Information Theory and Coding: From Distributed Systems to Blockchains

Mohammad Ali Maddah-Ali

Sharif University of Technology

The Story of Two Trends

Distributed Systems:
Distributing Tasks Among Parties

Scalability

Decentralized Systems:
Distributing Managements Among Parties

Transparency

Stablished Trend: Distributed Systems

Many Challenges:

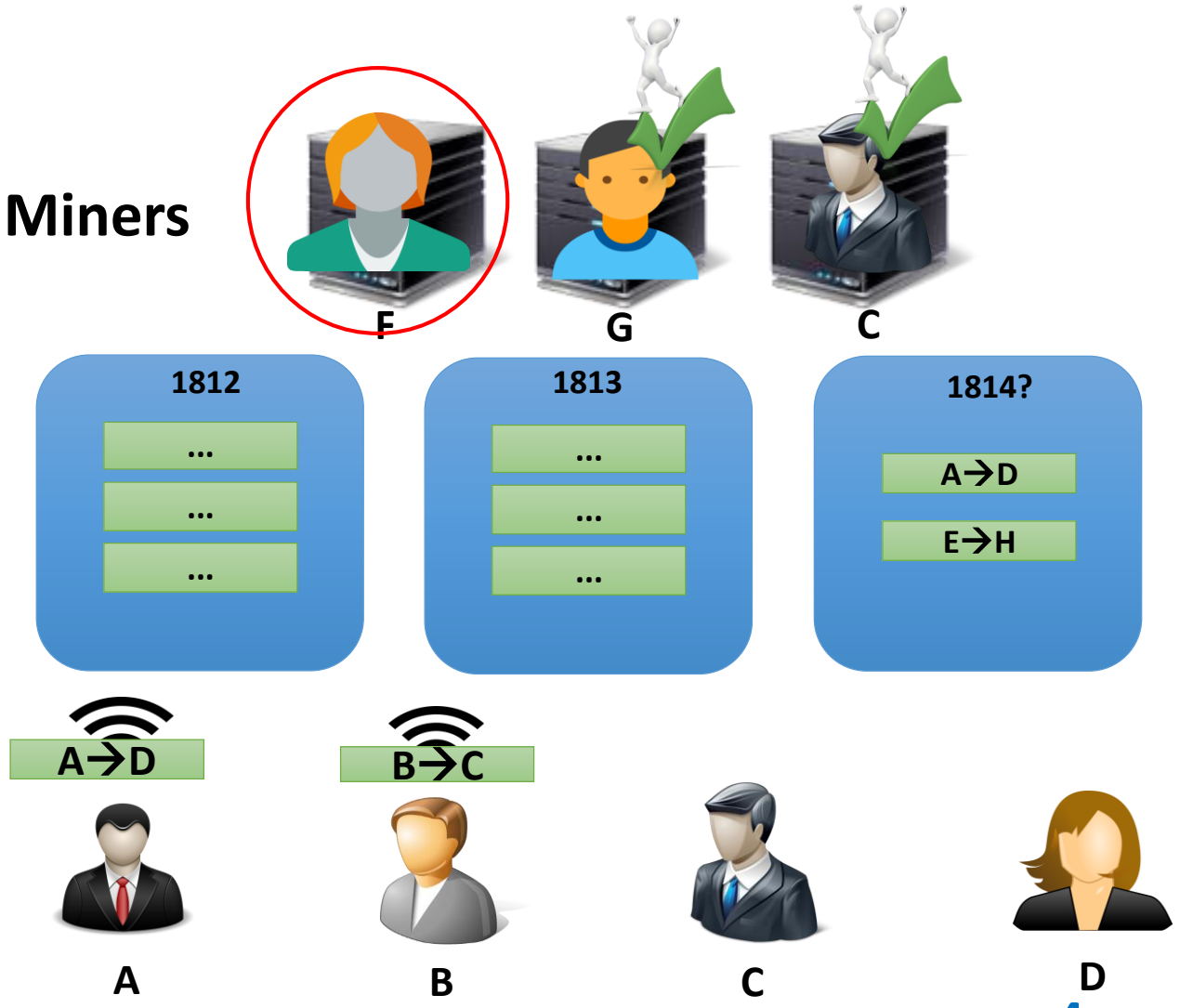
- Optimum interaction among
  - Storage
  - Computing
  - communication Resources
- Fault/Straggler Tolerance
- Privacy
- Synchronous vs Synchronous
- Scheduling
- …

Distributed Systems

Rising Trend: Decentralized Systems

Core Idea:

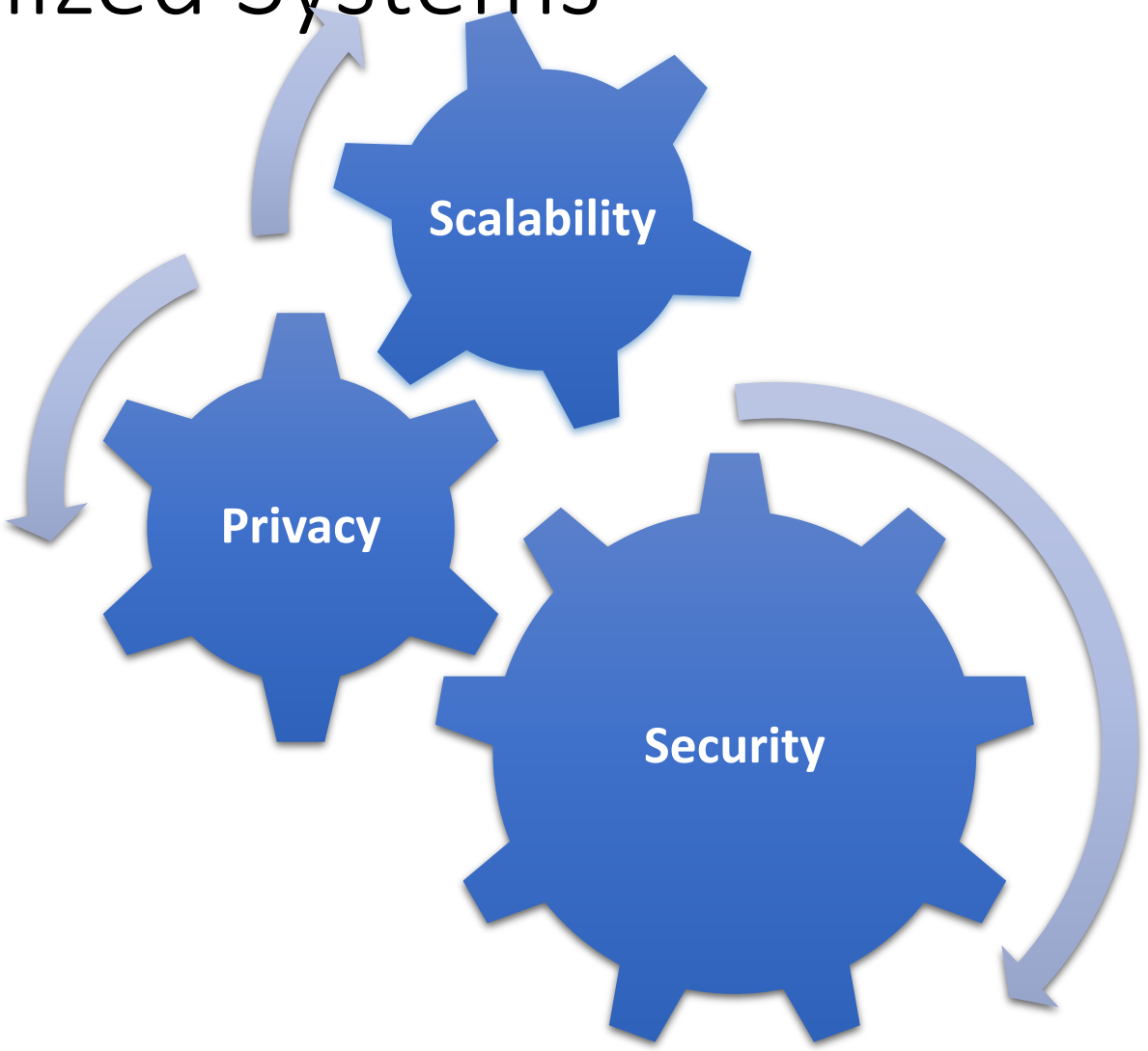
1. **One** miner **excuses** the Task
   - She is randomly selected
   - Chance is proportional to processing power
   - She is rewarded
2. **All** other miners **verify** his tasks
   - Voting right is proportional to Processing power
3. **All** keep a copy of database



Rising Trend: Decentralized Systems

Core Idea:

1. **One** miner **excuses** the Task
   - She is randomly selected
   - Chance is proportional to processing power
   - She is rewarded
2. **All** other miners **verify** his tasks
   - Voting right is proportional to Processing power
3. **All** keep a copy of database



Scalability vs Security

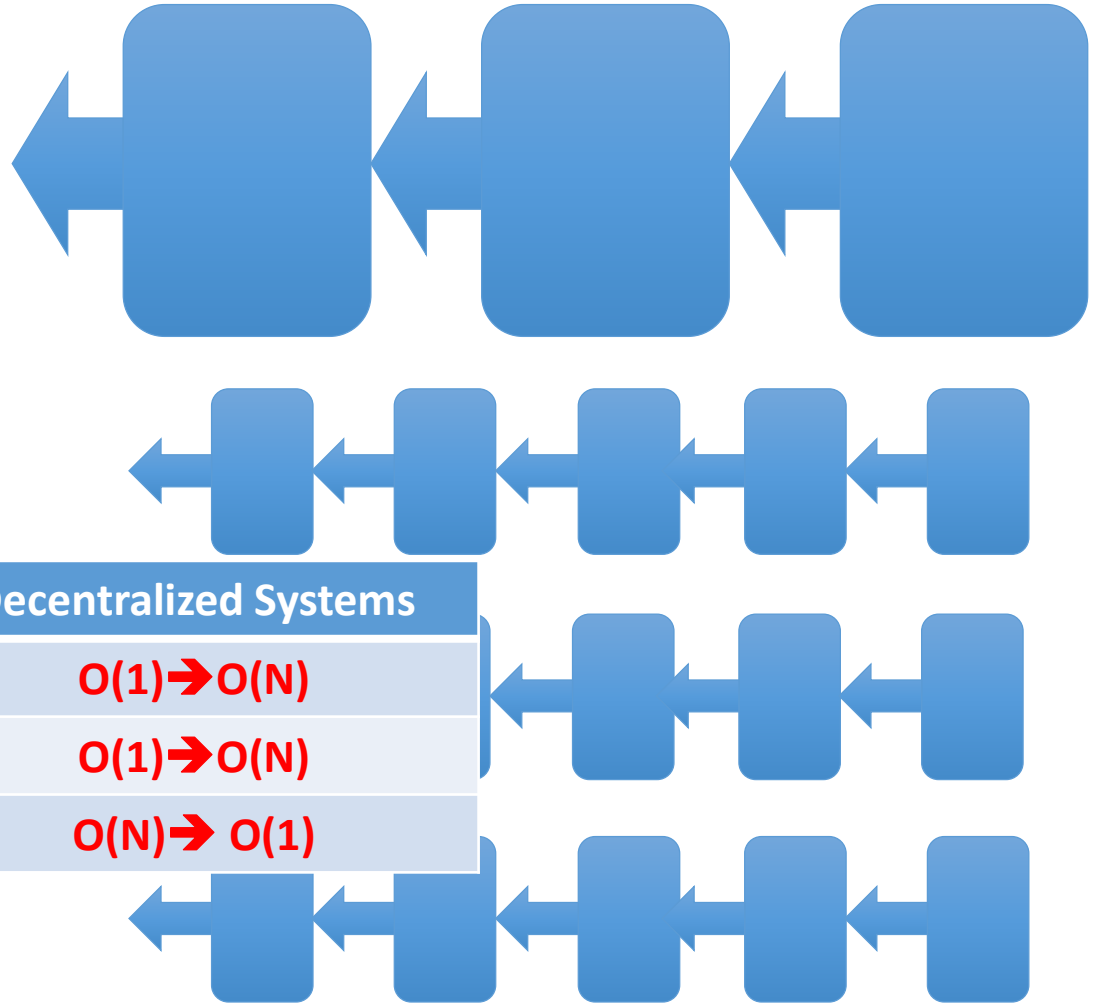
Contradiction!

	Distributed Systems	Decentralized Systems
Throughput/Processing	**O(N)**	**O(1)**
Storage	**O(N)**	**O(1)**
Security	**-**	**O(N)**

Sharding

Miners are randomly assigned to the shards

Transactions (tasks) are randomly assigned to the shards



	Distributed Systems	Decentralized Systems
Throughput/Processing	**O(N)**	**O(1)→O(N)**
Storage	**O(N)**	**O(1)→O(N)**
Security	**-**	**O(N)→ O(1)**

Sharding

Advantage:

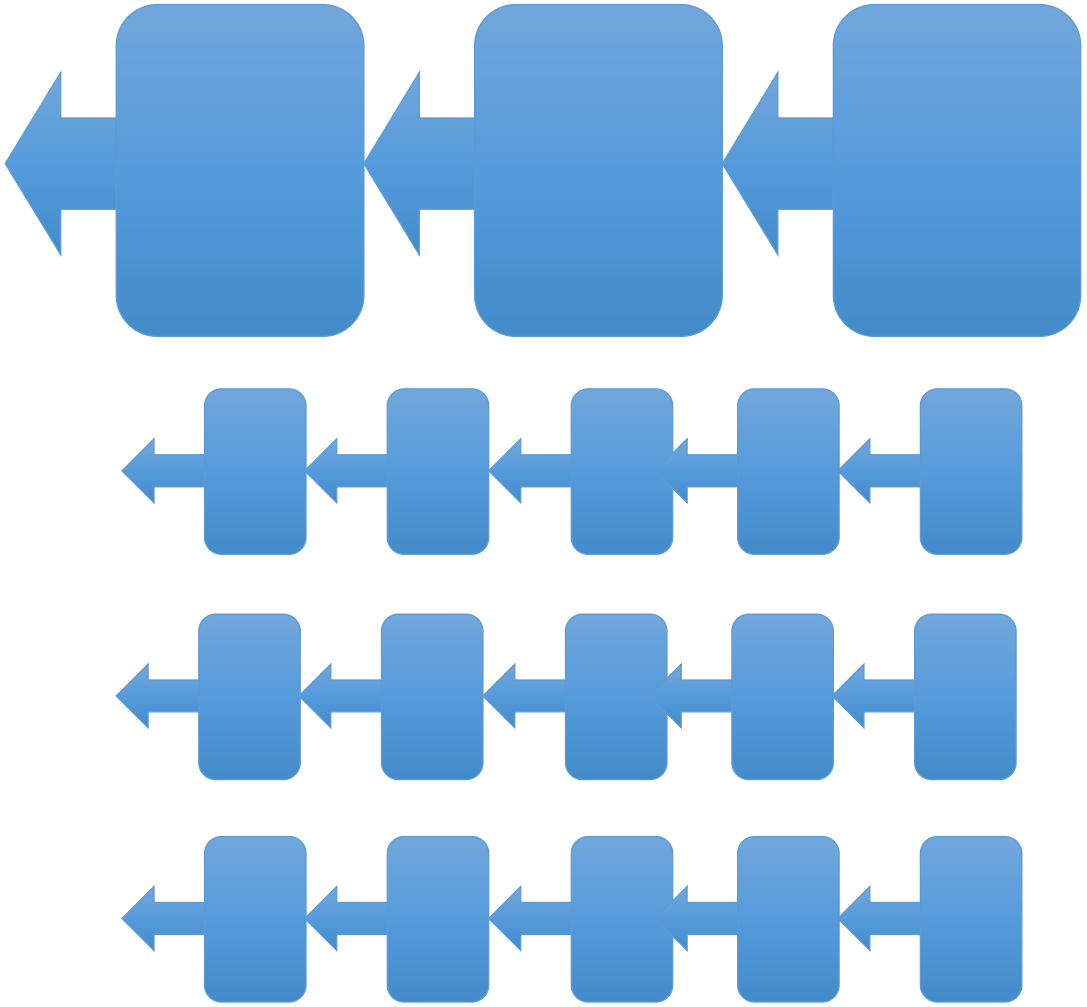
- Scalability

Challenge:

- Each shard is vulnerable to
  - ✓ Attacks
  - ✓ Faults

How to resolve this issue?

Coding Across Shards



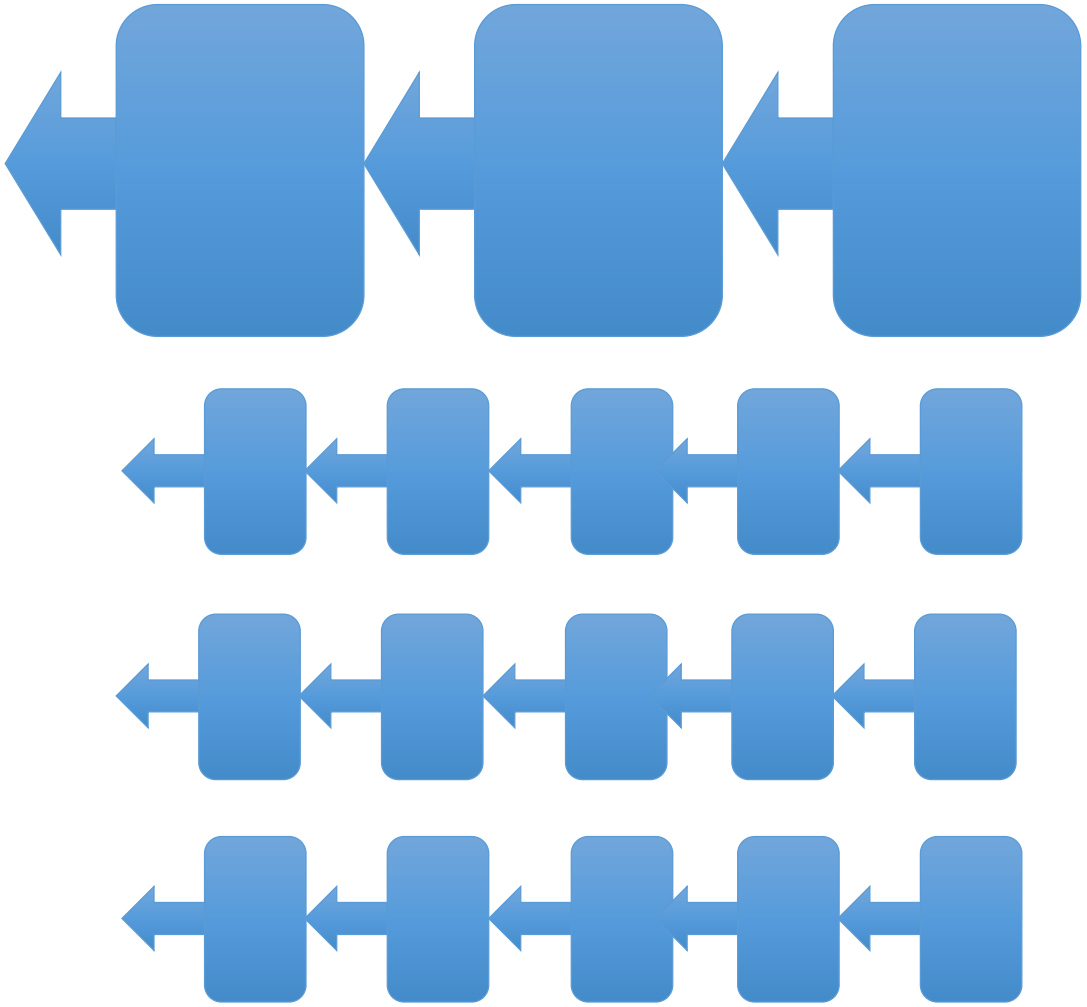
Coding in Sharding

Core Idea: Coding for Storage

✓ Each nodes stores some coded data of some other shards

$$\alpha_1 X_1 + \alpha_2 X_2 + \ldots$$

A Subset of nodes can recover the entire data
If many of nodes fail, still the data is protected



Li et. al. [2018]
Abadi & Maddah-Ali [2019]
Badihi and Maddah-Ali [2019]

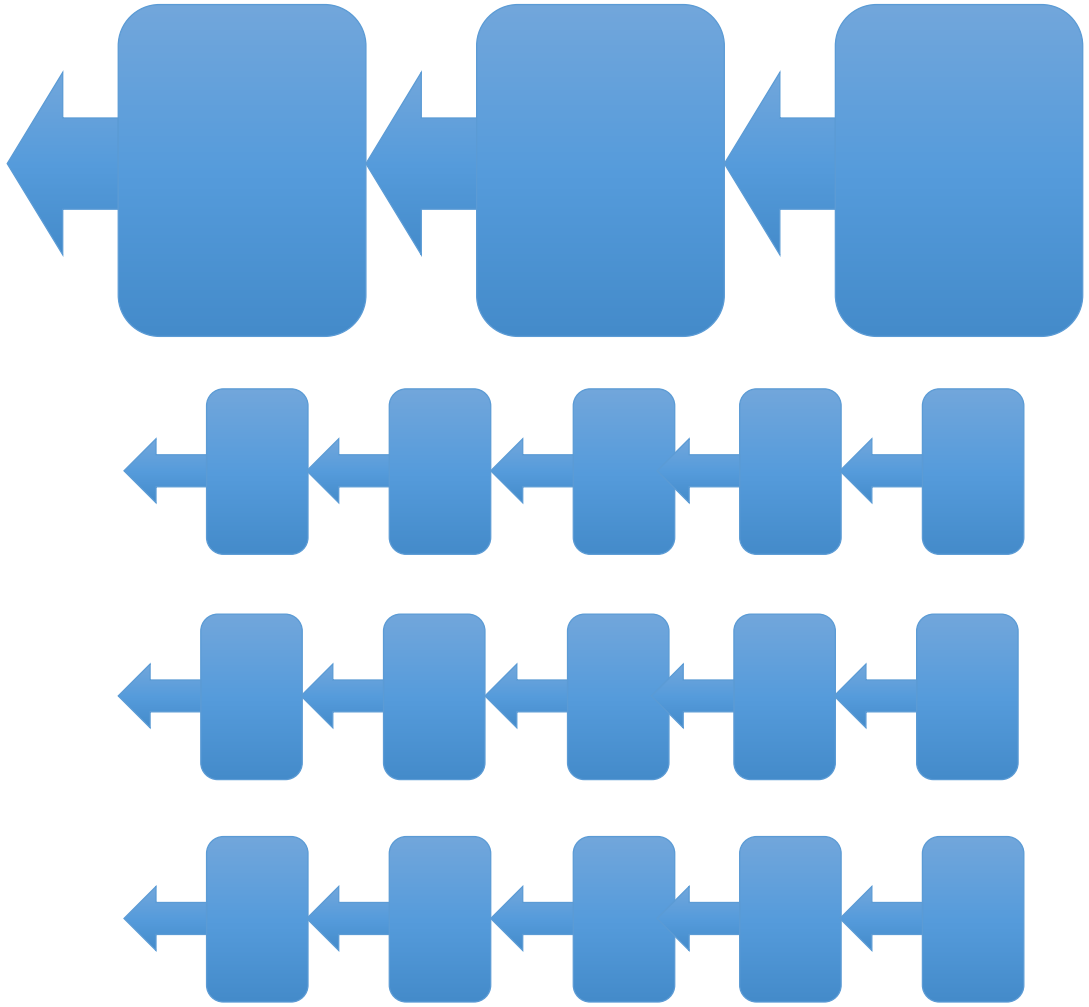
Coding in Sharding

Core Idea: Coding for Storage

- ✓ Each nodes stores some coded data of some other shards

Tradeoff Among

- ✓ Storage per Node
- ✓ Bandwidth Recovery
- ✓ Security Against Faults
- ✓ Power of Adversary

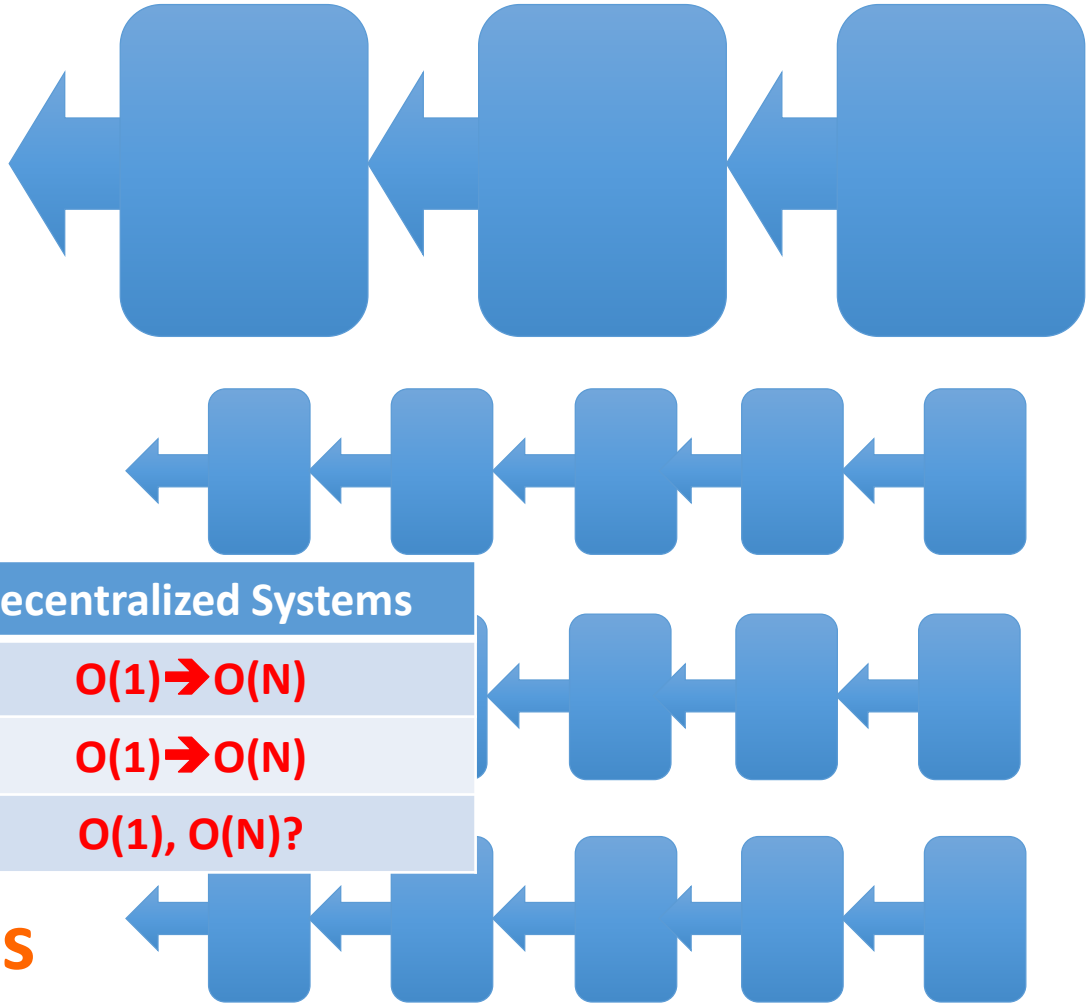


Li et. al. [2018]
Abadi & Maddah-Ali [2019]
Badihi and Maddah-Ali [2019]

Coding in Sharding

Core Idea: Coding for Storage

✓ Each nodes stores some coded data of some other shards



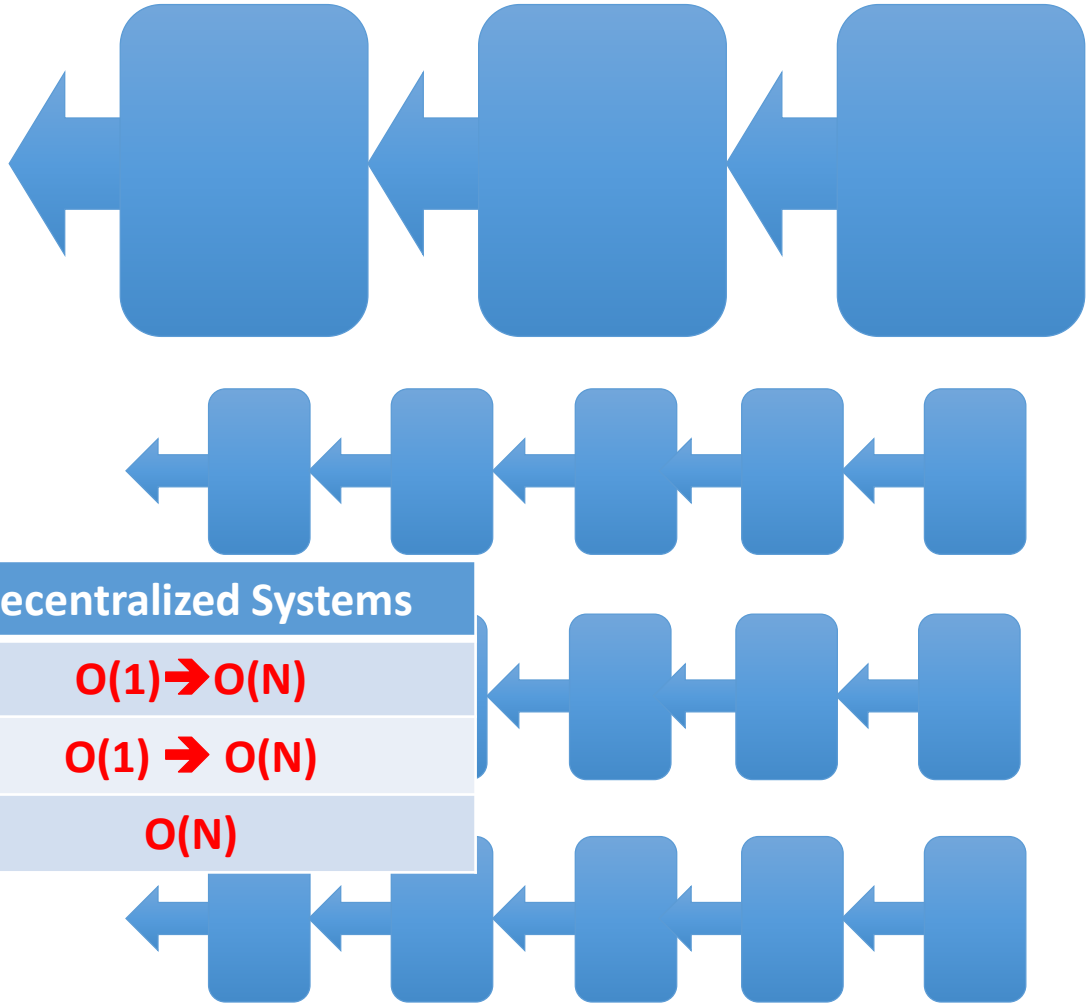
	Distributed Systems	Decentralized Systems
Throughput/Processing	**O(N)**	**O(1)→O(N)**
Storage	**O(N)**	**O(1)→O(N)**
Security	**-**	**O(1), O(N)?**

Challenge: Not everybody verifies everything!

Coding in Sharding

Core Idea:

- ✓ **Coding for Computing**
  - ✓ Coding that protect computing as well

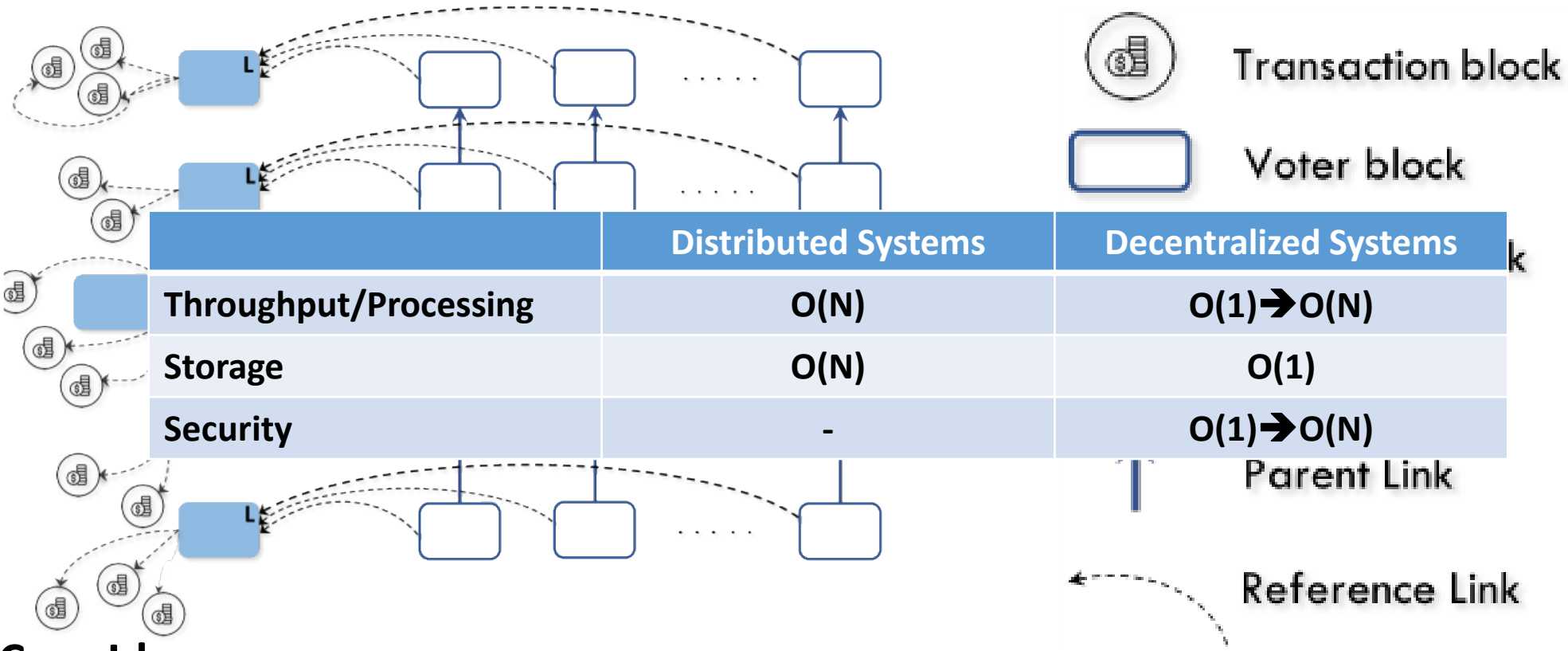


	Distributed Systems	Decentralized Systems
Throughput/Processing	**O(N)**	**O(1)➔O(N)**
Storage	**O(N)**	**O(1) ➔ O(N)**
Security	**-**	**O(N)**

$$\text{Verify}(\alpha_1 X_1 + \alpha_2 X_2 ...)$$

Li et. al. [2018]

Prism: Different Mentality



	Distributed Systems	Decentralized Systems
Throughput/Processing	O(N)	O(1)➔O(N)
Storage	O(N)	O(1)
Security	-	O(1)➔O(N)

Core Idea:

- ✓ **Sorting Transactions**
- ✓ **Majority voting**

Claim: Achieving Physical Limits of The Network

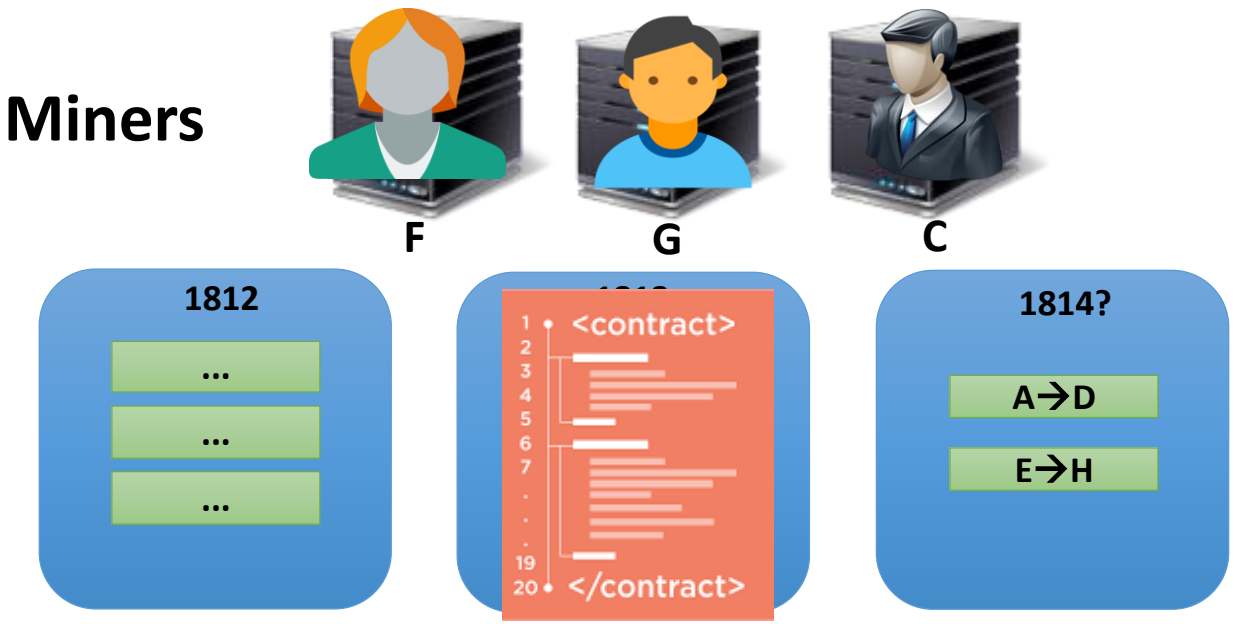
Bagaria et.al [2018]

Privacy vs Security

Smart Contracts

A Program

- Deployed on Blockchain
- Run each time by **a miner**
- The run is verified by **all other miners**



Expanding the applications of blockchains, beyond imagination

Zero-Knowledge Proofs

In Zero-Knowledge Proof:

Convincing a verifier you know the answer of a solution without revealing the solution

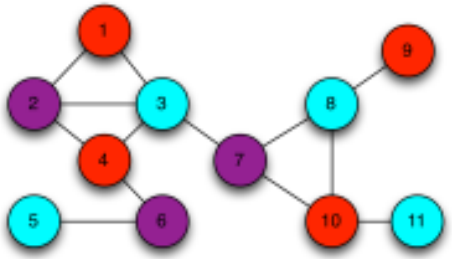
Graph 3 coloring

- **We want to prove that a graph, namely G, has a 3 coloring**

Solution:

The prover colors the graph (which he claims he knows how to)

Then he covers his solution with hats!



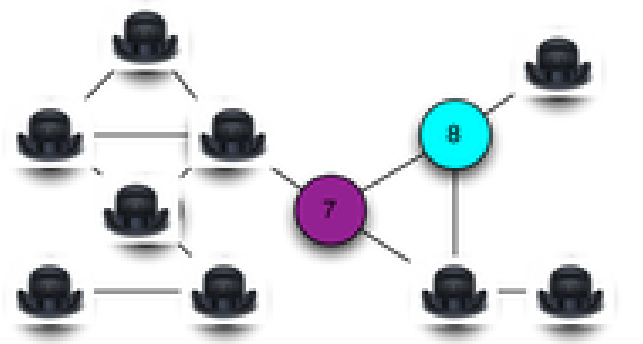
Zero-Knowledge Proofs

Graph 3 coloring

- **We want to prove that a graph, namely G, has a 3 coloring**

Solution (cntd.): now the verifier randomly picks an edge and the prover reveals the two vertices!

- **If the colors are the same the prover is convinced that the prover is lying**
- **If the colors are differe onvinced!**



Zero-Knowledge Snarks

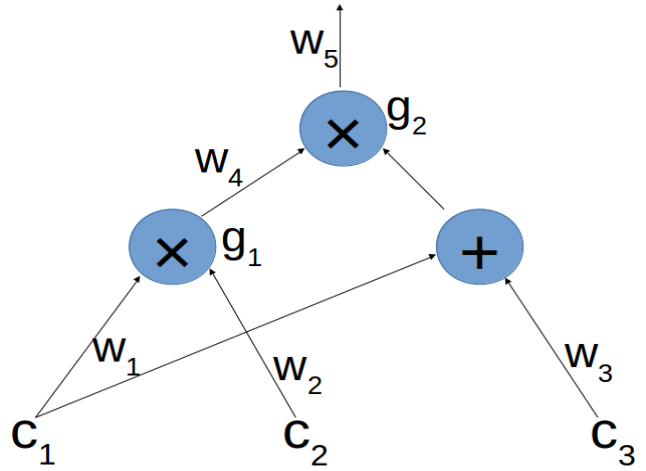
Task➔

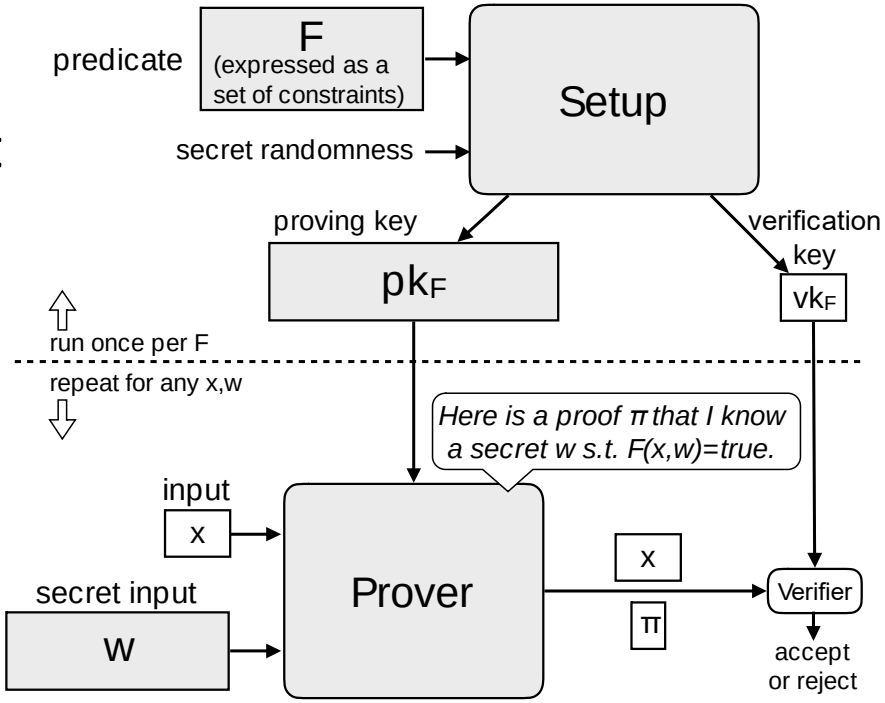
Arithmetic Circuit ➔

Rank 1 Constrains System ➔

Quadratic Arithmetic Program ➔

Verifying the polynomial at an encrypted secure point





Zcash I prove I have some money and can spend it!

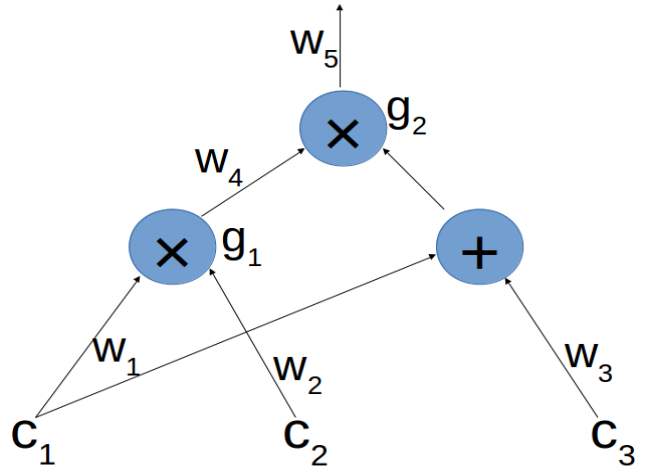
Challenge: Increasing computation load orderwise!

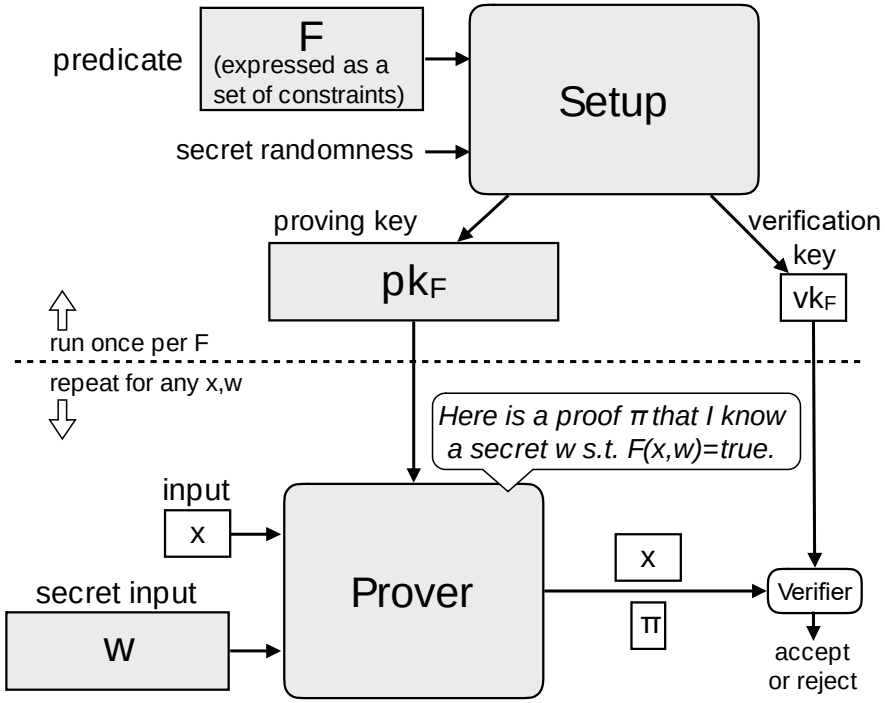
Zero-Knowledge Snarks

How: We need to distribute zero-knowledge proof to some untrusted servers!

Tool: Multi Party Computation

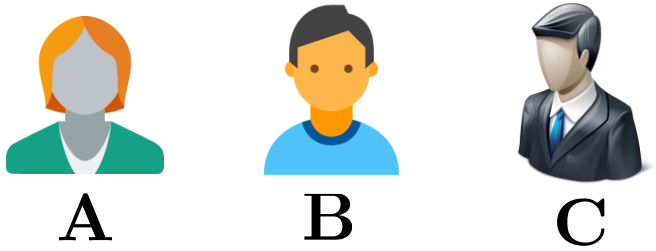
Challenge: Conventional Multiparty computation does not work!





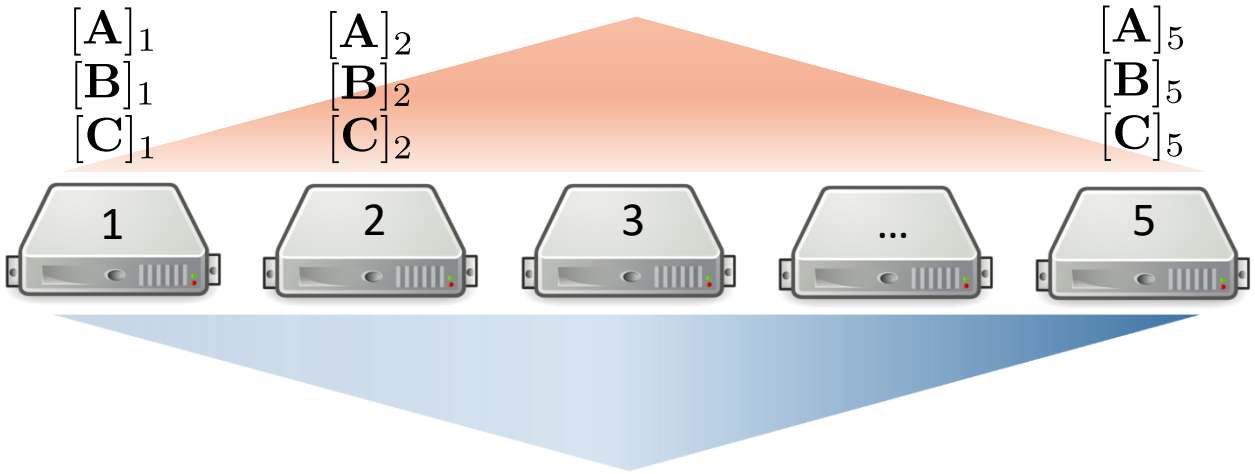
Multiparty Computation

Source



Huge Private Matrices

N Unreliable Servers



Servers can talk to each other.

Data Collector

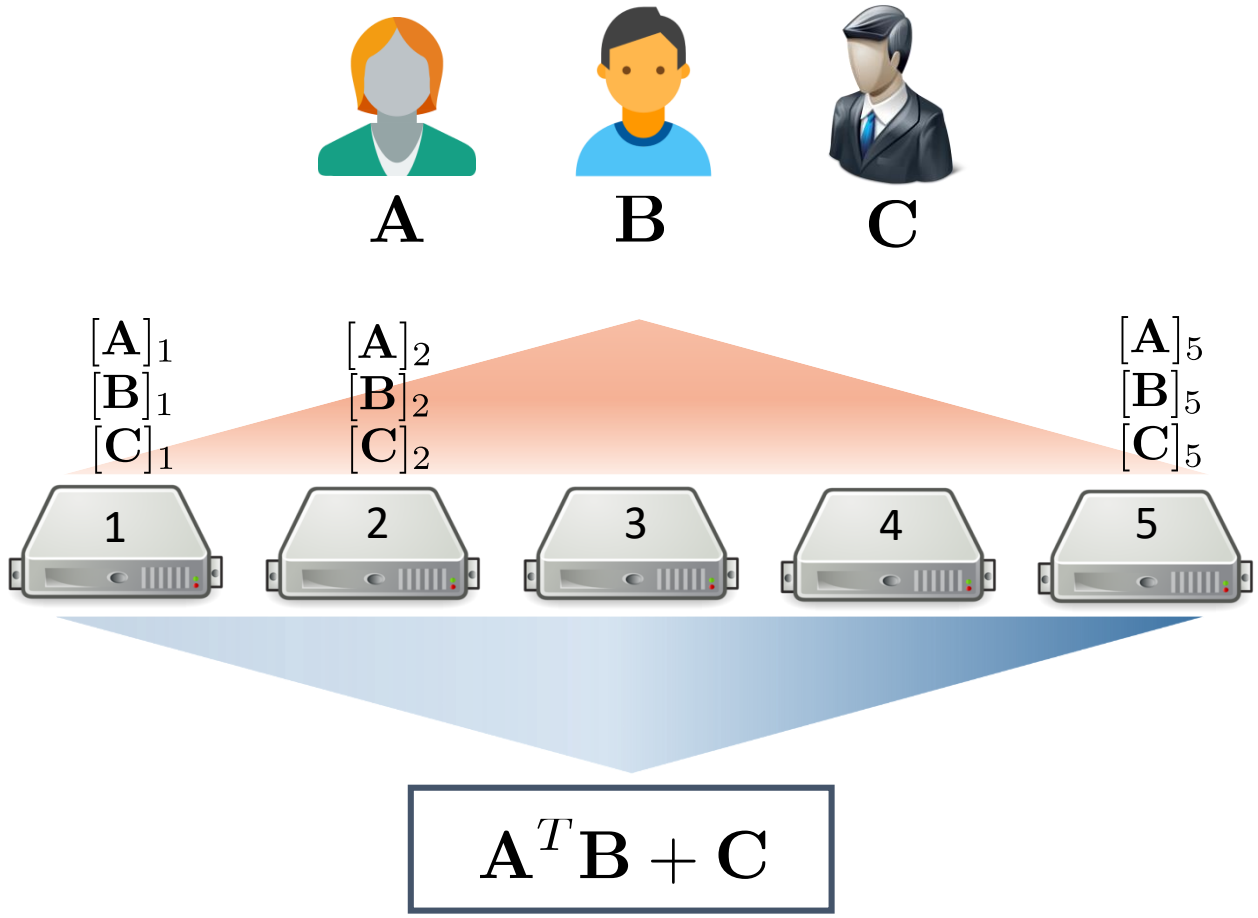
$$\mathbf{A}^T\mathbf{B} + \mathbf{C}$$

Servers Limitations: Storage

Servers are

- Limited storage
- Limited computing resource

$$\frac{\text{size}([\mathbf{A}]_n, [\mathbf{B}]_n, [\mathbf{C}]_n)}{\text{size}(\mathbf{A}, \mathbf{B}, \mathbf{C})} \leq \mu \leq 1$$

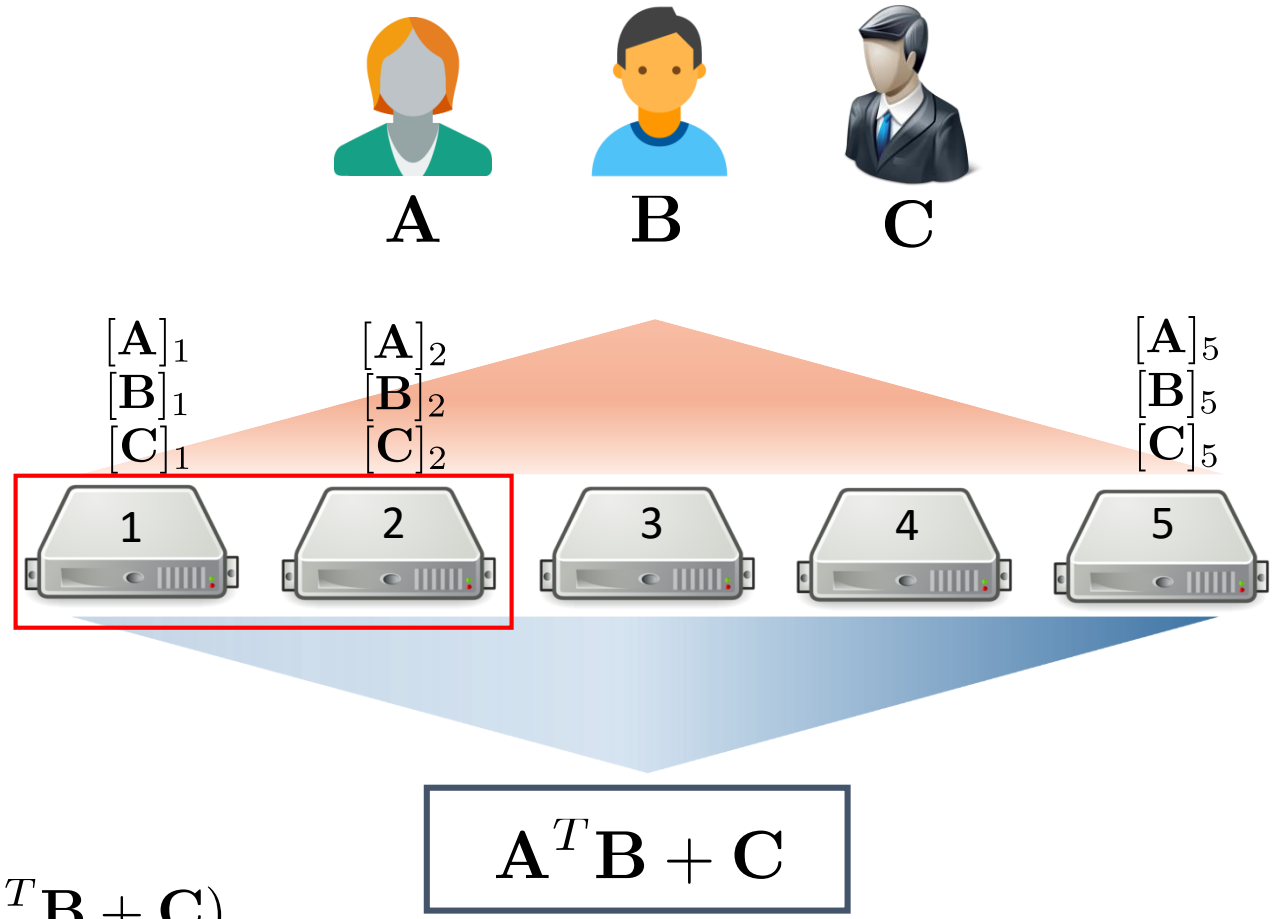


Servers Limitations: Semi-Honest

$\mu,\ t-1=2$

Servers maybe curious!

- Up to ***t-1*** of them may collude



Privacy Constraints:

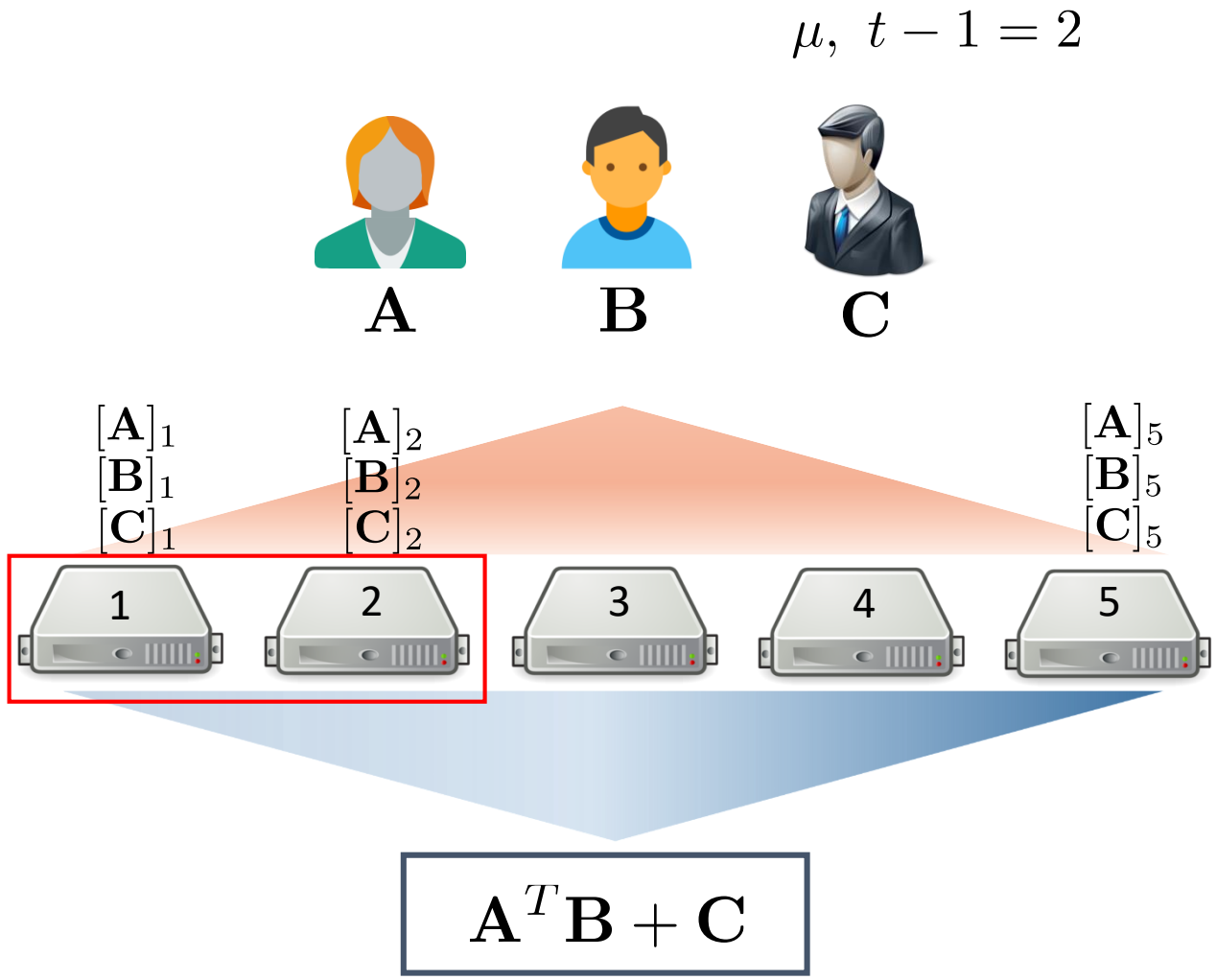
$$H(\mathbf{A} \mid \text{Inputs to } i \ \&\ j) = H(\mathbf{A})$$

$$H(\mathbf{A} \mid \mathbf{A}^T\mathbf{B}+\mathbf{C}, \text{Inputs to DC}) = H(\mathbf{A} \mid \mathbf{A}^T\mathbf{B}+\mathbf{C})$$

Objective

Objective:

- To use minimum number of servers, subject to
  - Storage limitations
  - Privacy constrains
  - Correctness



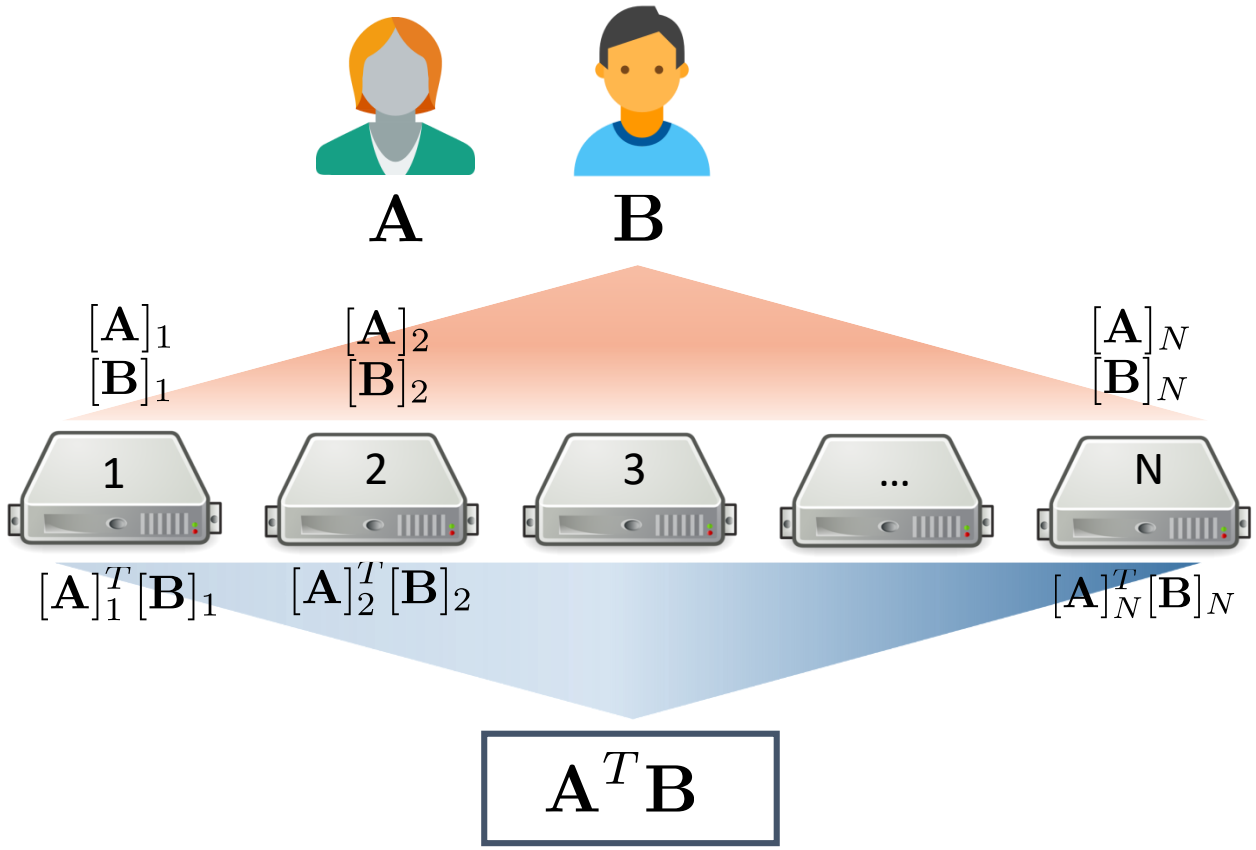
Result

Job Splitting+ Shamir's Sharing

$$N \geq \frac{1}{\mu^2}(2t-1)$$

Polynomial Sharing

$$N \geq 2\frac{1}{\mu^2} + 2t - 3$$



Semi-Honest Nodes: Akbari and Maddah-Ali [2018]
Adversarial Nodes: Hosseini, Maddah-Ali, Aref [2019]
MPC for ZK-SNARK: Rahimi, Maddah-Ali [2020]

Result

$$\mu = \frac{1}{10}, t = 100$$

Job Splitting+ Shamir's Sharing

$$N \geq \frac{1}{\mu^2}(2t-1)$$

$$N \approx 20000$$

Polynomial Sharing

$$N \geq 2\frac{1}{\mu^2} + 2t - 3$$

$$N \approx 400$$

Semi-Honest Nodes: Akbari and Maddah-Ali [2018]
Adversarial Nodes: Hosseini, Maddah-Ali, Aref [2019]
MPC for ZK-SNARK: Rahimi, Maddah-Ali [2020]

Conclusion

- Decentralized Systems raises many challenges
- Only few of them were reviewed
- Information Theory and Coding Techniques may offer effective solutions for those problems