80th Meeting of the IFIP 10.4 Working Group on Dependable Computing and Fault Tolerance

June 24, 2021

Security and Privacy for Distributed Optimization and Learning

Nitin Vaidya

Georgetown University

5-Hour Tutorial

disc.georgetown.domains → Talks

Secret to happiness is to
lower your expectations to the point
where they're already met

- Hobbes (paraphrased)

Bill Waterson

Goals

- Background
- Problem formulation
- Intuition

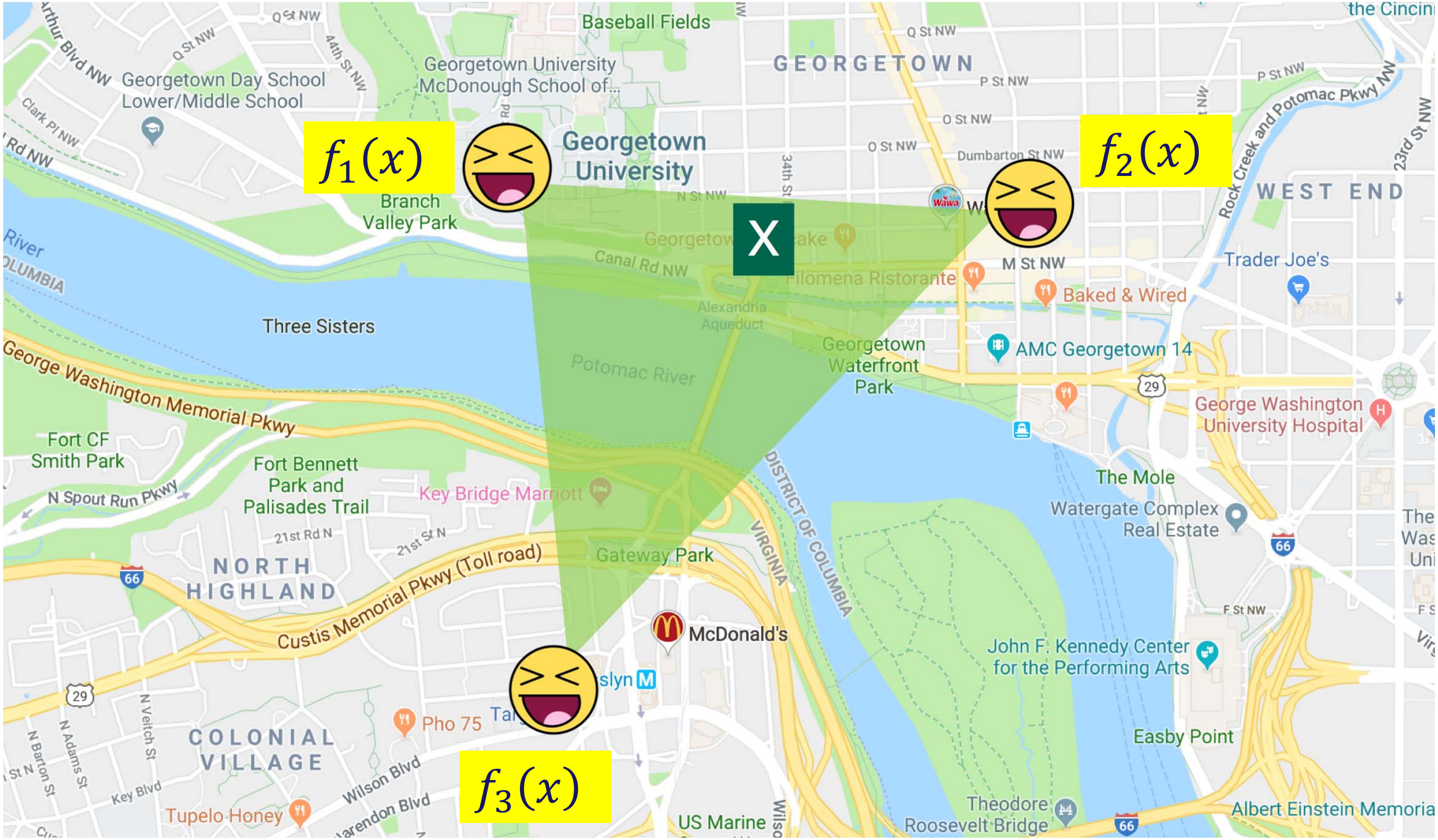
No theorems/proofs

Distributed Optimization

Rendezvous


Baseball Fields
GEORGETOWN
Georgetown Day School Lower/Middle School
Georgetown University McDonough School of...
Georgetown University
Foundry Branch Valley Park
Georgetown Cupcake
Canal Rd NW
Filomena Ristorante
M St NW
Baked & Wired
Trader Joe's
WEST END
Alexandria Aqueduct
Three Sisters
Potomac River
Georgetown Waterfront Park
AMC Georgetown 14
George Washington Memorial Pkwy
George Washington University Hospital
Fort CF Smith Park
Fort Bennett Park and Palisades Trail
Key Bridge Marriott
N Spout Run Pkwy
The Mole
Watergate Complex Real Estate
NORTH HIGHLAND
Custis Memorial Pkwy (Toll road)
Gateway Park
DISTRICT OF COLUMBIA
VIRGINIA
McDonald's
John F. Kennedy Center for the Performing Arts
COLONIAL VILLAGE
ROSSLYN
Pho 75
Easby Point
Wilson Blvd
Tupelo Honey
US Marine
Theodore Roosevelt Bridge
Albert Einstein Memoria
Rock Creek and Potomac Pkwy NW
Q St NW
P St NW
O St NW
N St NW
Dumbarton St NW
34th St NW
30th St NW
27th St NW
23rd St NW
F St NW
21st Rd N
21st St N
16th Rd N
Key Blvd
N Veitch St
N Adams St
N Barton St

Rendezvous

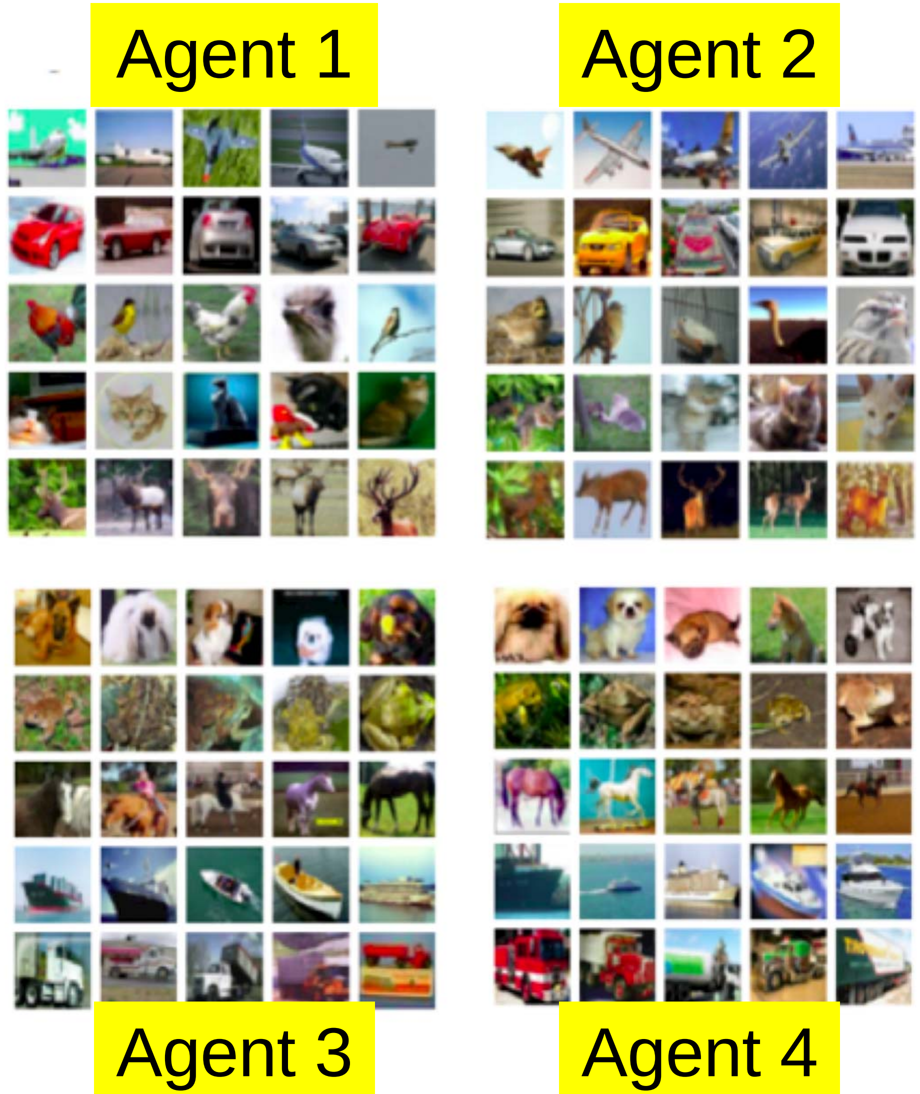


$f_1(x)$
$f_2(x)$
$f_3(x)$
X
Baseball Fields
GEORGETOWN
Georgetown University McDonough School of...
Georgetown Day School Lower/Middle School
Georgetown University
Branch Valley Park
WEST END
Canal Rd NW
M St NW
Trader Joe's
Filomena Ristorante
Baked & Wired
Alexandria Aqueduct
Three Sisters
Potomac River
Georgetown Waterfront Park
AMC Georgetown 14
George Washington Memorial Pkwy
George Washington University Hospital
Fort CF Smith Park
Fort Bennett Park and Palisades Trail
Key Bridge Marriott
The Mole
Watergate Complex Real Estate
N Spout Run Pkwy
NORTH HIGHLAND
Custis Memorial Pkwy (Toll road)
Gateway Park
DISTRICT OF COLUMBIA
VIRGINIA
McDonald's
John F. Kennedy Center for the Performing Arts
COLONIAL VILLAGE
Pho 75
Easby Point
Wilson Blvd
Tupelo Honey
Theodore Roosevelt Bridge
Albert Einstein Memoria
US Marine

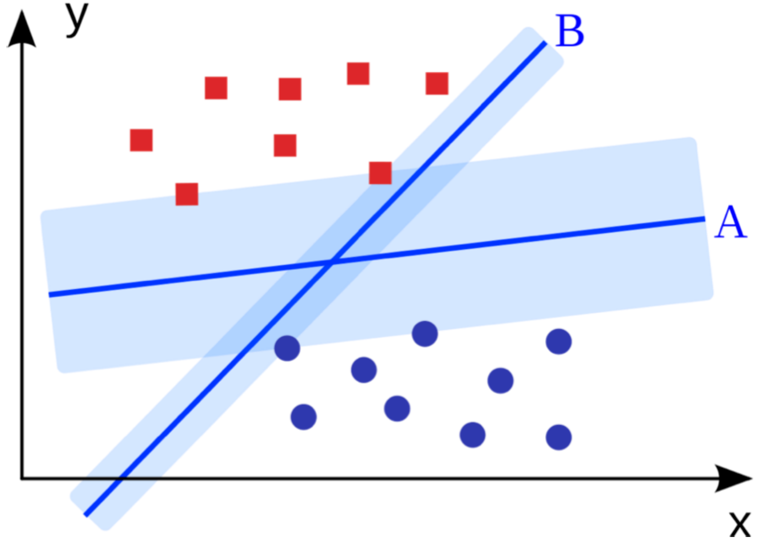
$$\mathrm{argmin} \sum f_i(x)$$

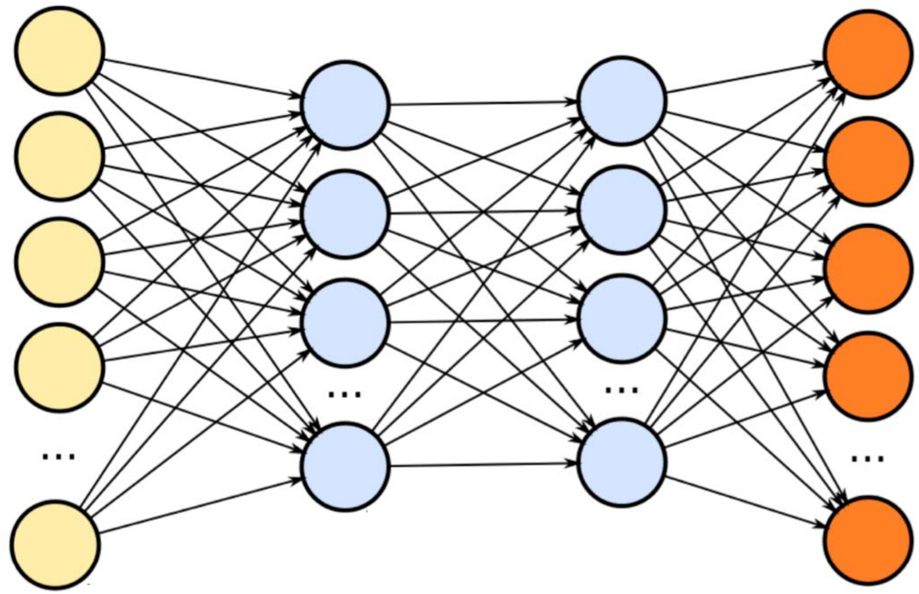
Machine Learning

- Data is distributed across different agents



- Data is distributed across different agents ➔ Collaborate to learn





Machine Learning

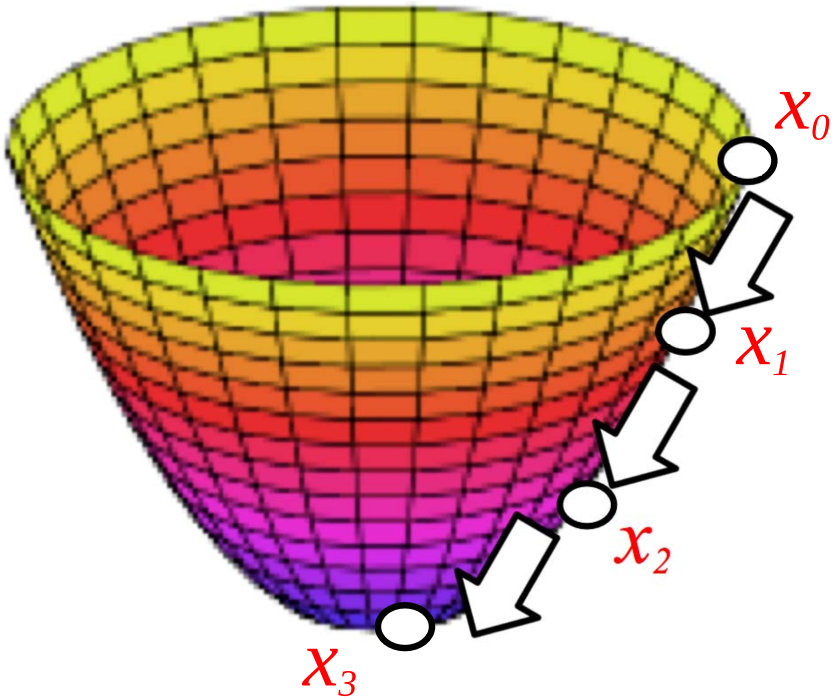
Minimize global loss

$\operatorname{argmin} \sum f_i(x)$

$$\mathrm{argmin} \sum f_i(x)$$

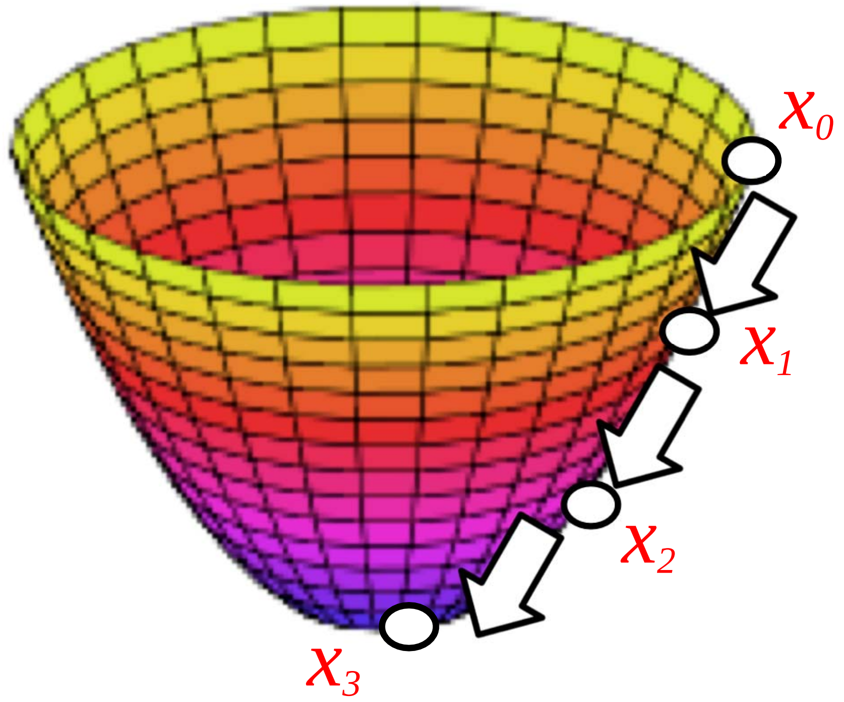
Gradient Method

$$f(x) = \sum f_i(x)$$



Gradient Method

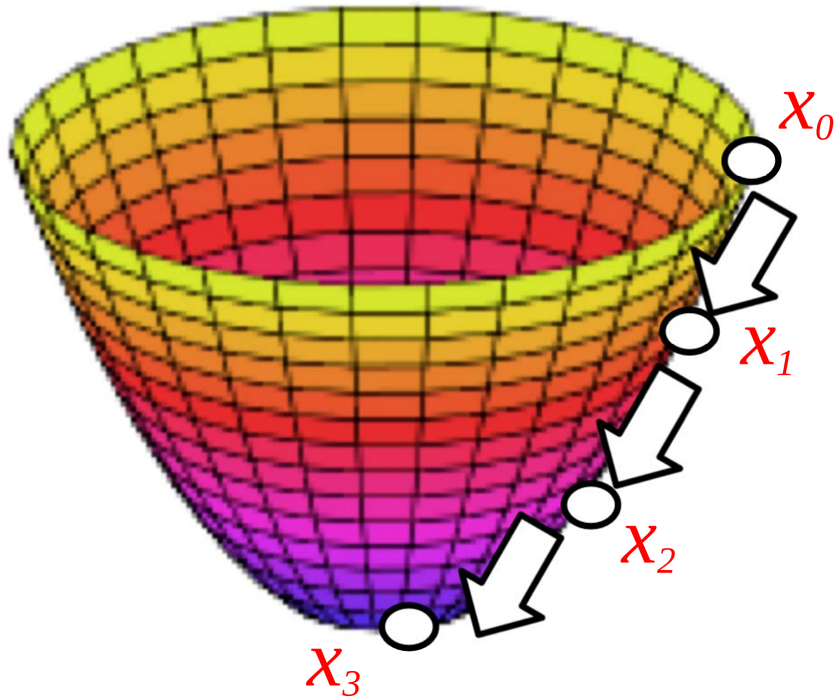
$$f(x) = \sum f_i(x)$$



$$x_{k+1} \leftarrow x_k - \lambda \sum_i \nabla f_i(x_k)$$

Gradient Method

$$f(x) = \sum f_i(x)$$



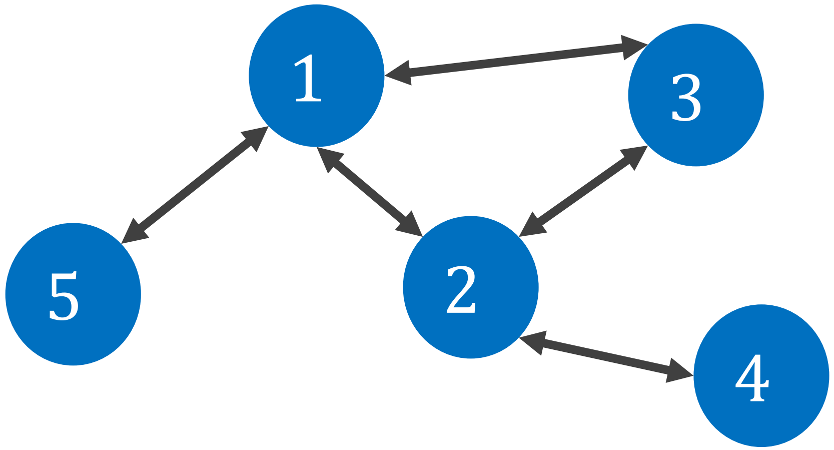
$$x_{k+1} \leftarrow x_k - \lambda \sum_i \nabla f_i(x_k)$$

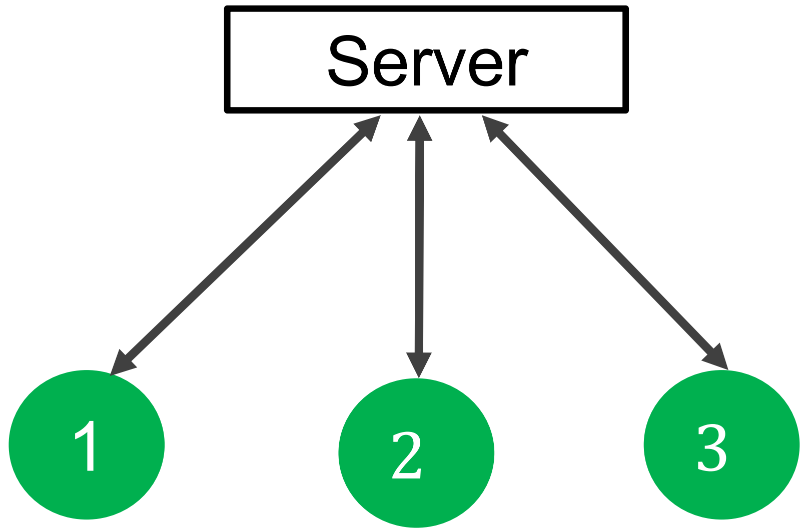
Distributed Optimization

- Each agent i knows own cost function $f_i(x)$
- Need to cooperate to minimize $\sum f_i(x)$

➔ Distributed algorithms

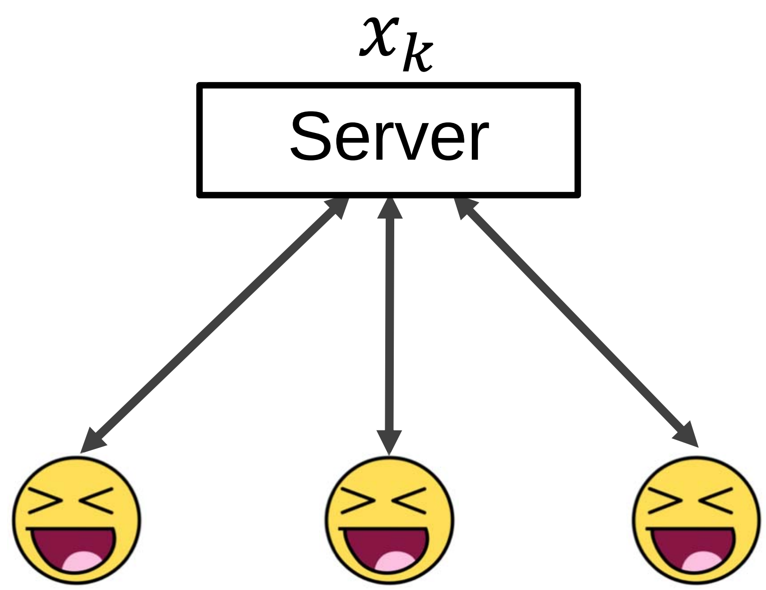
Architectures





Parameter Server

- Server maintains estimate x_k

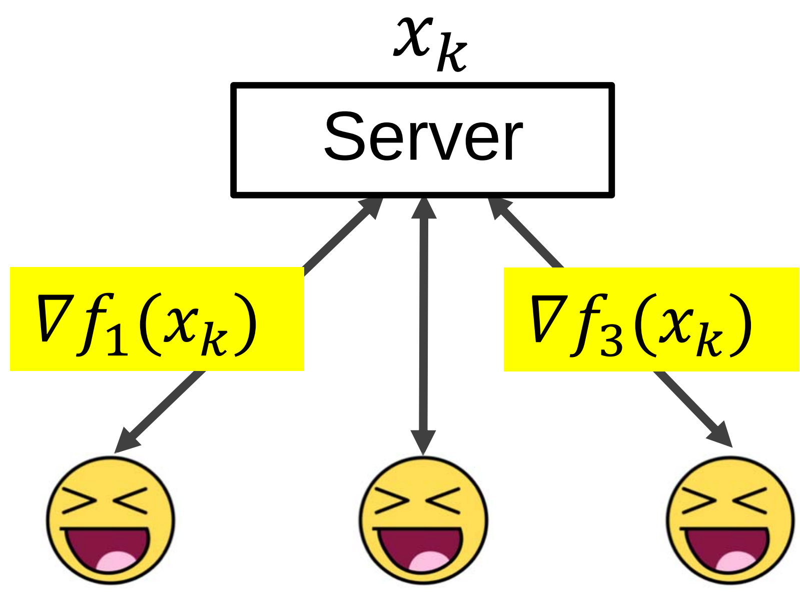


Parameter Server

- Server maintains estimate x_k

In each iteration

- Agent i
  - Receives x_k from server

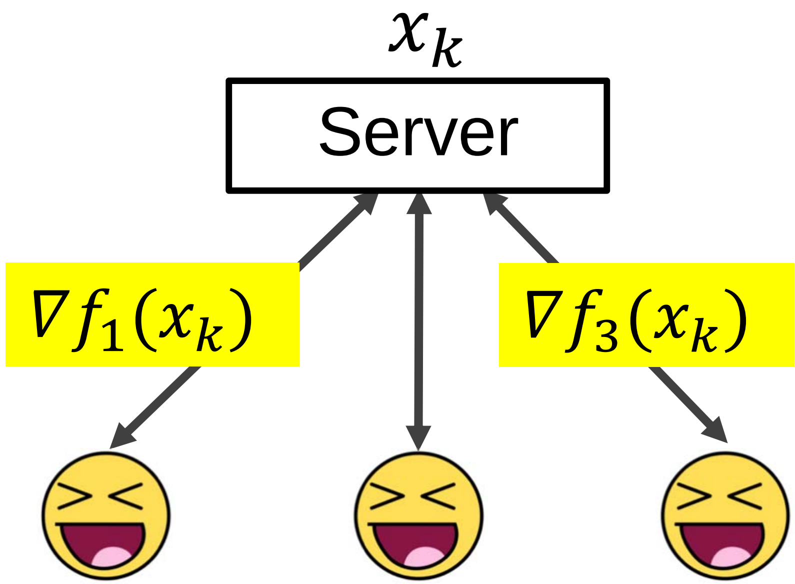


Parameter Server

- Server maintains estimate x_k

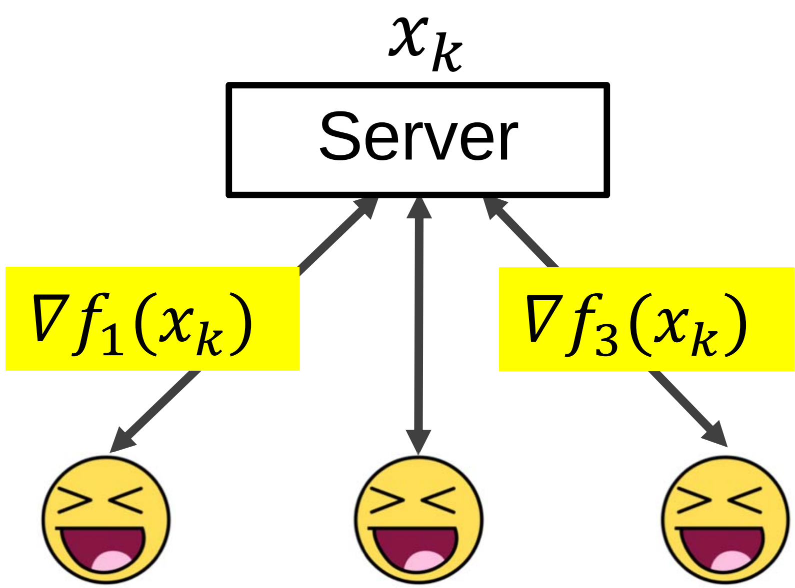
<u>In each iteration</u>

- Agent i
  - Receives x_k from server
  - Uploads gradient $\nabla f_i(x_k)$



Parameter Server

- Server maintains estimate x_k



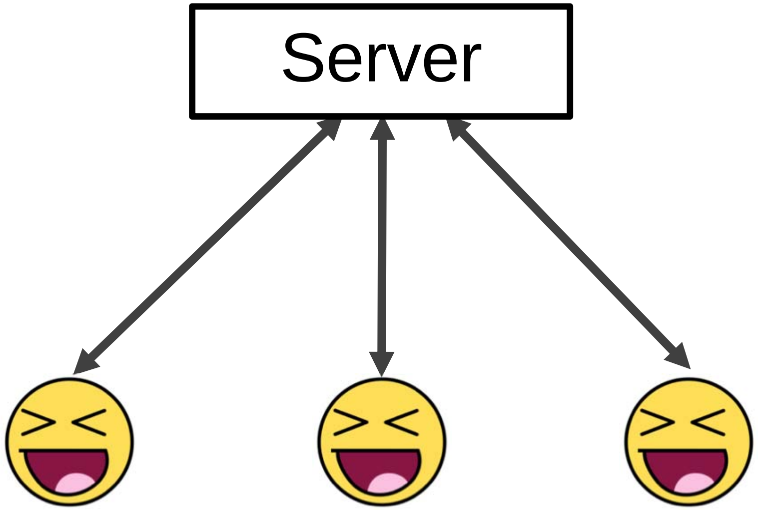
<u>In each iteration</u>

- Agent i
  - Receives x_k from server
  - Uploads gradient $\nabla f_i(x_k)$

- Server updates estimate

$$x_{k+1} \leftarrow x_k - \lambda \sum \nabla f_i(x_k)$$

Many Variations



... stochastic optimization

... asynchronous

... gradient compression

... acceleration

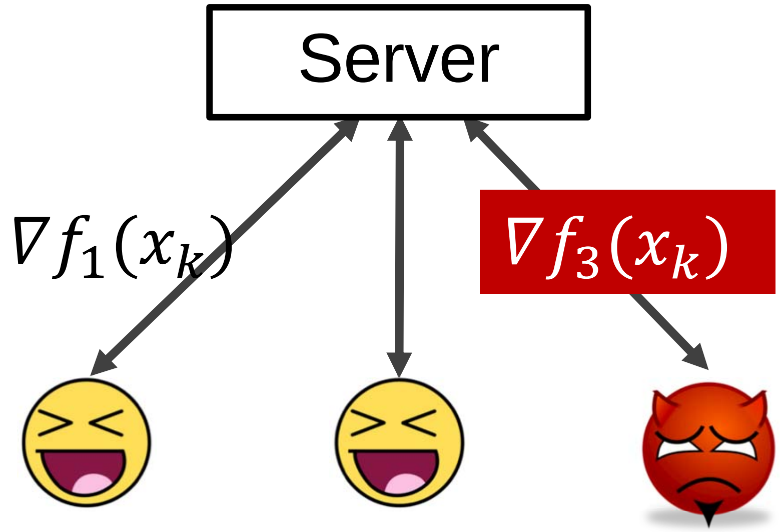
... shared memory

Challenges

Challenges

- Fault-tolerant distributed optimization

$$f_1(x) + f_2(x) + f_3(x)$$

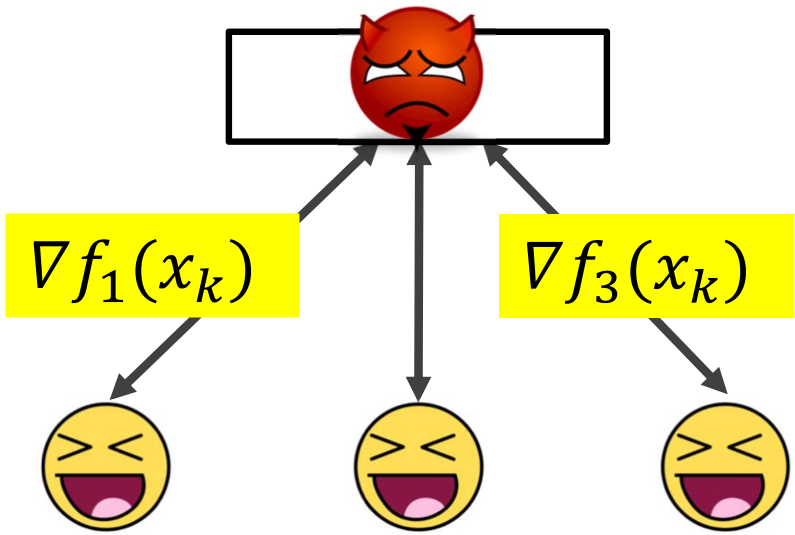


How to optimize if agents inject bogus information?

Challenges

- Privacy-preserving distributed optimization

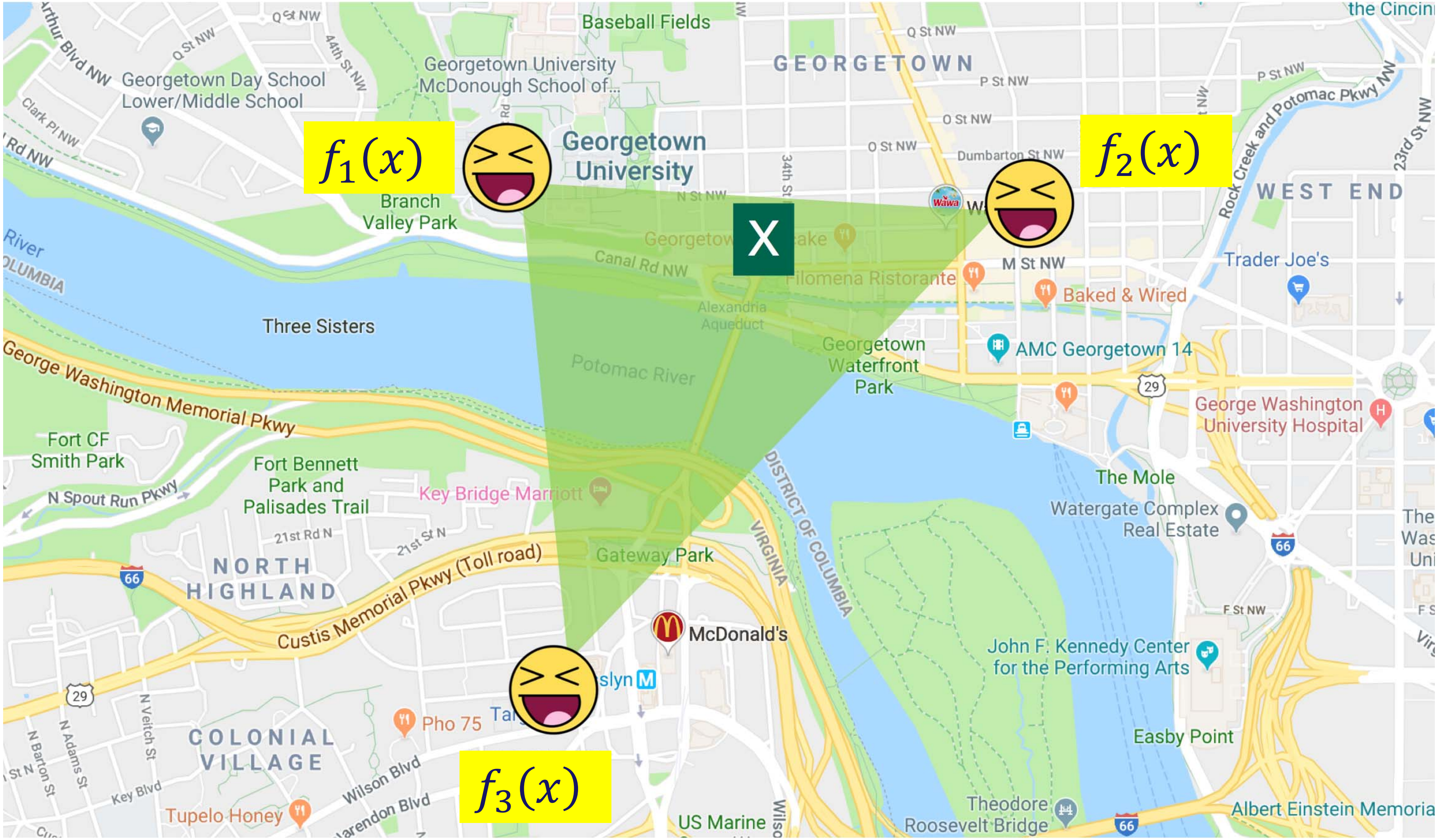
How to collaborate without revealing own cost function?



Secure / Fault-Tolerant Optimization

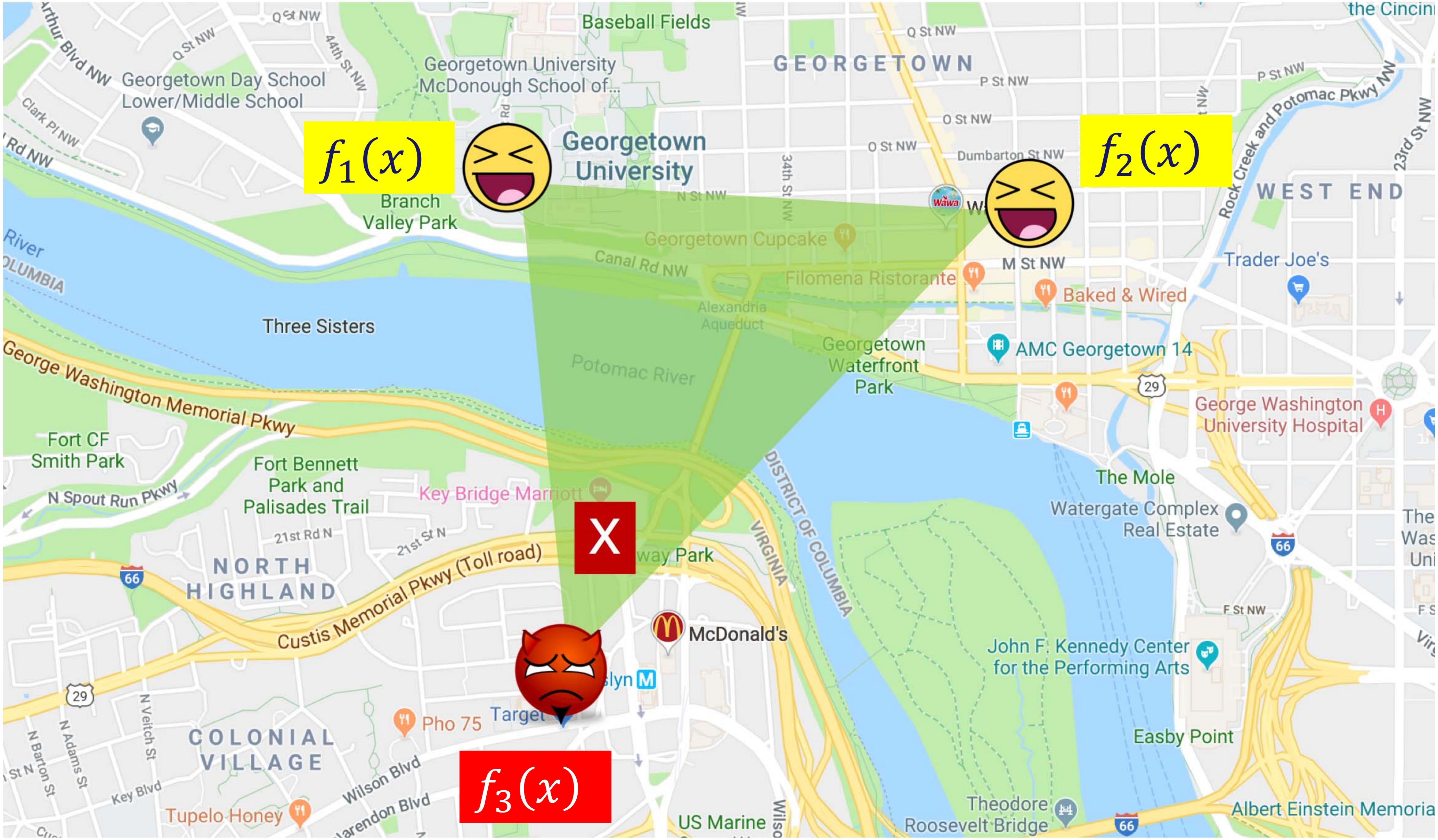
2015 …

Rendezvous



$f_1(x)$
$f_2(x)$
$f_3(x)$
X
Georgetown University
GEORGETOWN
WEST END
Three Sisters
Potomac River
George Washington Memorial Pkwy
Custis Memorial Pkwy (Toll road)
NORTH HIGHLAND
COLONIAL VILLAGE
Georgetown Waterfront Park
AMC Georgetown 14
George Washington University Hospital
The Mole
Watergate Complex Real Estate
John F. Kennedy Center for the Performing Arts
Key Bridge Marriott
Gateway Park
McDonald's
Theodore Roosevelt Bridge
Trader Joe's
Baked & Wired
Filomena Ristorante
Easby Point
Albert Einstein Memoria

Rendezvous



$f_1(x)$
$f_2(x)$
$f_3(x)$
X
Georgetown University
GEORGETOWN
WEST END
Potomac River
Three Sisters
George Washington Memorial Pkwy
Custis Memorial Pkwy (Toll road)
NORTH HIGHLAND
COLONIAL VILLAGE
Georgetown Waterfront Park
Key Bridge Marriott
McDonald's
George Washington University Hospital
Watergate Complex Real Estate
John F. Kennedy Center for the Performing Arts
Theodore Roosevelt Bridge

Machine Learning

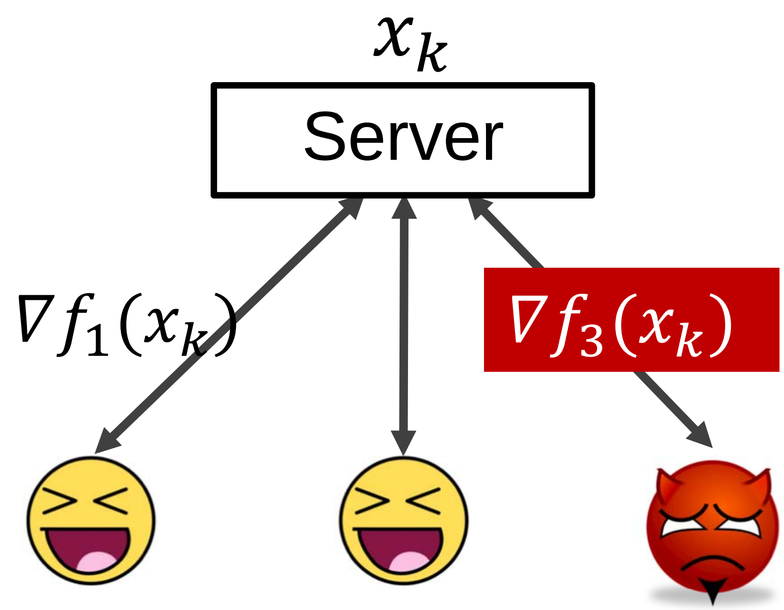
Faulty agent can adversely affect model parameters

Minimize global loss

$$\sum f_i(x)$$

Parameter Server

- Server maintains estimate x_k



<u>In each iteration</u>

- Agent i
  - Downloads x_k from server
  - Uploads gradient $\nabla f_i(x_k)$

- Server updates estimate

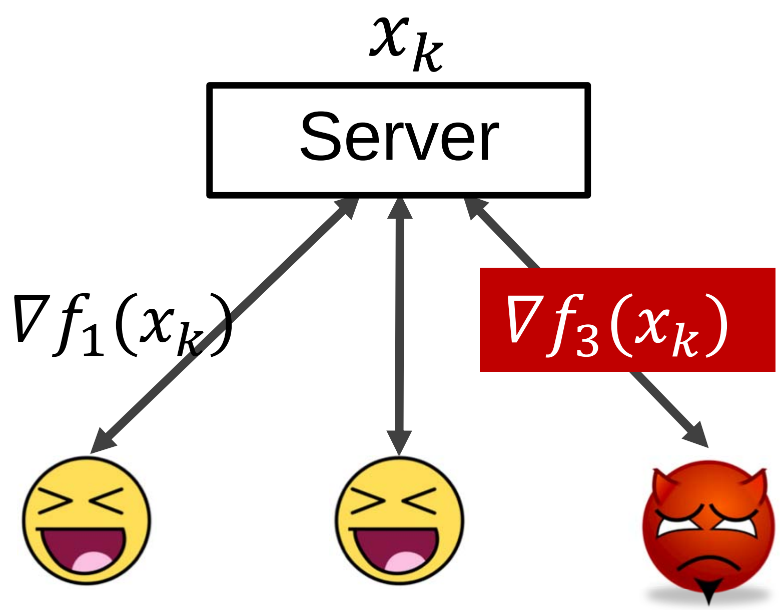
$$x_{k+1} \leftarrow x_k - \lambda \sum \nabla f_i(x_k)$$

Parameter Server

- Server maintains estimate x_k

<u>In each iteration</u>

- Agent i
  - Downloads x_k from server
  - Uploads gradient $\nabla f_i(x_k)$



- Server updates estimate

$$x_{k+1} \longleftarrow x_k - \lambda \, \text{Filtered−Gradient}$$

But what do we mean by fault-tolerance?

Fault-Tolerance

- Optimize over only good agents … set *G*

Fault-Tolerance

- Optimize over only good agents … set G

$$\operatorname{argmin} \sum_{i \in G} f_i(x)$$

$$\operatorname{argmin} \sum_{i \in G} f_i(x)$$

Is this achievable?

$$\operatorname{argmin} \sum_{i \in G} f_i(x)$$

Is this achievable?

It Depends

$$\operatorname{argmin} \sum_{i \in G} f_i(x)$$

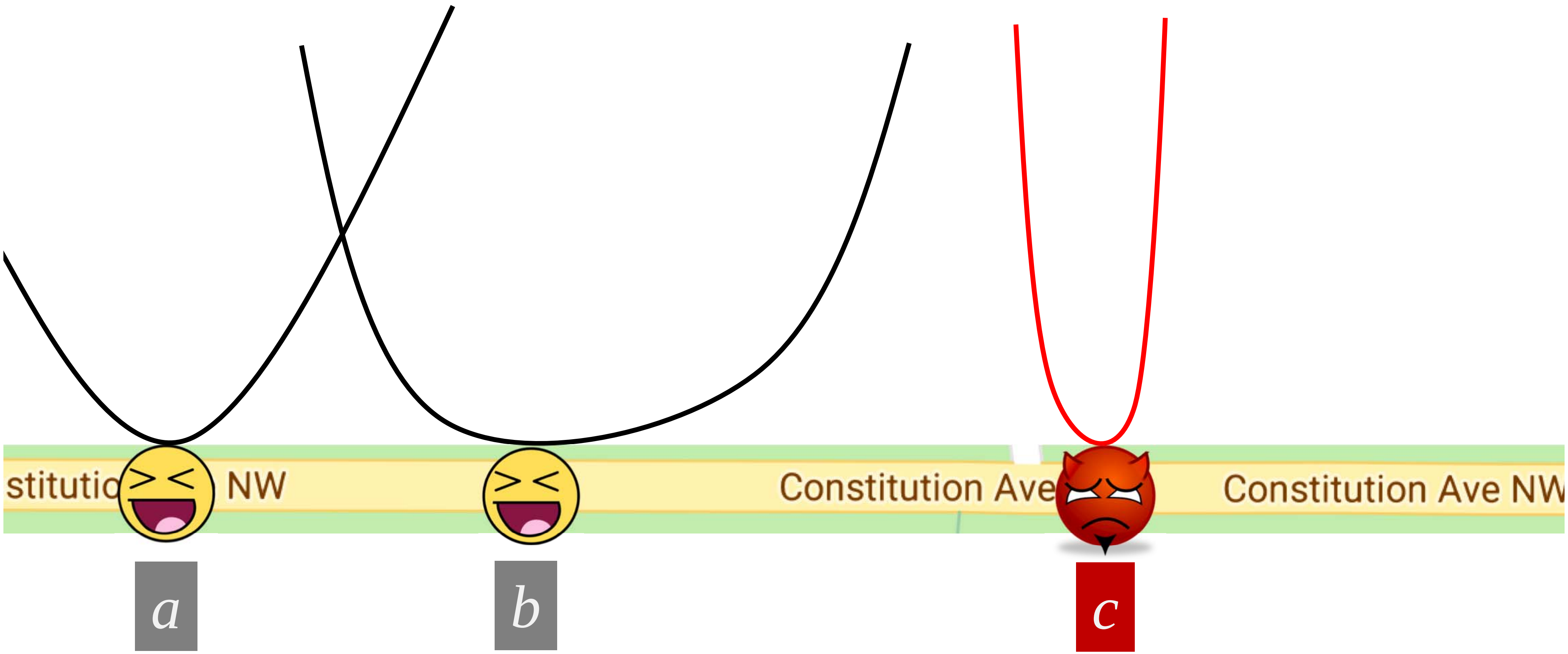
Is this achievable?

It Depends

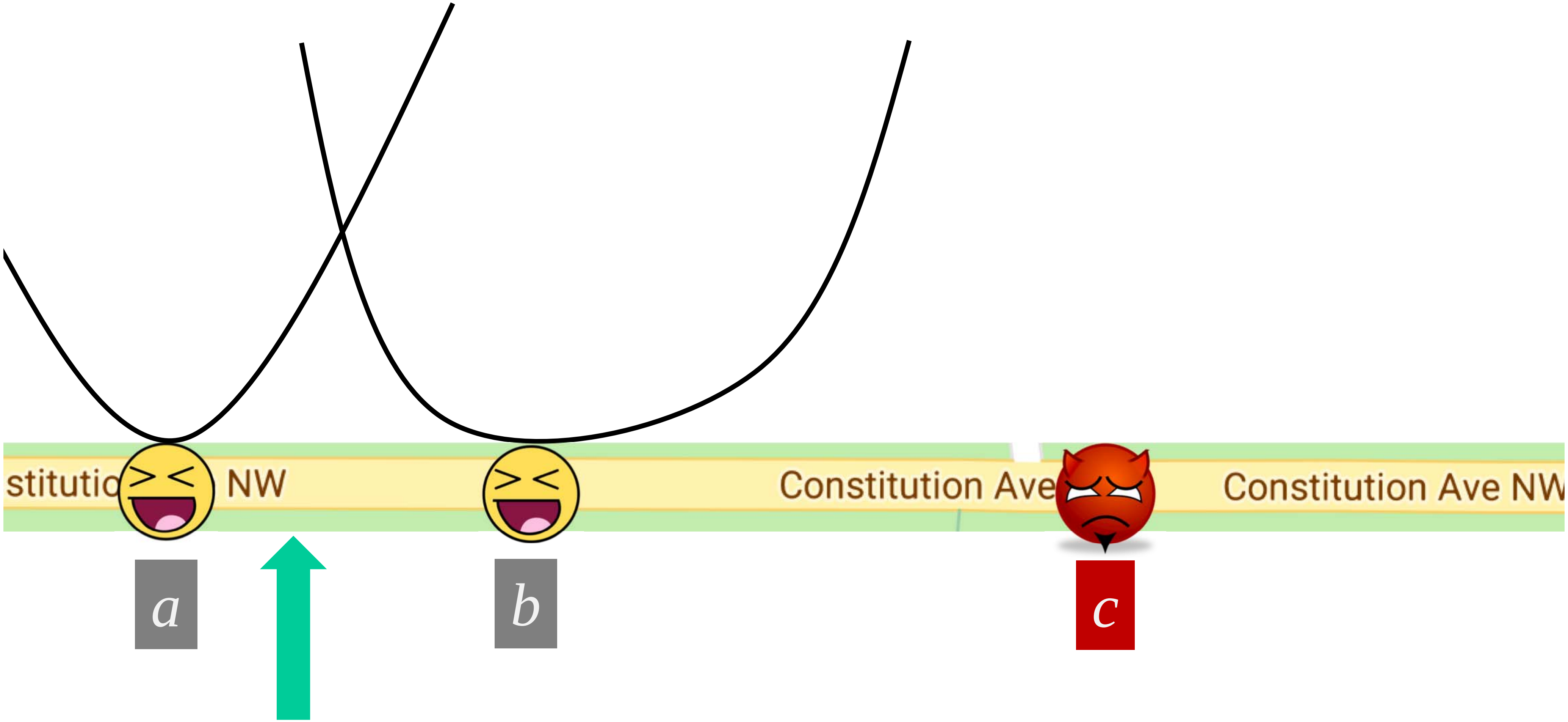
Independent functions

"Enough" redundancy

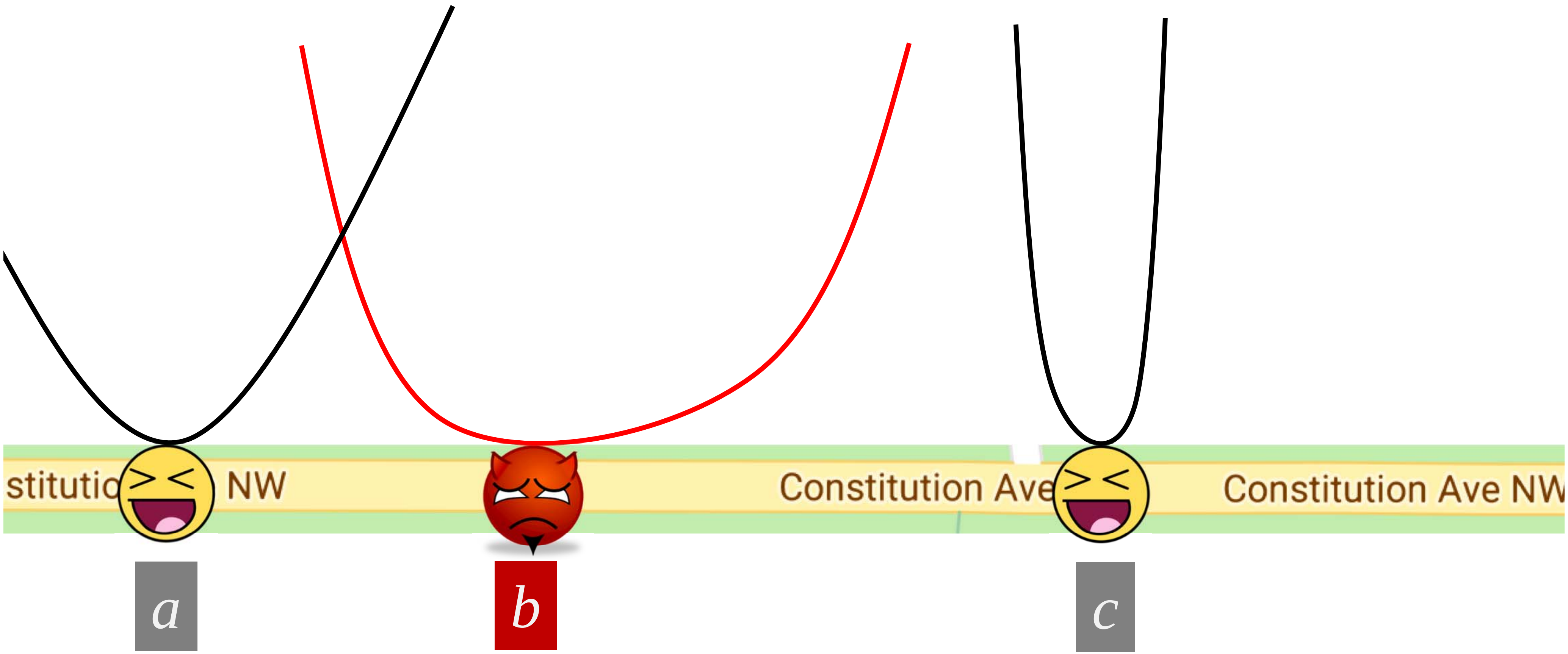
Independent Functions



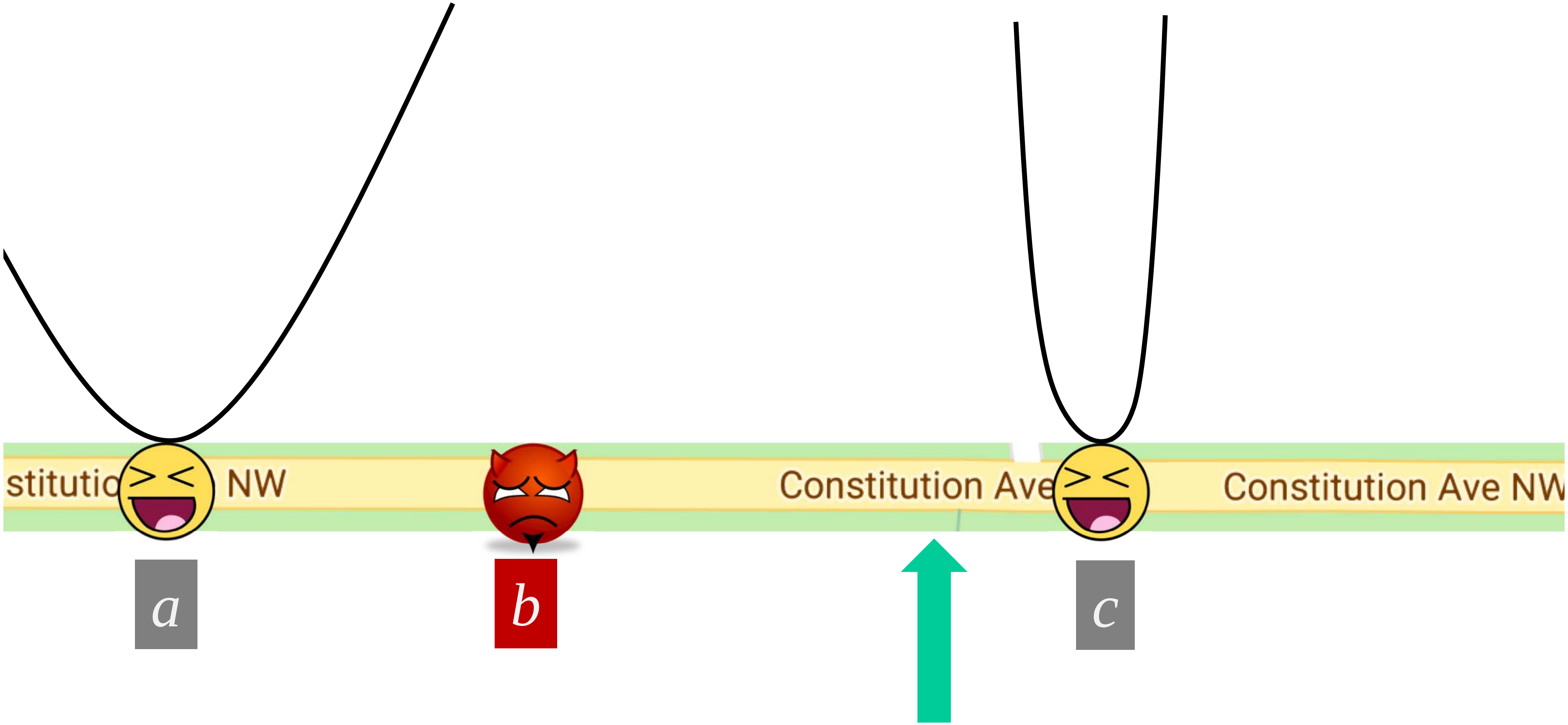
Independent Functions



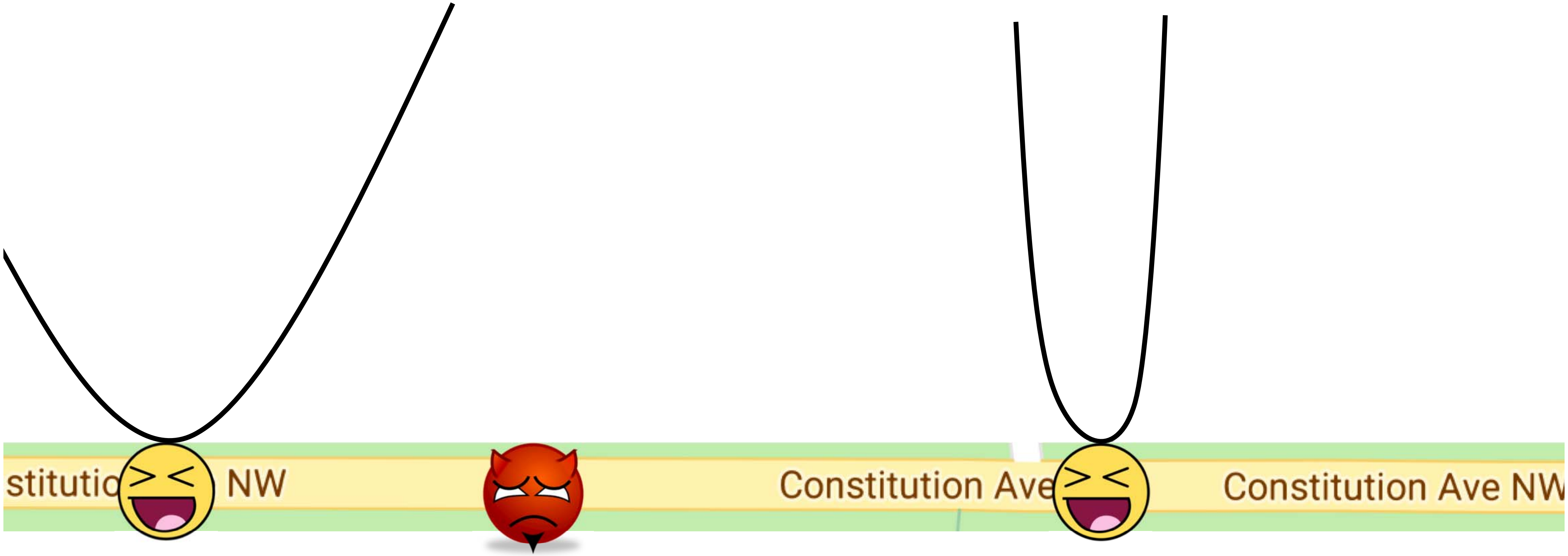
Independent Functions



Independent Functions



Independent Functions



Provably impossible to compute

$$\operatorname{argmin} \sum_{i \in G} f_i(x)$$

$$\operatorname{argmin} \sum_{i \in G} f_i(x)$$

Is this achievable?

Independent functions

"Enough" redundancy

Approximate

Exact

$$\operatorname{argmin} \sum_{i \in G} f_i(x)$$

Is this achievable?

Independent functions

"Enough" redundancy

Approximate

Exact

An Example of Redundancy

n agents

t bad agents

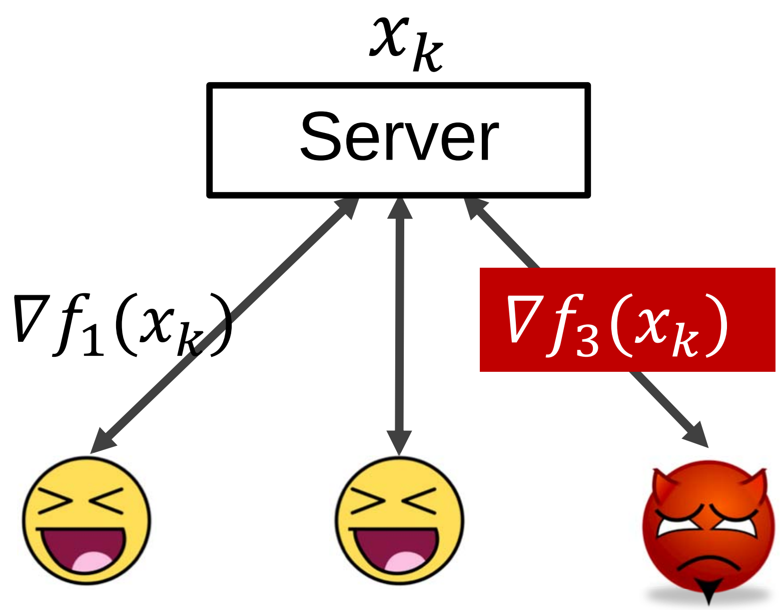
- Aggregate cost of ANY $n-2t$ agents has argmin identical to desired $\text{argmin} \sum_{i \in G} f_i(x)$

Parameter Server

- Server maintains estimate x_k

<u>In each iteration</u>

- Agent i
  - Downloads x_k from server
  - Uploads gradient $\nabla f_i(x_k)$



- Server updates estimate

$$x_{k+1} \leftarrow x_k - \lambda \, \text{Filtered–Gradient}$$

Norm Filter

- Clip the largest t norms to equal $t+1^{\text{th}}$ norm

$$|\nabla f_1(x_k)| = 1$$

$$|\nabla f_2(x_k)| = 3$$

$$|\nabla f_3(x_k)| = 2$$

Norm Filter

- Clip the largest t norms to equal $t+1$th norm

$$|\nabla f_1(x_k)| = 1$$

$$|\nabla f_2(x_k)| = 3$$

$$|\nabla f_3(x_k)| = 2$$

Filtered gradient = $\nabla f_1(x_k) + \frac{2}{3}\nabla f_2(x_k) + \nabla f_3(x_k)$

Norm Filter

- Clip the largest t norms to equal $t+1$th norm

$$|\nabla f_1(x_k)| = 1$$

$$|\nabla f_2(x_k)| = 3$$

$$|\nabla f_3(x_k)| = 2$$

Filtered gradient = $\nabla f_1(x_k) + \frac{2}{3}\nabla f_2(x_k) + \nabla f_3(x_k)$

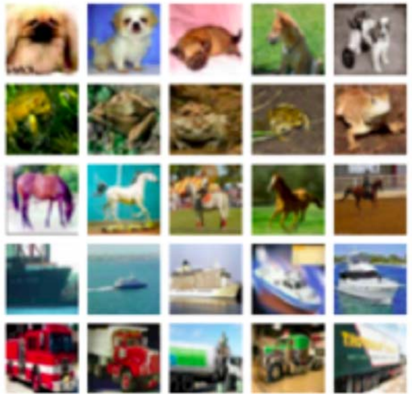
Exact optimum computed despite faulty agents

Another Example of Redundancy

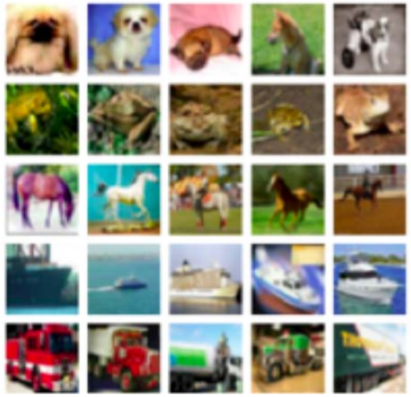
Another Example of Redundancy

- Machine learning

Agent 1



Agent 2



- Agents draw samples from identical data distribution

- Filter on stochastic gradients

Relaxing Redundancy Requirements

Ideas also extend to a relaxed notion of

"enough redundancy"

$$\underset{}{\operatorname{argmin}} \sum_{i \in G} f_i(x)$$

Is this achievable?

Independent functions

“Enough” redundancy

Approximate

Exact

How to Approximate $\operatorname{argmin} \sum_{i \in G} f_i(x)$?

How to Approximate $\operatorname{argmin} \sum_{i \in G} f_i(x)$?

Several alternatives for approximation

How to Approximate $\operatorname{argmin} \sum_{i \in G} f_i(x)$?

$$\operatorname{argmin} \sum_{i \in G} f_i(x) = \operatorname{argmin} \sum_{i \in G} \frac{1}{|G|} f_i(x)$$

How to Approximate $\arg\min \sum_{i \in G} f_i(x)$?

$$\arg\min \textstyle\sum_{i \in G} f_i(x) = \arg\min \textstyle\sum_{i \in G} \frac{1}{|G|} f_i(x)$$

- Ideal goal: Equal weight for all non-faulty agents

How to Approximate $\operatorname{argmin} \sum_{i \in G} f_i(x)$?

$$\operatorname{argmin} \sum_{i \in G} f_i(x) = \operatorname{argmin} \sum_{i \in G} \frac{1}{|G|} f_i(x)$$

- Ideal goal: Equal weight for all non-faulty agents
- Approximation: Unequal weights

Results

- For each faulty agent,
  a good agent may be ignored → Weight 0

Results

- For each faulty agent,
  a good agent may be ignored ➔ Weight 0

- But remaining $(n - 2t)$ good agents get "almost" uniform weight

n agents,
up to t faulty

Results

- For each faulty agent,
  a good agent may be ignored ➔ Weight 0

- But remaining $(n - 2t)$ good agents get "almost" uniform weight

n agents,
up to t faulty

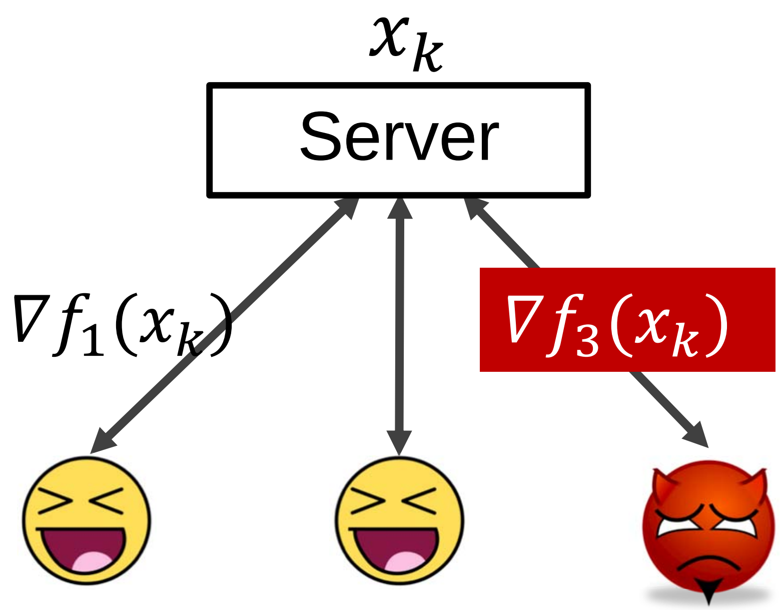
- Bad agents all get weight = 0

Parameter Server

- Server maintains estimate x_k

In each iteration

- Agent i
  - Downloads x_k from server
  - Uploads gradient $\nabla f_i(x_k)$



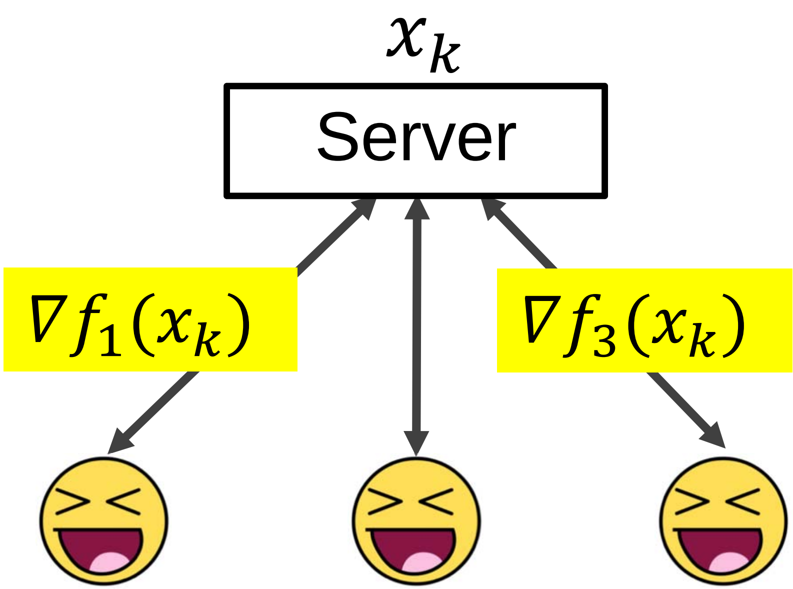
- Server updates estimate

$$x_{k+1} \leftarrow x_k - \lambda \, \text{Filtered-Gradient}$$

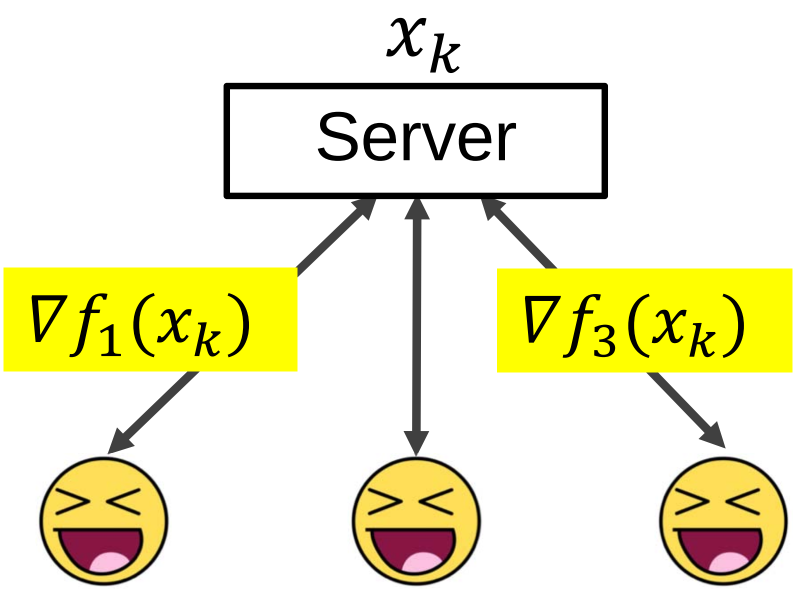
Privacy-Preserving Optimization

2016 …

Communication Leaks Information



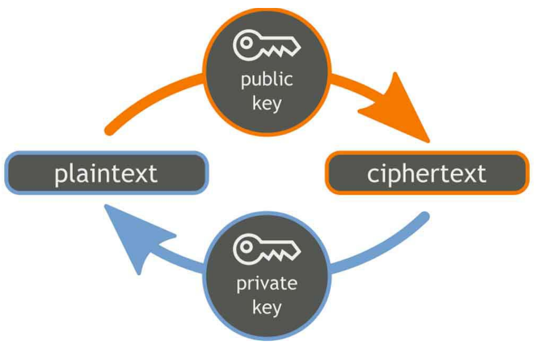
Communication Leaks Information



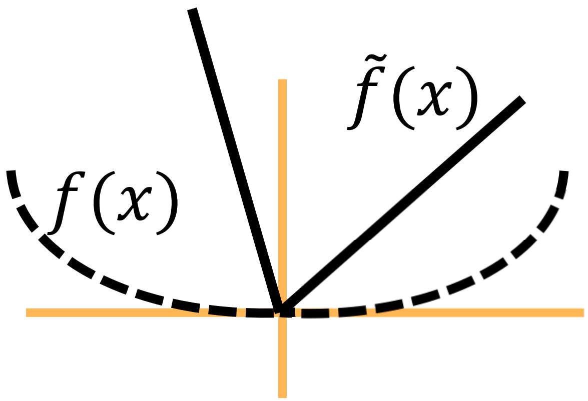
Server can use gradients to infer polynomial cost functions (up to a constant)

Related Work

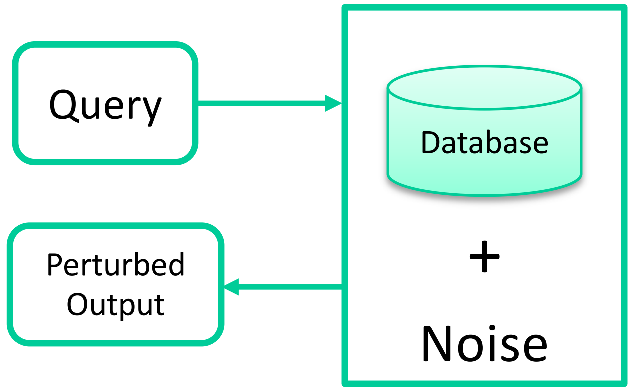
Cryptographic Methods



Transformation Methods



Differential Privacy

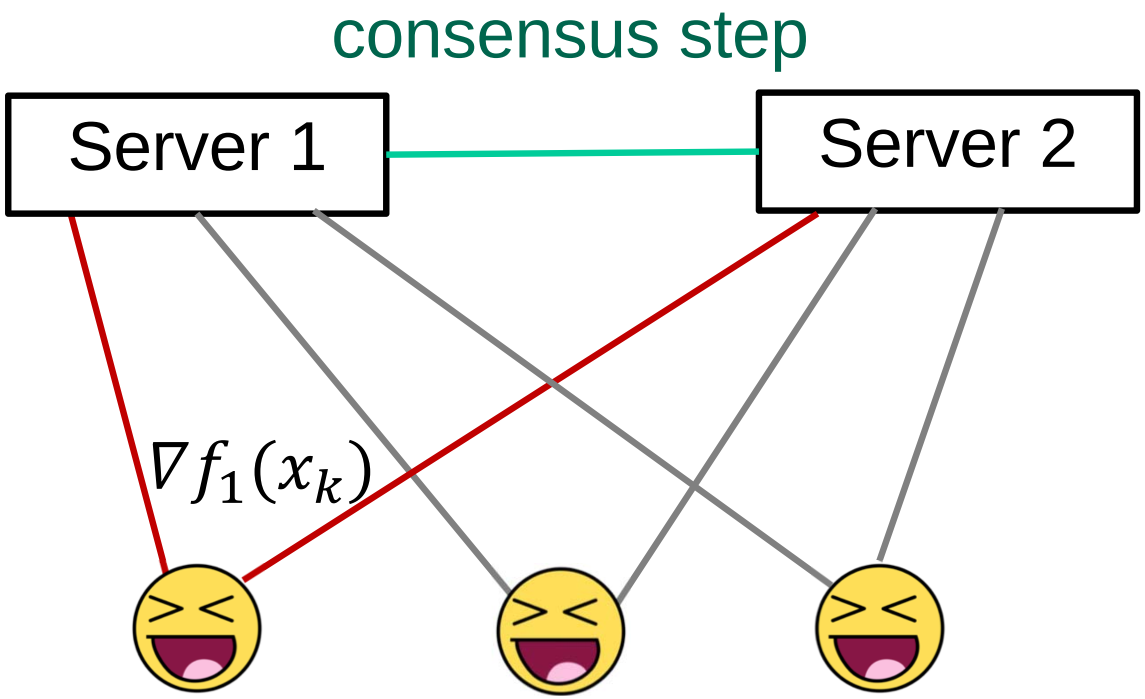


Our Approach

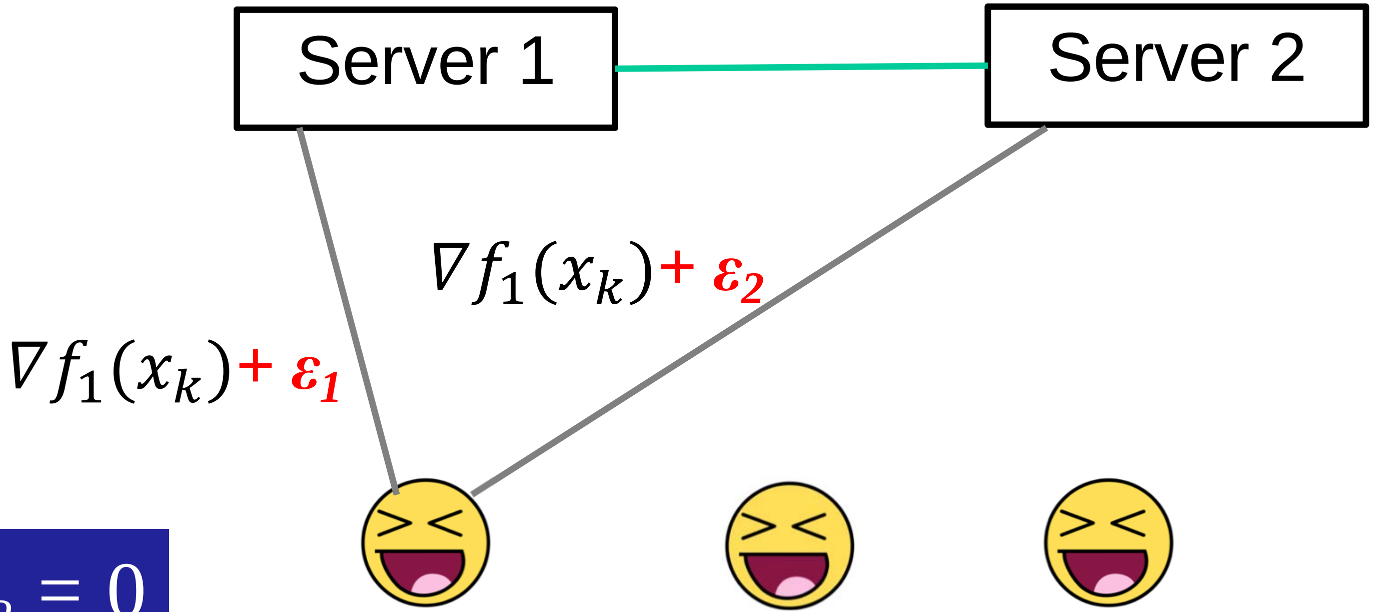
- Motivated by secret sharing & differential privacy

➔ Add cancellable noise

Multiple Parameter Servers

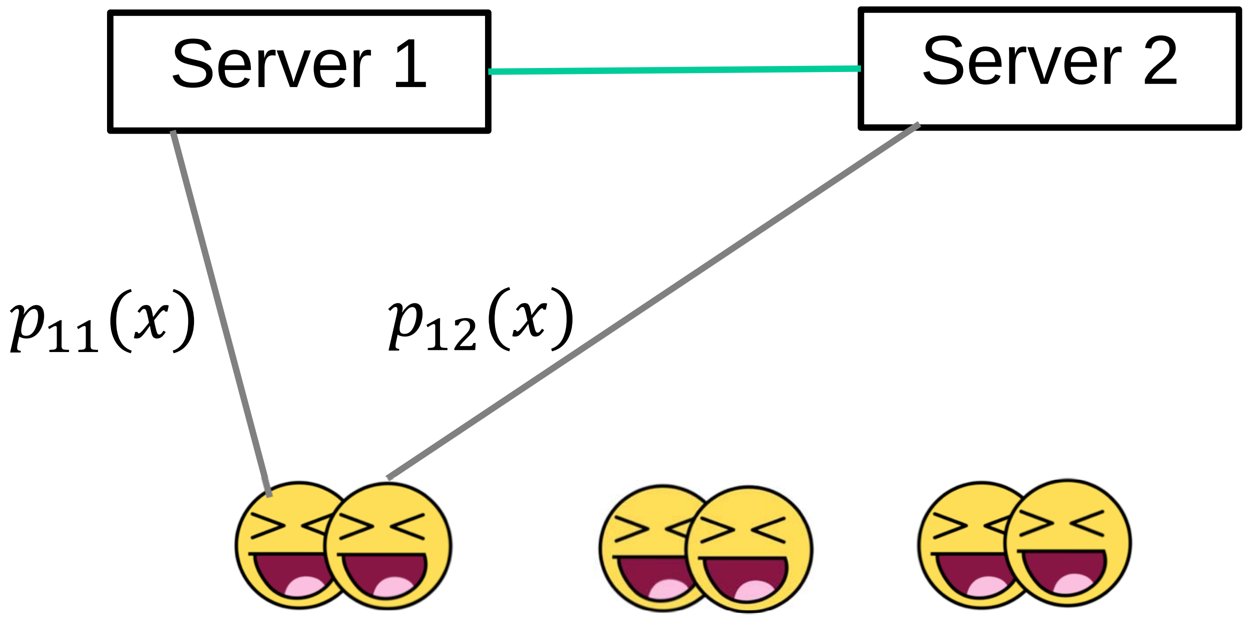


Improving Privacy



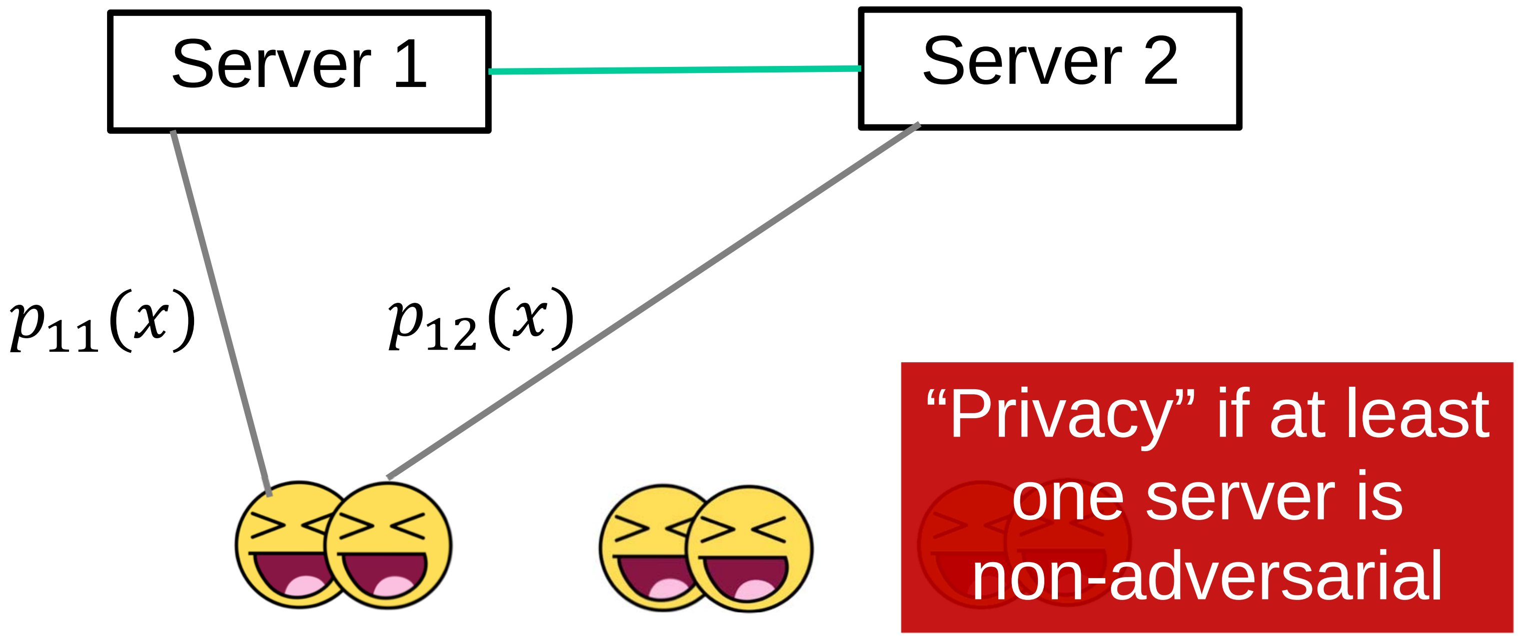
$\varepsilon_1 + \varepsilon_2 = 0$
over time
and space

Convex Sum of Non-Convex Functions



$$p_{11}(x) + p_{12}(x) = f_1(x)$$

Convex Sum of Non-Convex Functions



$$p_{11}(x) + p_{12}(x) = f_1(x)$$

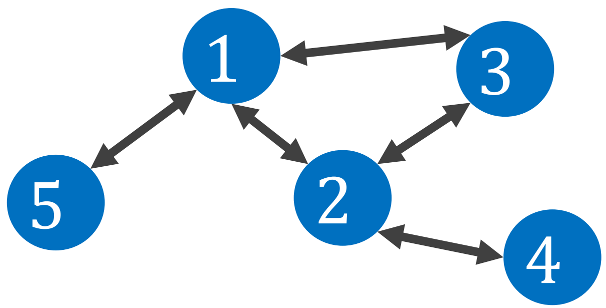
Summary: argmin $\sum f_i(x)$

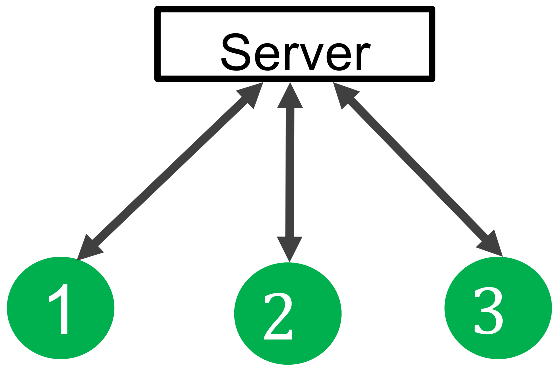
Fault-tolerance

➔ Gradient filters

Privacy

➔ Cancellable noise





Acknowledgments

- Lili Su
- Shripad Gade
- Nirupam Gupta
- Shuo Liu
- Connor Lu
- Dimitrios Pylorof

Thanks!

A longer tutorial at

disc.georgetown.domains ➔ Talks

Results
(Scalar x)

Can output

$$\operatorname{argmin} \sum_{i \in G} \alpha_i \, f_i(x)$$

for some weights (α_i)

where at least $(n - 2t)$ good agents have weight

$$\alpha_i \geq \frac{1}{2(n - 2t)}$$