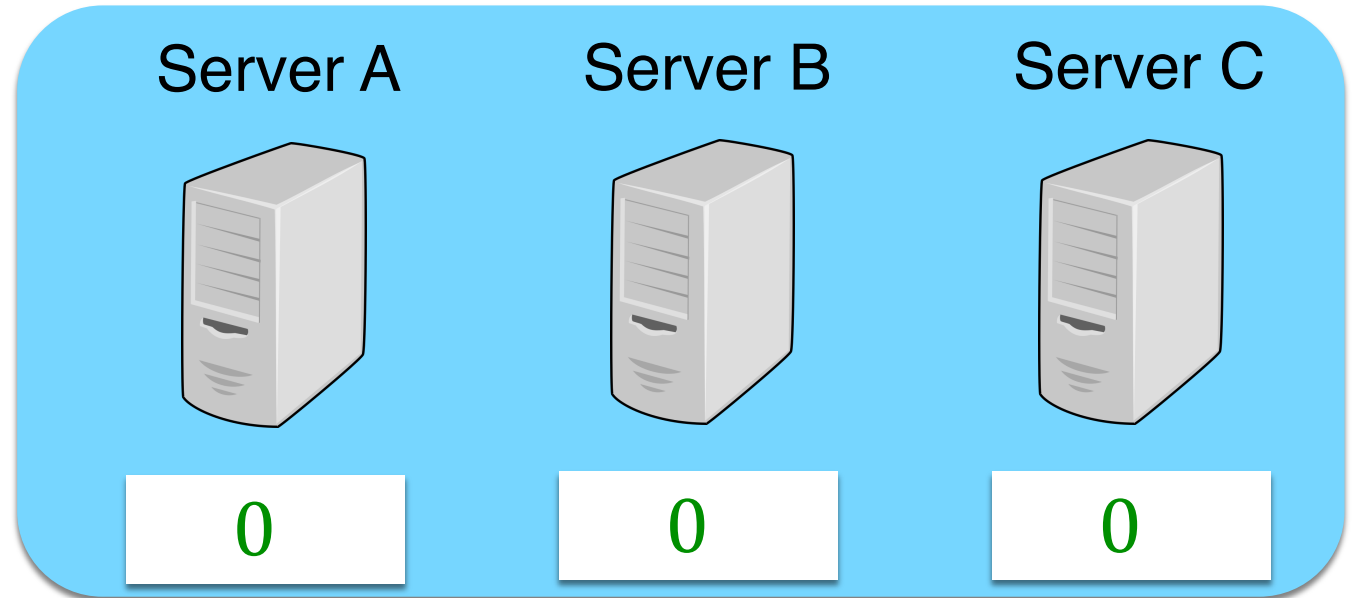
Send one share to each server.

# Straw-man scheme

Private sums without  
disruption resistance

[C88], [BGW88], ...

[KDK11] [DFKZ13] [PrivEx14] ...



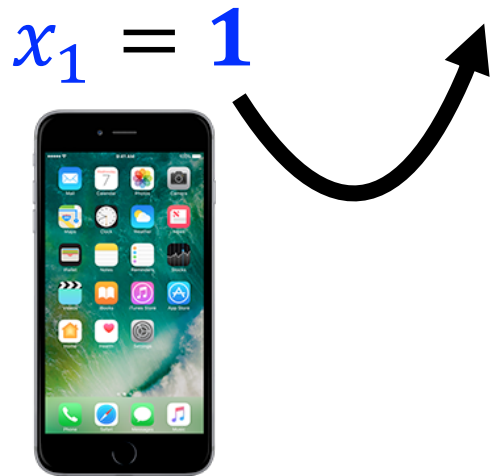
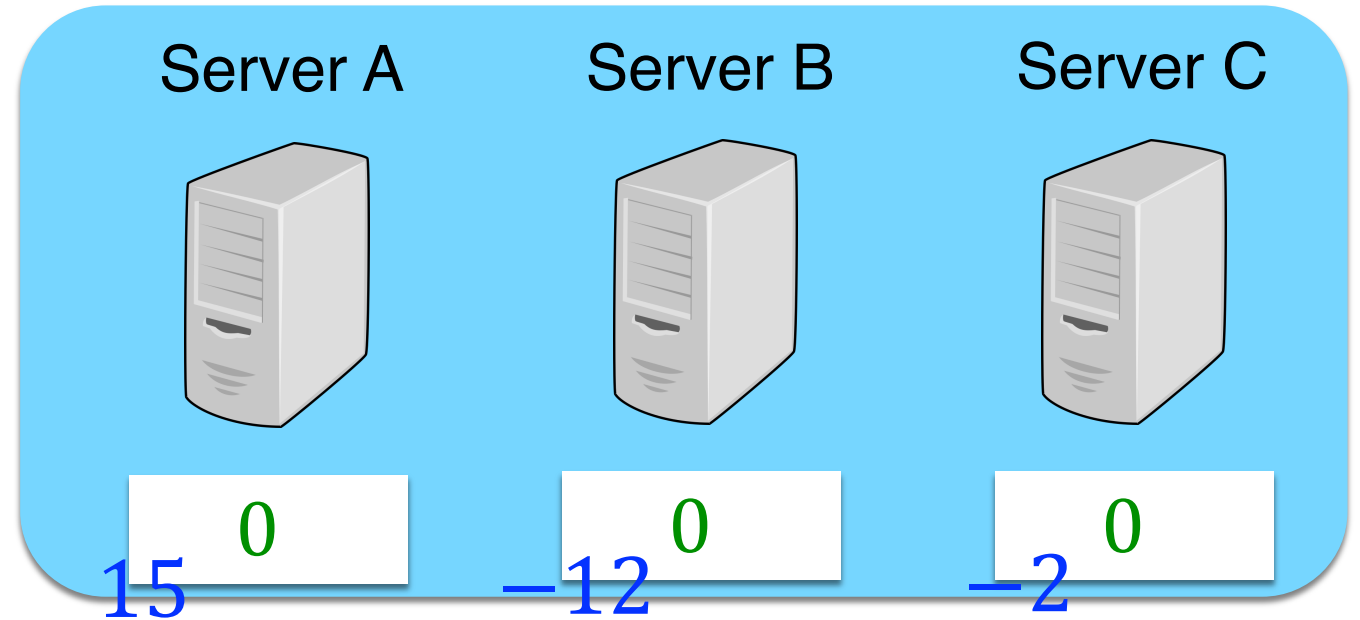
Pick three random “shares” that sum to  $x_1 = 1$ .

$$1 = 15 + (-12) + (-2) \quad (\text{mod } p)$$

Send one share to each server.

# Straw-man scheme

Private sums without  
disruption resistance



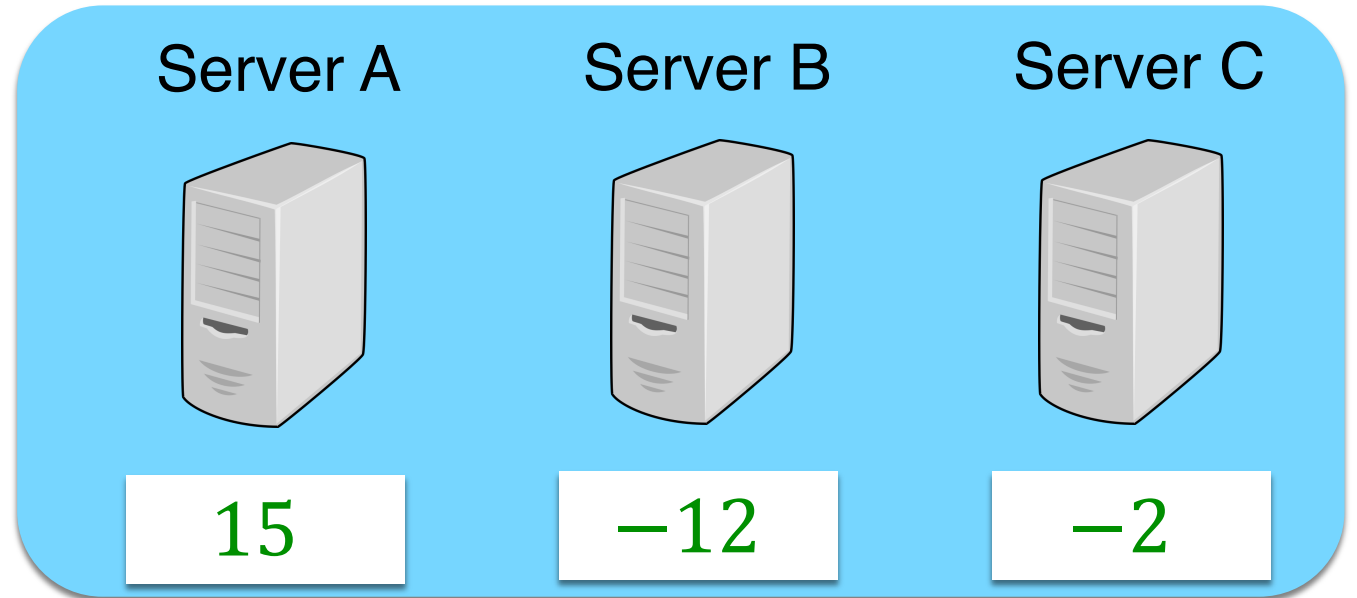
Pick three random “shares” that sum to  $x_1 = 1$ .

$$1 = 15 + (-12) + (-2) \quad (\text{mod } p)$$

Send one share to each server.

# Straw-man scheme

Private sums without  
disruption resistance



$$x_1 = 1$$



Pick three random “shares” that sum to  $x_1 = 1$ .

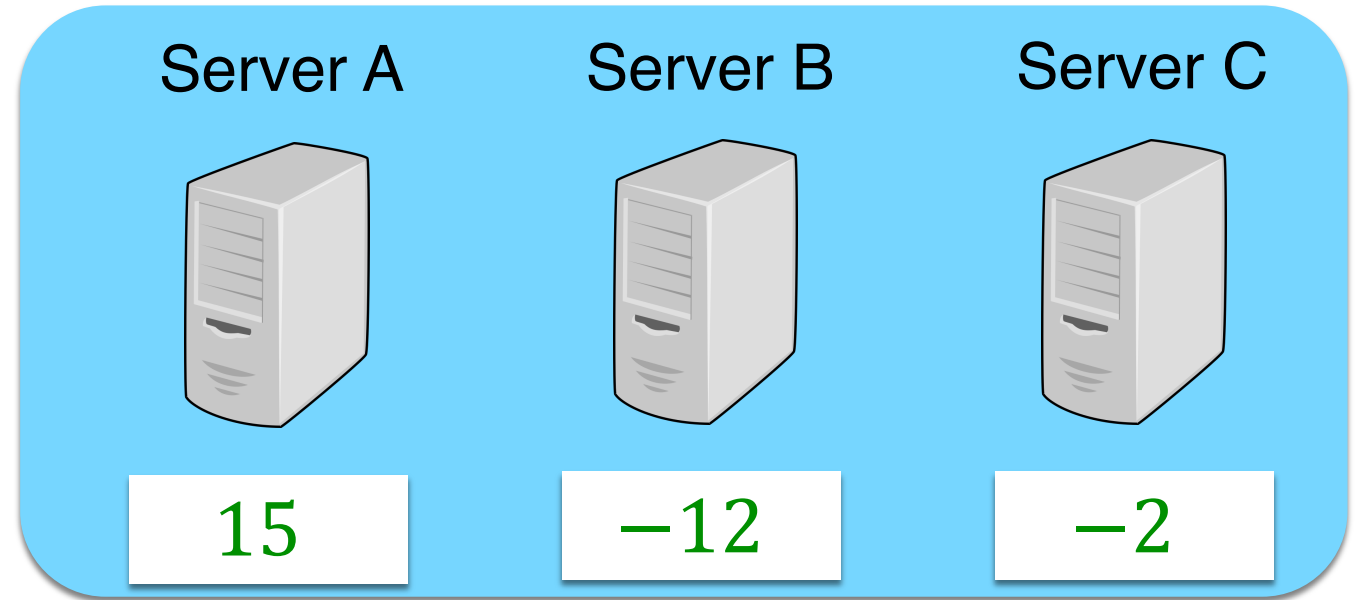
$$1 = 15 + (-12) + (-2) \quad (\text{mod } p)$$

Send one share to each server.



# Straw-man scheme

Private sums without  
disruption resistance

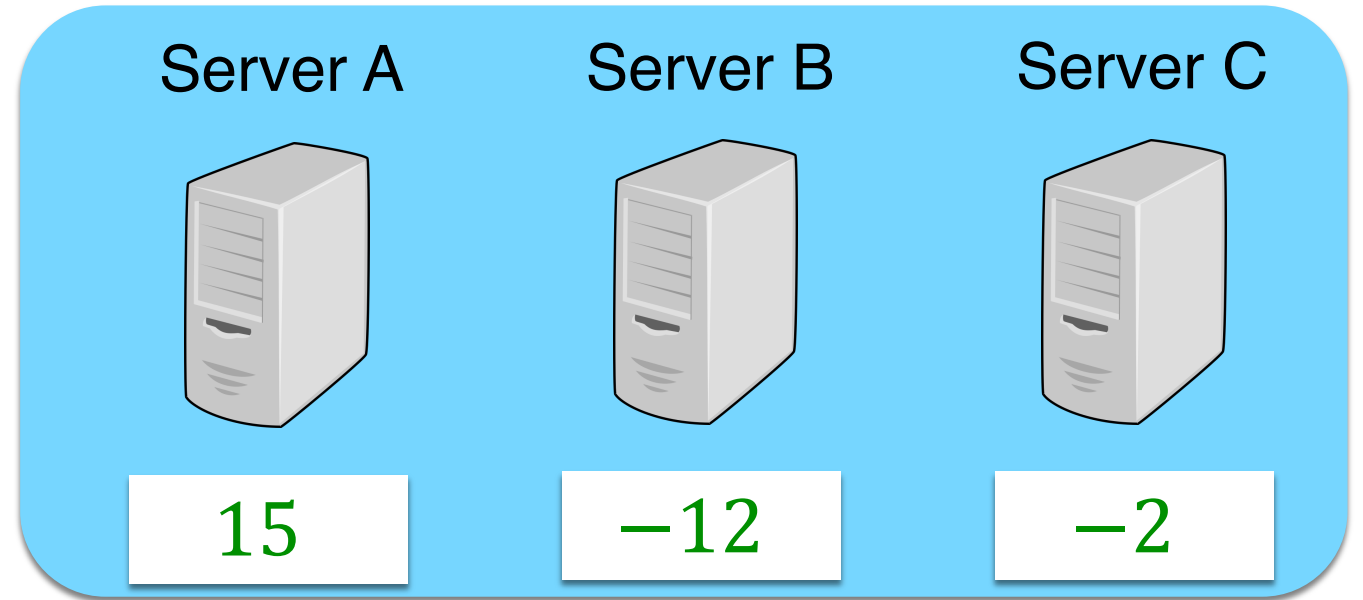


$$x_2 = 0$$



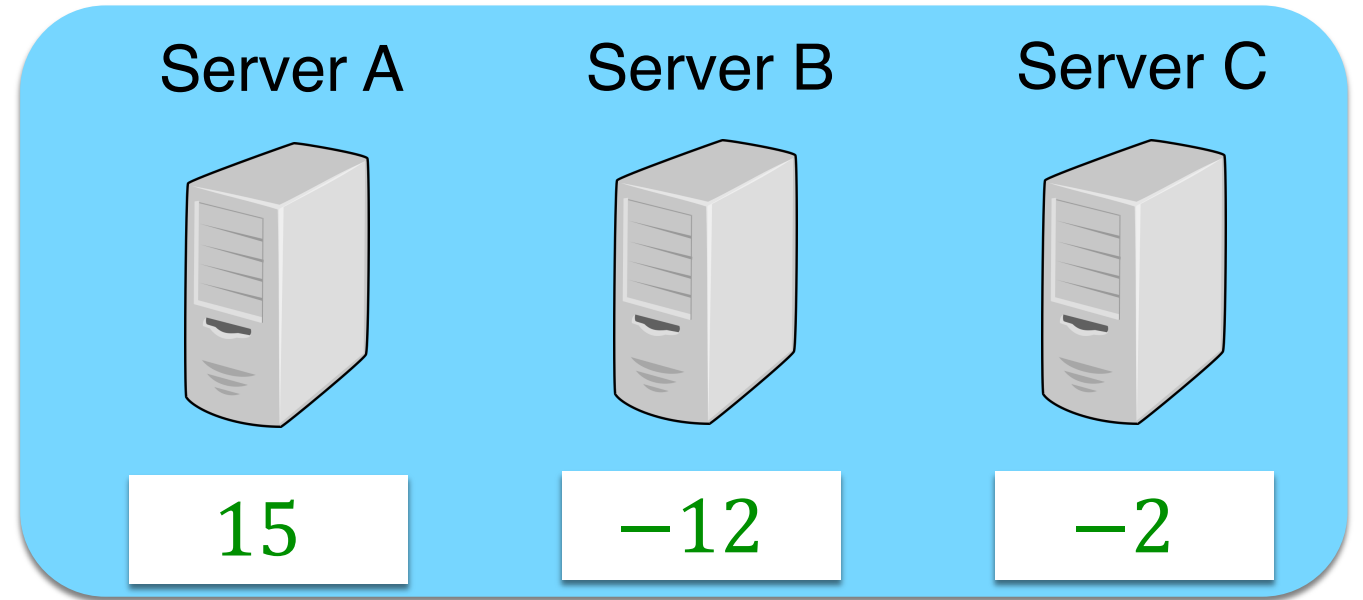
# Straw-man scheme

Private sums without  
disruption resistance

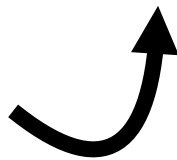


# Straw-man scheme

Private sums without  
disruption resistance

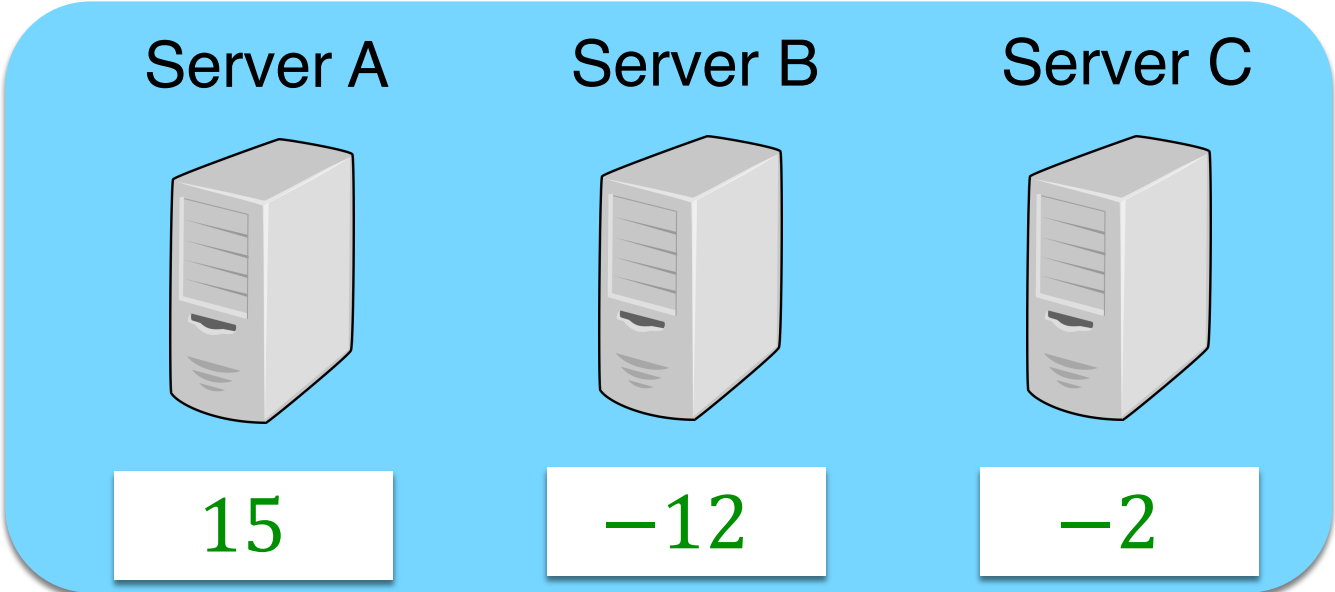


$$x_2 = 0 = (-10) + 7 + 3$$

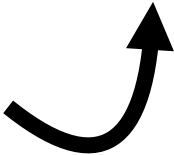


# Straw-man scheme

Private sums without  
disruption resistance

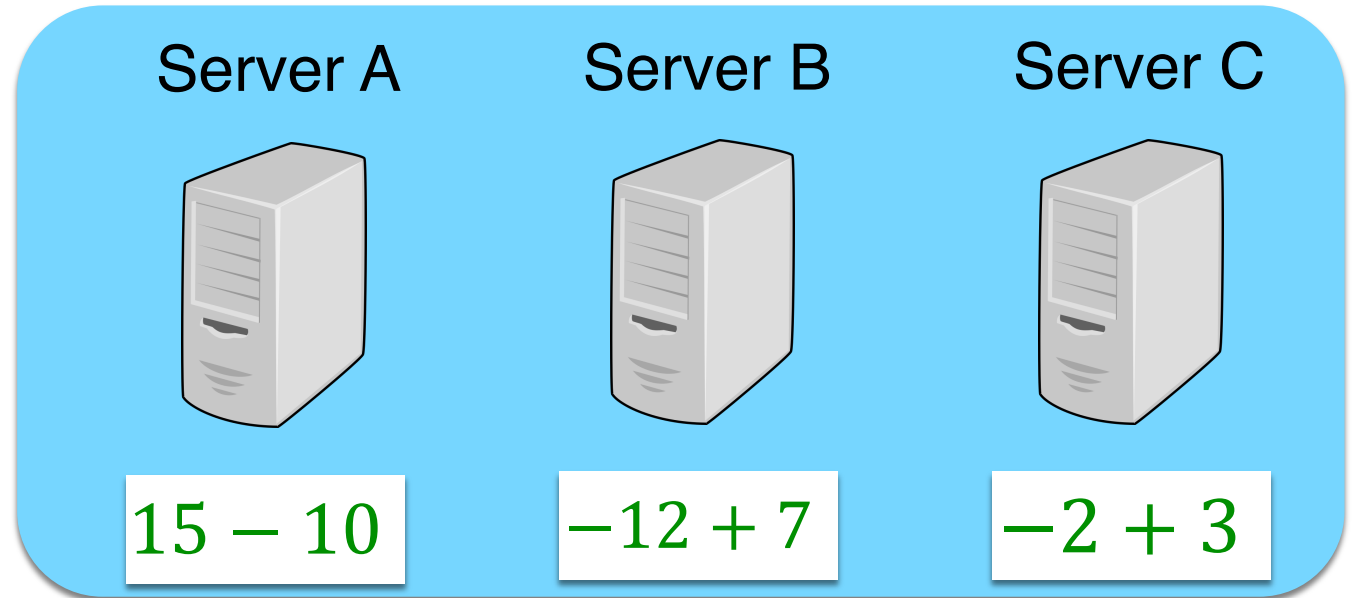


$x_2 = 0$       -10      7      3



# Straw-man scheme

Private sums without  
disruption resistance

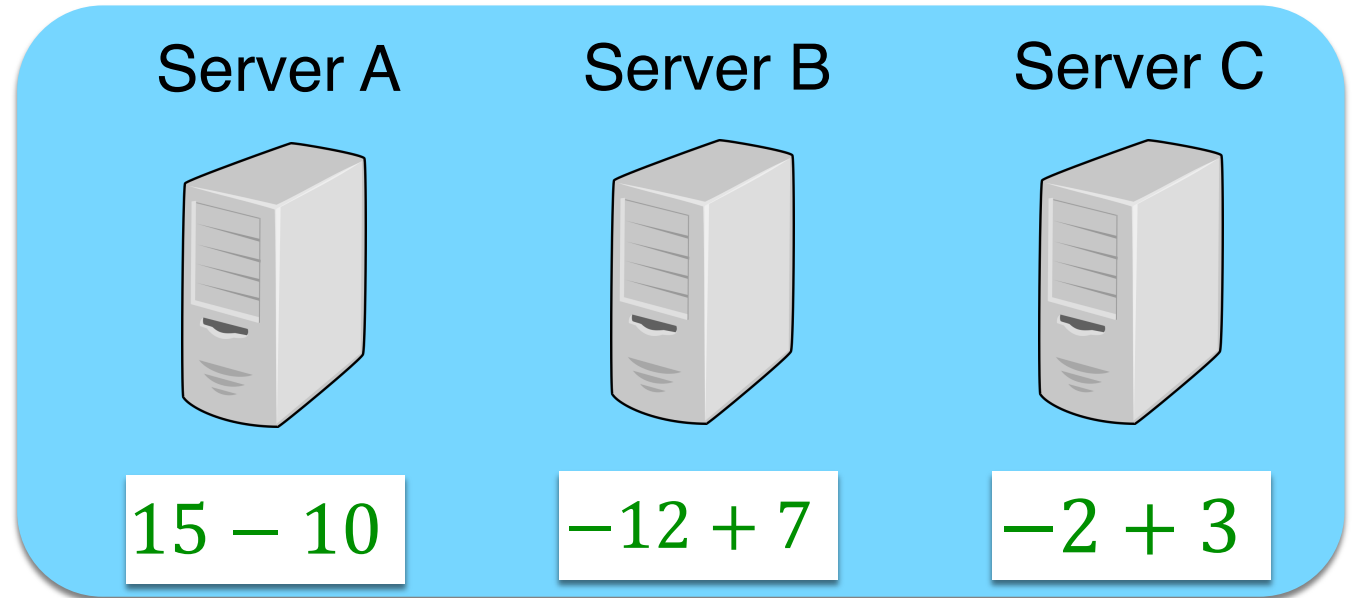


$$x_2 = 0$$



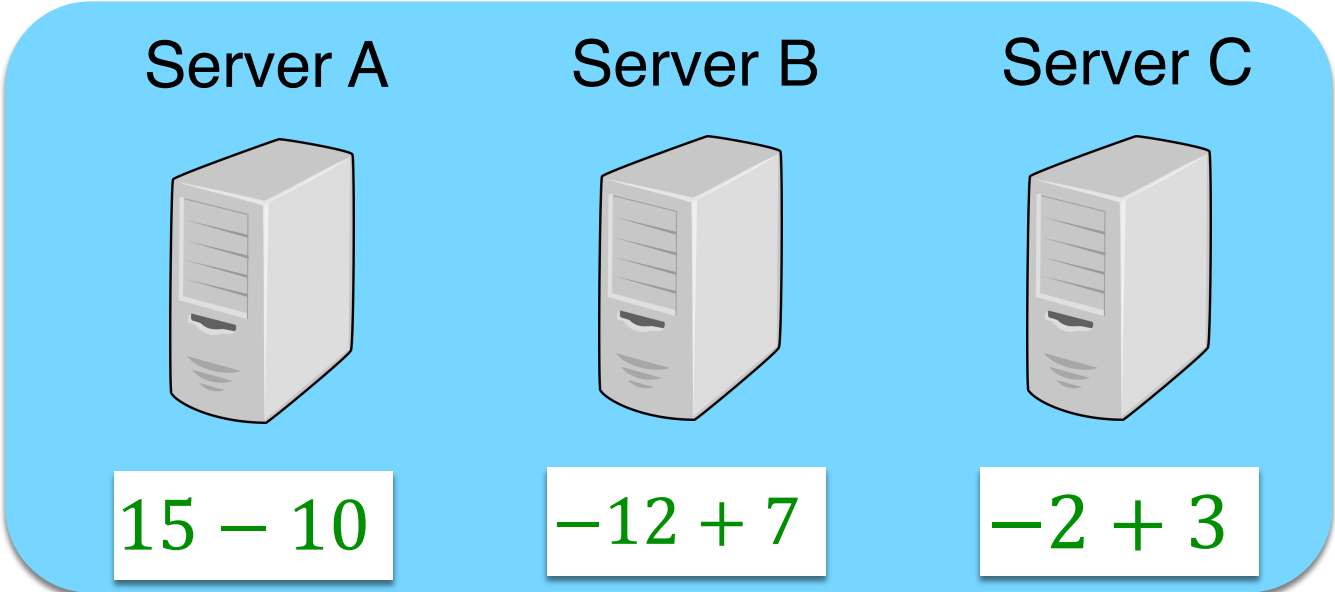
# Straw-man scheme

Private sums without  
disruption resistance



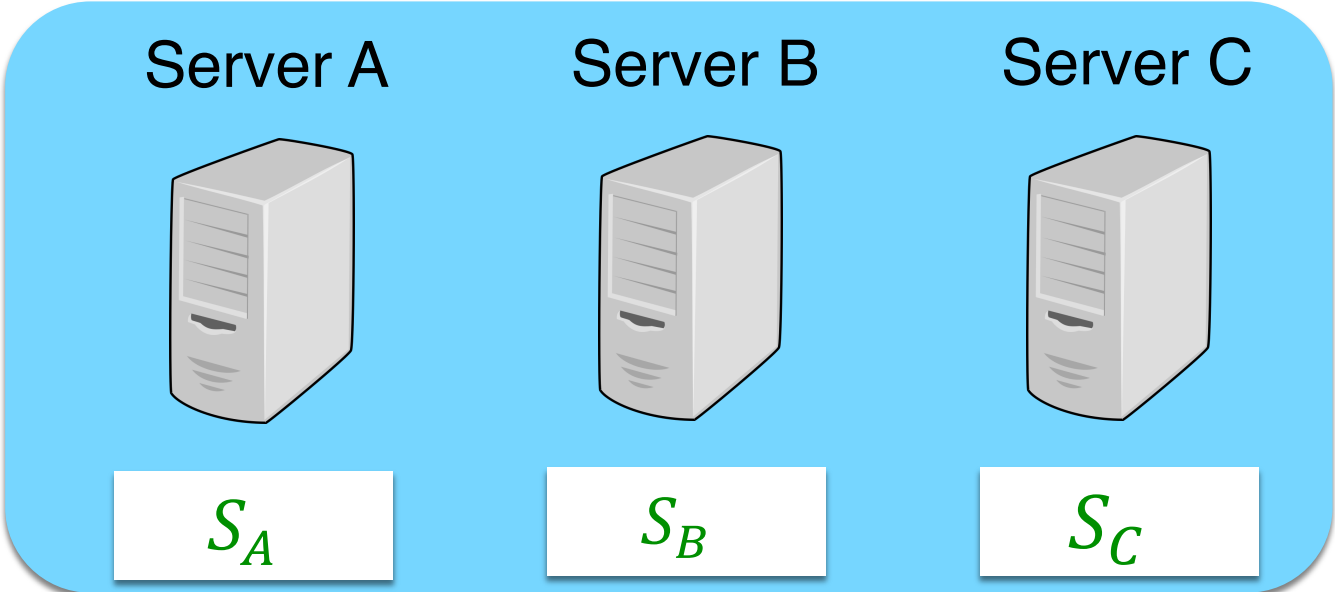
# Straw-man scheme

Private sums without  
disruption resistance



# Straw-man scheme

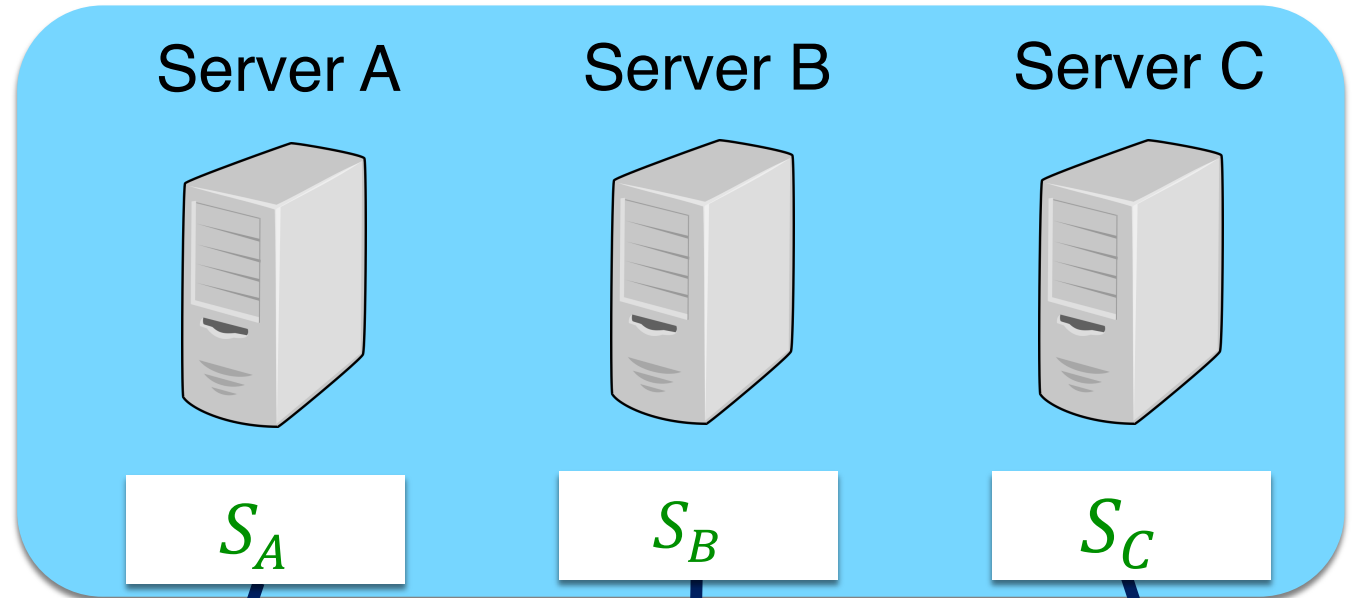
Private sums without  
disruption resistance





# Straw-man scheme

Private sums without  
disruption resistance

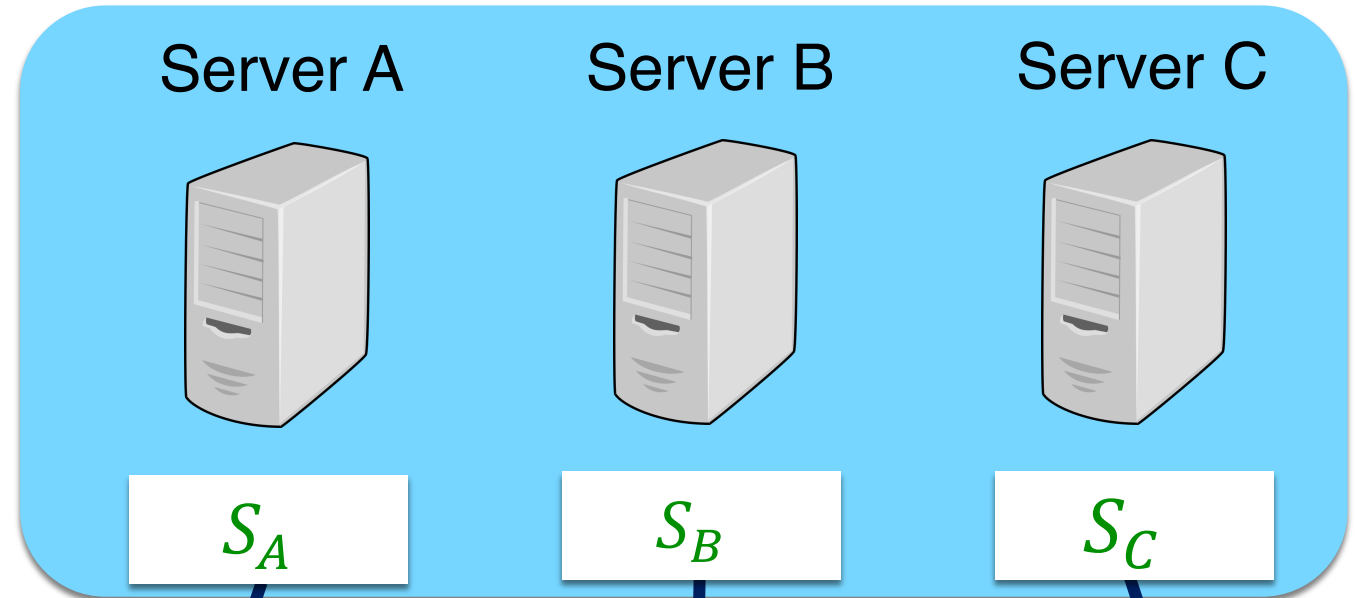


$$\begin{aligned} & (15 - 10 + \dots) + (-12 + 7 + \dots) + (-2 + 3 + \dots) \\ &= x_1 + x_2 + x_3 + \dots \end{aligned}$$

Servers learn the sum of the clients' values and nothing else.

# Straw-man scheme

Private sums without  
disruption resistance



$$(15 - 10 + \dots) + (-12 + 7 + \dots) + (-2 + 3 + \dots) \\ = x_1 + x_2 + x_3 + \dots$$

e.g., learn that 58,329 users  
blocked trackers from fb.com...  
don't learn which users did

Servers learn the sum of the  
clients' values and nothing else.

# Private sums: Straw-man scheme



**Correctness.**

Servers learn the sum of the  $x_i$ s



**$f$ -Privacy.**

Attacker must compromise all servers to learn more than sum of  $x_i$ s



**Efficiency.**

No heavy cryptographic operations

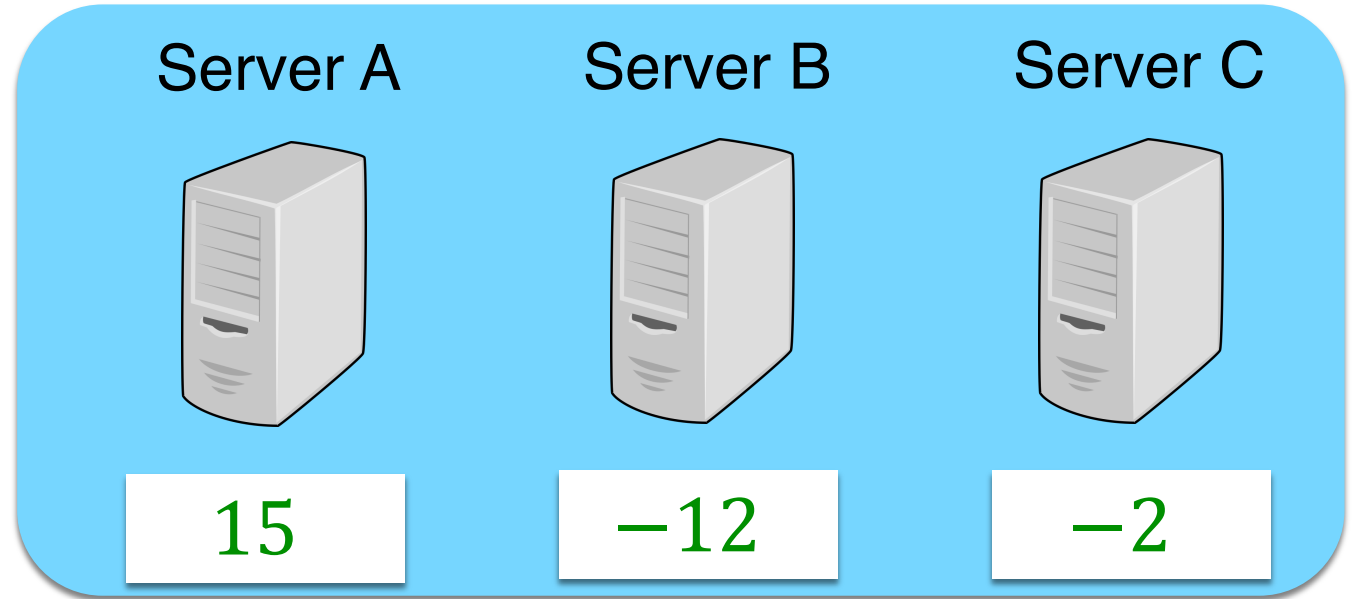


**Disruption  
resistance.**

**One malicious client can  
corrupt the output.**

# Straw-man scheme

One malicious client can corrupt output



$$x_2 = -53$$

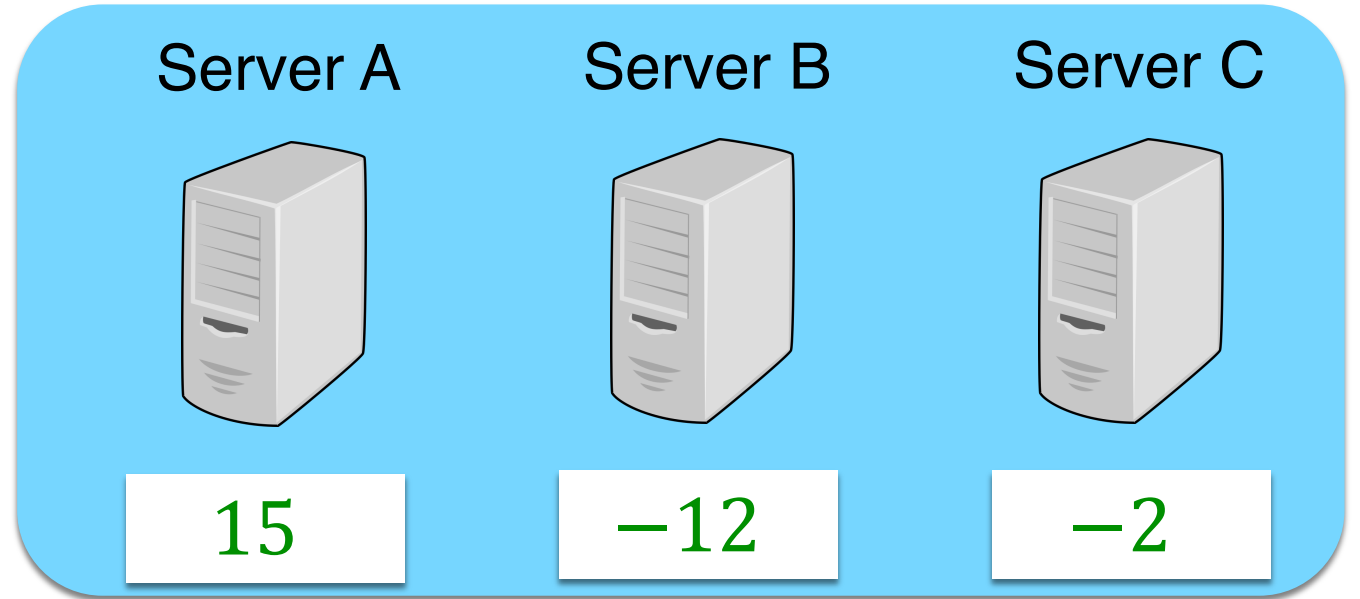


Should be a value in the set  $\{0,1\}$

Evil ad network

# Straw-man scheme

One malicious client can corrupt output



$$x_2 = -53 = -19 + -16 + -18$$

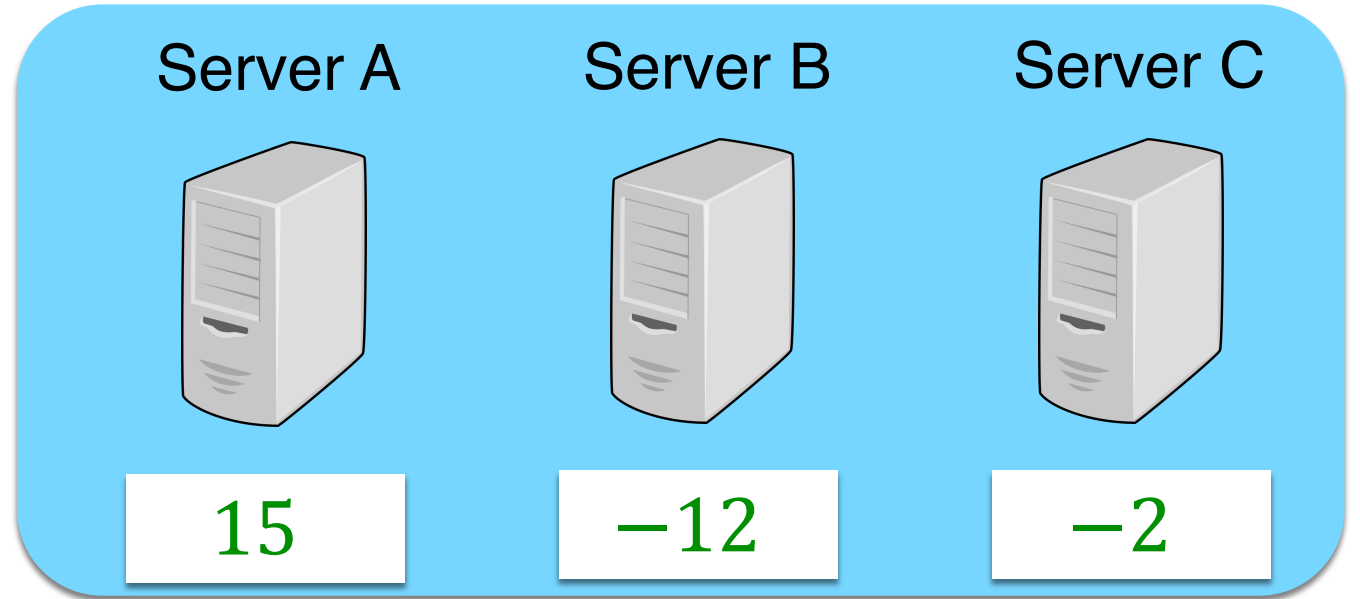


Should be a value in the set {0,1}

Evil ad network

# Straw-man scheme

One malicious client can corrupt output



$$x_2 = -53 \quad -19 \quad -16 \quad -18$$

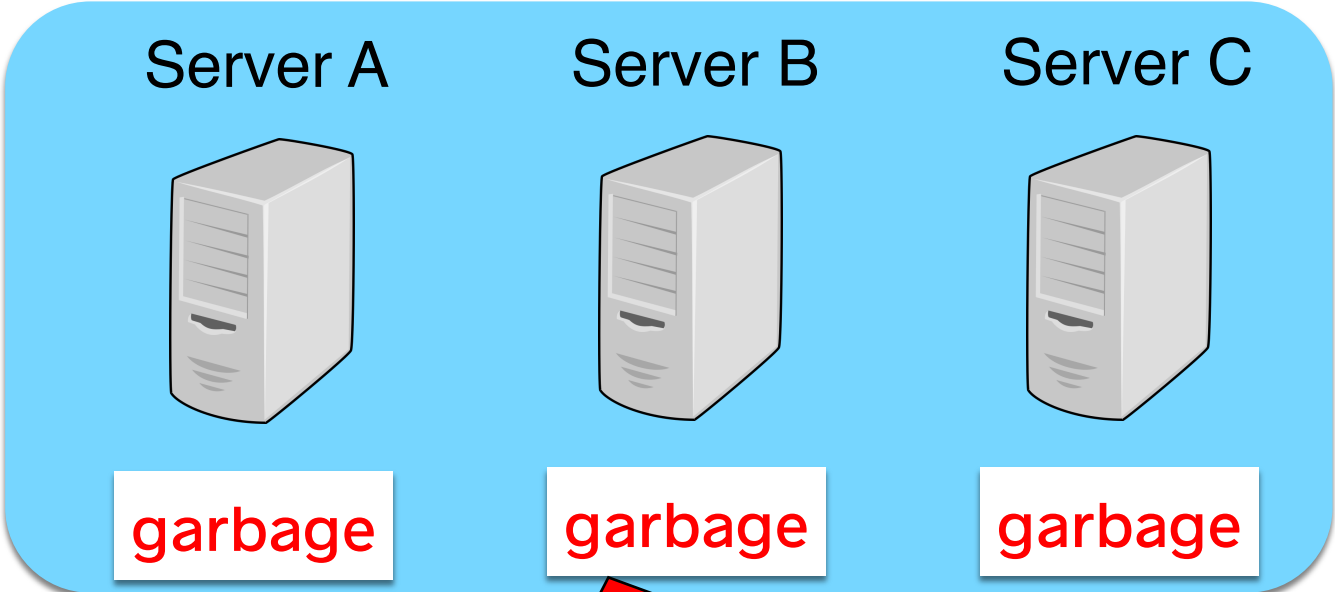


Should be a value in the set  $\{0,1\}$

Evil ad network

# Straw-man scheme

One malicious client can corrupt output



One malicious client can corrupt the output.



Evil ad network

# Powerful but costly tools...



## Multiparty computation

[GMW87], [BGW88]



# Powerful but costly tools...



**Multiparty  
computation**

[GMW87], [BGW88]



**Traditional  
zero-knowledge  
proofs**

[GMR89]

# Powerful but costly tools...



**Multiparty  
computation**

[GMW87], [BGW88]



**Traditional  
zero-knowledge  
proofs**

[GMR89]



**New tool: Proof on  
secret-shared data**

# Techniques for providing disruption resistance

Testing that a length- $n$  vector (e.g., data for  $n$  trackers) consists of secret-shared 0/1 integers.

	Public-key ops.		Communication		Slow-down
	Client	Server	C-to-S	S-to-S	
Dishonest-maj. MPC	0	$\tilde{\Theta}(n)$	0	$\tilde{\Theta}(n)$	5,000× at server
GGPR-style zkSNARK	$\tilde{\Theta}(n)$	$\tilde{O}(1)$	$\tilde{O}(1)$	$\tilde{O}(1)$	500× at client
Discrete-log-based NIZK	$\tilde{\Theta}(n)$	$\tilde{\Theta}(n)$	$\tilde{\Theta}(n)$	$\tilde{\Theta}(n)$	50× at server

(Table hides log factors.)

# Techniques for providing disruption resistance

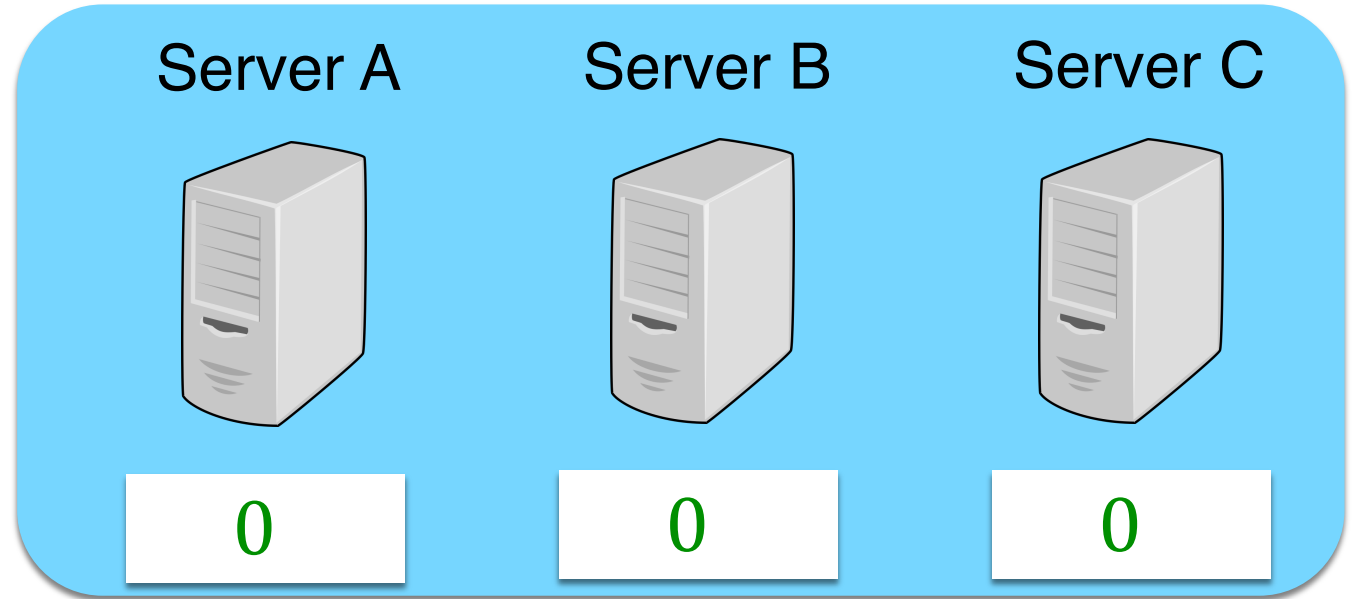
Testing that a length- $n$  vector (e.g., data for  $n$  trackers) consists of secret-shared 0/1 integers.

	Public-key ops.		Communication		Slow-down
	Client	Server	C-to-S	S-to-S	
Dishonest-maj. MPC	0	$\tilde{\Theta}(n)$	0	$\tilde{\Theta}(n)$	5,000× at server
GGPR-style zkSNARK	$\tilde{\Theta}(n)$	$\tilde{O}(1)$	$\tilde{O}(1)$	$\tilde{O}(1)$	500× at client
Discrete-log-based NIZK	$\tilde{\Theta}(n)$	$\tilde{\Theta}(n)$	$\tilde{\Theta}(n)$	$\tilde{\Theta}(n)$	50× at server
<b>Prio</b> (latest version)	0	0	$\tilde{O}(1)$	$\tilde{O}(1)$	1×

(Table hides log factors.)

# Contribution:

Prevent disruption using proofs on secret-shared data



$\mathbf{x} \in \{0,1\}^n$   
Data for  $n$  domains



$[\mathbf{x}]_A$

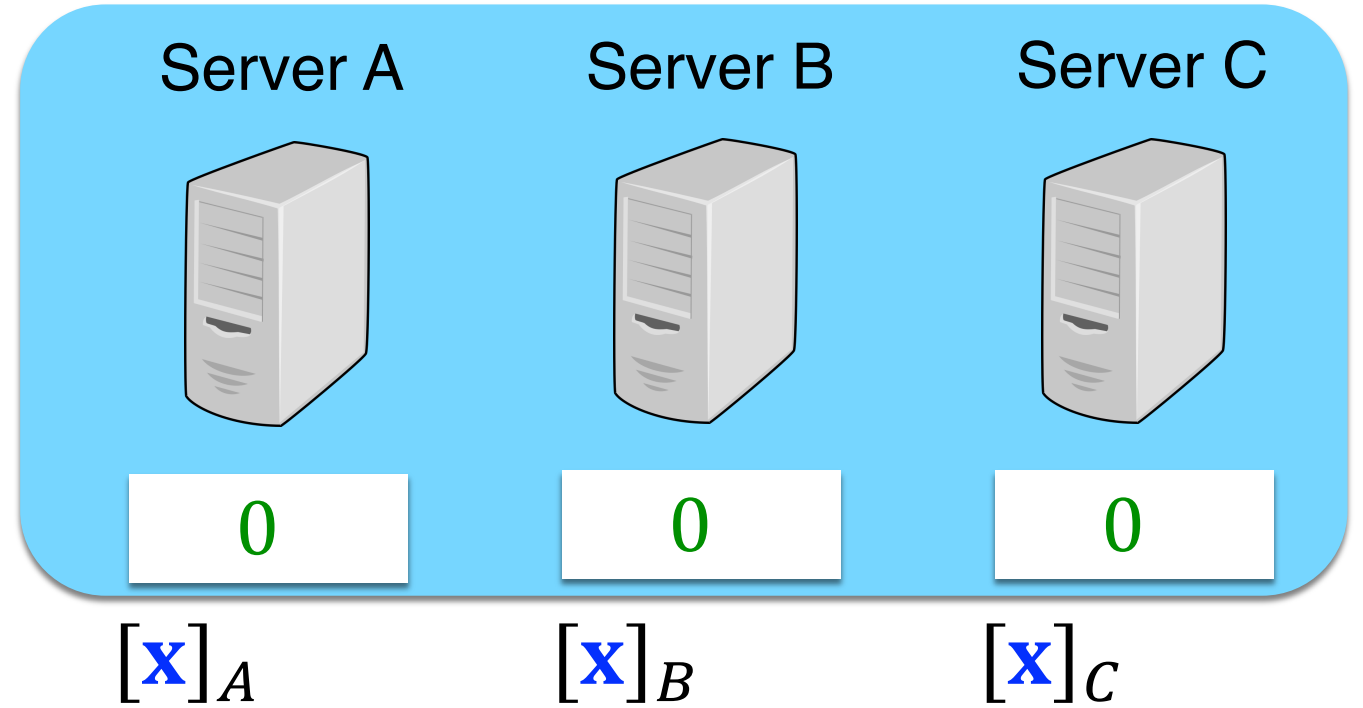
$[\mathbf{x}]_B$

$[\mathbf{x}]_C$

Dimension- $n$  vectors  
of integers mod  $p$ .  
(i.e., in  $\mathbb{Z}_p^n$ )

# Contribution:

Prevent disruption using proofs on secret-shared data



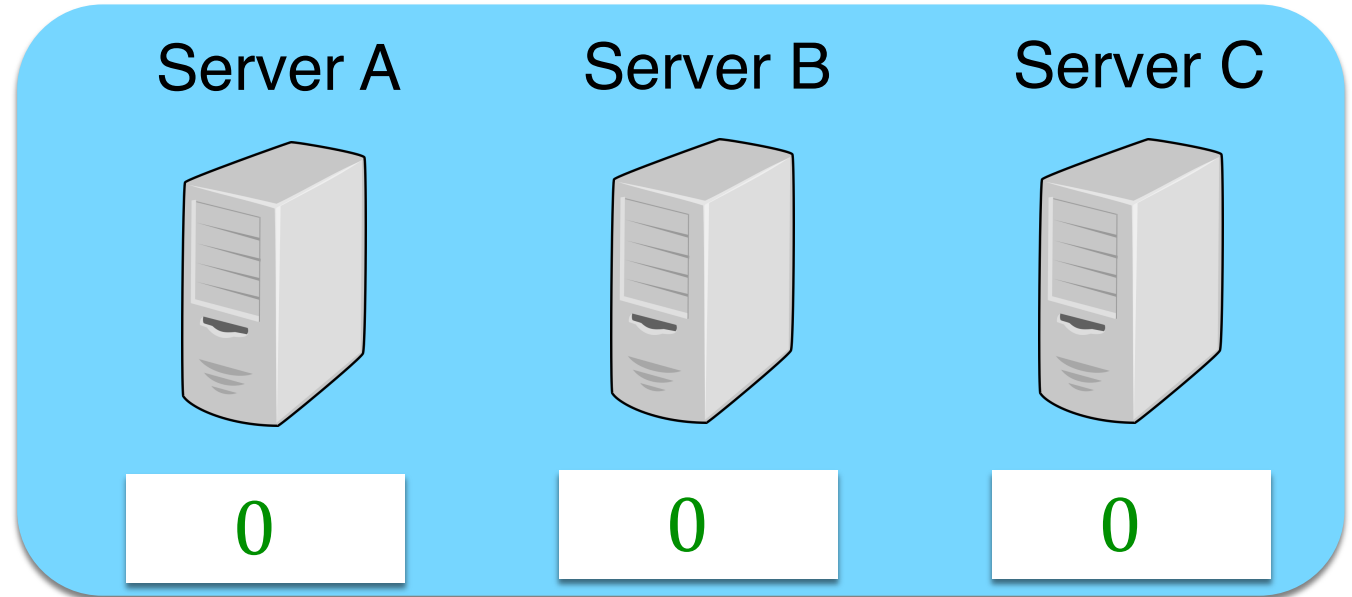
$$x \in \{0,1\}^n$$

Data for  $n$  domains



# Contribution:

Prevent disruption using proofs on secret-shared data



$[\mathbf{x}]_A$

$[\mathbf{x}]_B$

$[\mathbf{x}]_C$

Want to be convinced that  
 $\mathbf{x} = [\mathbf{x}]_A + [\mathbf{x}]_B + [\mathbf{x}]_C \in \{0,1\}^n \in \mathbb{Z}_p^n$

$$\mathbf{x} \in \{0,1\}^n$$

Data for  $n$  domains

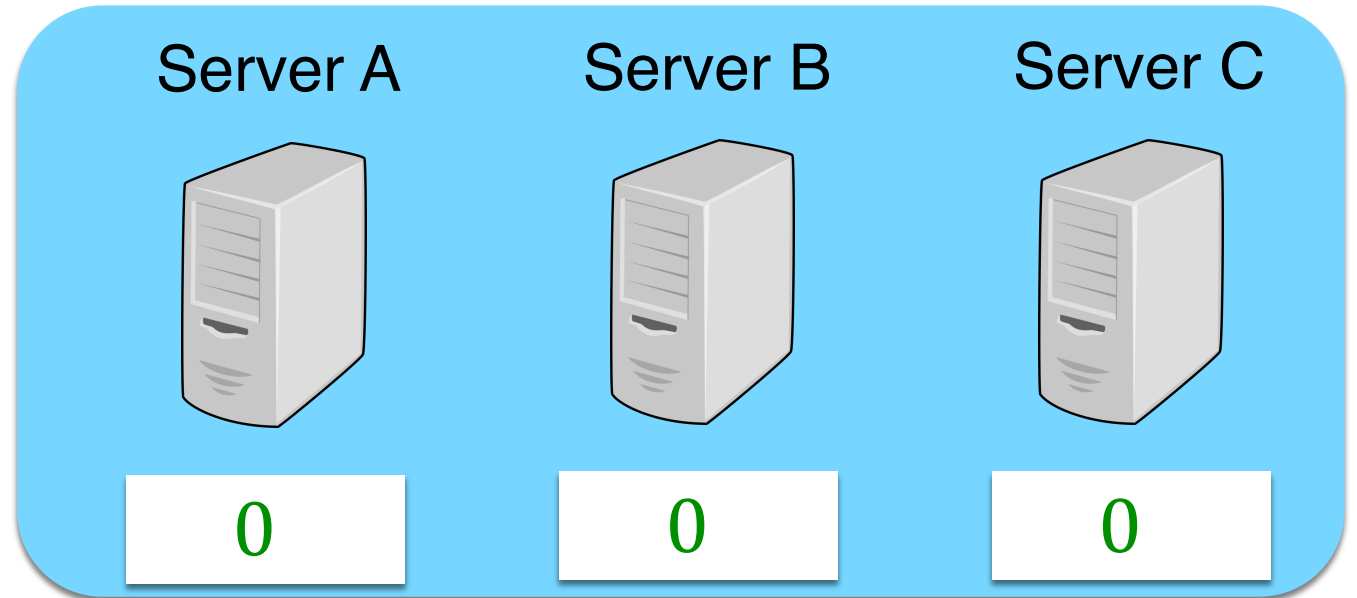


# Contribution:

Prevent disruption using proofs on secret-shared data

$$\mathbf{x} \in \{0,1\}^n$$

Data for  $n$  domains



$$[\mathbf{x}]_A$$

$$[\mathbf{x}]_B$$

$$[\mathbf{x}]_C$$

Want to be convinced that

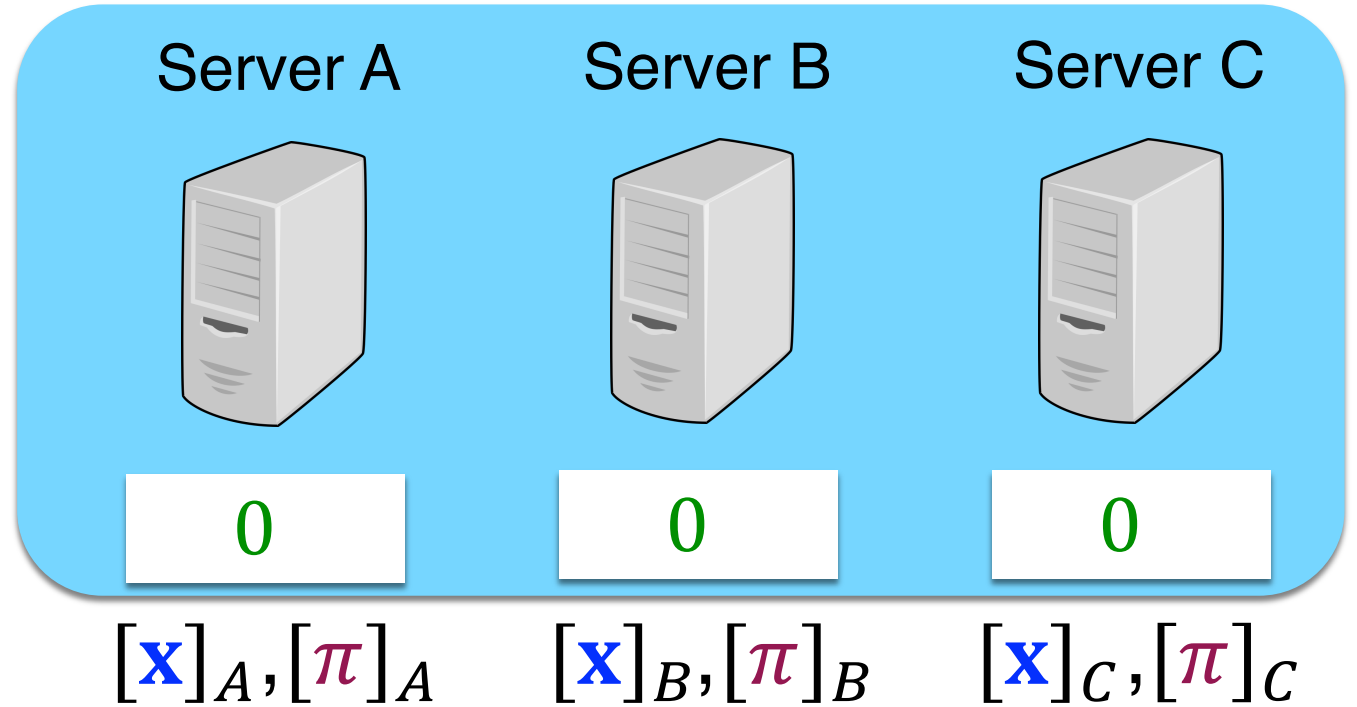
$$\mathbf{x} = [\mathbf{x}]_A + [\mathbf{x}]_B + [\mathbf{x}]_C \in \{0,1\}^n \in \mathbb{Z}_p^n$$

More generally, that  $\text{Valid}(\mathbf{x})$  holds,  
for some predicate  $\text{Valid}$



# Contribution:

Prevent disruption using proofs on secret-shared data



$$\mathbf{x} \in \{0,1\}^n$$

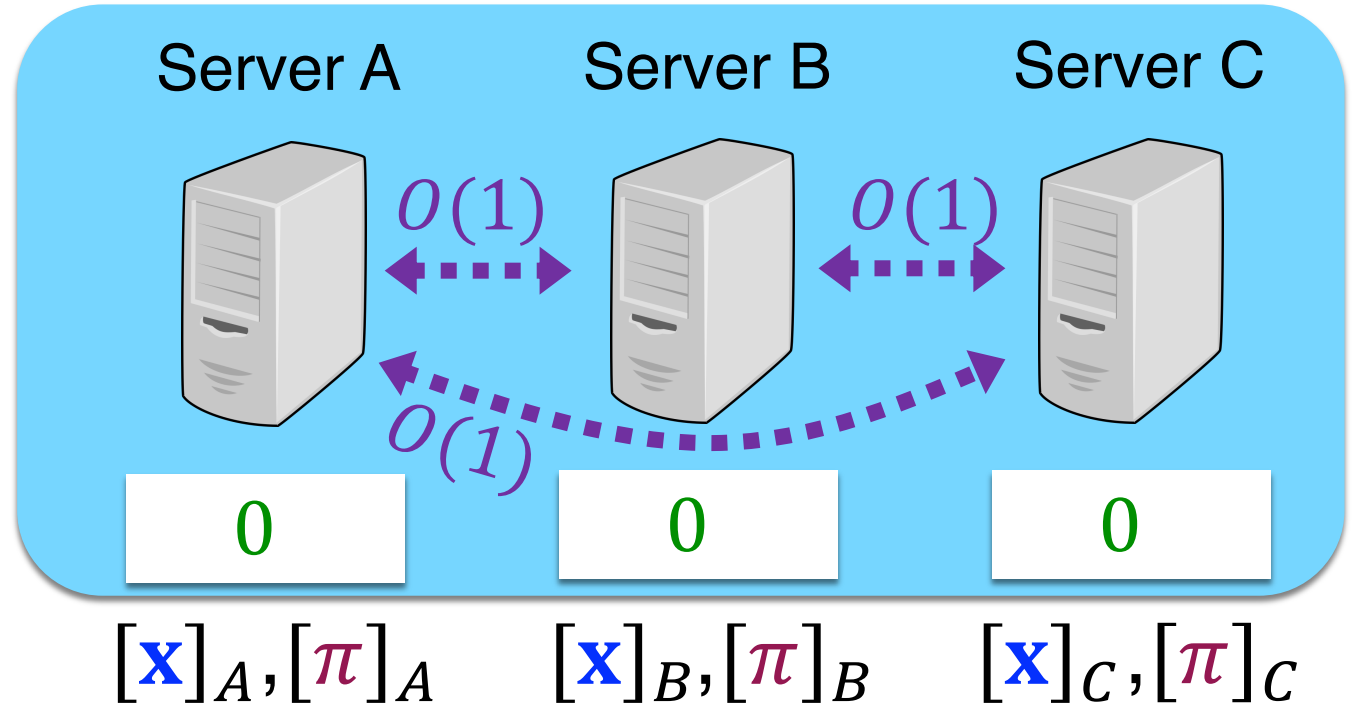
Data for  $n$  domains



- Client sends proof to servers that **Valid(x)** holds
  - For our example, **Valid(x)** = " $\mathbf{x} \in \{0,1\}^n$ "
  - Servers exchange  $O(1)$  bytes to check proof
- **Prevents disruption in Prio**
  - Servers reject invalid client submissions

# Contribution:

Prevent disruption using proofs on secret-shared data



$$\mathbf{x} \in \{0,1\}^n$$

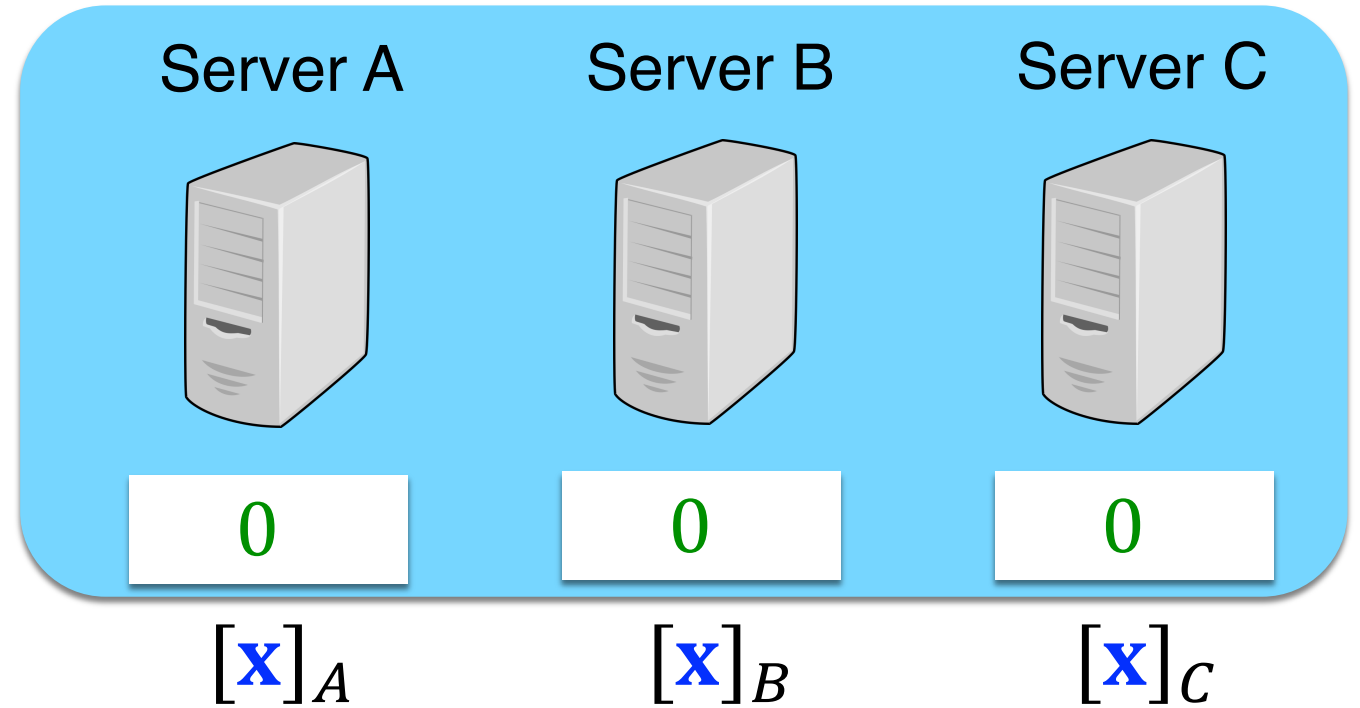
Data for  $n$  domains



- Client sends proof to servers that **Valid(x)** holds
  - For our example, **Valid(x)** = " $\mathbf{x} \in \{0,1\}^n$ "
  - Servers exchange  $O(1)$  bytes to check proof
- **Prevents disruption in Prio**
  - Servers reject invalid client submissions

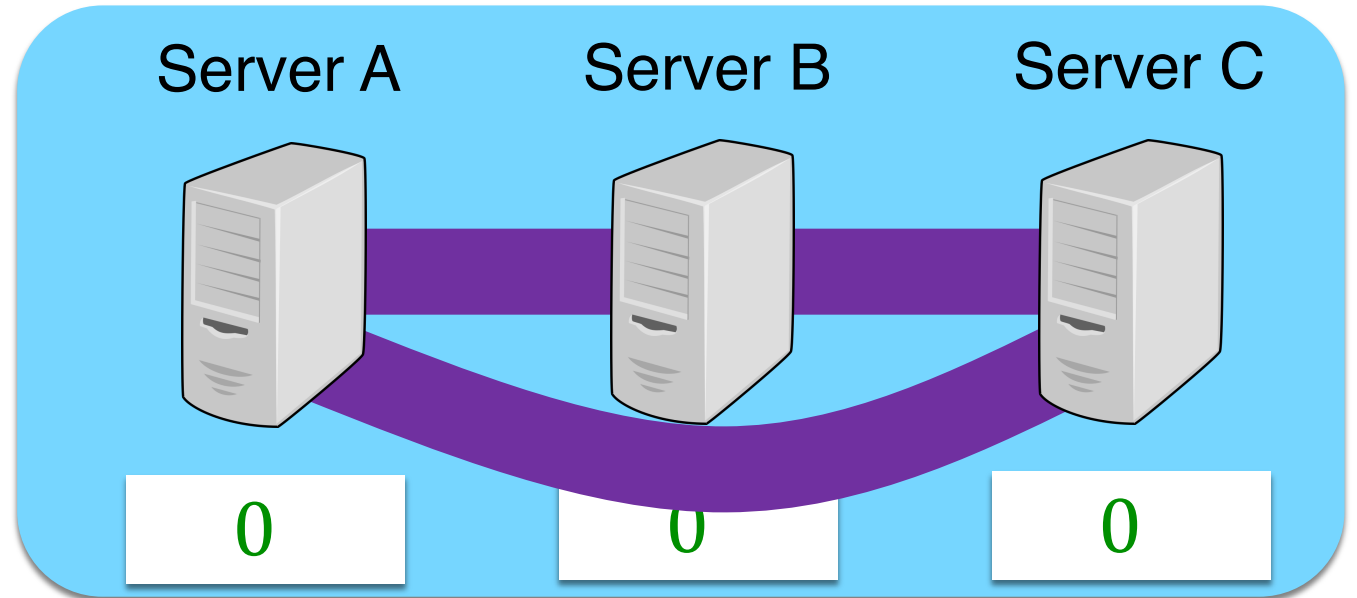
# How to construct a proof on secret-shared data\*

\*simplified



# How to construct a proof on secret-shared data\*

\*simplified



$[x]_A$

$[x]_B$

$[x]_C$

Could use secure multi-party computation to check that  $\text{Valid}(x)$  holds

[GMW87], [BGW88], ...

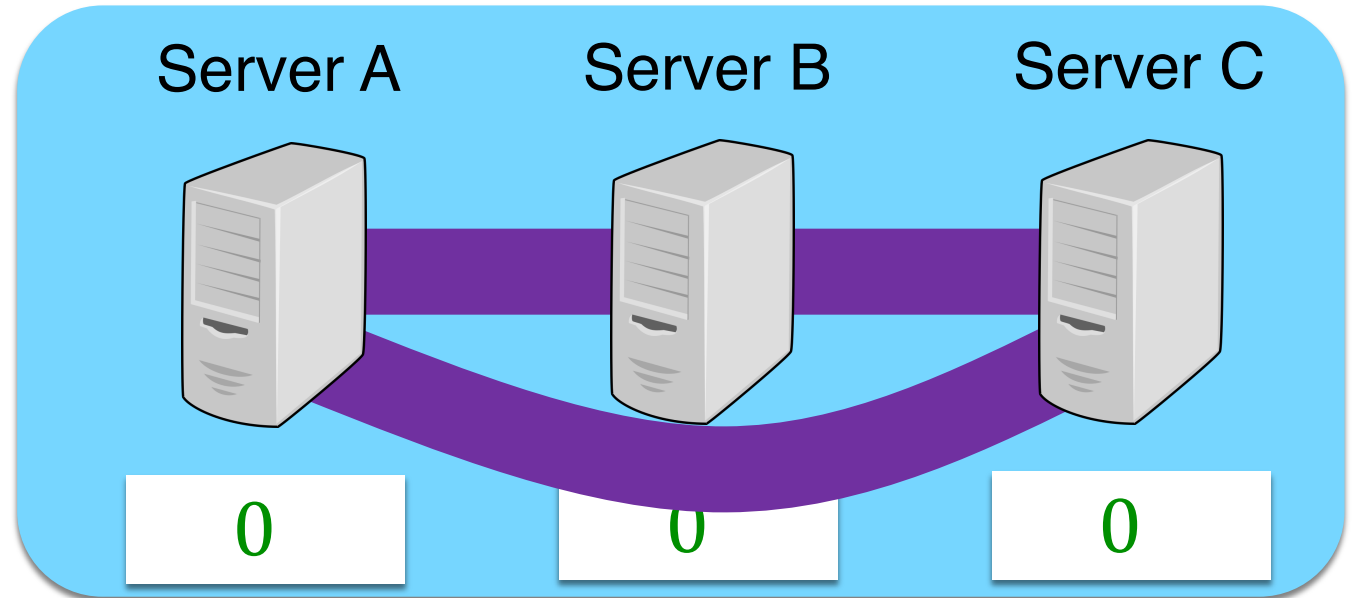
# How to construct a proof on secret-shared data\*

\*simplified



$$\mathbf{x} \in \{0,1\}^n$$

Data for  $n$  domains



$[\mathbf{x}]_A$

$[\mathbf{x}]_B$

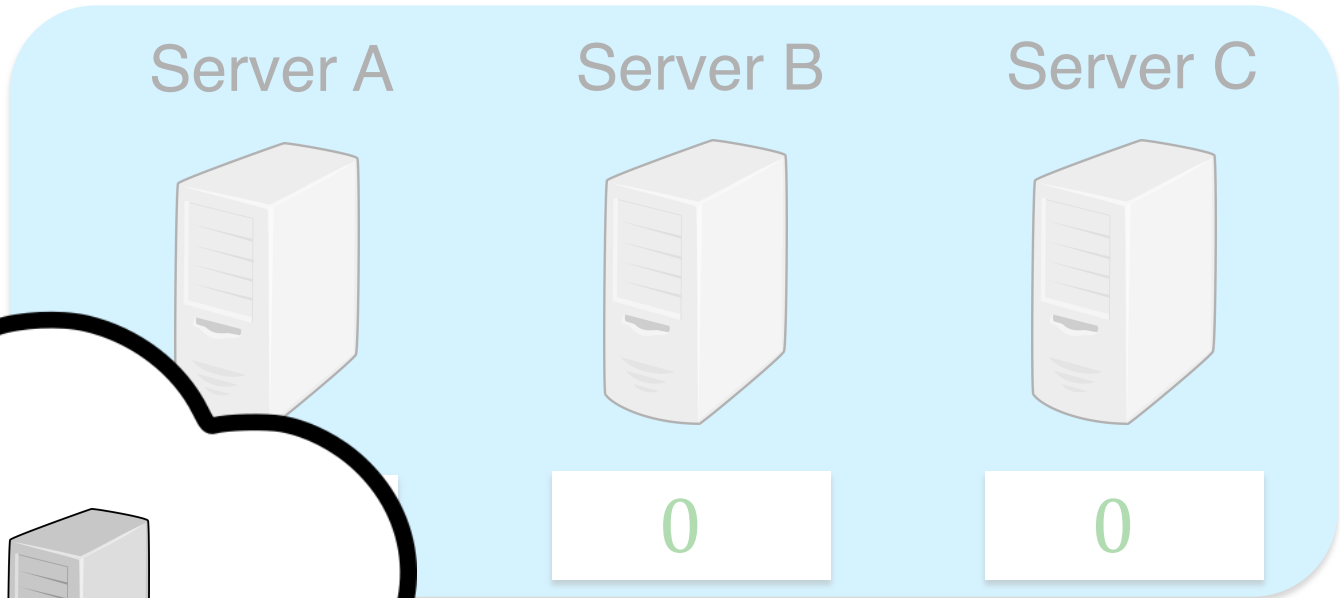
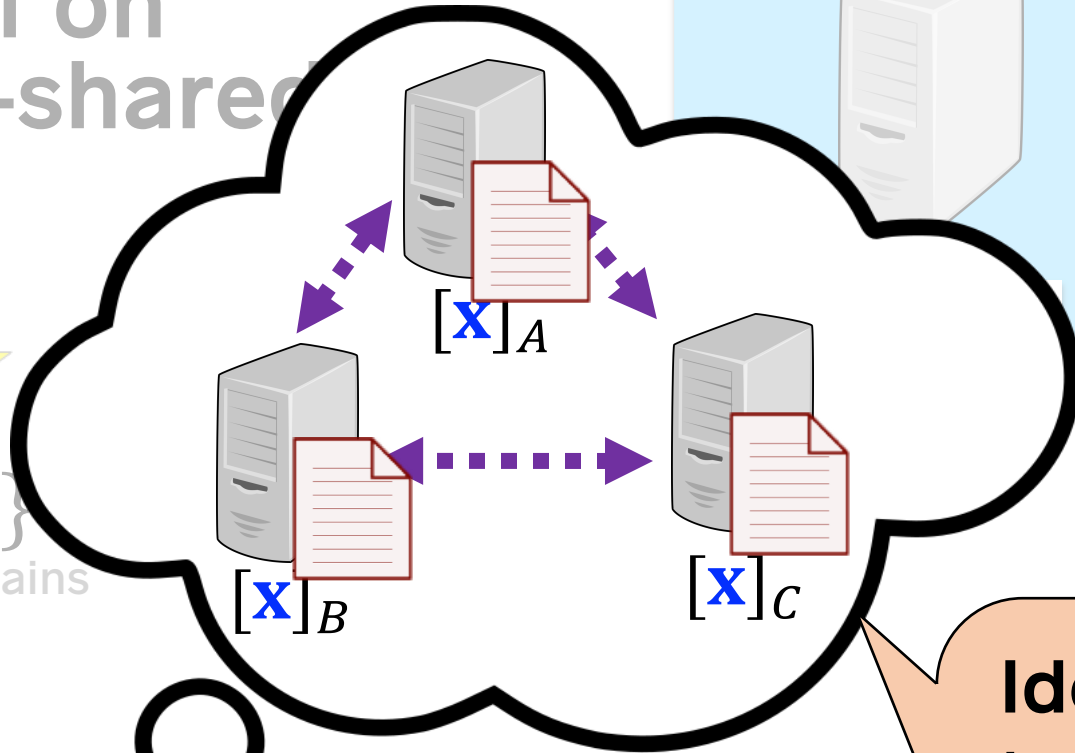
$[\mathbf{x}]_C$

Could use secure multi-party computation to check that  $\text{Valid}(\mathbf{x})$  holds

[GMW87], [BGW88], ...

# How to construct a proof on secret-shared

$x \in \{0,1\}$   
Data for  $n$  domains



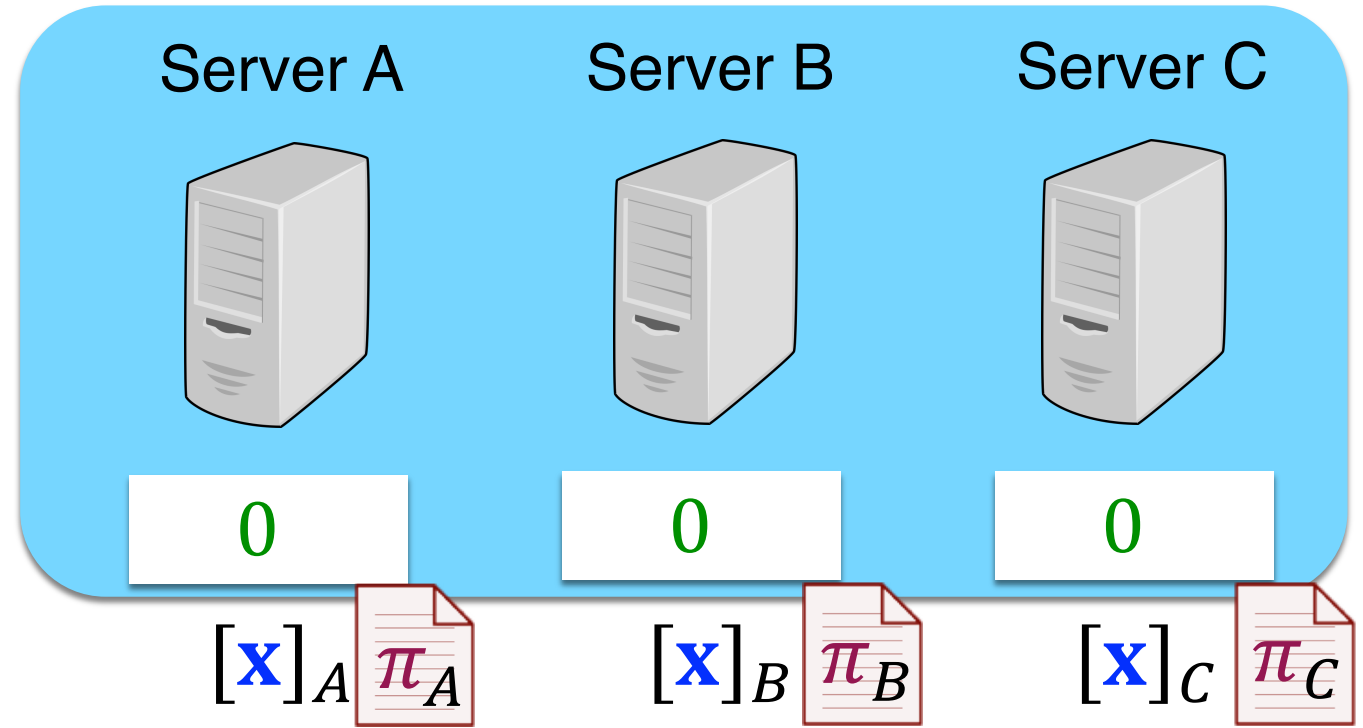
$[x]_B$

$[x]_C$

**Idea:** Client generates transcripts that servers would have observed in a multi-party computation of **Valid**( $x$ ).

See also [IKOS07]

# How to construct a proof on secret-shared data



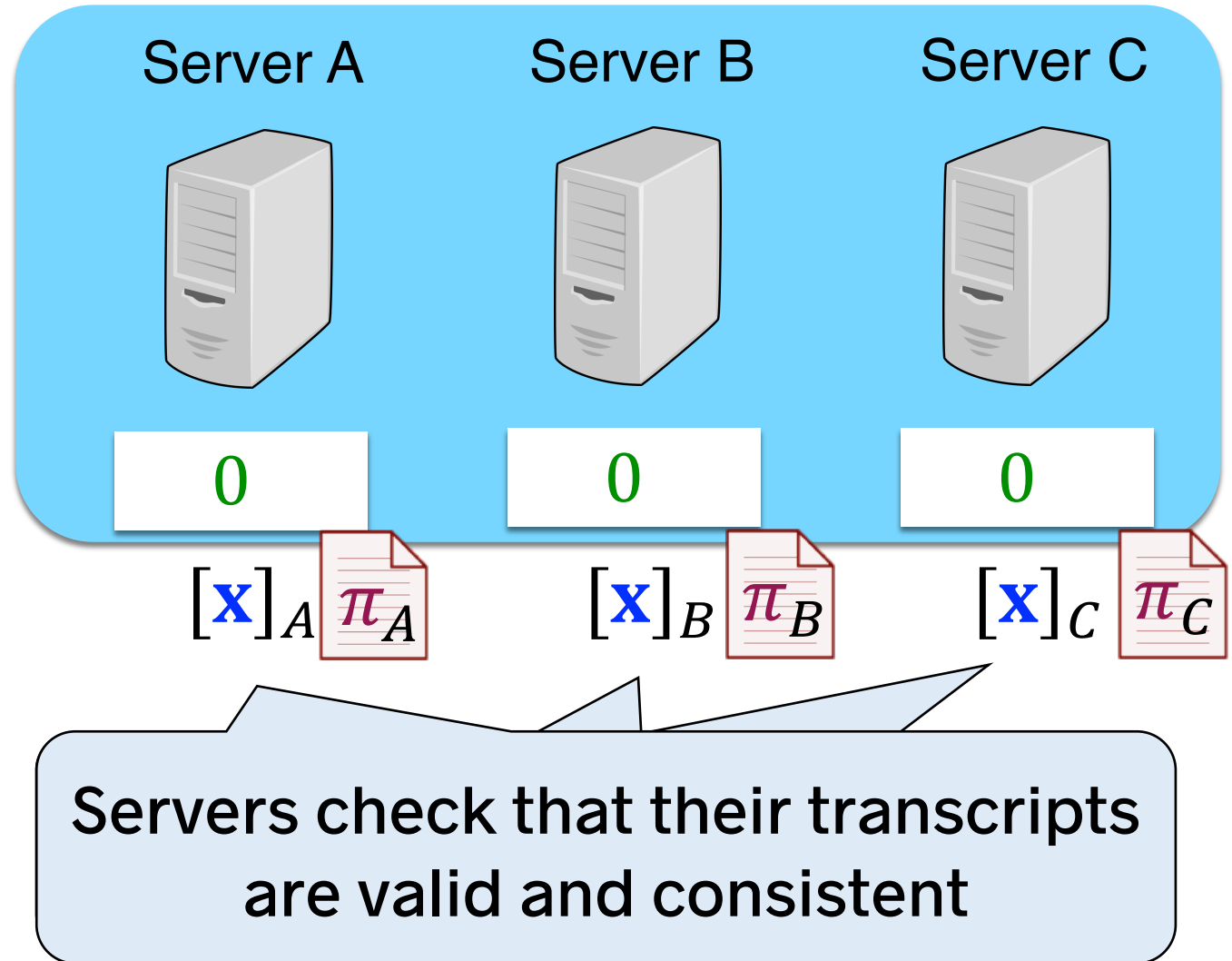
$$\mathbf{x} \in \{0,1\}^n$$

Data for  $n$  domains



# How to construct a proof on secret-shared data

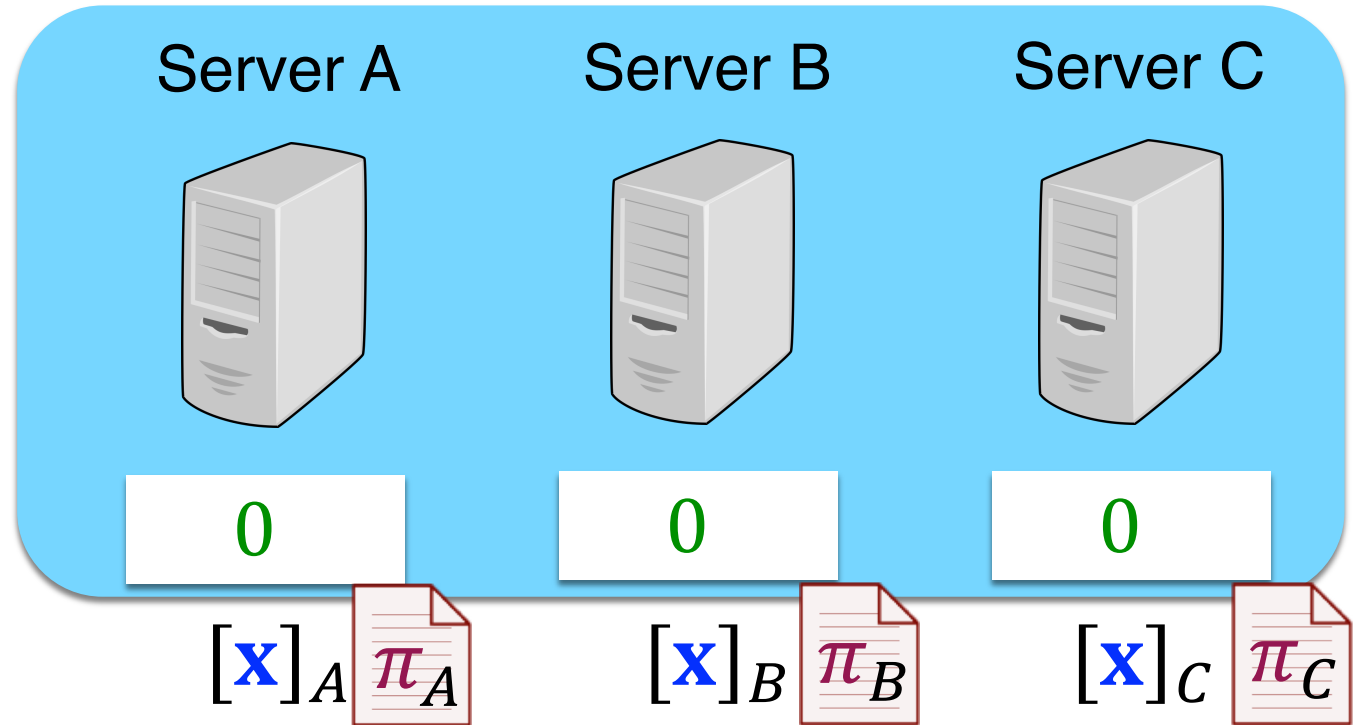
$\mathbf{x} \in \{0,1\}^n$   
Data for  $n$  domains





# How to construct a proof on secret-shared data

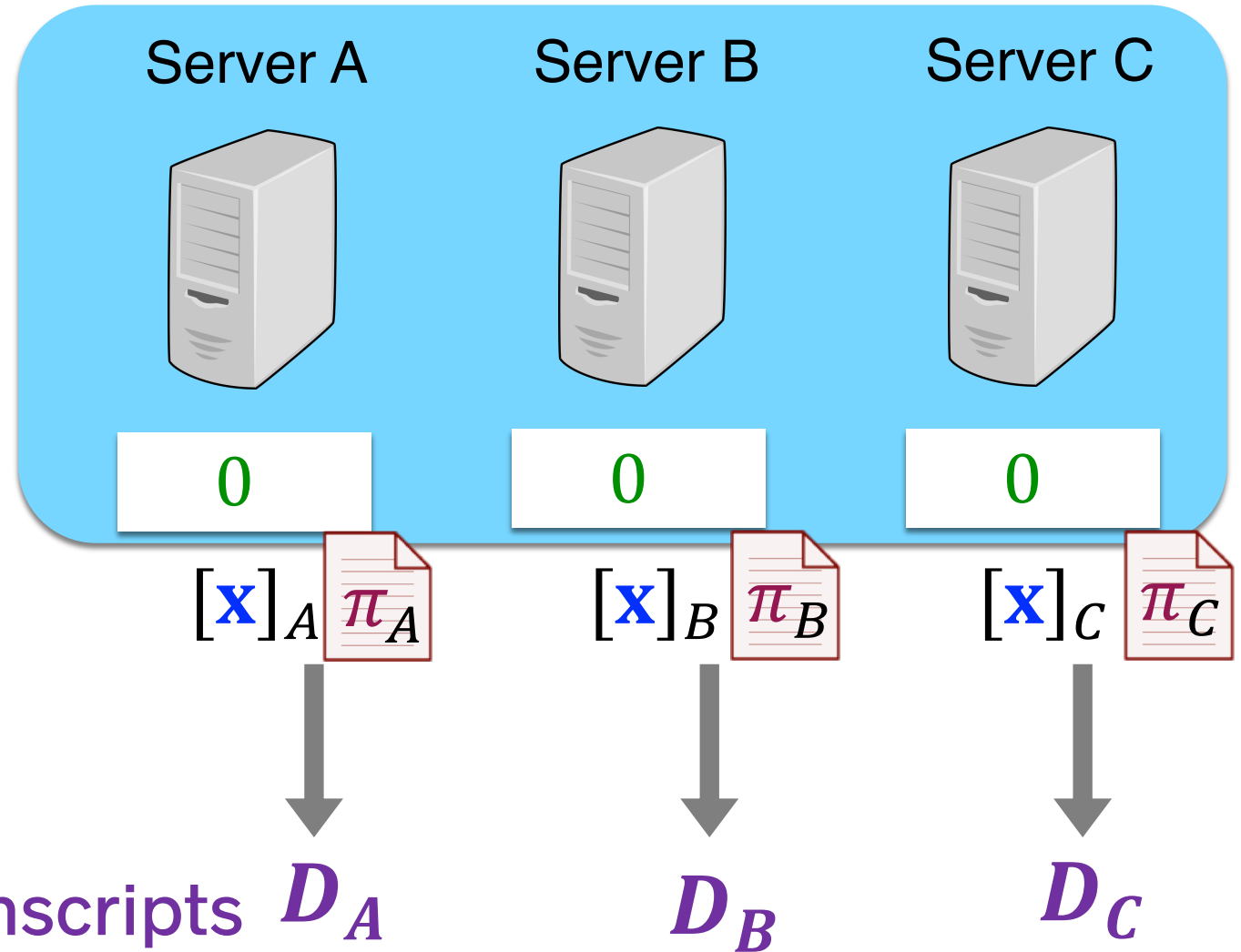
$\mathbf{x} \in \{0,1\}^n$   
Data for  $n$  domains



Servers check that their transcripts are valid and consistent

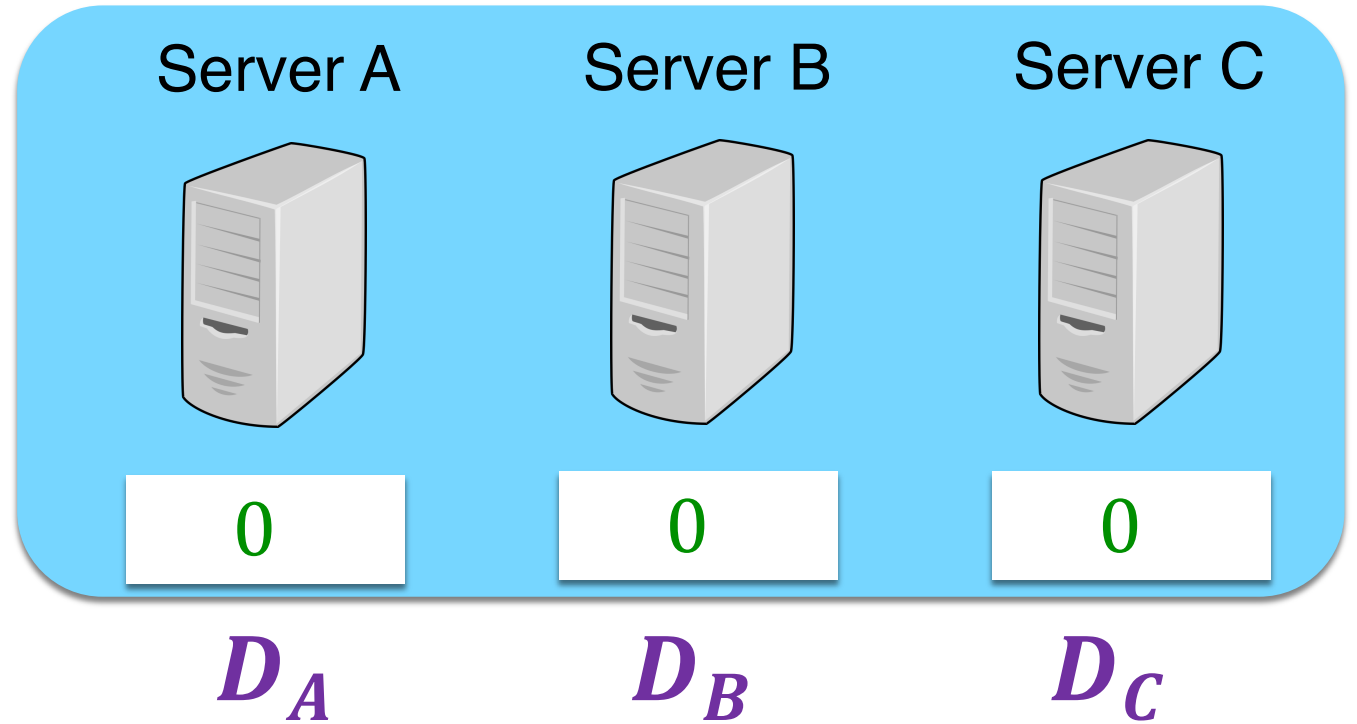
Checking a transcript is much easier than generating one.

# How to construct a proof on secret-shared data



“Randomized digest” of transcripts  $D_A$   
(Leak nothing about client's value  $x$ )

# How to construct a proof on secret-shared data



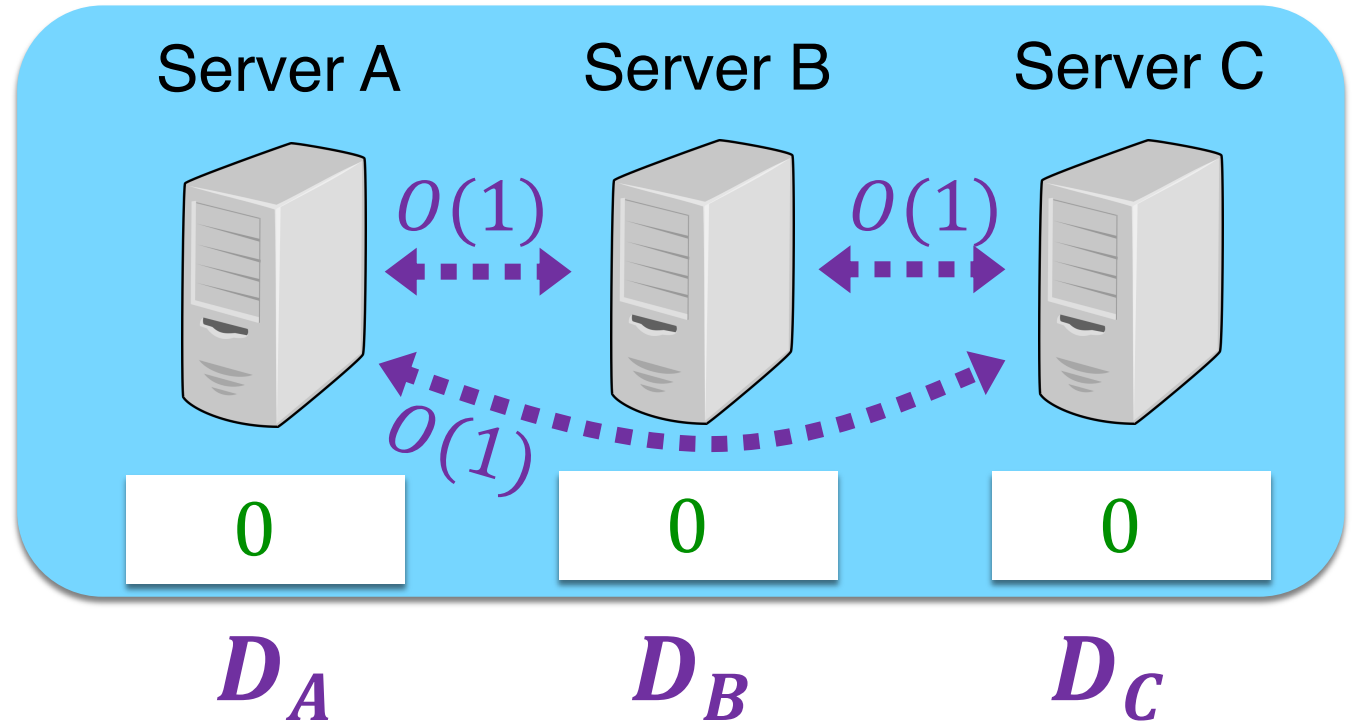
If  $x$  is well formed:  $D_A + D_B + D_C = 0$

If  $x$  is malformed:  $D_A + D_B + D_C \neq 0$  with high probability

Servers publish  $D_A, D_B, D_C$  and check that they sum to 0.

→ Servers accept  $x$  if so.

# How to construct a proof on secret-shared data



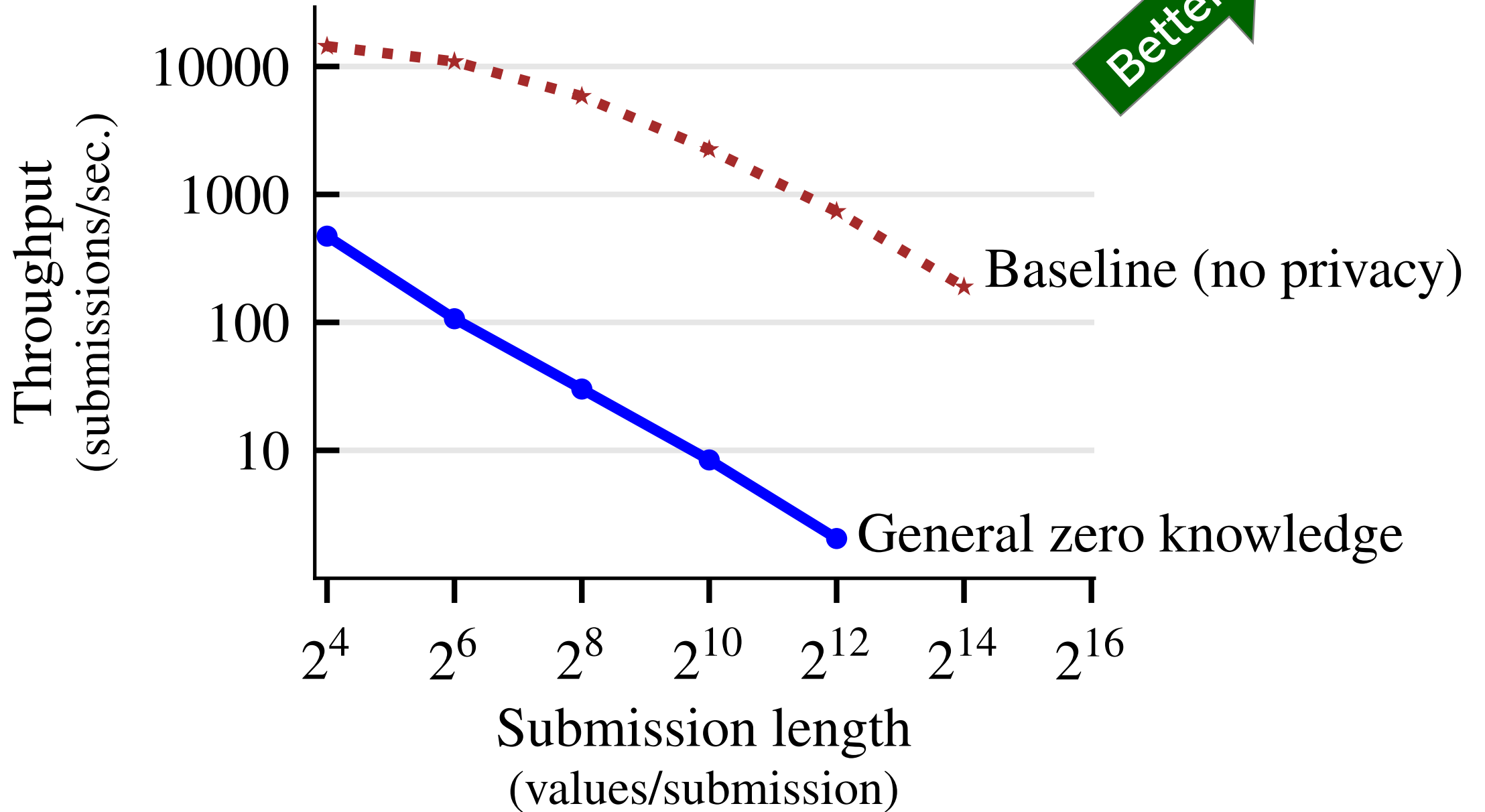
If  $x$  is well formed:  $D_A + D_B + D_C = 0$

If  $x$  is malformed:  $D_A + D_B + D_C \neq 0$  with high probability

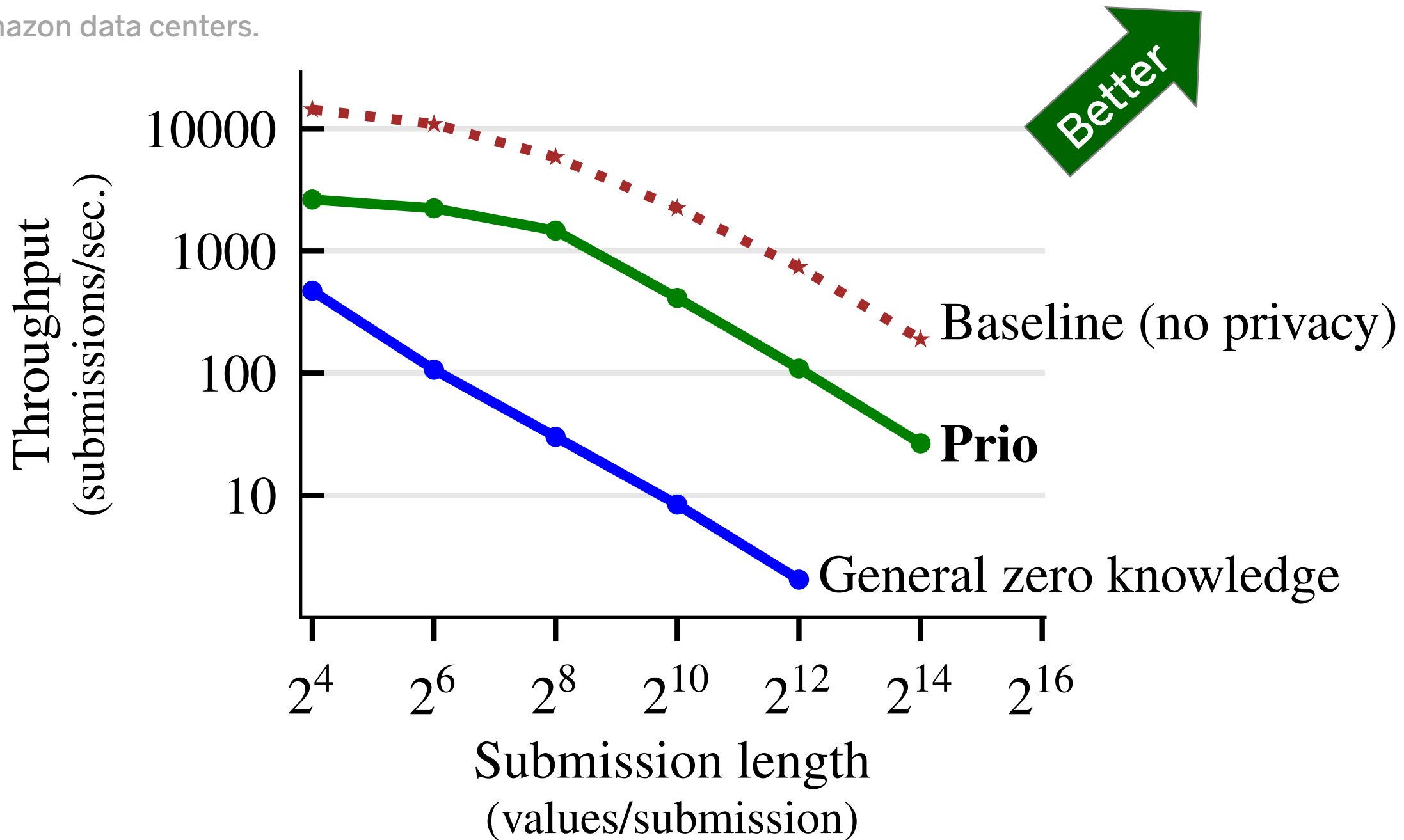
Servers publish  $D_A, D_B, D_C$  and check that they sum to 0.

→ Servers accept  $x$  if so.

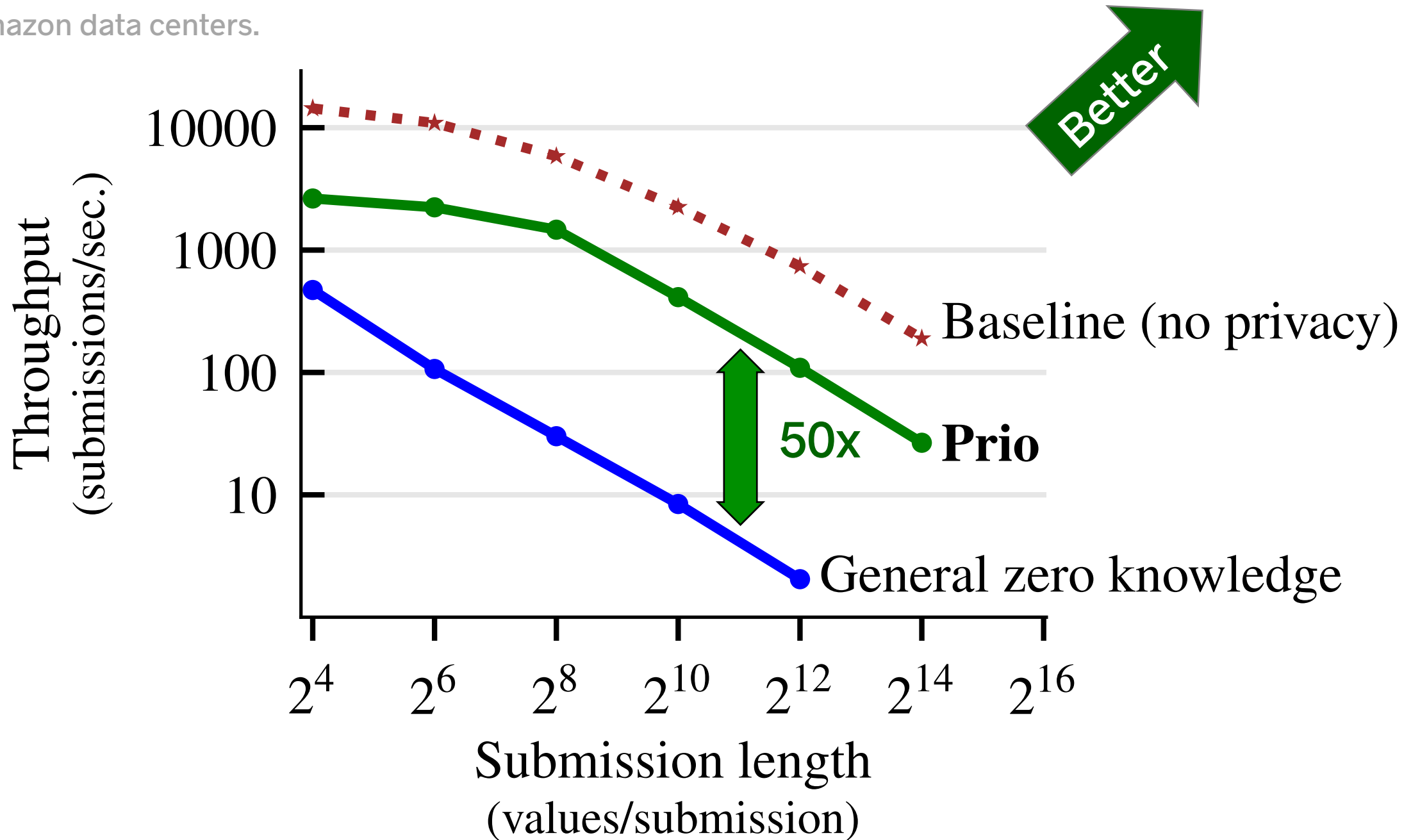
Five-server cluster in five Amazon data centers.



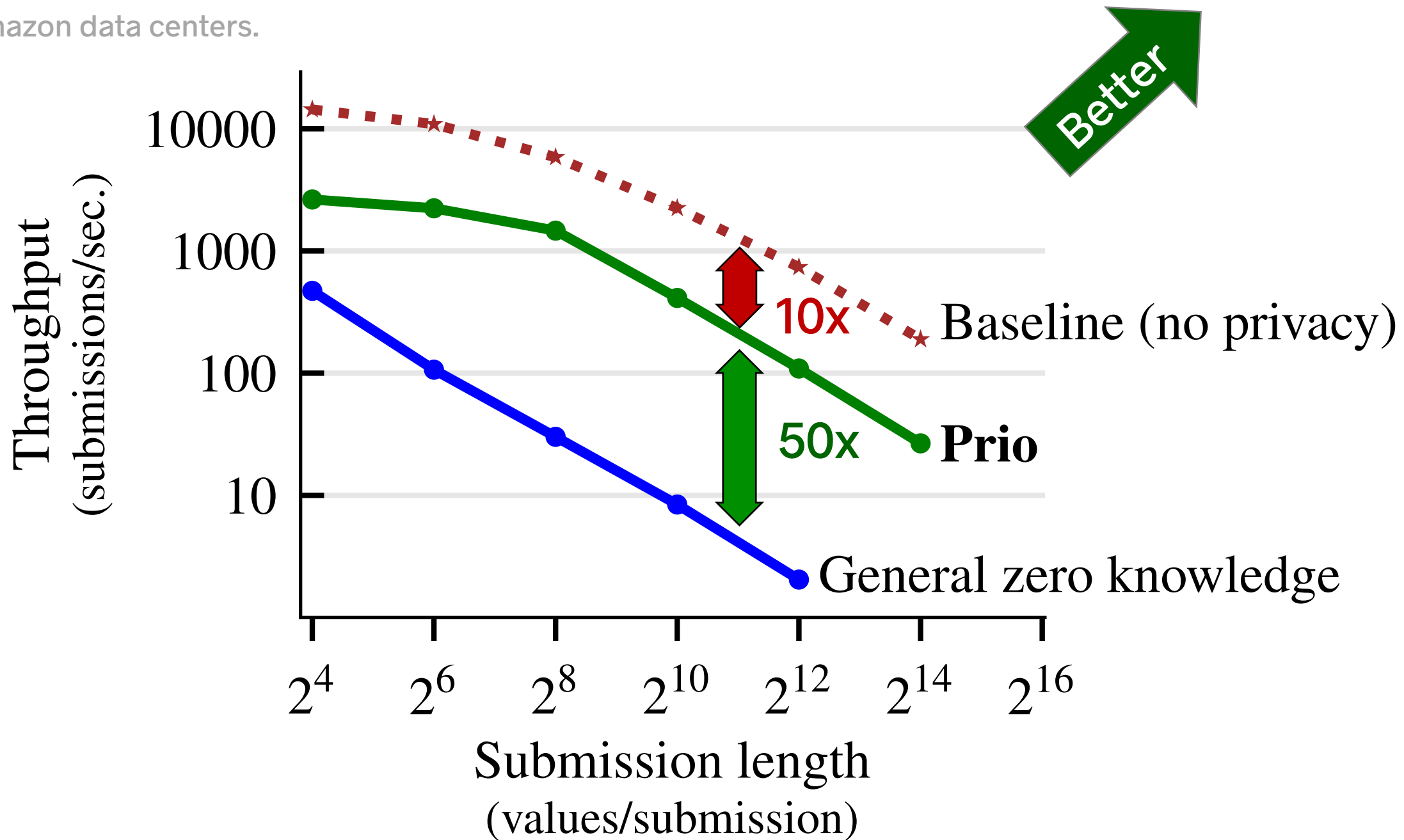
Five-server cluster in five Amazon data centers.



Five-server cluster in five Amazon data centers.



Five-server cluster in five Amazon data centers.





# Prio supports a range of aggregation functions

- **Average**
- **Variance** [PBBL11]
- **Most popular** (approx.) [MDD16]
- **Min and max** (approx.)
- **Quality of arbitrary regression model** ( $R^2$ )
- **Least-squares regression**
- **Gradient descent step**  
[BIKMMPRSS17]



# Firefox Deployment

# Firefox deployment



Uses libprio, a C library we wrote that implements Prio

- [github.com/mozilla/libprio](https://github.com/mozilla/libprio) – 3.5k LoC
- Encoding a length-1024 data packet: 35ms in Firefox browser  
(more optimizations possible)
- Python bindings to simplify server-side data analysis

Pilot phase, 11/2018-now

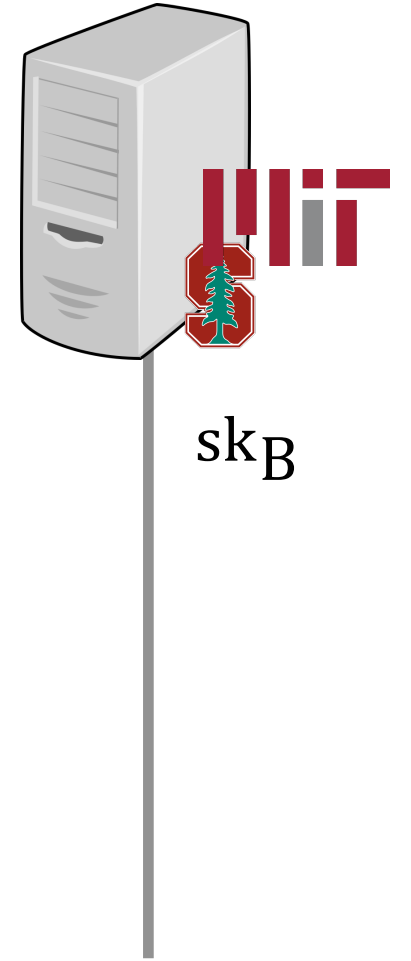
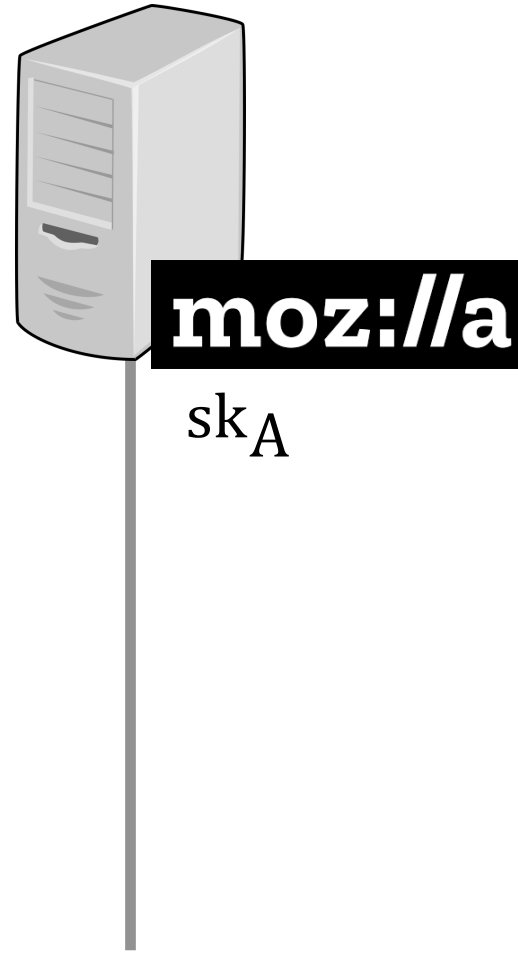
- Implemented in Firefox, but Mozilla currently runs all servers
- Enabled by default only in the “Nightly” build

Next step: **Move second server to external org.** (In progress)

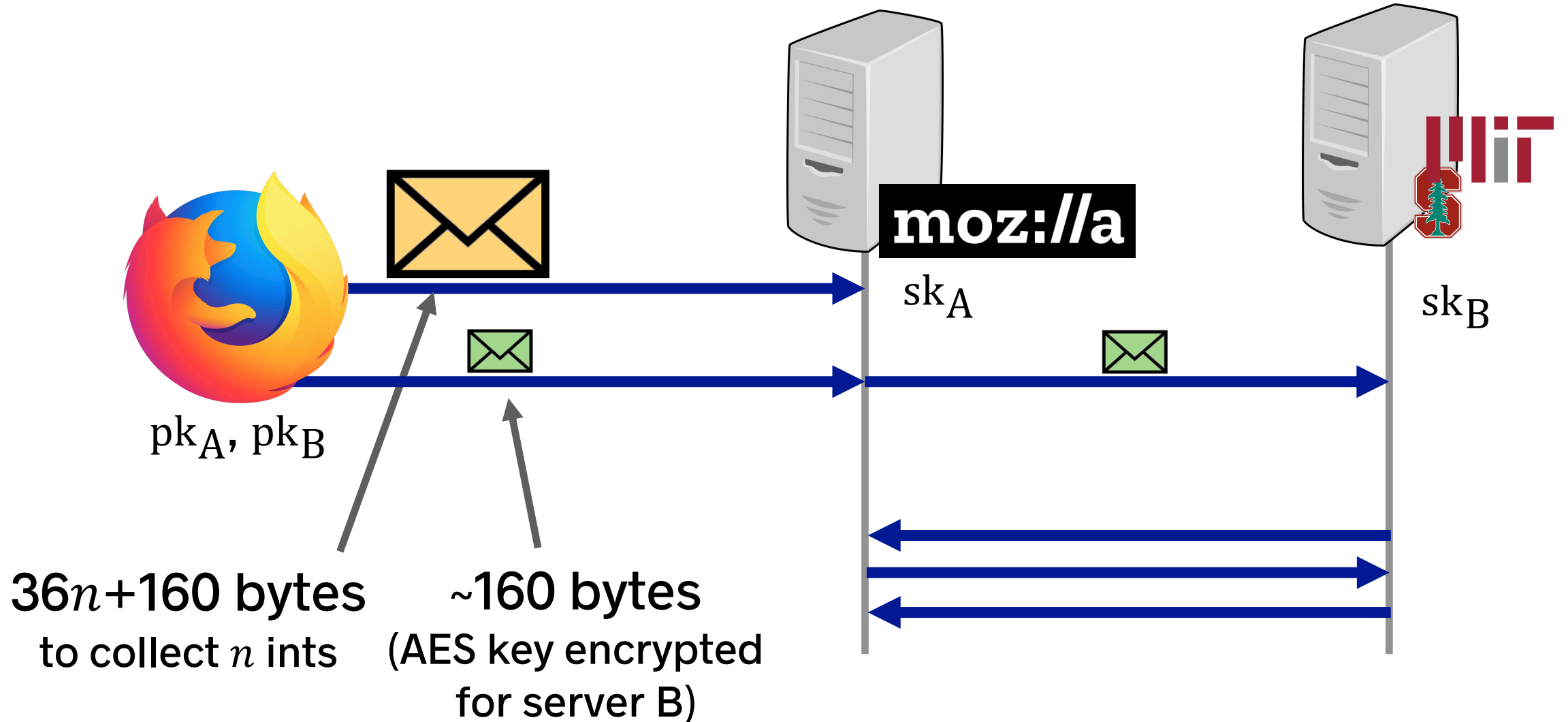
# Firefox deployment



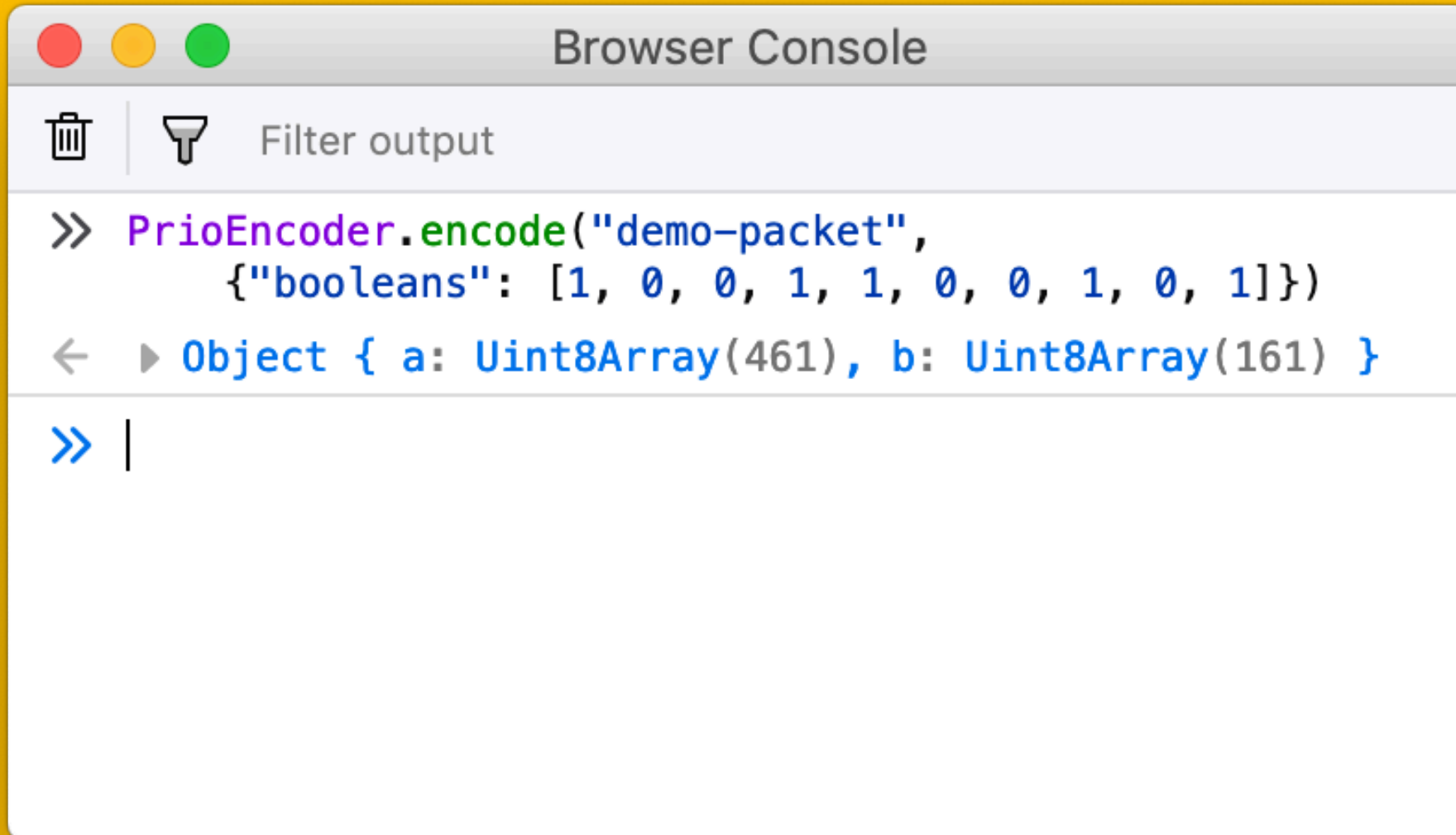
$pk_A, pk_B$



# Firefox deployment



In Firefox, set preference `devtools.chrome.enabled`, then in browser console...



The screenshot shows a 'Browser Console' window with a title bar containing three colored window control buttons (red, yellow, green) and the text 'Browser Console'. Below the title bar is a toolbar with a trash icon, a funnel icon, and the text 'Filter output'. The main area of the console contains the following text:

```
>> PrioEncoder.encode("demo-packet",  
    {"booleans": [1, 0, 0, 1, 1, 0, 0, 1, 0, 1]})  
← ▶ Object { a: Uint8Array(461), b: Uint8Array(161) }  
  
>> |
```

In Nightly, set `pref. telemetry.origin_telemetry_test_mode.enabled`, browse for a while, then visit `about:telemetry`.

## Origin Telemetry Find in Origin Telemetry

Firefox Origin Telemetry encodes data before it is sent so that Mozilla can count things, but not know whether or not any given Firefox contributed to that count. ([learn more](#))

origin	count
news.google.com	1
doubleclick.net	4
bluekai.com	2
amazon-adsystem.com	3
www.google.com	1
scorecardresearch.com	1



Copy Summary

View ▾

**Closed**

Bug 1543712 Opened 9 months ago Closed 6 months ago

## Enable Origin Telemetry

### ▼ Categories

Product: Core ▾

Component: Privacy: Anti-Tracking ▾

Type:  task

Priority: P2

### ▼ Tracking

Status: RESOLVED FIXED

Milestone: mozilla69

Tracking Flags:	Tracking	Status
firefox69	---	<a href="#">fixed</a>





[Docs](#) » [Toolkit](#) » [Telemetry](#) » [Data collection](#) » [Origin Telemetry](#)

[View page source](#)

---

## Origin Telemetry

*Origin Telemetry* is an experimental Firefox Telemetry mechanism that allows us to privately report origin-specific information in aggregate. In short, it allows us to get exact counts of how *many* Firefox clients do certain things on specific origins without us being able to know *which* clients were doing which things on which origins.

As an example, Content Blocking would like to know which trackers Firefox blocked most frequently. Origin Telemetry allows us to count how many times a given tracker is blocked without being able to find out which clients were visiting pages that had those trackers on them.



moz://a



Download Firefox

Search Mozilla Hacks

# Testing Privacy-Preserving Telemetry with Prio



By [Robert Helmer](#), [Anthony Miyaguchi](#),  
[Eric Rescorla](#)

Posted on [October 29, 2018](#) in [Firefox](#) and [Privacy](#) [Share This](#)

Building a browser is hard; building a good browser inevitably requires gathering a lot of data to make sure that things that work in the lab work in the field. But as soon as you gather data, you have to make sure you protect user privacy. We're always looking at ways to improve the security of our data collection, and lately we've been experimenting with a really cool technique called Prio.



*Posted on 2019-04-26 in mozilla*

## Firefox Origin Telemetry: Putting Prio in Practice

Prio is neat. It allows us to learn counts of things that happen across the Firefox population without ever being able to learn which Firefox sent us which pieces of information.

For example, Content Blocking will soon be using this to count how



# Deployment stats

- **Initially, collecting data on ~2,500 blacklist rules**  
fb.com, google-analytics.com, adwords.google.com, ...
- **Data collected on 0.014% of pageloads for 1% of clients**
- **Expect to process ~200m telemetry submissions per day**
  - Submission from client every 24h or on shutdown
  - = Tens of gigabytes of data per day to the second server



# The second server

- **Prio requires 2+ non-colluding servers, maintained ideally**
    - by independent organizations,
    - on independent infrastructure (not both on AWS), and
    - in different countries (under independent legal jurisdictions).
  - **Serious non-technical challenge, but reasons for optimism**
    - Infrastructure costs are modest
    - $\exists$  multiple candidate orgs with privacy-centric mission
    - If Org2 uses Prio, Mozilla can be the “second server” for Org2
- Mozilla is working to sign up a partner org in 2020.



# You can help!

[github.com/mozilla/libprio/](https://github.com/mozilla/libprio/)

## Small things

- Add support for aggregating a wider range of data types
- Implement client- and server-side optimizations
- Implement differential-privacy features

## Big things

- Rewrite parts of libprio in Rust
- Be the external org that runs the second server

→ Eligible for Mozilla's bug-bounty program. ←

# Conclusion

- Prio is a new system for privacy-preserving telemetry
- Firefox is using Prio to collect data to improve the browser's new tracking-protection feature
- Deployment is ongoing!
  - Ask if you're interested in helping out.

Henry Corrigan-Gibbs (EPFL & MIT CSAIL), [henrycg@csail.mit.edu](mailto:henrycg@csail.mit.edu)

Dan Boneh (Stanford), Gary Chen, Steven Englehardt, Robert Helmer, Chris Hutten-Czapski, Anthony Miyaguchi, Eric Rescorla, and Peter Saint-Andre (Mozilla)

Details: [bugzilla.mozilla.org/show\\_bug.cgi?id=1543712](https://bugzilla.mozilla.org/show_bug.cgi?id=1543712)

Code: [github.com/mozilla/libprio/](https://github.com/mozilla/libprio/)

Paper: [crypto.stanford.edu/prio/](https://crypto.stanford.edu/prio/)

# Conclusion

- Prio is a new system for privacy-preserving telemetry
- Firefox is using Prio to collect data to improve the browser's new tracking-protection feature
- Deployment is ongoing!
  - Ask if you're interested in helping out.

Henry Corrigan-Gibbs (EPFL & MIT CSAIL), [henrycg@csail.mit.edu](mailto:henrycg@csail.mit.edu)

Dan Boneh (Stanford), Gary Chen, Steven Englehardt, Robert Helmer, Chris Hutten-Czapski, Anthony Miyaguchi, Eric Rescorla, and Peter Saint-Andre (Mozilla)

Details: [bugzilla.mozilla.org/show\\_bug.cgi?id=1543712](https://bugzilla.mozilla.org/show_bug.cgi?id=1543712)

Code: [github.com/mozilla/libprio/](https://github.com/mozilla/libprio/)

Paper: [crypto.stanford.edu/prio/](https://crypto.stanford.edu/prio/)

**Thank you!**



