Data Asset for Collaborative Intelligence

Feida Zhu[1] , Huiwen Liu[2], Xin Mu[3] , Jian Pei[4]

1. Singapore Management University, fdzhu@smu.edu.sg
2. Singapore Management University, hwliu.2018@phdcs.smu.edu.sg
3. Peng Cheng National Laboratory, Shenzhen, China, mux@pcl.ac.cn
4. Simon Fraser University, jpei@cs.sfu.ca

Tutorial Outline

- Data Asset: What and Why
- Data Asset Core Components
- Data Asset Governance for Decentralized Collaborative Intelligence
- Data Asset Ecosystems
- Challenges and Future Directions

Tutorial Outline

- Data Asset: What and Why
  - Background and Motivation
  - Definition & History
- Data Asset Core Components
- Data Asset Governance for Decentralized Collaborative Intelligence
- Data Asset Ecosystems
- Challenges and Future Directions

Data Asset: What and Why

- Data has been studied primarily as a computational resource at the discretion of a centralized entity in control
  - Research and engineering have been focused on the efficiency and effectiveness of the computation of data
  - Questions are seldom raised as the following
    - Who has the right to own, access, use, and benefit from, the data?
    - What are the cost of data?
    - How to value data?
    - How to allocate the financial benefit from the monetization of data among all contributing parties?
    - How to incentivize data sharing and trading?

Data Asset: What and Why

- Here are the problems:
  - Data, dubbed as “new oil”, is increasingly important for all businesses in almost all industries, private and public sectors alike
    - By 2022, 60% of global GDP will be digitised, with the World Economic Forum predicting that some 60 - 70% of new value will be “based on data-driven digitally enabled networks and platforms” in the coming years.
  - Most businesses do not have sufficient data they need within their own ecosystem
  - Data from different communities across different silos can complement one another to unleash the true power of big data

[1] International Data Corporation (30 October 2018) ‘By 2023 Nearly Every Enterprise Will Be “Digital Native” in an Increasingly Digitized Global Economy’
[2] World Economic Forum (December 2018) ‘Our Shared Digital future: Building an Inclusive, Trustworthy and Sustainable Digital Society’

Data Asset: What and Why

- What happens when data start to flow

a **Example: Improve customer experience, resulting in new income**

Information exchanged was used to better understand customer needs and deliver more relevant solutions, experience and offers to a combined customer base.

- A bank and telco (both a Data Provider and Data Consumer) were looking to engage in two-way bilateral data sharing to improve customer experience and advance business outcomes across both entities.
- The data exchanged was used to better understand customer needs and deliver more relevant solutions, experience and offers to a combined customer base.
- The partnership enabled both parties and their partners to anticipate and better respond to customers' evolving motivations and preferences, and offer customers value that went beyond a single entity or industry.

b **Example: Improve overall efficiency over supply chain, resulting in cost reduction**

Real-time data on current stock is provided to customers, allowing them better visibility over inventory and optimise goods turnaround. Relevant customer data is used to optimised delivery route in real-time.

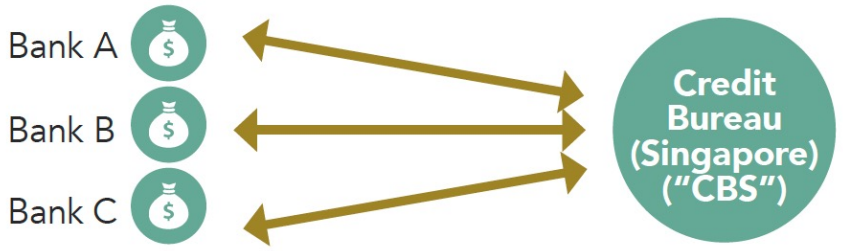
- A local SME launched an integrated platform for real-time orders and inventory management.
- This eliminates the unnecessary stocking of slow-moving or soon-expiring stock across warehouses and increases the turnaround of goods. Additionally, it has led to greater accuracy and productivity through minimising data entry, email and phone communication.
- The provider also use relevant customer data to optimise delivery routes real-time. Customers benefit from having access to real-time delivery details, proof of delivery and feedback on delivery services in real-time.

[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Data Asset: What and Why

- What happens when data start to flow

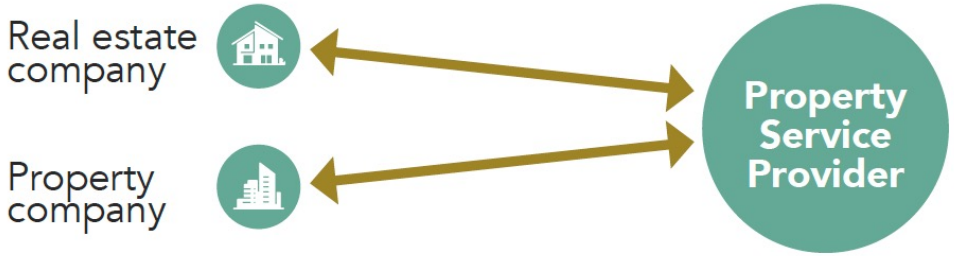
c **Example: Provide comprehensive information for overall market efficiency for the sector**



Banks disclose and obtain credit-related information to mitigate consumer credit risk through information pooling from CBS to provide credit providers determine the likelihood of the customer repaying, enhancing their risk assessment capabilities.

- CBS is a joint venture between the Association of Banks in Singapore (**"ABS"**) and Infocredit Holding Pte Ltd. The Banking Act allowed CBS members (mainly banks) to disclose and obtain credit related information.
- To do this, CBS aggregates credit-related information amongst participating members and presents a more complete risk profile of a customer to credit providers, helping credit providers make better lending decisions quickly and more objectively.

d **Example: Provide comprehensive information for public good**



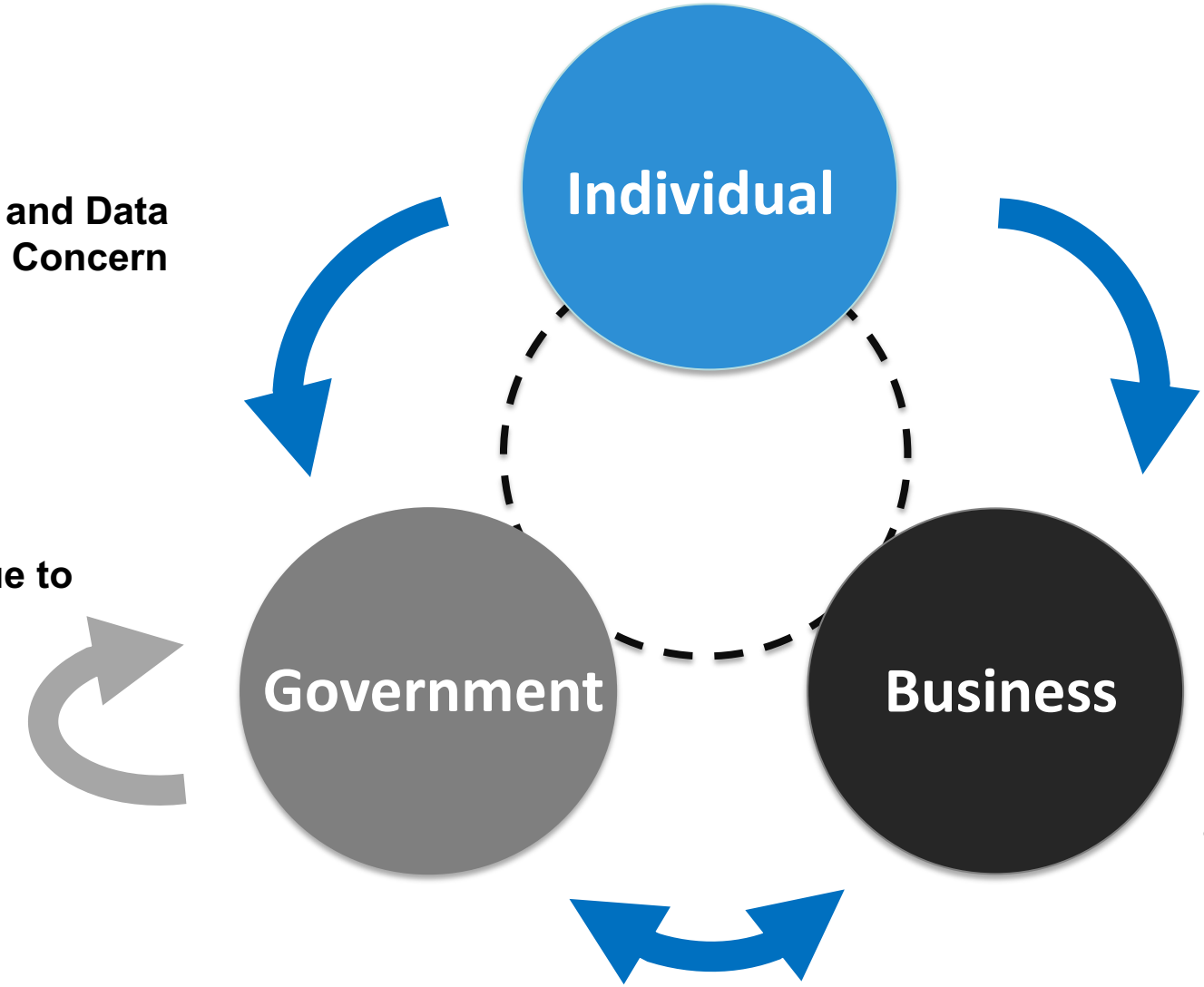
Provider aggregates property valuation data from various real estate and property companies throughout Singapore to provide a real-time, comprehensive view of property market valuations.

- Data shared is free of charge, in return the real estate/ property companies are able to then access and obtain data from the data service. Consumers can also use the information to make informed decisions on their property. This creates a more efficient real estate market overall.

[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Data Asset: What and Why

- Data Ecosystem and Bottleneck I: Among Different Parties

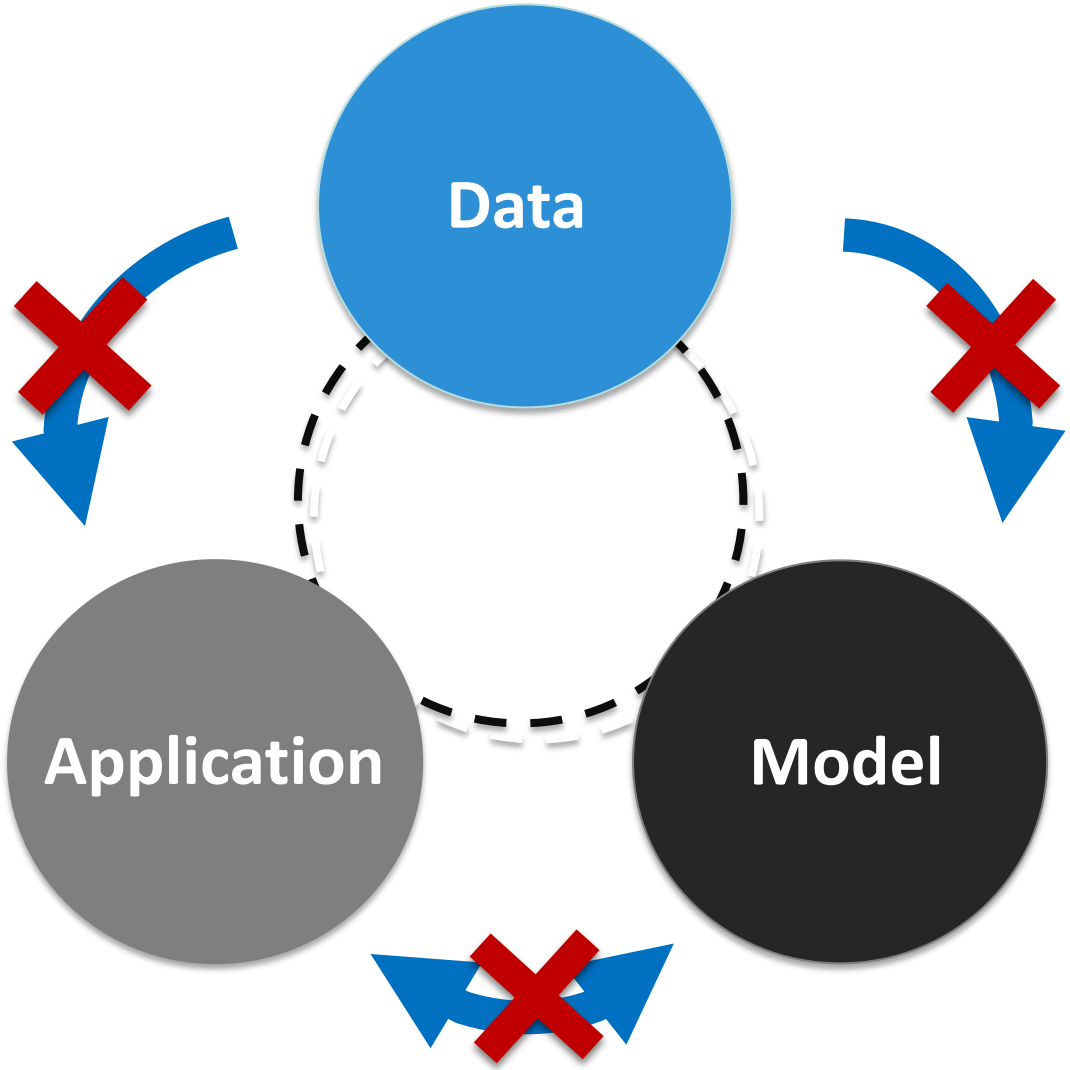


- **Unfair Value Allocation** with Users as data contributors left out
- **Weak Data Control**
  Users have little knowledge of and control over their data
- **Poor Privacy Protection**
  Low transparency and insufficient measures
- **Privacy and Data Leakage Concern**
- **Data Wall Due to Policy and Regulation**
- **Data Isolation among Corporates For Security Concerns**
- **Low Incentivization for Data Sharing**

Data Asset: What and Why

- Data Ecosystem and Bottleneck II: Among Different Components

- **Low-Quality Data from Questionable Sources**
- **Inaccurate User Insights from Fragmented Data Silos**



- **Hard to access real user data for model design and training**

Inferior Data Intelligence Due to Inaccessibility to Advanced Models

Data Asset: What and Why

Where is the Root of the Problem?

It's NOT in "Data Intelligence"

It's in "Data Governance"

To govern, data must be established as an "Asset"

Data Asset: What and Why

- Attributes of asset (Accounting Standards)
  - Assets are expected to bring economic benefits or service potential to accounting entities.
  - Assets should be the resources owned or controlled by the accounting entity.
  - Assets are formed by past transactions or events of accounting entities.
  - The economic benefits related to the resource are likely to flow into the enterprise.
  - The cost or value of the resource can be reliably measured.

Posner E. Sequence as explanation: The international politics of accounting standards[J]. Review of International Political Economy, 2010, 17(4): 639-664.
Zhu Y, Zhong N, Xiong Y. Data explosion, data nature and dataology[C]//International Conference on Brain Informatics. Springer, Berlin, Heidelberg, 2009: 147-158.
Emmanuel I, Stanier C. Defining big data[C]//Proceedings of the International Conference on Big Data and Advanced Wireless Technologies. 2016: 1-6.

Data Asset: What and Why

- Attributes of data
  - Physical attributes
    - The physical properties of data assets refer to the fact that data assets exist in binary form in storage media, occupy physical space, and are tangible.
  - Existence attributes
    - The existence property of a data asset is its readability.
  - Information attributes
    - The information attribute of a data asset is its value.

Posner E. Sequence as explanation: The international politics of accounting standards[J]. Review of International Political Economy, 2010, 17(4): 639-664.
Zhu Y, Zhong N, Xiong Y. Data explosion, data nature and dataology[C]//International Conference on Brain Informatics. Springer, Berlin, Heidelberg, 2009: 147-158.
Emmanuel I, Stanier C. Defining big data[C]//Proceedings of the International Conference on Big Data and Advanced Wireless Technologies. 2016: 1-6.

Data Asset: What and Why

- Historical connotations of data assets
  - Information assets---Data of value or potential value that has been or should be recorded (1994).
  - Digital assets---Anything such as text or media that is formatted as bit code and has the right to use (2006).
  - Data assets---Data is an asset, and companies should treat data as corporate assets (2009).

KPMG/IMPACT. Information as an Asset: The Board Agenda[J]. 1994.
Waddington P. Information as an asset: the invisible goldmine[J]. Business Information Review, 1995, 12(1): 26-36.
van Niekerk A. A METHODOLOGICAL APPROACH TO MODERN DIGITAL ASSET MANAGEMENT: AN EMPIRICAL STUDY[C]//Allied Academies International Conference. International Academy for Case Studies. Proceedings. Jordan Whitney Enterprises, Inc, 2006, 13(1): 53.
Toygar A, Rohm Jr C E, Zhu J. A new asset type: digital assets[J]. Journal of International Technology and Information Management, 2013, 22(4): 7.
Genders R, Steen A. Financial and estate planning in the age of digital assets: A challenge for advisers and administrators[J]. Financial Planning Research Journal, 2017: 75-80.
Fisher T. The data asset: How smart companies govern their data for business success[M]. John Wiley & Sons, 2009.

Data Asset: What and Why

- Properties for Data Asset
  - Identifiable and definable
    - data assets may be made up of specific files or specific tables or records within a database

    Form
  - Promise probable future economic benefits
    - the data asset must have a useful application. Identifying productive uses for data is often necessary to assign value to the asset

    Value
  - Under the organisation's control
    - the organisation must also have rights to use the data in a way consistent with its rights under applicable law and any contractual licensing arrangements, while also protecting the data and restricting access to it by others.

    Control

[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Data Asset: What and Why

- Data Asset Definition:
  - A “Data Asset” is information in the form of data with measurable value and confirmable right to own and control.

Data Asset: What and Why

- Data Asset Governance:
  - Governance refers to the policies, mechanisms, procedures and operational processes to guide and enforce the maintenance of the integrity and robustness of data asset ecosystem

Data Asset Governance

VALUE RIGHT CONTROL

Tutorial Outline

- Data Asset: What and Why
- Data Asset Core Components
  - Value
  - Right
  - Control
- Data Asset Governance for Decentralized Collaborative Intelligence
- Data Asset Ecosystems
- Challenges and Future Directions

Tutorial Outline

- Data Asset: What and Why
- Data Asset Core Components
  - Value
    - Data Intelligence --- How to bring value out of data? (not covered in this tutorial)
    - Data Pricing --- How to value data? (covered in Part A)
  - Right
  - Control
- Data Asset Governance for Decentralized Collaborative Intelligence
- Data Asset Ecosystems
- Challenges and Future Directions

Tutorial Outline

- Data Asset: What and Why
- Data Asset Core Components
  - Value
  - Right
    - Type of Right
    - Right Confirmation
    - Right Enforcement
  - Control
- Data Asset Governance for Decentralized Collaborative Intelligence
- Data Asset Ecosystems
- Challenges and Future Directions

Data Asset Governance: Right

- Types of Right
  - Ownership
  - Access
  - Use
  - Retention
  - Disclosure
  - Transfer
  - Benefit
  - Disposal

Data Asset Governance: Right

- Right Confirmation
  - By Law
    - For corporate data
      - Intellectual property law
    - For personal data
      - PDPA, GDPR, etc.
  - By Agreement
    - Bilateral
    - Multilateral
    - Decentralized

Data Asset Governance: Right

- Right Confirmation
  - By Law

1.	Singapore Personal Data Protection Act 2012 (Act No. 26 of 2012)	Singapore
2.	Singapore Cybersecurity Act 2018 (Act No. 9 of 2018)	Singapore
3.	IMDA Data Protection Certification Trustmark Certification Criteria	Singapore
4.	Personal Data Protection Commission Guide to Anonymisation	Singapore
5.	Monetary Authority of Singapore Guidelines on Outsourcing Risk Management	Singapore
6.	Monetary Authority of Singapore Principles to Promote Fairness, Ethics, Accountability and Transparency in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector	Singapore
7.	Association of Banks Singapore's Outsourced Service Providers Guidelines	Singapore
8.	Bank Negara Malaysia Policy Document on Outsourcing	Malaysia
9.	Malaysia Personal Data Protection Act 2010 (Act 709)	Malaysia
10.	Vietnam Law on Cybersecurity 2019 (No. 24/2018/QH14)	Vietnam
11.	Thailand Cybersecurity Act 2019	Thailand
12.	Government Regulation No. 82 of 2012 Concerning the Electronic System and Transaction Operation	Indonesia

[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Data Asset Governance: Right

- Right Confirmation
  - By Law

13.	Hong Kong Monetary Authority SA-2 Supervisory Policy Manual on Outsourcing	HongKong
14.	California Consumer Privacy Act of 2018	USA
15.	Australian Computer Society Data Sharing Frameworks Technical White Paper	Australia
16.	Australia Privacy Act 1988 (Act No.119, 1988)	Australia
17.	EU General Data Protection Regulation (Regulation (EU) 2016/679)	EU
18.	European Commission Study on Data Sharing Between Companies in Europe	EU
19.	National Health Service General Data Protection Regulation Guidance	UK
20.	EU-U.S. Privacy Shield (2016)	EU and USA
21.	Asia-Pacific Economic Cooperation Privacy Framework (2005)	APEC
22.	Organisation for Economic Co-Operation and Development Revised Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (2013)	OECD
23.	The FAIR Guiding Principles for Scientific Data Management and Stewardship	NA

[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Data Asset Governance: Right

- Right Confirmation by Agreement
  1. Grant of the licence/permissions to use the data for the intended purpose
  2. Restrictions to the permitted use of the data (if any), such as territorial or time limitations, exclusivity or commercialization rights
  3. Warranties or other assurances provided in relation to the Data Provider's rights in the data
  4. Allocation of liability for contract breaches and other liabilities between the parties, as well as indemnification and other remedies when breaches occur
  5. Confidentiality
  6. Term/duration of the agreement
  7. Governing law and resolving disputes.

[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Data Asset Governance: Right

- Right Enforcement Elements
  - Authorization
    - Consent
    - Withdrawal
  - Transparency
    - Purpose
    - Notification
  - Integrity
    - Accuracy
    - Consistency
    - Protection

Data Asset Governance: Right

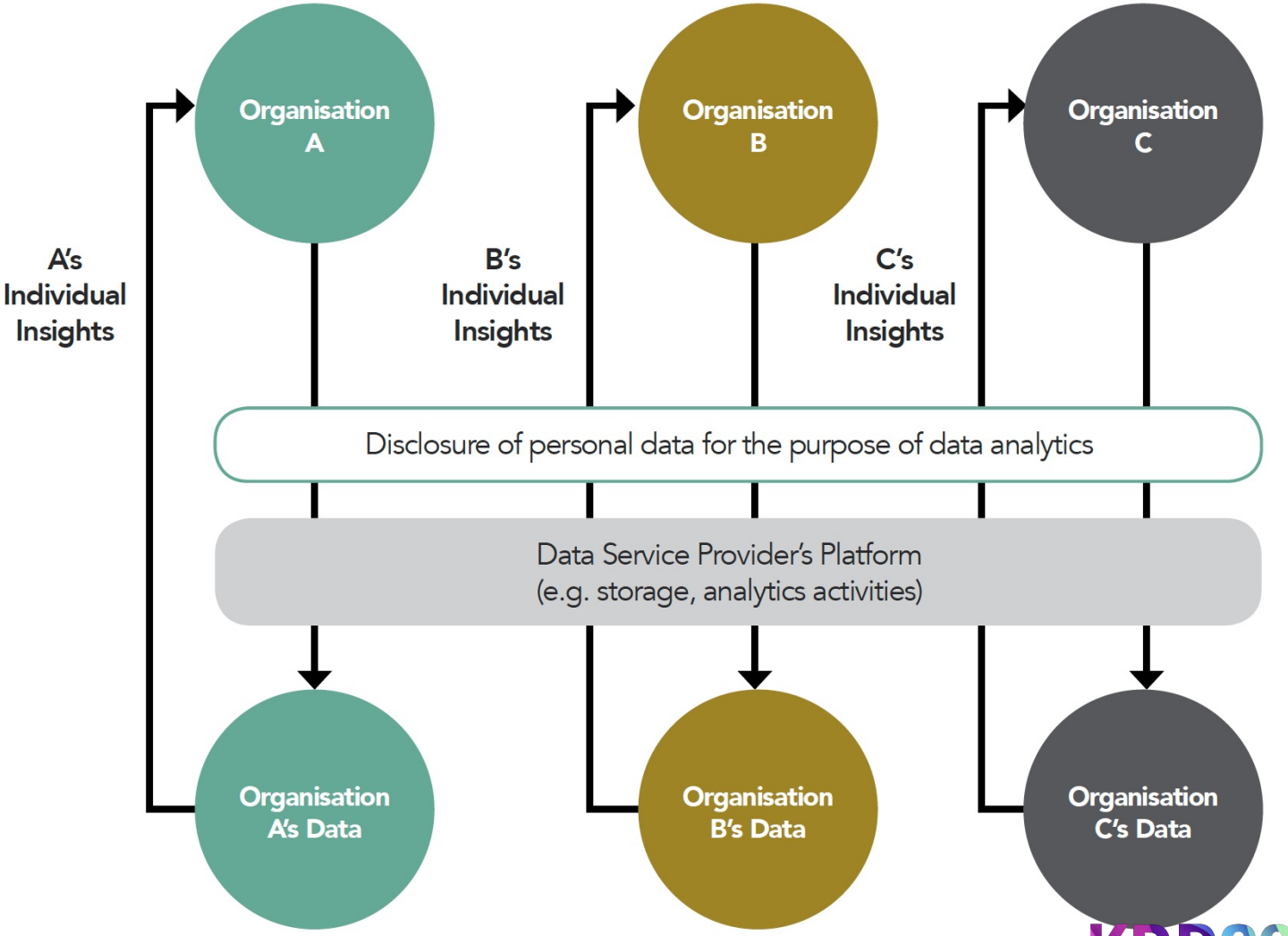
- Right Enforcement Elements
  - Authorization
    - Consent:
      - Fresh Consent

- As Organisation ABC wants to share the data with Organisation XYZ for a new purpose, ABC must notify the individuals of the new purpose and obtain their consent.[6]
- If there are any potential risks to the individuals as a result of sharing the personal data, Organisation ABC should highlight these risks to the individuals when obtaining their consent.
- Organisation ABC must allow the individuals to withdraw their consent if they no longer want their personal data to be shared for this purpose.

[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Data Asset Governance: Right

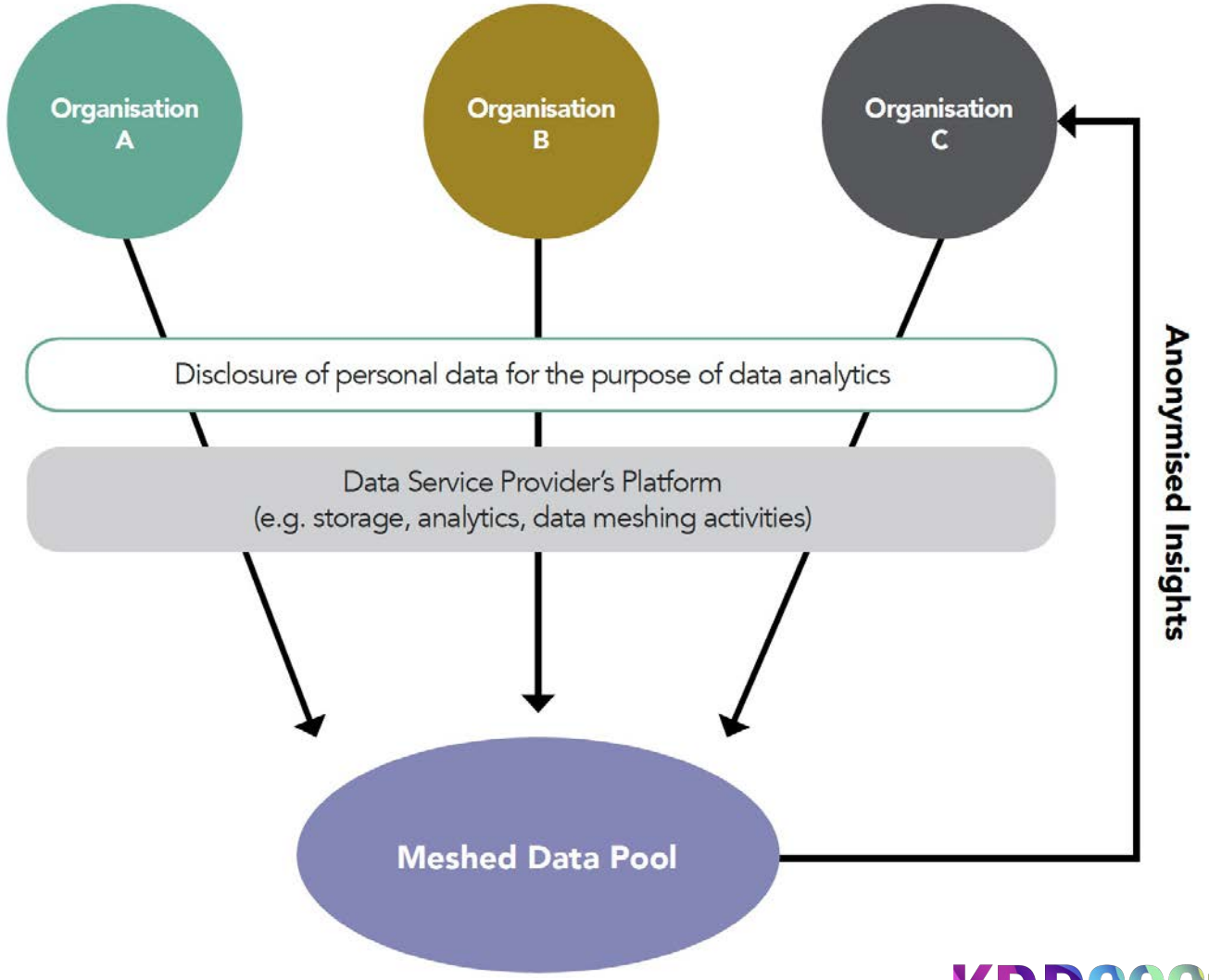
- Right Enforcement Elements
  - Authorization
    - Consent:
      - Fresh Consent



[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Data Asset Governance: Right

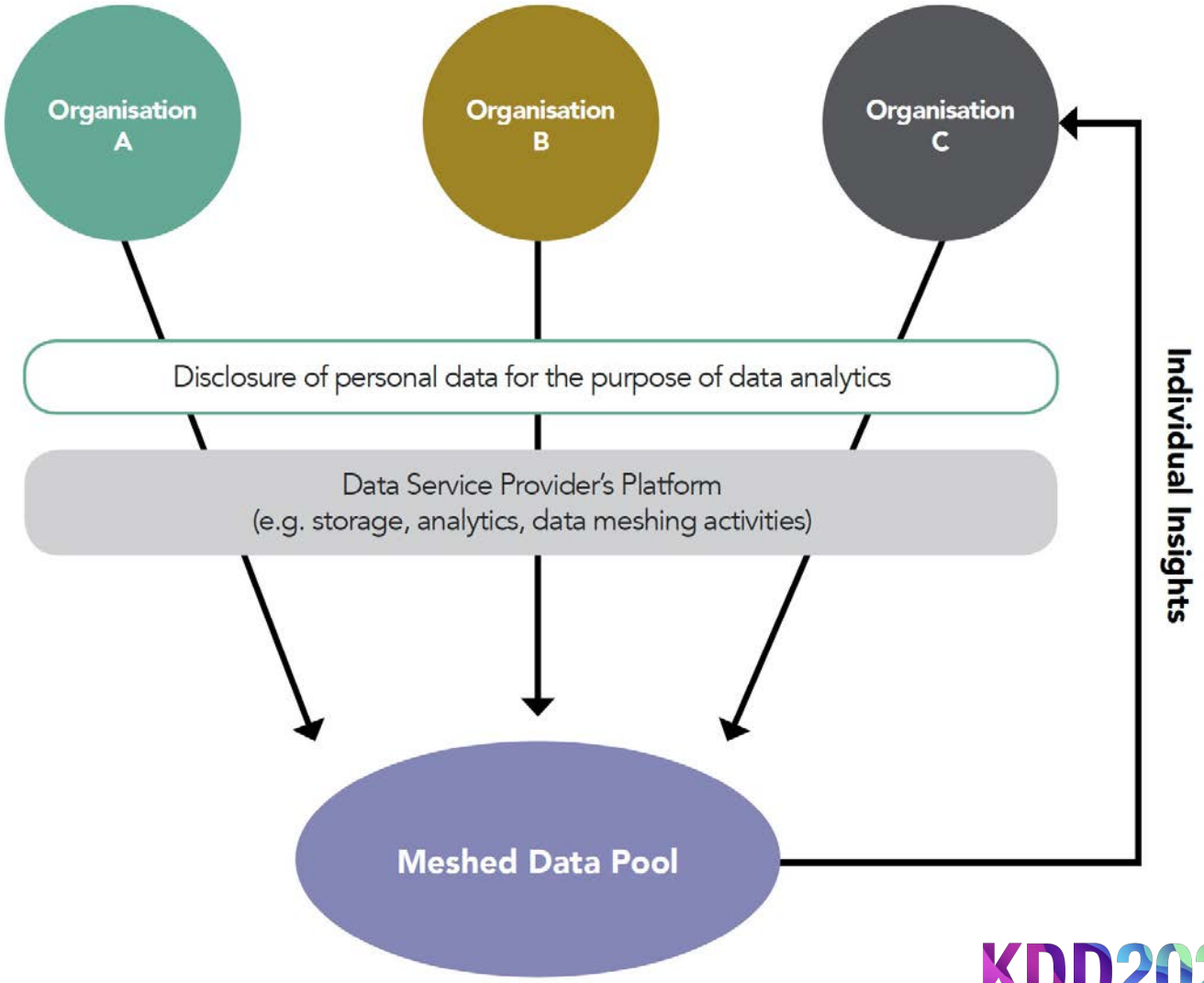
- Right Enforcement Elements
  - Authorization
    - Consent:
      - Fresh Consent



[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Data Asset Governance: Right

- Right Enforcement Elements
  - Authorization
    - Consent:
      - Fresh Consent



[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Data Asset Governance: Right

- Right Enforcement Elements
  - Authorization
    - Consent:
      - Dynamic
      - Iterative

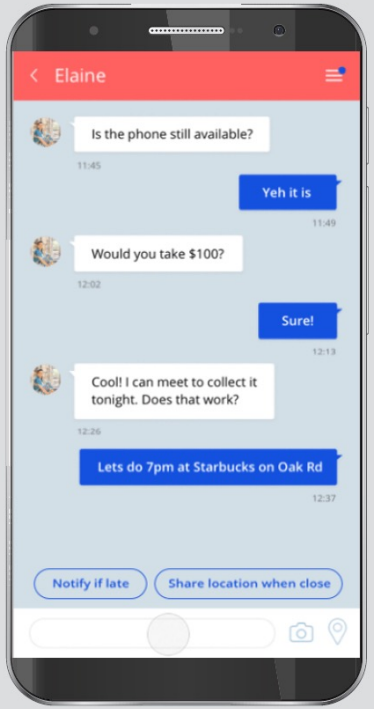
Example: Dynamic Consent

Objective:
Enable individuals to share personal data (in this instance, location data) intuitively in social apps (e.g. chat groups) when meeting up, depending on strength of relationship.

Design Features:
For close friends: **Automatic sharing** of location information via mini-map, pin-pointing an individual's location.

For acquaintances: **Consent is sought** to either disclose ETA and location only when in close proximity or to send a simple notification to the other party.

Extracted from Singapore Design Jam (Nov 2018), co-hosted by IMDA and TTC Labs



[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Data Asset Governance: Right

- Right Enforcement Challenges
  - Ownership
    - Legal Challenges
    - Technical Challenges
  - Access
    - Mode of Delivery
  - Use
  - Retention
  - Disclosure
  - Benefit
  - Disposal

	Wire	Removable Storage Media	Wi-Fi	Remote Access/ VPN	Object Storage URL / SFTP	API	Distributed Ledger
Continuous Access	✓		✓	✓	✓	✓	✓
High Volume of Data	✓			✓	✓	✓	
High Speed of Transfer	✓		✓	✓	✓	✓	
Highly Sensitive Data	✓					✓	✓
Affordability		✓	✓		✓	✓	
Secure by Design	✓						✓

[1] https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Tutorial Outline

- Data Asset: What and Why
- Data Asset Core Components
  - Value
  - Right
  - Control
    - Operations
- Data Asset Governance for Decentralized Collaborative Intelligence
- Data Asset Ecosystems
- Challenges and Future Directions

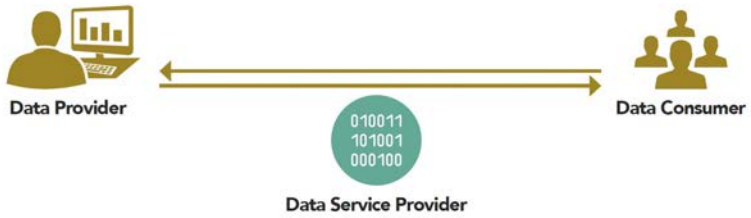
Data Asset Governance: Control

- Asset Operation Types
  - Sharing
  - Trading
  - Pledging
  - Leasing

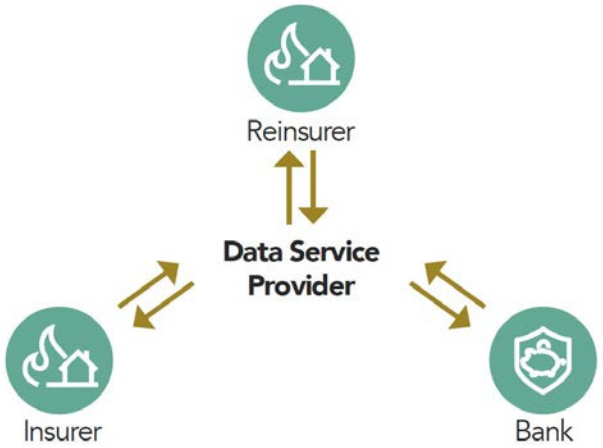
Data Asset Governance: Control

- Asset Operation Types
  - Sharing Mode

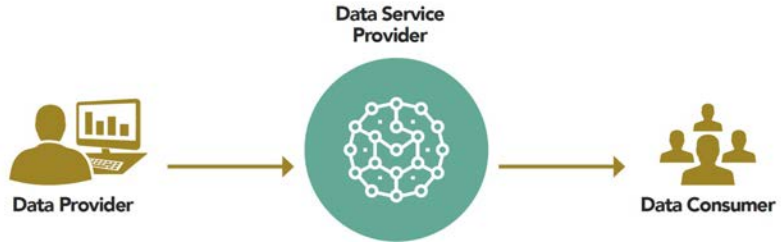
Bilateral



Multilateral



Decentralized



Tutorial Outline

- Data Asset: What and Why
- Data Asset Core Components
- Data Asset Governance for Decentralized Collaborative Intelligence
  - Governance principles
    - Trust
    - Incentive
  - "Trust" for data asset governance for decentralized collaborative intelligence
  - "Incentive" for data asset governance for decentralized collaborative intelligence
- Data Asset Ecosystems
- Challenges and Future Directions

Data Asset Governance: Governance principles

TRUST + INCENTIVE

- Transparency
- Minimalism
- Accessibility
- Standardisation
- Fairness & Ethics
- Accountability
- Security & Integrity

- Data economy design
- Data pricing
- Value allocation
- Tokenomics

Data Asset Governance: Governance principles

- Governance Principles
  - Transparency
    - the openness of all parties involved in data asset operation to make available all information that is necessary for the successful delivery of the data asset operation partnership.
  - Minimalism
  - Accessibility
  - Standardisation
  - Fairness & Ethics
  - Accountability
  - Security & Integrity

Data Asset Governance: Governance principles

- Governance Principles
  - Transparency
  - Minimalism
    - Data asset operation right should be granted only for the portion that is necessary for the designated task
  - Accessibility
  - Standardisation
  - Fairness & Ethics
  - Accountability
  - Security & Integrity

Data Asset Governance: Governance principles

- Governance Principles
  - Transparency
  - Minimalism
  - Accessibility
    - the ability of parties to access the data they need, when they need it
  - Standardisation
  - Fairness & Ethics
  - Accountability
  - Security & Integrity

Data Asset Governance: Governance principles

- Governance Principles
  - Transparency
  - Minimalism
  - Accessibility
  - Standardisation
    - To apply consistent legal, technical and other measures to data asset operation partnerships
  - Fairness & Ethics
  - Accountability
  - Security & Integrity

Data Asset Governance: Governance principles

- Governance Principles
  - Transparency
  - Minimalism
  - Accessibility
  - Standardisation
  - Fairness & Ethics
    - To go beyond meeting data protection and technical or security standards or regulatory requirements. It extends to the need to apply ethical standards to the creation and use of data asset systems and frameworks, starting from the initial design phase.
  - Accountability
  - Security & Integrity

Data Asset Governance: Governance principles

- Governance Principles
  - Transparency
  - Minimalism
  - Accessibility
  - Standardisation
  - Fairness & Ethics
  - Accountability
    - To demonstrate compliance with data protection laws and other rules specific to the data asset operation partnership, and that each party has robust governance structures in place, and a corporate culture that encourages employees to take responsibility for the handling of data
  - Security & Integrity

Data Asset Governance: Governance principles

- Governance Principles
  - Transparency
  - Minimalism
  - Accessibility
  - Standardisation
  - Fairness & Ethics
  - Accountability
  - Security & Integrity
    - the implementation of measures and mechanisms designed to securely protect and safeguard information and data to enable a secure environment for data asset operation.

Tutorial Outline

- Data Asset: What and Why
- Data Asset Core Components
- Data Asset Governance for Decentralized Collaborative Intelligence
  - Governance principles
  - "Trust" for data asset governance for decentralized collaborative intelligence
    - Agreement
    - Accounting
    - Auditing
    - Privacy
  - "Incentive" for data asset governance for decentralized collaborative intelligence
- Data Asset Ecosystems
- Challenges and Future Directions

"Trust" for data asset governance -- Agreement

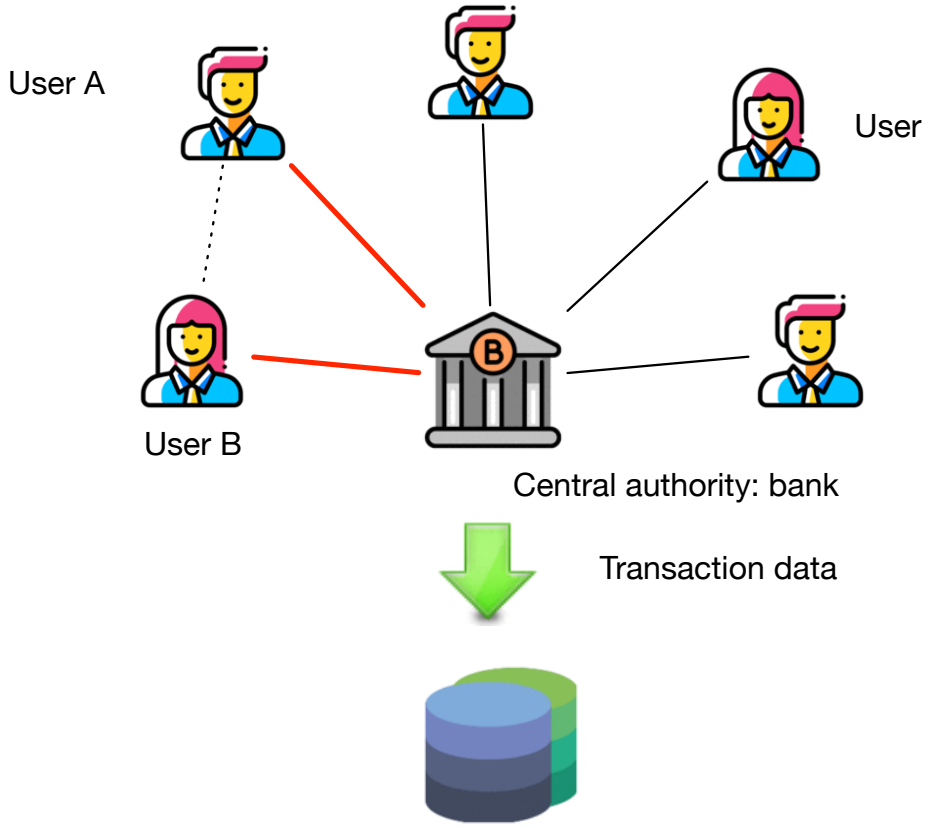
- Backgrounds
- Consensus Problem and Development
- Extended Consensus Definition
- Evaluative Framework
- Consensus Evaluation
- Conclusion

Outline

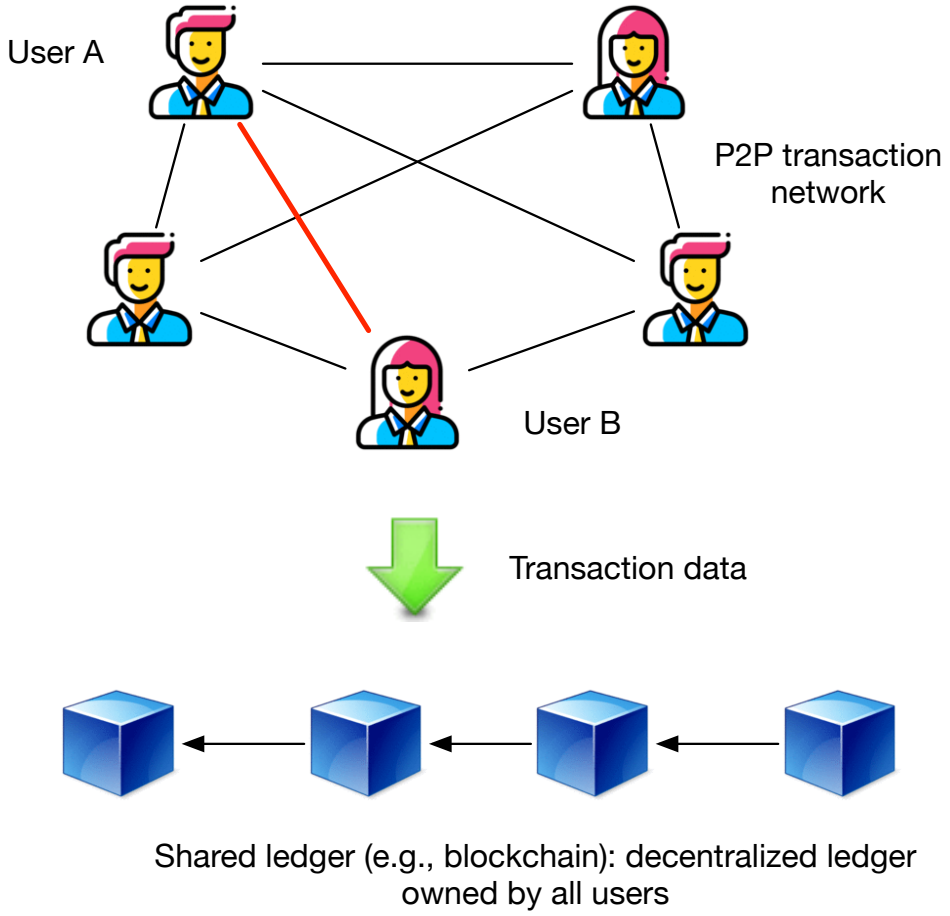
- Backgrounds
  - Distributed Ledger Technology (DLT)
    - Trust crisis: from the centralized trading mode to decentralized trading mode
    - Consensus: core component of distributed ledger system for decentralized trading mode
- Consensus Problem and Development
- Extended Consensus Definition
- Evaluative Framework
- Consensus Evaluation
- Conclusion

Distributed ledger technology (DLT)

- Two solutions for “trust crisis” in transactions



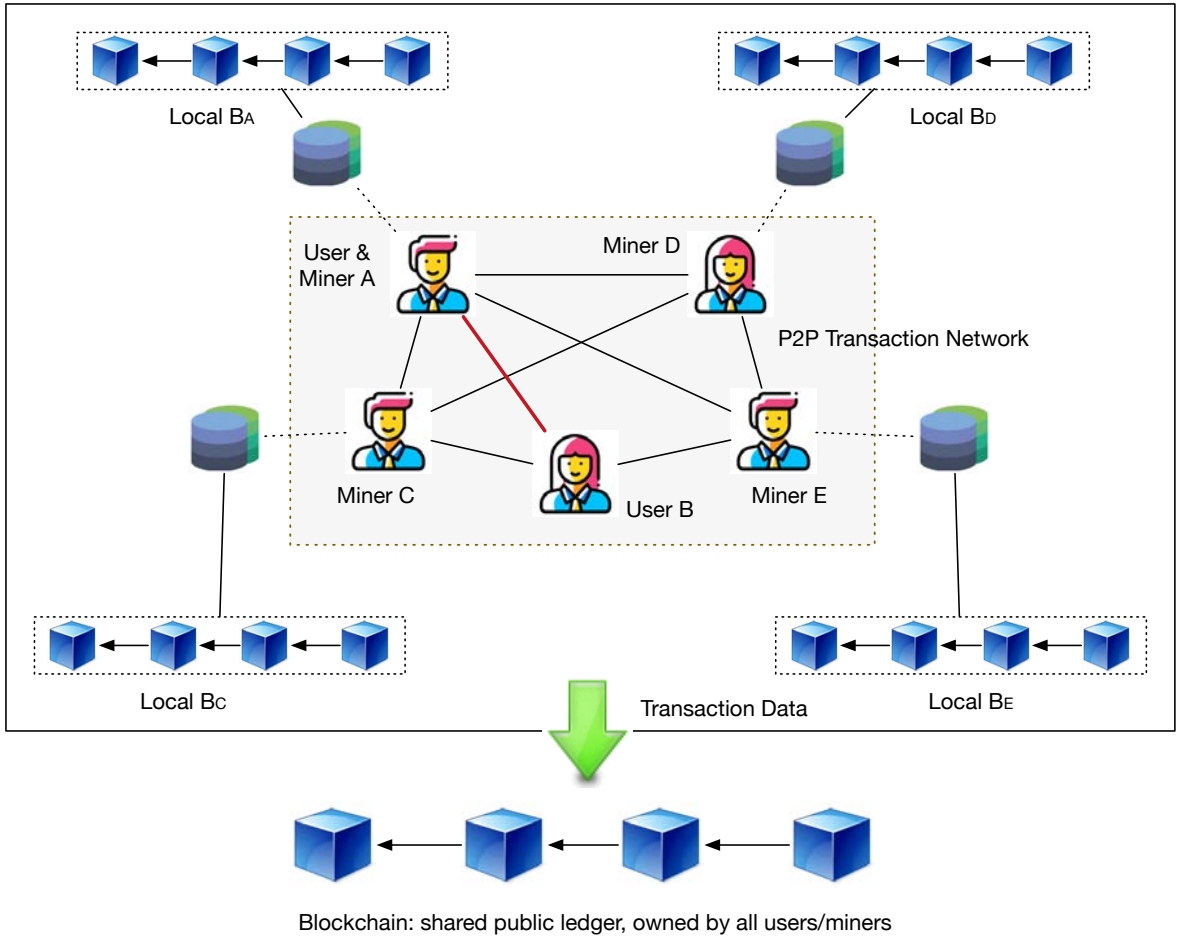
Traditional solution: centralized trading mode



Emerging solution: decentralized trading mode

Distributed ledger system (DLS)

- All miners maintain the synchronized ledger



System structures:

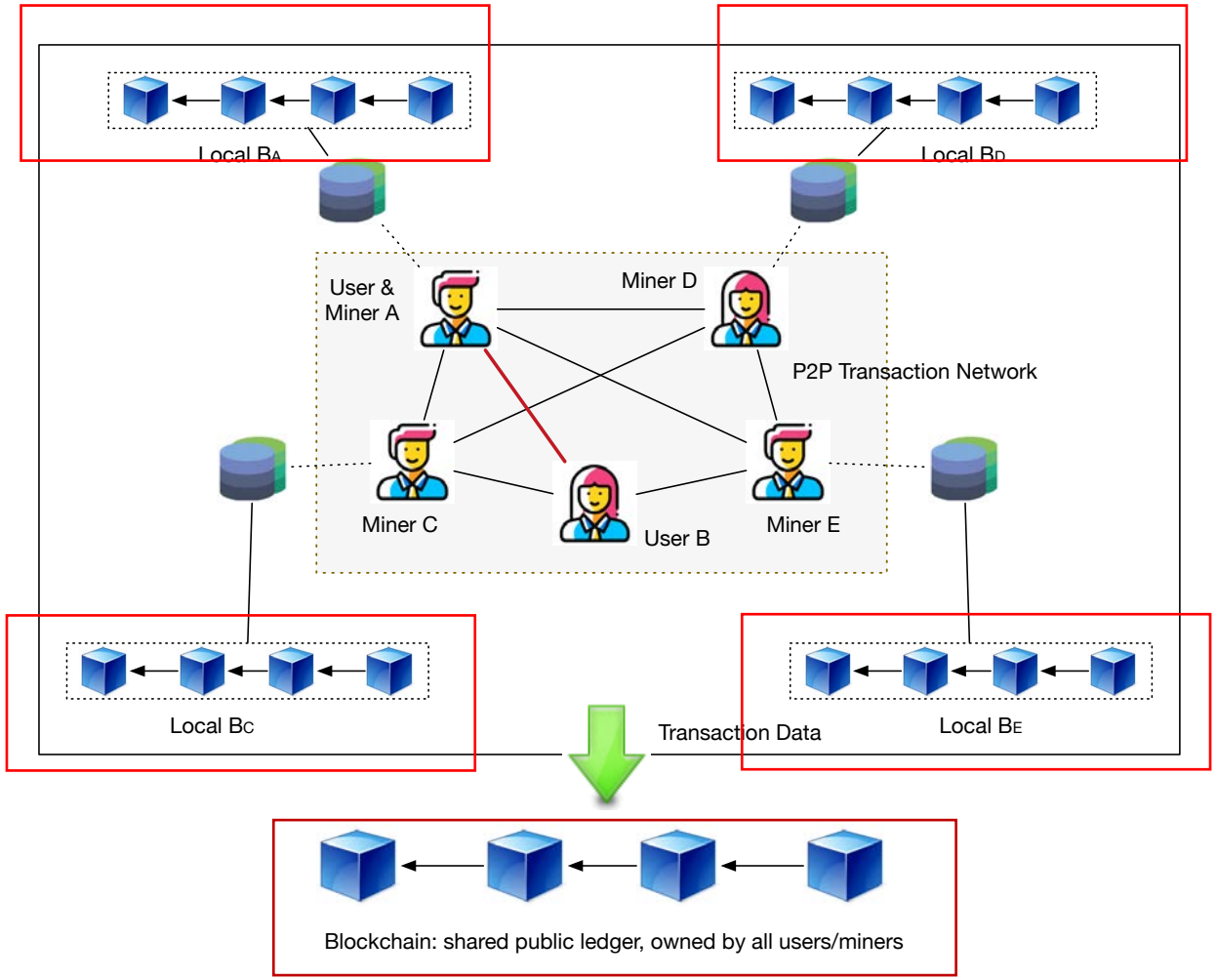
1. Users and maintainers in the system are connected by the P2P network
2. Each maintainer has its own independent storage in different forms (e.g., blockchain)

Most importantly, all miners need to maintain the correctness and synchronization of the ledger

But why?

Consensus in distributed ledger system

- Why is reaching agreement for distributed ledger important?



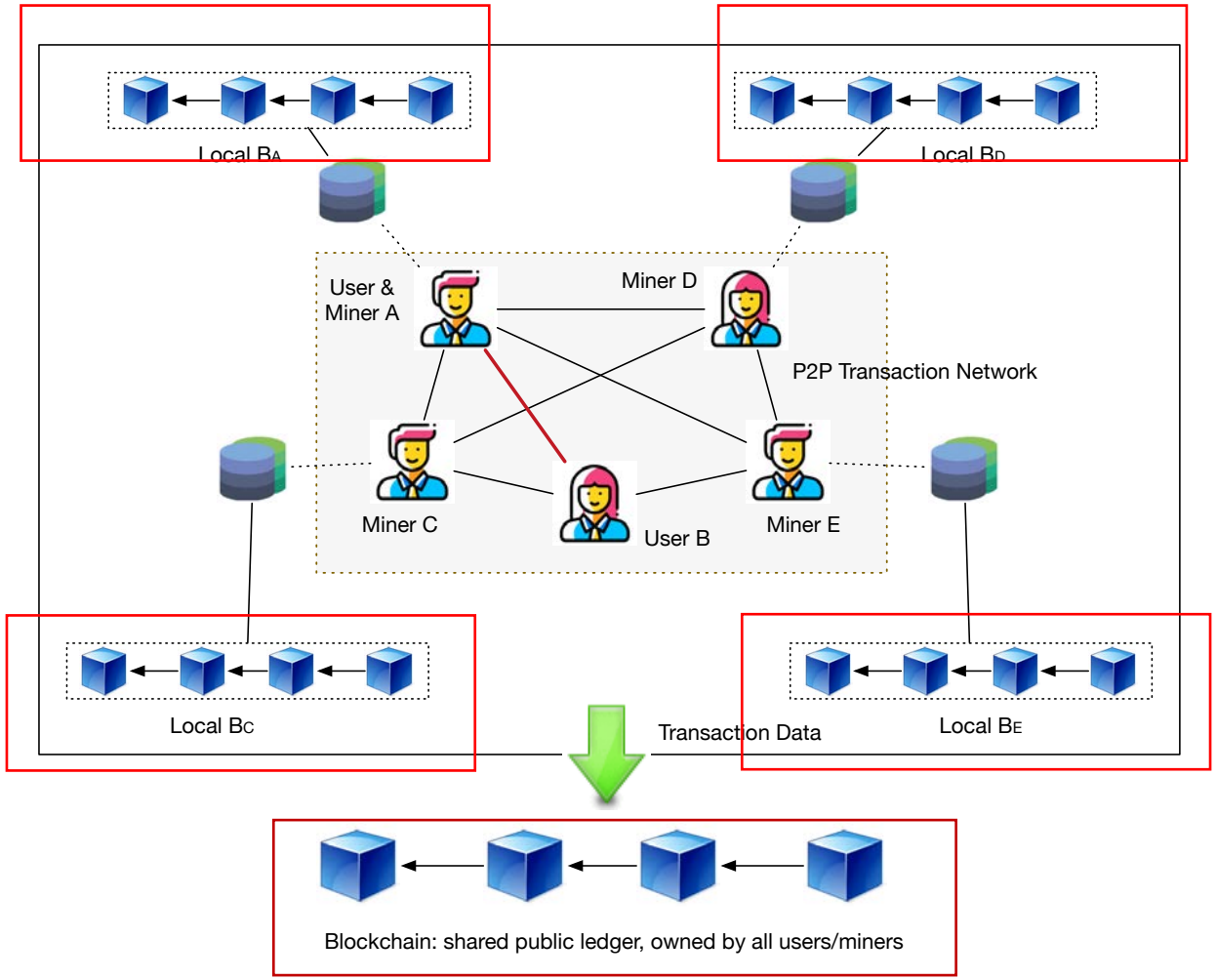
Distributed ledger system has shared ledger, and each node in the system has a local view of the ledger, must agree (roughly) on what shared ledger is.

But how?

KDD2021

Consensus in distributed ledger system

- How to reach agreement of distributed ledger?



Macro:

How to guarantee the correctness of transactions and consistency of storage (i.e., $B_A = B_B = ... = B_D$)?

Micro:

1. Who packages the block?
2. How to verify the block and transactions in the block?
3. How to guarantee the consistency of independent storage of each user?
4. ……

Consensus

Serves to make sure a valid agreement is reached among a group of distributed nodes continuously

Consensus in distributed ledger system

- Why is reaching agreement for distributed ledger hard?
  - Nodes die
  - Nodes lie
  - Nodes sleep (and wake up)
  - Nodes don’t hear all messages
  - Nodes hear messages incorrectly
  - Groups of nodes split into cliques (partition)
  - ……

More formally, these are known as failure models

Classic consensus definition

- An algorithm achieves consensus if it satisfies the following conditions:
  - Agreement: all non-faulty processes decide on the same output value.
  - Termination: all non-faulty processes eventually decide on some output value.

Basic underlying model

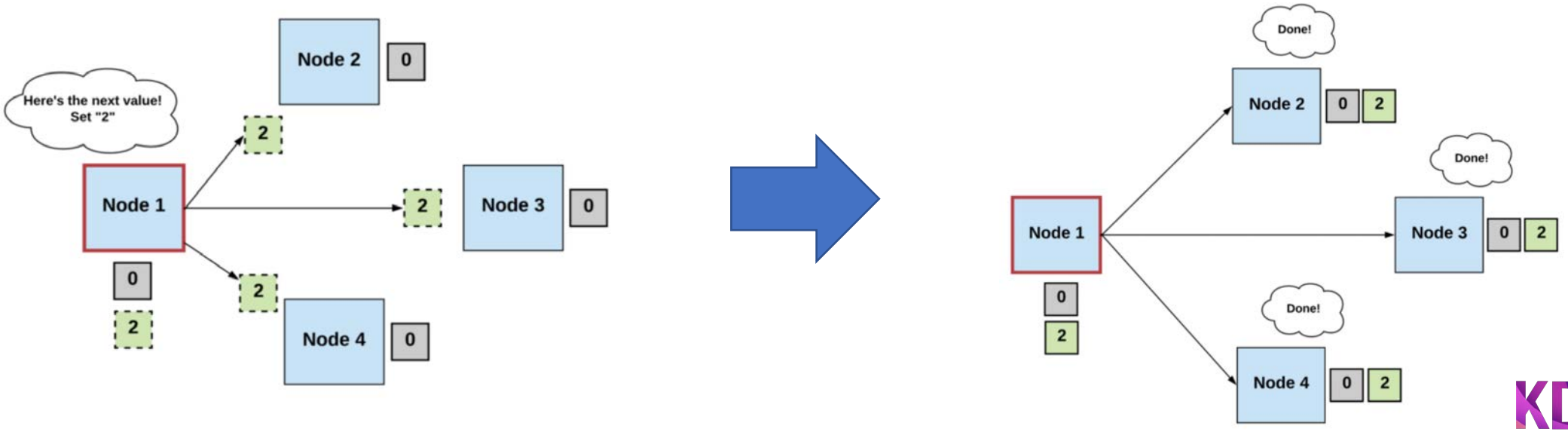
- Network reliability (reliable vs. unreliable)
  - Reliable: all messages are eventually delivered intact exactly once.
- Timing model (synchronous vs. asynchronous)
  - Synchronous: communicating message delays and process delays are bounded, enabling communication in synchronous rounds.

Basic consensus process

- Step 1: Propose
  - Processes elect a single process (i.e., leader/coordinator) to make decisions, and the leader proposes the next valid output value.
  - The non-faulty processes listen to the value being proposed by the leader, validate it, and propose it as the next valid value.
- Step 2: Decide
  - The non-faulty processes must come to a consensus on a single correct output value. If it receives a threshold number of identical votes which satisfy some criteria, then the processes will decide on that value.
  - Otherwise, the process starts over.

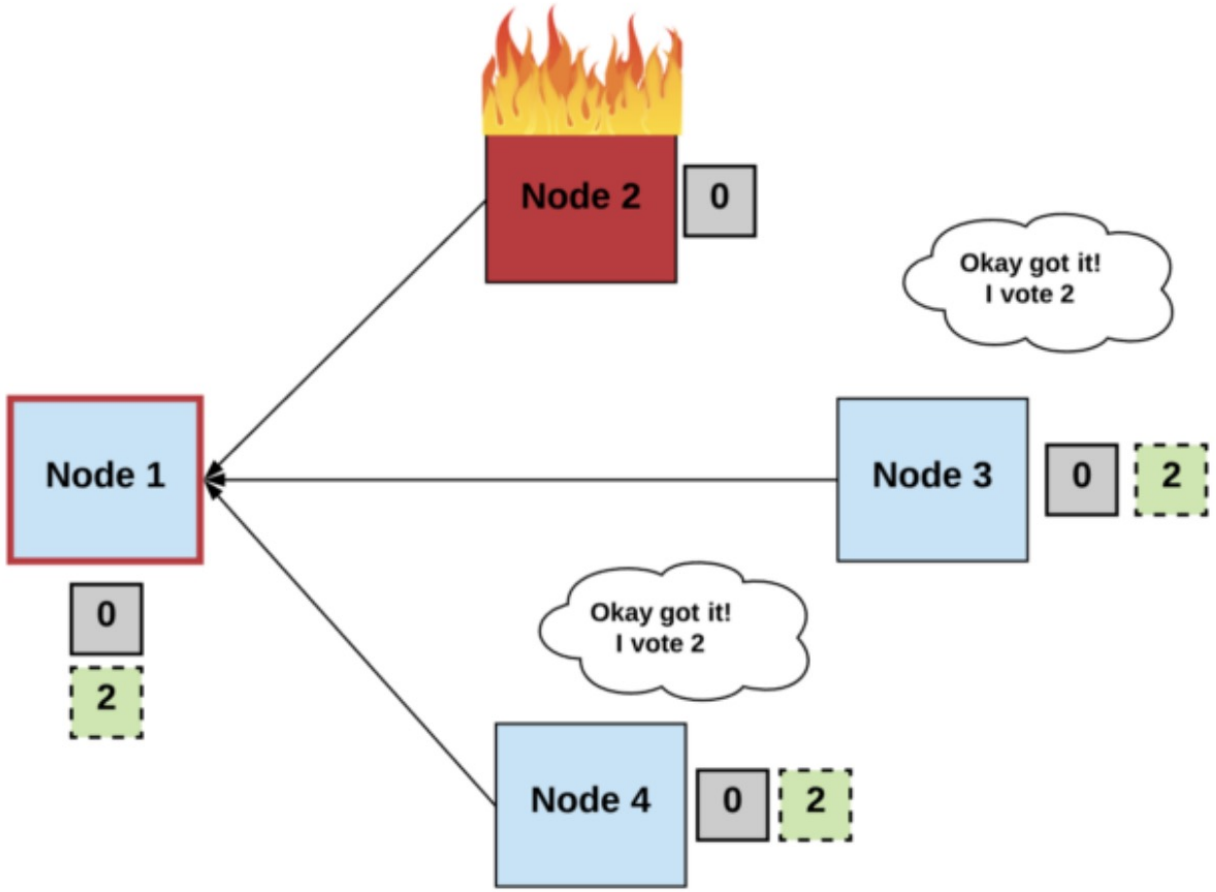
Consensus in the simplest setting

- Synchronous, reliable networks, free of faults
- The solution is trivial with one round of proposal messages.
- Intuition: all processes receive the same values sent by other processes.
- Step 1. At the beginning of each round, each Pi proposes value
- Step 2. At end of round, each Pi decides from received values.



Faults in practical distributed system

- It’s impossible to have a system free of faults



Faults in distributed system

- Failure model (Benign vs. Byzantine)
  - Benign faults:
    - Error is self-evident and components do not undergo incorrect state transition during failure
    - Examples:
      - Fail-stop: faulty nodes stop and do not send messages.
      - omission fault, crash fault, timing fault, data out-of-bound
  - Byzantine faults:
    - faulty nodes may send arbitrary messages.

Byzantine Generals Problem

- We imagine that several divisions of the Byzantine army are camped outside an enemy city, each division commanded by its own general.
- The generals can communicate with one another only by messengers.

[1] Lamport L, Shostak R, Pease M. The Byzantine generals problem[M]//Concurrency: the Works of Leslie Lamport. 2019: 203-226.

Byzantine Generals Problem

- After observing the enemy, they must decide upon a common plan of action.
- However, some of the generals may be traitors, trying to prevent the loyal generals from reaching an agreement.

[1] Lamport L, Shostak R, Pease M. The Byzantine generals problem[M]//Concurrency: the Works of Leslie Lamport. 2019: 203-226.

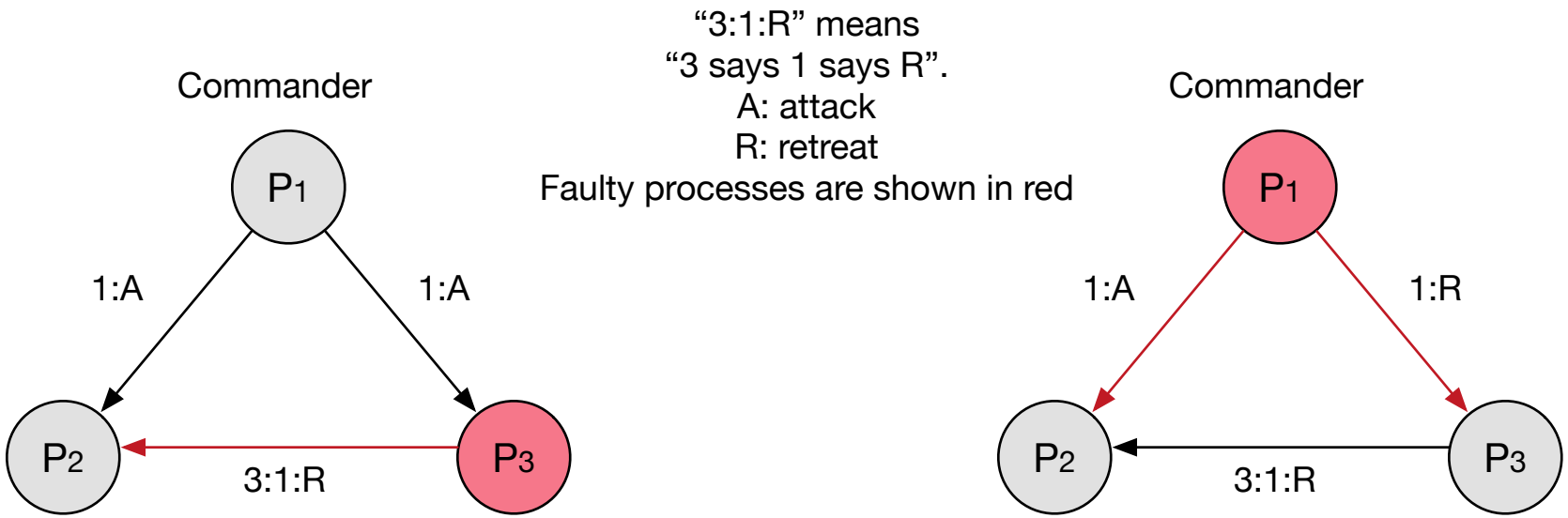
Byzantine Generals Problem

- Generals = Nodes/Processes
- The abstract problem
  - Each division of Byzantine army is directed by its own general.
  - There are n generals, some of which are traitors.
  - All armies are camped outside enemy castle, observing enemy.
  - Communicate with each other (private) by messengers.
- Requirements:
  - G1: All loyal generals decide upon the same plan of action
  - G2: A small number of traitors cannot cause the loyal generals to adopt a bad plan
- Note: We do not have to identify the traitors.

Naïve solution

- i^{th} general sends v_i to all other generals
- To deal with two requirements:
  - All generals combine their information $v_1, v_2, \ldots, v_{n-1}$ in the same way
  - Majority$(v_1, v_2, \ldots, v_{n-1})$, ignore minority traitors
- Naïve solution does not work:
  - Traitors may send different values to different generals.
  - Loyal generals might get conflicting values from traitors
- Requirement: Any two loyal generals must use the same value of v(i) to decide on the same plan of action.

Impossibility with three Byzantine generals



P2 cannot distinguish who is traitor and get the correct result

[Lamport 1982]

Intuition: No matter which process (commander P1 or subordinate P3) is the traitor, subordinate P2 cannot distinguish and cannot get the correct result.

Assumptions

- System model:
  - n processors, at most m are faulty
  - fully connected, message passing
  - receiver always knows the identity of the sender
  - reliable communication, only processors fail (byzantine)
  - the value communicated is 0 or 1 (A or R)
- Synchronous computation: processes run in a lock step manner.
  - In each step a process receives one or more messages, performs a computation, and sends one or more messages to other processes
  - A process knows all messages it expects to receive in a round.

Assumptions

- Byzantine faults: process can behave arbitrarily. They can change the contents of a message before it relays the message to other processes, i.e. it can lie about what it received from another process.
- Performance Aspects: number of rounds (time) and number of messages

Lamport's 1982 result, generalized by Pease

- The Lamport/Pease result shows that consensus is impossible:
  - with byzantine faults,
  - if one-third or more processes fail ($N \leq 3m$),
    - Lamport shows it for 3 processes, but Pease generalizes to N.
  - even with synchronous communication.
- Intuition: a node presented with inconsistent information cannot determine which process is faulty.
- Good news: consensus can be reached if $N > 3m$, regardless of fault type.

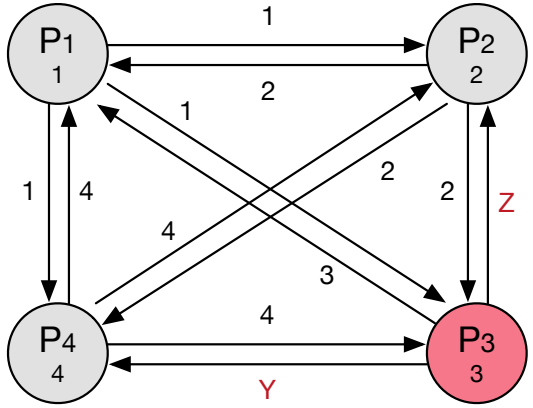
[1] Lamport L, Shostak R, Pease M. The Byzantine generals problem[M]//Concurrency: the Works of Leslie Lamport. 2019: 203-226.

Ex. (m=1, n=3m+1=4), assume P3 is faulty

- OM(1)
  - P1 sends 1 to P2, P3, P4.
  - Complexity: n-1 messages

	P1	P2	P3	P4
P2		2		
P3			3	
P4				4
ICV				

	P1	P2	P3	P4
P1	1			
P3			Z	
P4				4
ICV				



	P1	P2	P3	P4
P1	1			
P2		2		
P3			Y	
ICV				

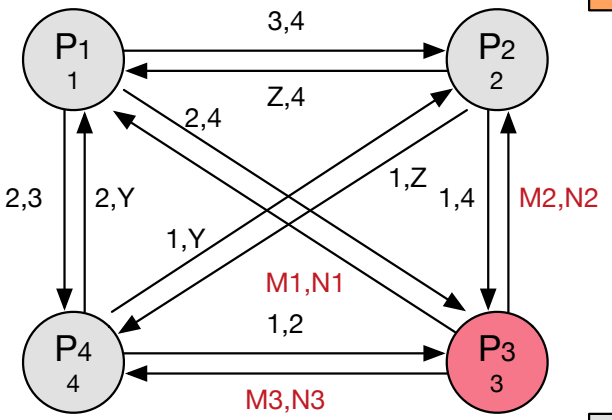
	P1	P2	P3	P4
P1	1			
P2		2		
P4				4
ICV				

Ex. (m=1, n=3m+1=4), assume P3 is faulty

- OM(0):
  - Each P_i acts as a source process, sends its value to each other $P_j, j \neq i$.
  - Complexity: (n-1)(n-2) messages

	P1	P2	P3	P4
P2		2	Z	4
P3		M1	3	N1
P4		2	Y	4
ICV				

	P1	P2	P3	P4
P1	1		3	4
P3	M2		Z	N2
P4	1		Y	4
ICV				



	P1	P2	P3	P4
P1	1	2	3	
P2	1	2	Z	
P3	M3	N3	Y	
ICV				

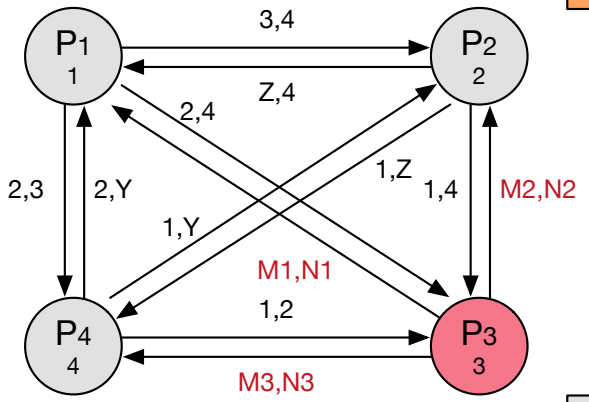
	P1	P2	P3	P4
P1	1	2		4
P2	1	2		4
P4	1	2		4
ICV				

Ex. (m=1, n=3m+1=4), assume P3 is faulty

- Decide
  - P1, P2, P4: (1, 2, NIL, 4), consensus completes successfully

	P1	P2	P3	P4
P2		2	Z	4
P3		M1	3	N1
P4		2	Y	4
ICV	1	2	NIL	4

	P1	P2	P3	P4
P1	1		3	4
P3	M2		Z	N2
P4	1		Y	4
ICV	1	2	NIL	4

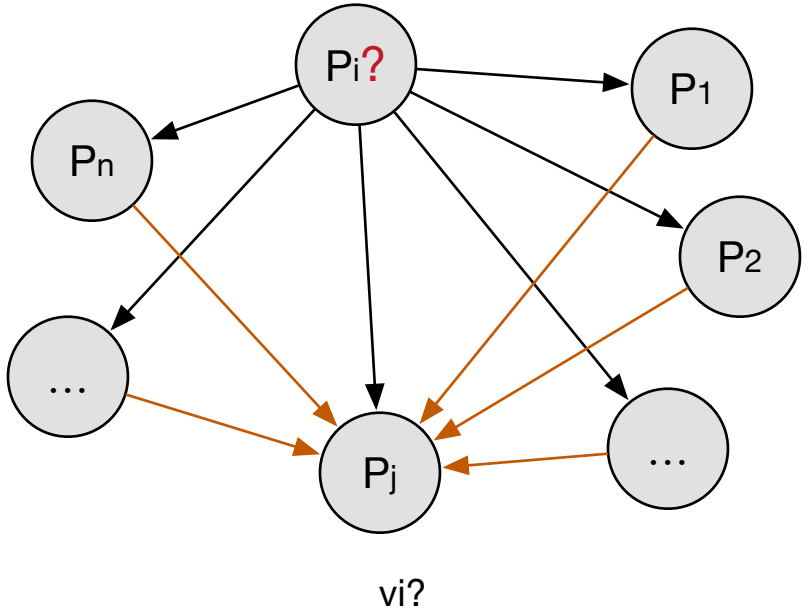


	P1	P2	P3	P4
P1	1	2	3	
P2	1	2	Z	
P3	M3	N3	Y	
ICV	1	2	NIL	4

	P1	P2	P3	P4
P1	1	2		4
P2	1	2		4
P4	1	2		4
ICV				

The procedure for general $n \geq 3m + 1$

- For P_j , how to get the correct value of P_i $(j \neq i)$
  - Always take the majority of the values received, does it work?



1. If P_i is the honest process
Always can get the correct majority

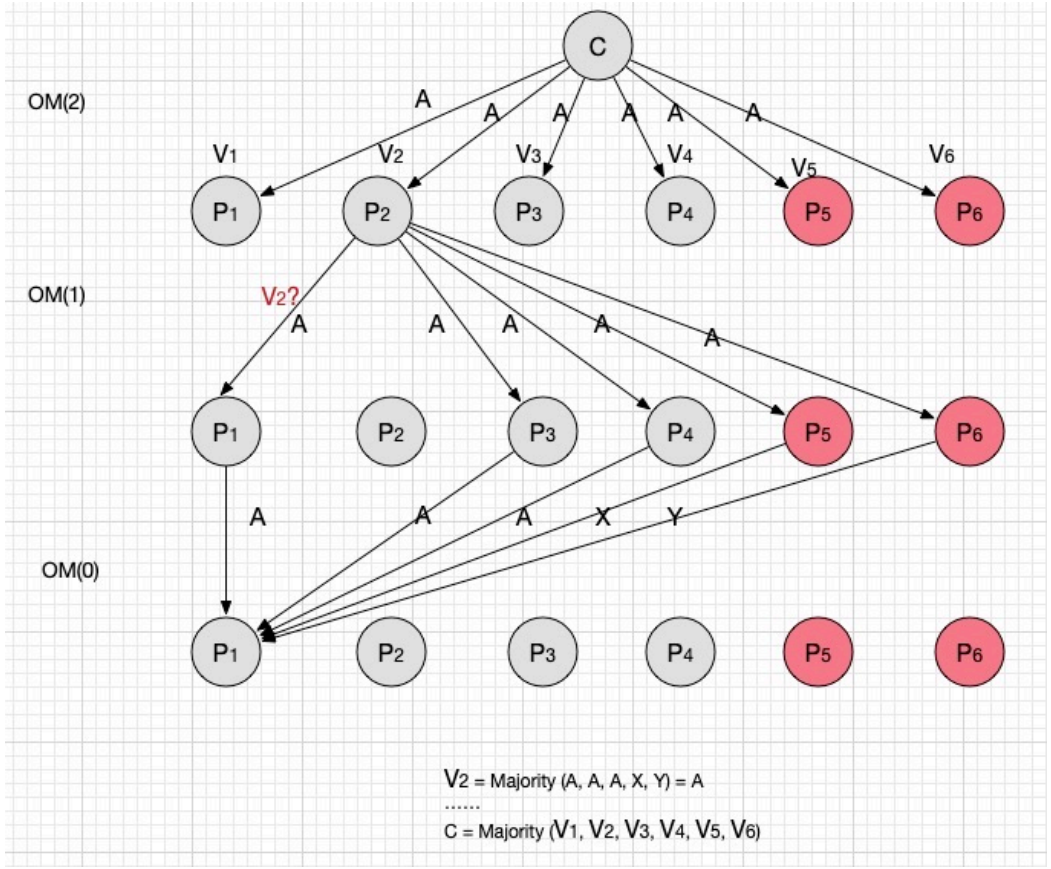
2. If P_i is the dishonest process

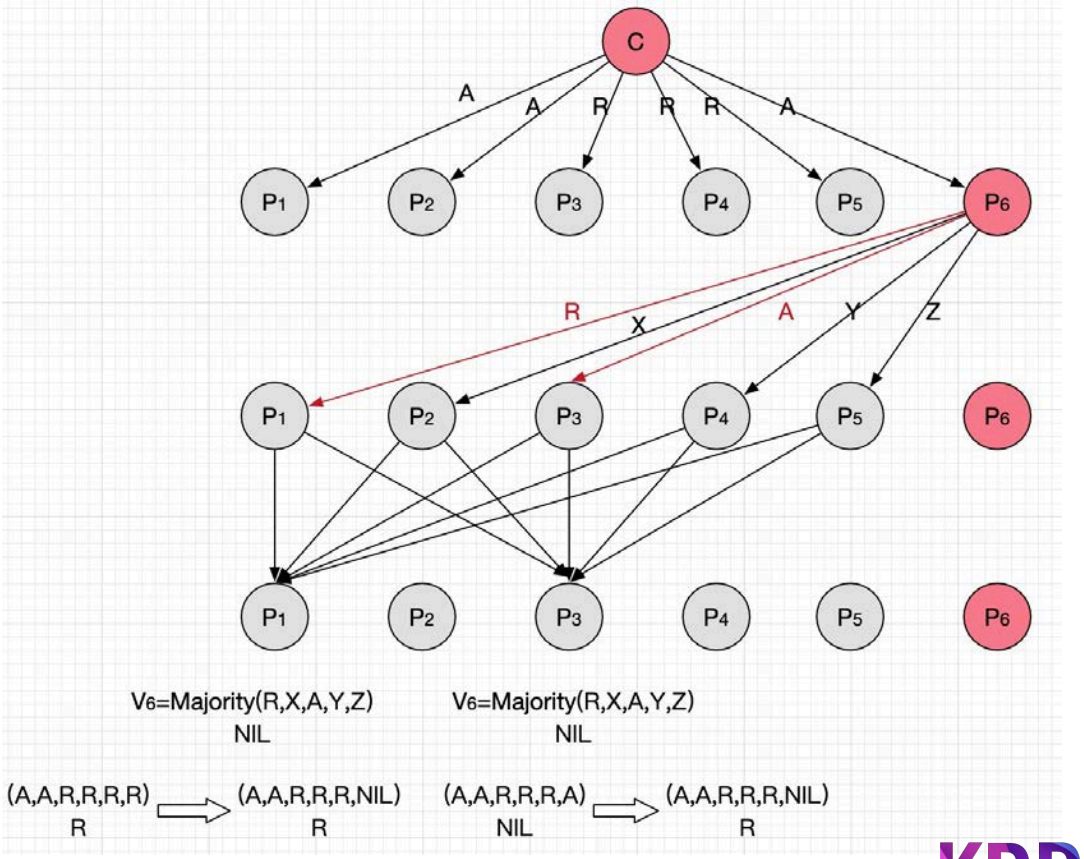
1) If majority exists, loyal processes would get the consistent result

1) If majority does not exisit, loyal processes would identify the faulty process and use the default result instead

The procedure for general $n \geq 3m + 1$

- For P_j , how to get the correct value of P_i $(j \neq i)$
  - Always take the majority of the values received, does it work?





Complexity analysis

- m=1, n=3m+1=4
  - Number of rounds: 2
    - First round: send their own values
    - Second round: transfer the received value from others
  - Number of messages: (n-1) + (n-1)(n-2)

General complexity analysis

- In general ($n \geq m + 1$):
  - Number of rounds: $m + 1$
    - The first round is for exchanging the self values.
    - The others m are for “processor X told me...”
  - Number of messages:
    - $(n-1) + (n-1)(n-2) + \cdots + (n-1)(n-2) \ldots (n-m)$
  - Complexity: O(n^m), extremely inefficient

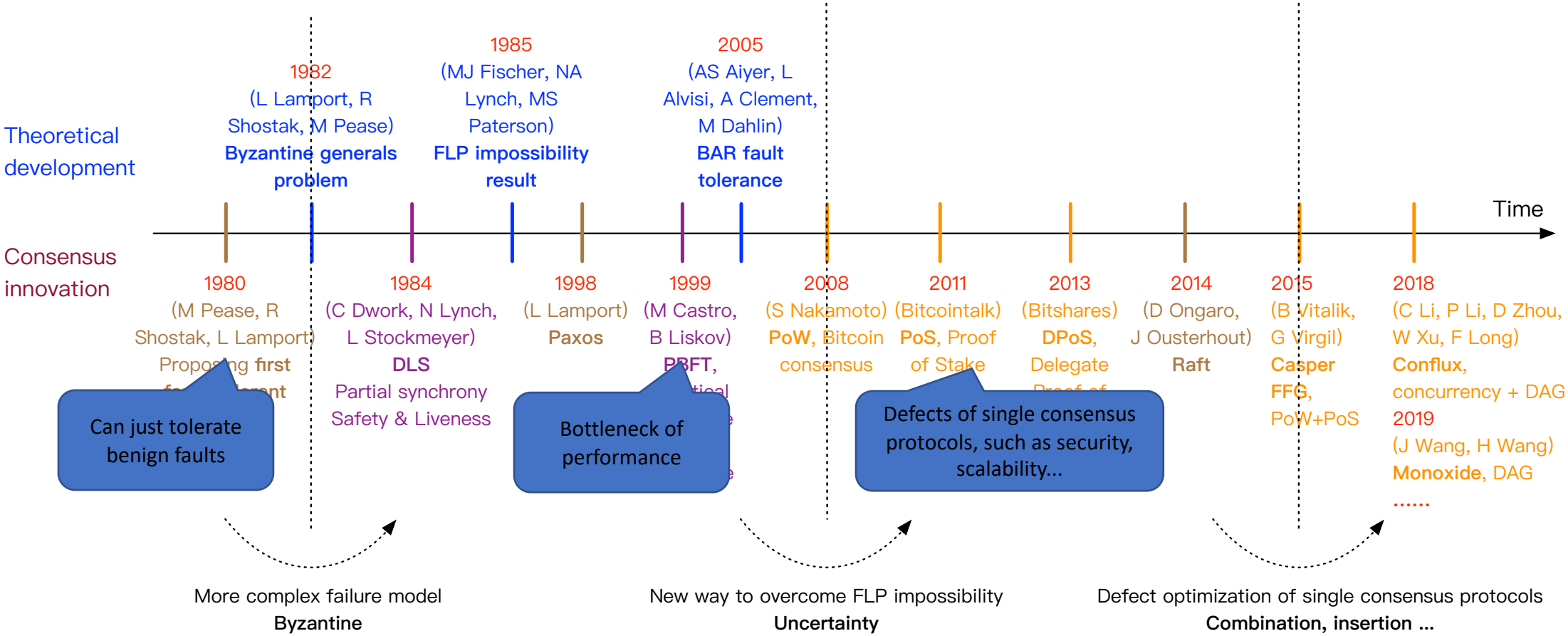
Summary: Byzantine Faults

- A solution exists if fewer than one-third are faulty ($n > 3m$).
- It works only if communication is synchronous.
- Like fail-stop consensus, the algorithm requires $m + 1$ rounds.
- The algorithm is very expensive and therefore impractical.
  - Number of messages is exponential in the number of rounds

Fischer-Lynch-Patterson (FLP Impossibility Result 1985)

- It is impossible to have a deterministic protocol that solves consensus in a message-passing asynchronous system in which at most one process may fail by crashing.
  - Intuition: a “failed” process may just be slow, and can rise from the dead at exactly the wrong time.
  - Incompatibility of a consensus: Asynchrony, determinism and fault-tolerance
- Necessity of a practical distributed ledger system: fault-tolerance
- Ways to overcome the FLP impossibility result
  - Synchronous assumptions
  - Randomness in protocol design
  - Hybrid: synchronous assumptions + randomness in protocol design

Timeline of consensus development



- Theoretical development
- Innovation of simple fault-tolerant consensus
- Innovation of Byzantine fault-tolerant consensus
- Innovation of BAR fault-tolerant consensus

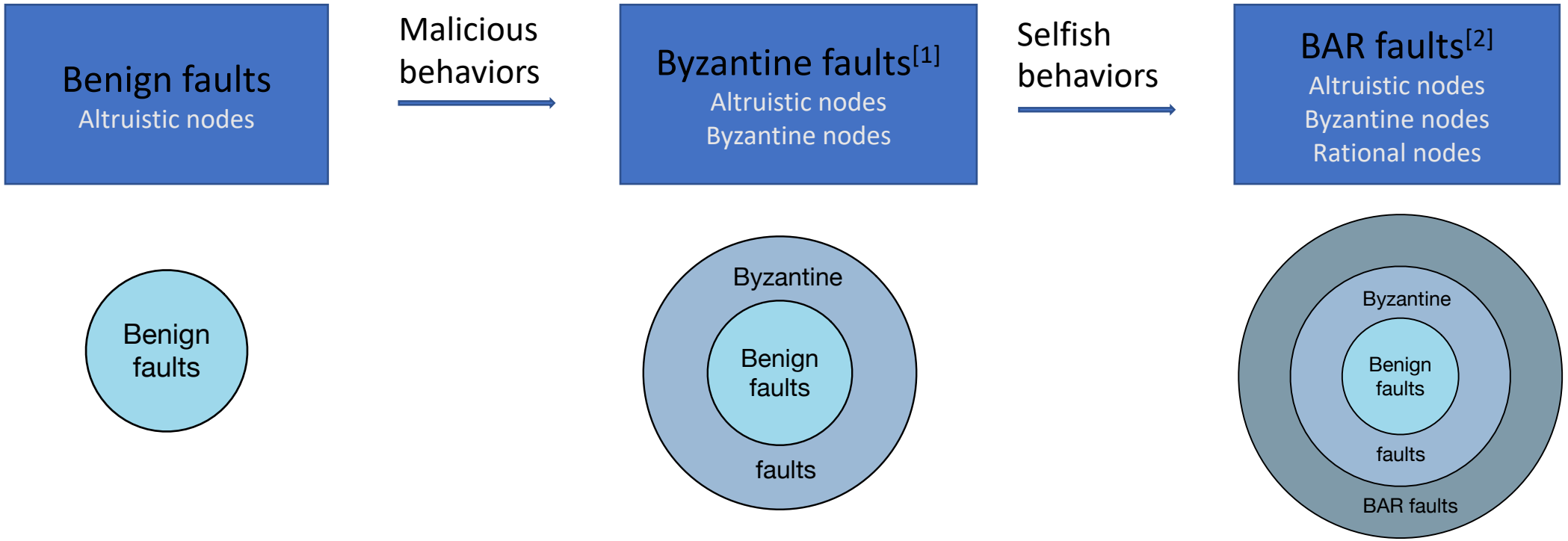
KDD2021

Extended failure model

- BAR faults: more complicated behaviors of attackers
  - Byzantine: Byzantine nodes aim to harm the system with malicious behavior all the time
  - Altruistic: Honest nodes always follow the protocol.
  - Rational: Rational nodes only follow the protocol if it benefits them.

Extended failure model

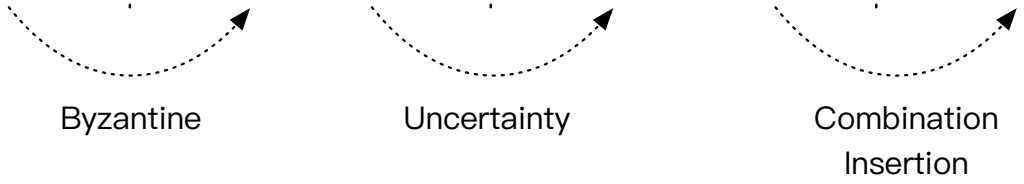
- Specify the faults that consensus can tolerate



[1] Lamport L, Shostak R, Pease M. The Byzantine generals problem[M]//Concurrency: the Works of Leslie Lamport. 2019: 203-226.
[2] Aiyer A S, Alvisi L, Clement A, et al. BAR fault tolerance for cooperative services[C]//Proceedings of the twentieth ACM symposium on Operating systems principles. 2005: 45-58.

Four stages of consensus development

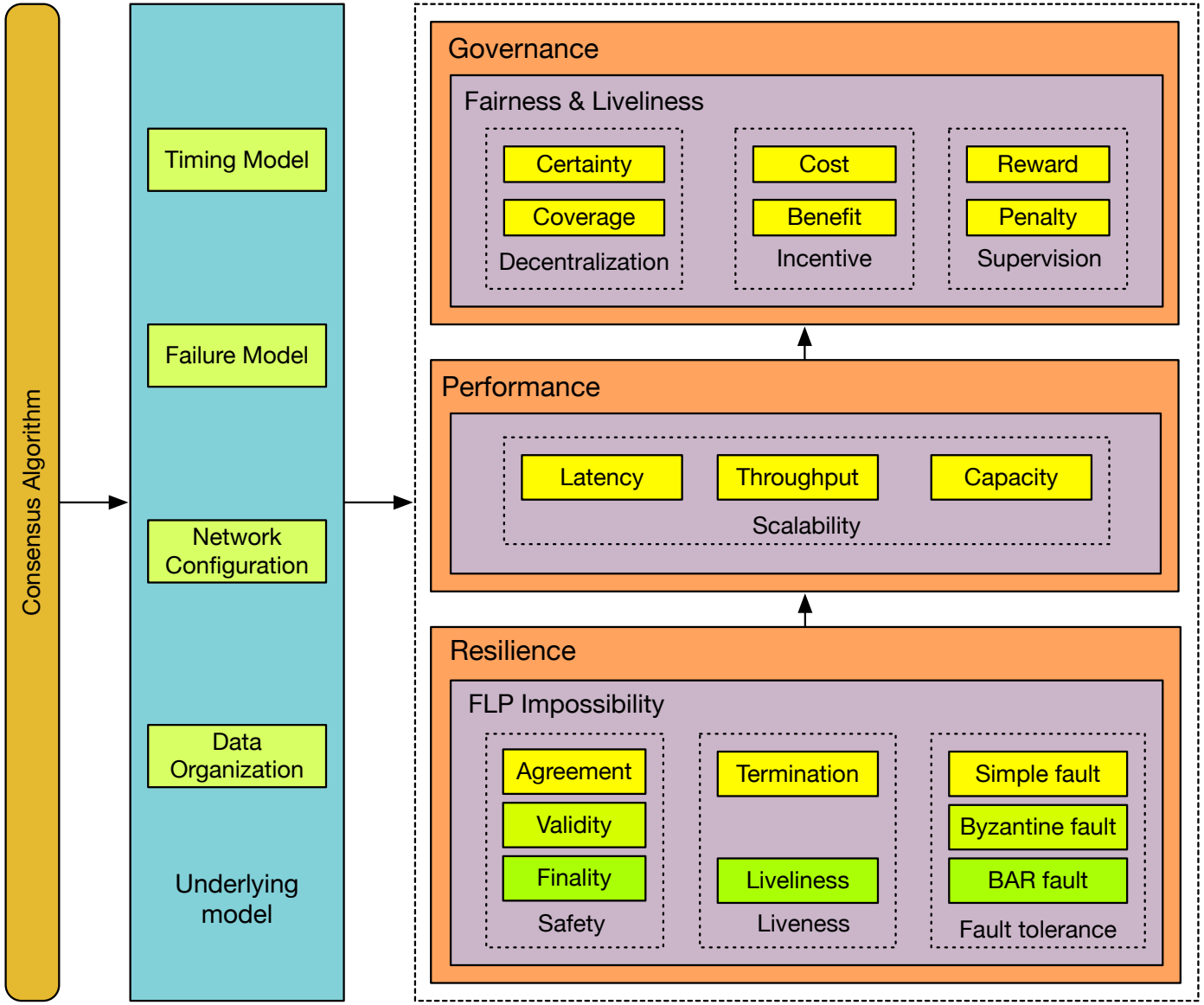
Stages	1st stage	2nd stage	3rd stage	4th stage
Fault-tolerance	Benign faults	Byzantine faults	BAR faults	BAR faults
Ways to overcome FLP result	Timeouts (synchronous assumptions)	Partial synchrony (synchronous assumptions)	Randomness in protocol design	Hybrid: synchronous assumptions + randomness in protocol design
Reliability	Agreement	Agreement, Validity, Safety	Agreement, Validity, Finality, Safety	Agreement, Validity, Finality, Safety
Sustainability	Termination	Termination, Liveness	Termination, Liveliness, Liveness	Termination, Liveliness, Liveness



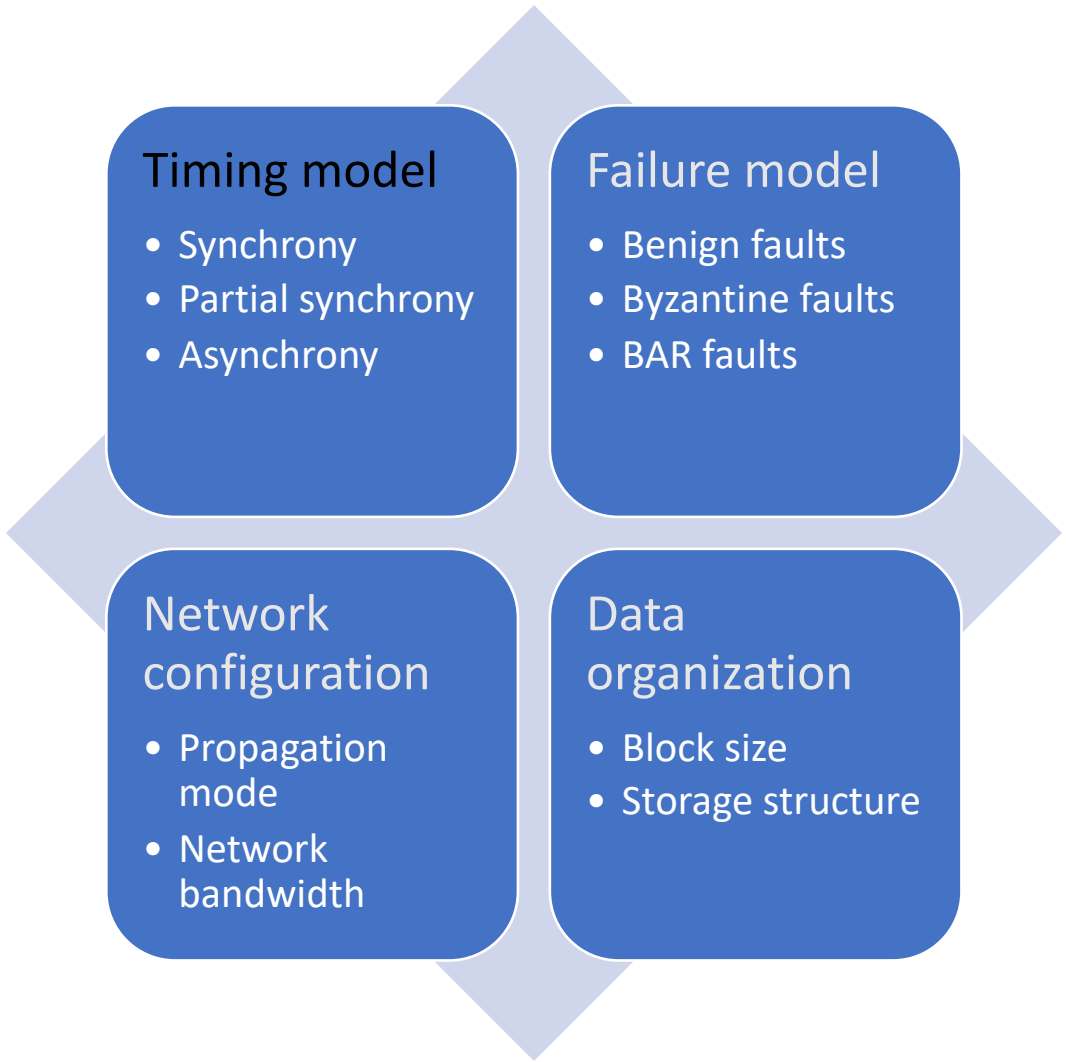
Extended consensus definition

- An algorithm achieves consensus if it satisfies the following two conditions:
  - Safety
    - Agreement: all non-faulty processes decide on the same output value
    - Validity: the decided value must be one of the non-Byzantine inputs
    - Finality: Once a consensus is recorded, it should be immutable
  - Liveness
    - Termination: all non-faulty nodes eventually decide on some output value
    - Liveliness: all non-faulty nodes are available to decide new values continuously

Overview

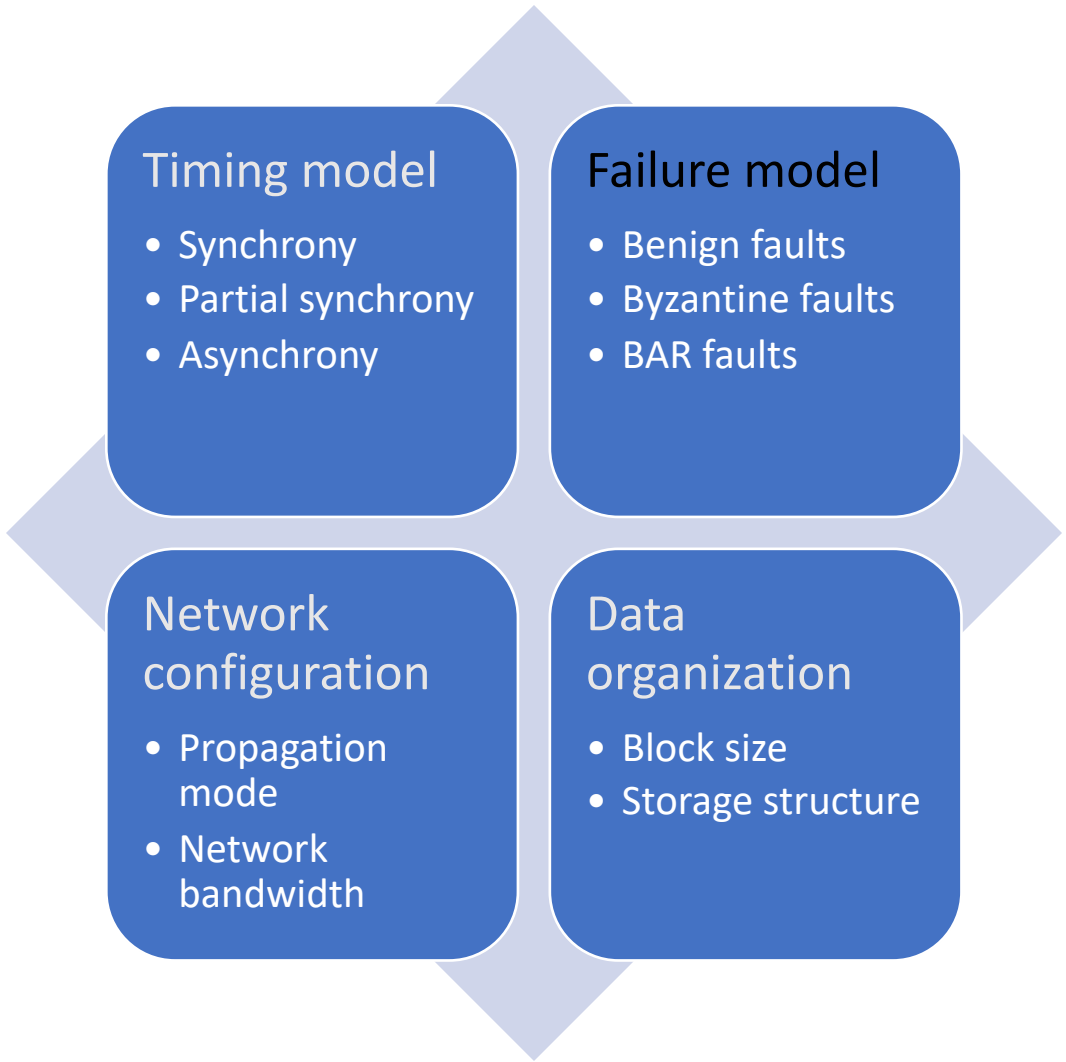

Consensus Algorithm
Timing Model
Failure Model
Network Configuration
Data Organization
Underlying model
Governance
Fairness & Liveliness
Certainty
Coverage
Decentralization
Cost
Benefit
Incentive
Reward
Penalty
Supervision
Performance
Latency
Throughput
Capacity
Scalability
Resilience
FLP Impossibility
Agreement
Validity
Finality
Safety
Termination
Liveliness
Liveness
Simple fault
Byzantine fault
BAR fault
Fault tolerance

Underlying model: timing model



- Specifies the delay of message-passing
  - Synchrony: a known upper bound on the delay of messages
  - Partial synchrony: an unknown upper bound on the delay of messages
  - Asynchrony: no upper bound on the delay of messages

Underlying model: failure model



- Specifies the faults that consensus can tolerate
  - Benign faults: faults of processes are self-evident
    - Crash faults: fail-stop faults of processes/nodes
    - Omission faults: interrupt faults of message-passing
    - Timing faults: processes' response lies outside the specified time interval
  - Byzantine faults: processes behave maliciously
    - General malicious behaviors of Byzantine nodes
  - BAR faults: Byzantine, Altruistic and Rational
    - Malicious behaviors of Byzantine nodes
    - Attacks by rational nodes out of selfish interest

Attacks for BAR faults tolerance

- General attacks in distributed ledger environment

Attacks	Target	Description
Nothing-at-Stake[1]	PoS	A situation where someone loses nothing when behaving badly, but stands to gain everything
Selfish-mining[2]	PoW	Miners selectively withhold mined blocks and only gradually publish them
Eclipse attacks[3]	General	Attackers tend to create a logical partition in the network

- Root-cause: double-spending

Attacks	Description
Double-spending[4]	Attackers tend to use the same tokens to issue two (or more) transactions

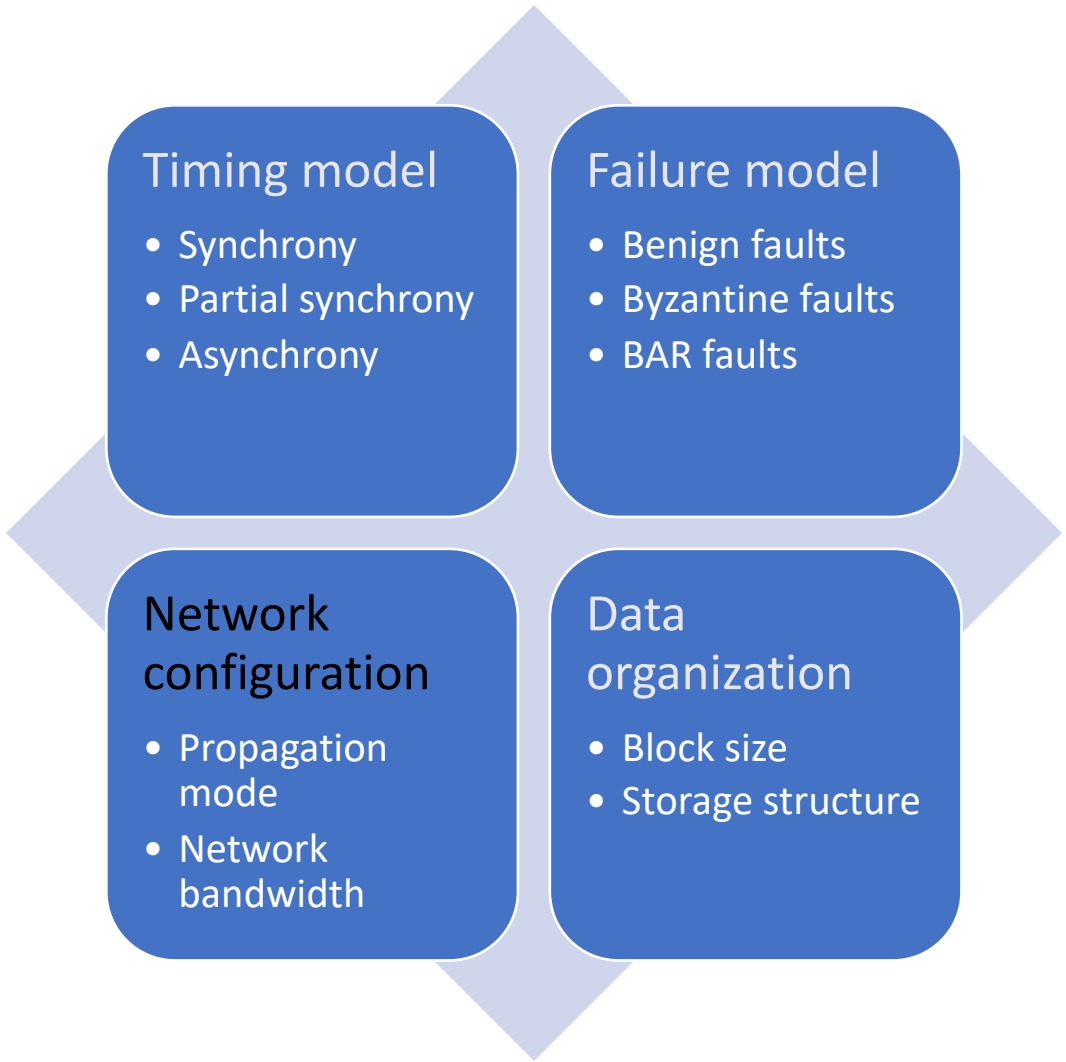
[1] Saleh F. Blockchain without waste: Proof-of-stake[J]. Available at SSRN 3183935, 2019.

[2] Sapirshtein A, Sompolinsky Y, Zohar A. Optimal selfish mining strategies in bitcoin[C]//International Conference on Financial Cryptography and Data Security. Springer, Berlin, Heidelberg, 2016: 515-532.

[3] Wüst K, Gervais A. Ethereum eclipse attacks[R]. ETH Zurich, 2016.

[4] Karame G O, Androulaki E, Capkun S. Double-spending fast payments in bitcoin[C]//Proceedings of the 2012 ACM conference on Computer and communications security. ACM, 2012: 906-917.

Underlying model: network configuration

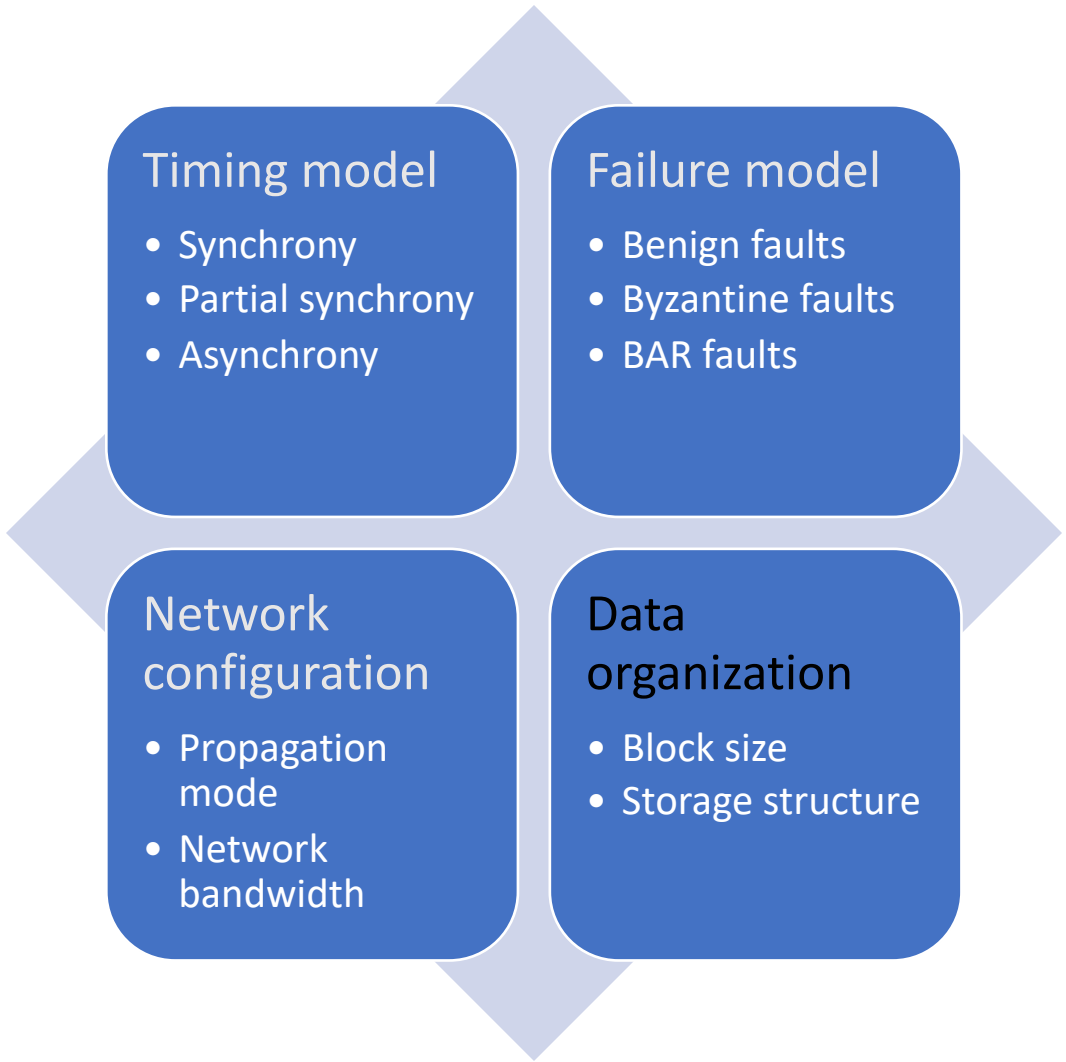


- Specifies networking environment consensus runs on
  - Propagation mode

Network overlay	Propagation mode	Propagation complexity
Full connection	Multicast	$O(C)$
Partial connection	Broadcast	$O(\log(N))$

  - Network bandwidth
    - Capacity of transmission

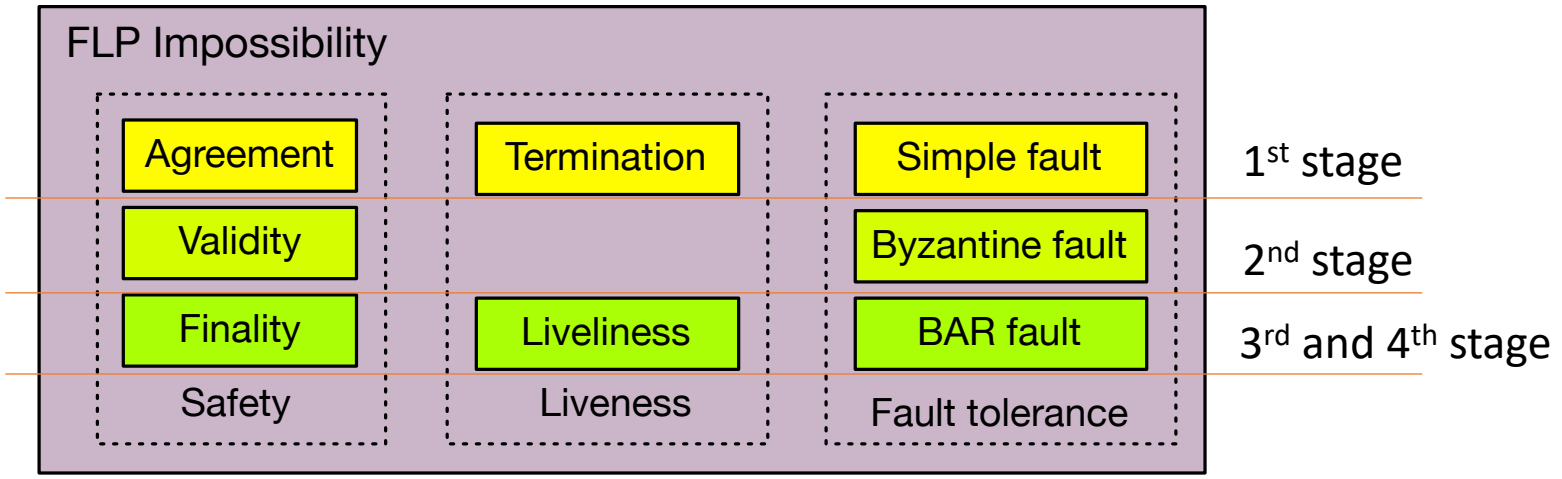
Underlying model: data organization



- Specifies data storage
  - Block size: capacity of block
  - Storage structure:
    - DAG, Blockchain

Evaluative model: resilience

- Evaluative dimensions
  - Safety
  - Liveness
  - Fault tolerance



Safety	Agreement	Non-faulty processors must decide on the same value
	Validity	The decided value must be one of the non-Byzantine inputs
	Finality	Once a consensus is recorded, it should be immutable
Liveness	Termination	Processors must decide the value in bounded time
	liveliness	Processors must be available to decide new values continuously
Fault tolerance		Determines the type of fault that can be tolerated and the rate of failure nodes tolerated by the system to maintain the safety (finality in DLT-based system) and liveness

Evaluative model: performance

- Evaluative dimensions

Scalability	Latency	Confirmation speed of transactions
	Throughput	Number of transactions processed per second (TPS)
	Capacity	The space utilization efficiency of memory

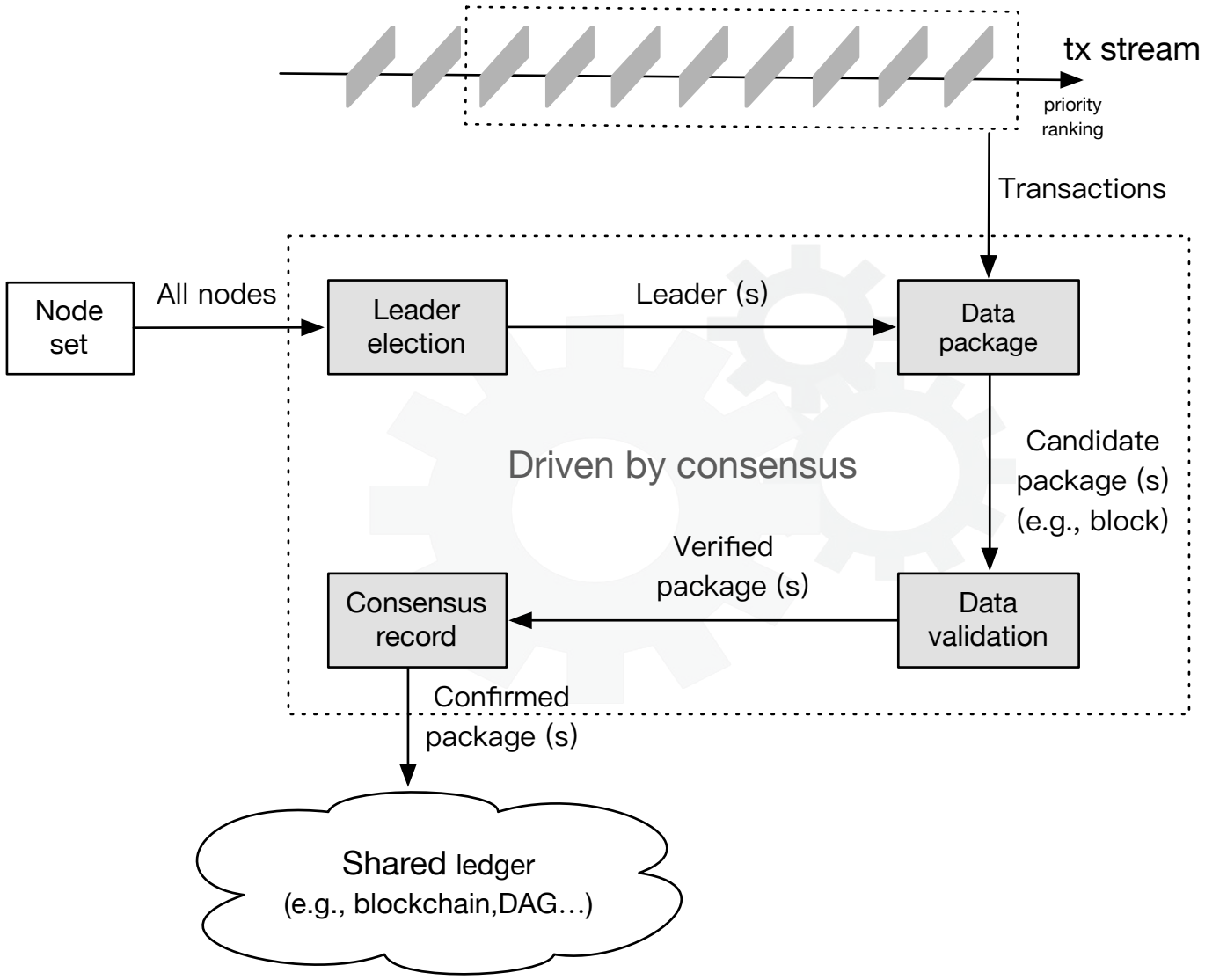
- Scalability
  - Extensibility of the number of nodes in the system
  - Constrained by latency, throughput and capacity simultaneously

Evaluative model: governance

- Evaluative dimensions

Decentralization	Certainty	The determinism in leader node selection
	Coverage	The ratio of consensus participating nodes to all nodes (e.g., full or partial)
Incentive	Cost	The resource or token cost for participating in consensus
	Benefit	System earnings for the miners or validators after they successfully complete a round of consensus
Supervision	Reward	System rewards for the supervision and reporting of the miners or validators
	Penalty	Punishment of miners or validators for their intentional attacks on the consensus

Extended consensus process



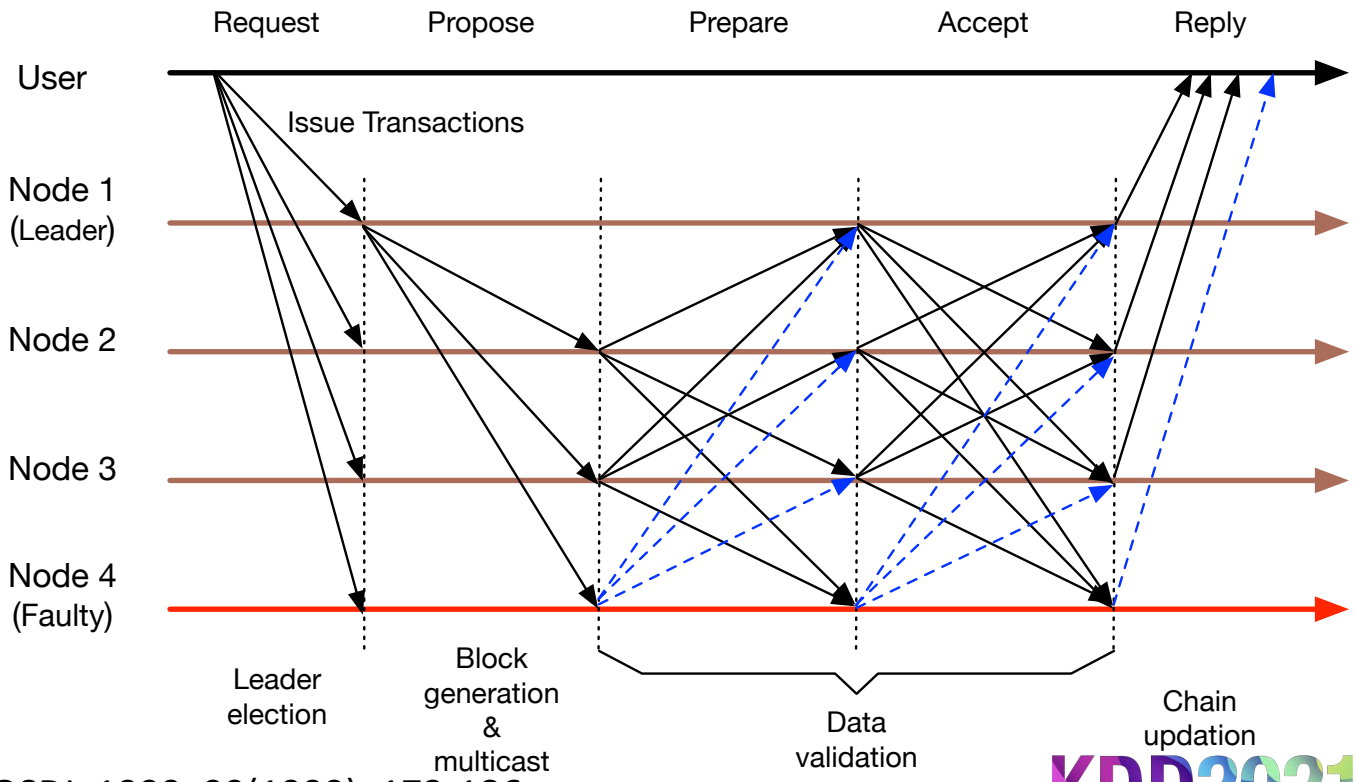
Consensus evaluation

- BFT-style consensus (2nd stage): PBFT
  - Byzantine fault-tolerant distributed consensus
  - Deterministic consensus in partially synchronous system
    - Applicable to consortium or permissioned system

Practical Byzantine Fault Tolerance(PBFT)

- Motivation
  - Design an efficient Byzantine fault-tolerant consensus
  - Reduce the complexity from traditional exponential to polynomial
- Process

Leader election	Round robin View change for the offline/malicious proposers
Data package	Proposer package the block and multicast to validators through direct P2P connection
Data validation	Verify and multicast
Consensus record	Commit the block and reply the request of the client



[1] Castro M, Liskov B. Practical Byzantine fault tolerance[C]//OSDI. 1999, 99(1999): 173-186.

PBFT Evaluation

- Underlying model

Timing model	Failure model	Network configuration	Data organization
Partially synchronous	Byzantine faults	Full connection	Not specified

PBFT Evaluation

- Evaluation: resilience
  - Fault-tolerance assumptions: $n \geq 3f + 1$ by number of nodes

Safety			Liveness
Agreement	Validity	Finality	Termination
Deterministic Round-robin, view change mechanism and cross validation	Deterministic View change mechanism and cross-validation	Deterministic Once agreement reached, result cannot be modified	Deterministic by synchrony assumptions

  - Way to overcome FLP impossibility
    - Use synchrony assumption to overcome the FLP impossibility

PBFT Evaluation

- Evaluation: governance
  - Because the PBFT is a traditional and deterministic consensus algorithm, there is no profit competition, so we don’t consider more about the governance problem.

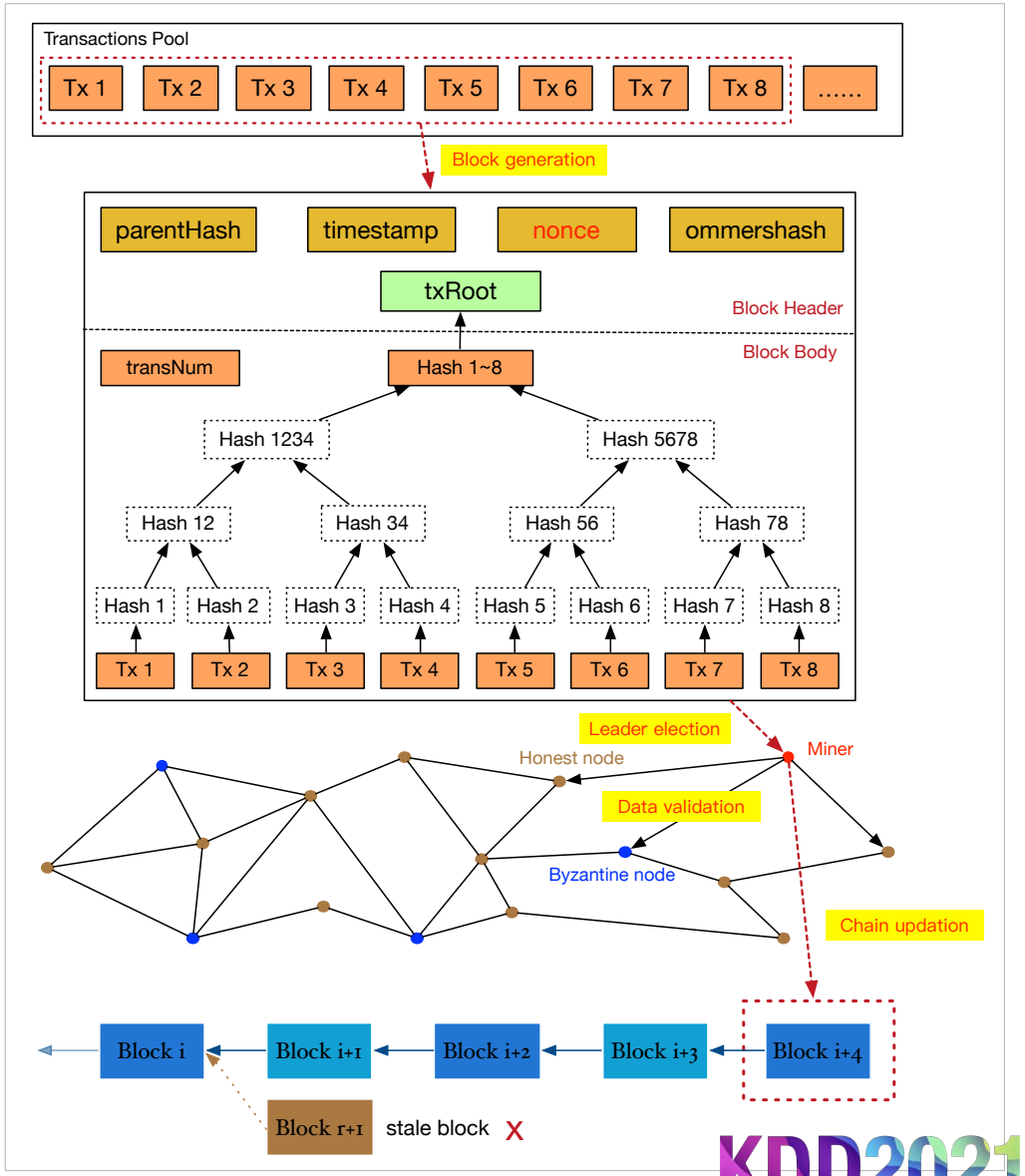
Consensus evaluation

- PoX-style consensus (3rd stage): PoW -> PoS
  - BAT fault-tolerant distributed consensus
  - Non-deterministic consensus in an asynchronous (by timing model) system
    - Applicable to permissionless system

PoW

- Motivation
  - Limited scalability in the 2nd stage
  - Large-scale application of consensus algorithms
- Process

Leader election	Computing competition Puzzle: $H(x\|\|nonce) \leq Difficulty$
Data package	The first miner finding the “nonce” can package the complete block
Data validation	Verify and broadcast blocks
Consensus record	Commit the block



[1] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system[J]. 2008.

PoW Evaluation

- Underlying Model

Timing model	Failure model	Network configuration	Data organization
Asynchronous	BAR faults	Partial connection	Blockchain

- Typical attacks of rational nodes
  - Double-spending
  - Eclipse attacks
  - Selfish-mining

PoW Evaluation

- Evaluation: resilience

Safety			Liveness
Agreement	Validity	Finality	Termination
Probabilistic Depends on the speed of message transmission in the P2P network	Probabilistic Depends on computing power of Byzantine nodes	Temporary Due to PoW protocol design	Deterministic Depends on Adjustment of puzzle difficulty

- Clarification on PoW overcoming FLP impossibility result
  - PoW does not achieve the type of consensus as constrained by FLP theorem.
  - PoW indeed includes non-determinism to mitigate attacks.

PoW Evaluation

- Evaluation: governance

Decentralization		Incentive		Supervision	
Certainty	Coverage	Cost	Benefit	Reward	Penalty
Probabilistic	Full	Consumption of electricity	Mining rewards	No reward	waste of electricity due to useless mining of stale blocks

- Analysis of fairness & liveliness
  - Randomness of hash function $H(\cdot)$ guarantees the fairness
  - Incentive from tokenomics + fairness guarantees liveliness

PoW Evaluation

- Anti-attacking

Attacks	Anti-attacking
Double-spending	Probabilistic depends on the computing power of attackers
Eclipse attack	Network connection initiation Enable an adversary to carry out 51% attacks with less than 51% computing power
Selfish-mining	Probabilistic depends on the computing power of attackers

Proof of Stake[1] (PoS)

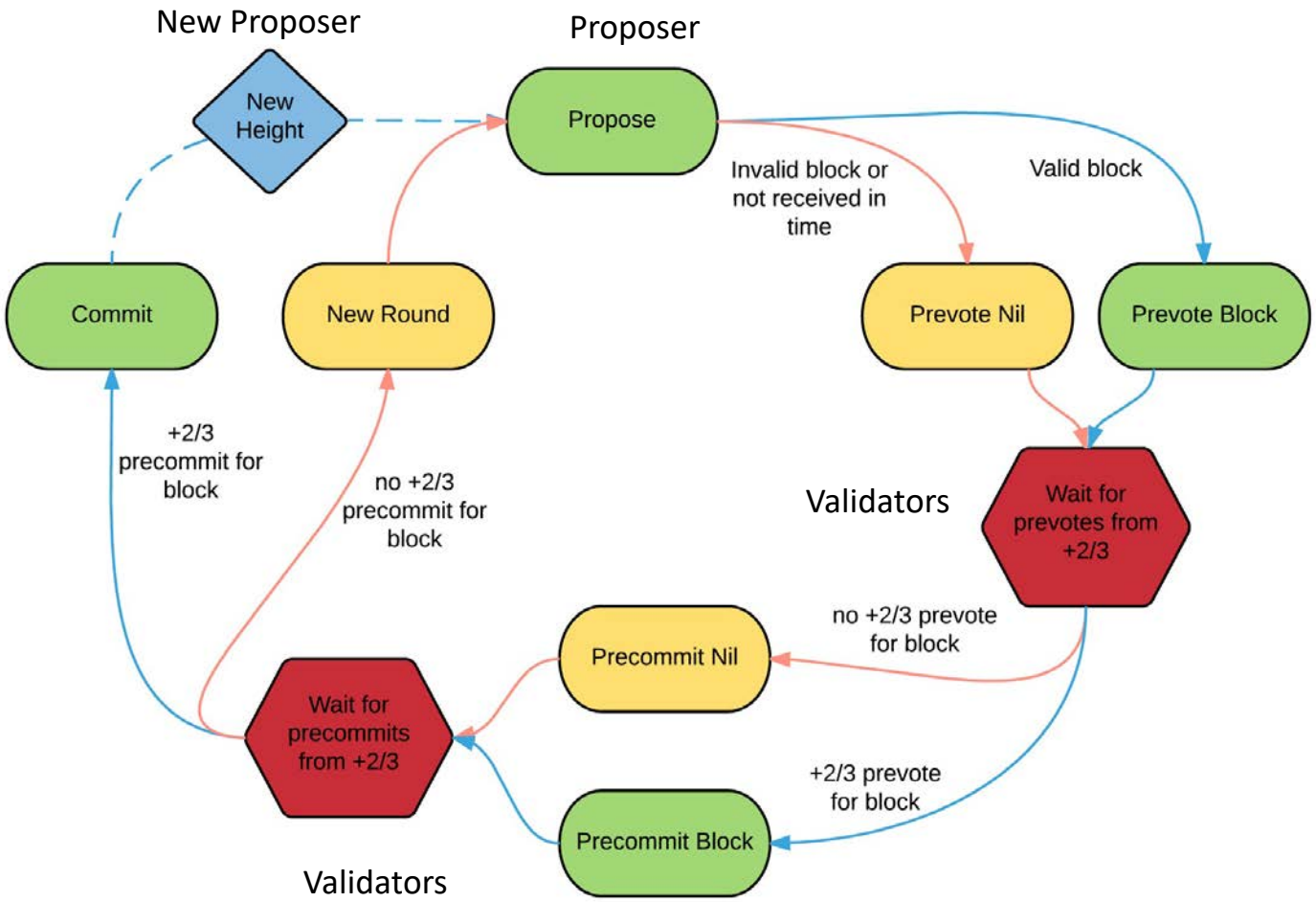
- Concept: proposals are made by and voted on those who can prove ownership of some stake of tokens in the network
- Motivation (contrast with PoW):
  - Save resources, mainly electricity power
- Implementation (Hybrid):
  - BFT-style PoS: Tendermint[2] (PBFT + PoS)
  - PoX-style PoS: Casper CFG[3] (PoW + BFT-style PoS)

[1] Saleh F. Blockchain without waste: Proof-of-stake[J]. Available at SSRN 3183935, 2019.
[2] Buchman E. Tendermint: Byzantine fault tolerance in the age of blockchains[D]. , 2016.
[3] Buterin V, Griffith V. Casper the friendly finality gadget[J]. arXiv preprint arXiv:1710.09437, 2017.

Tendermint: PBFT + PoS

- Motivation
  - Provide a secure consensus with accountability guarantees
  - PoS: voting power measured by stake, vs. nodes number in PBFT
- Process

New Height	Leader election	Round robin, simplified processing for offline/malicious proposers (directly skipped)
Rounds	Data package	Propose a new block and then gossip to the other validators
	Data validation	Two steps of voting to verify: Prevote & Precommit
Commit	Consensus record	Commit the block



[1] Buchman E. Tendermint: Byzantine fault tolerance in the age of blockchains[D]. , 2016.

Tendermint Evaluation

- Underlying model

Timing model	Failure model	Network configuration	Data organization
Partially synchronous	BAR faults	Partial connection	Blockchain

Tendermint Evaluation

- Evaluation: resilience
  - Fault-tolerance assumptions: $n \geq 3f + 1$ by voting power

Safety			Liveness
Agreement	Validity	Finality	Termination
Deterministic BFT-style consensus	Deterministic Voting validation	Deterministic BFT-style consensus, once agreement reached, result cannot be modified	Deterministic by synchrony assumptions

  - Way to overcome FLP impossibility
    - Use the synchrony assumptions to overcome the FLP impossibility

Tendermint Evaluation

- Evaluation: governance

Decentralization		Incentive		Supervision	
Certainty	Coverage	Cost	Benefit	Reward	Penalty
Deterministic	Partial by token deposit	Deposit for participating in consensus	System rewards for BFT vote	Partial deposit rewards for supervision	Whole deposit for any attacking behavior

Accountability of BFT Vote:
-> Deposit: validators must bond some stake in order to participate in consensus
-> Penalty: system can burn deposit of any attacker

- Analysis of fairness and liveliness
  - Use supervision to guarantee the fairness
  - Incentive from tokenomics + partial coverage by token deposit + fairness guarantees liveliness

Tendermint Evaluation

- Anti-attacking

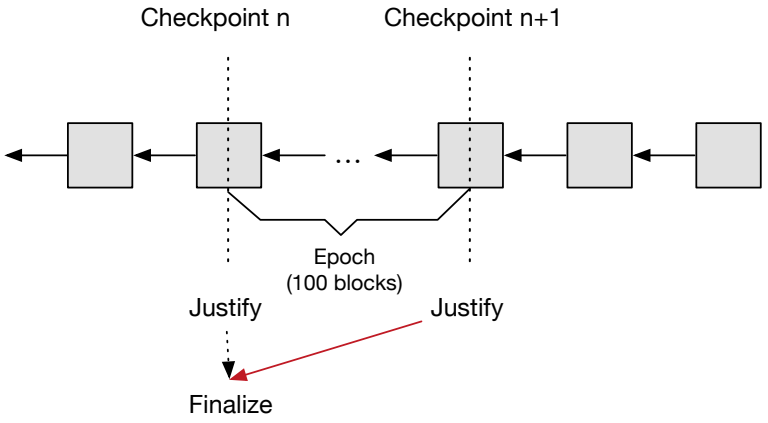
Attacks	Anti-attacking
Double-spending	Absolute: deterministic finality
Eclipse attacks	Probabilistic: depend on the offline/partition ratio If $1/3$ or more of the validators are offline or partitioned, the network may halt altogether
Nothing-at-stake	Absolute: clock mechanism guarantee that only one block can be proposed at the same height

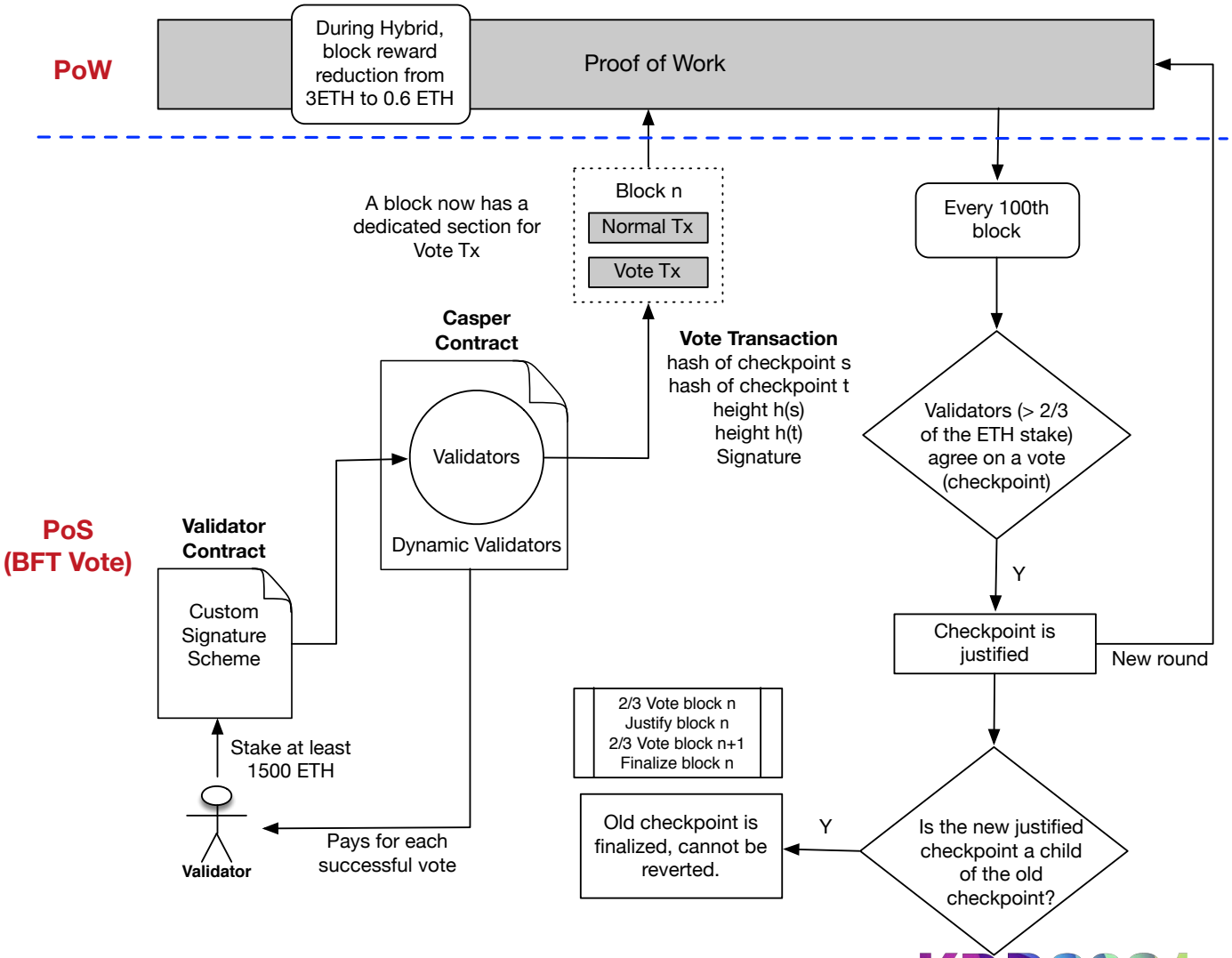
Tendermint Evaluation

- Evaluation: governance
  - Because the Tendermint is a deterministic consensus algorithm, there is no profit competition, so we don’t consider more about the governance problem.

Casper FFG: PoW + BFT-style PoS

- Motivation
  - Deterministically finalize the block of PoW chain
- Process
  - Designed as BFT-style PoS existing on top of the PoW chain
  - Three states of each checkpoint:
    - committed, justified and finalized





[1] Buterin V, Griffith V. Casper the friendly finality gadget[J]. arXiv preprint arXiv:1710.09437, 2017.

Casper FFG Evaluation

- Underlying Model

Timing model	Failure model	Network configuration	Data organization
Asynchronous Depends on PoW Partially synchronous Depends on BFT vote	BAR faults	Partial connection	Blockchain

  - Typical attacks of rational nodes
    - Double-spending
    - Eclipse attacks
    - Nothing-at-Stake
    - Selfish-mining

Casper FFG Evaluation

- Evaluation: resilience
  - Fault-tolerance assumptions
    - PoW: probabilistic
    - BFT vote: $n \geq 3f + 1$ by voting power

Safety			Liveness
Agreement	Validity	Finality	Termination
Probabilistic Depends on latency in PoW Deterministic After BFT vote	Probabilistic Depends on computing power of Byzantine nodes in PoW	Temporary Due to PoW protocol design Deterministic After BFT vote	Deterministic Depends on PoW

  - Clarification on Casper FFG overcoming FLP impossibility result
    - Similarly to the PoW, Casper FFG does not achieve the type of consensus as constrained by FLP theorem.
    - But because of the BFT vote, contrast with PoW, it has the eventual finality

Casper FFG Evaluation

- Evaluation: governance

Decentralization		Incentive		Supervision	
Certainty	Coverage	Cost	Benefit	Reward	Penalty
Probabilistic depends on PoW Deterministic depends on BFT vote	Full depends on PoW Partial depends on BFT vote	1) Consumption of electricity power of PoW 2)Deposit of BFT vote	1) Mining rewards of PoW 2) Voting rewards of BFT vote	"Finder's fee" of submitter of the slashing condition of BFT vote	1) Useless mining of stale block of PoW 2) Punishment of deposit for malicious behaviors BFT vote

  - Analysis of fairness and liveliness
    - Randomness guarantees the fairness of PoW
    - Incentive from tokenomics + fairness guarantees liveliness of PoW
    - Supervision guarantees the fairness of BFT vote
    - Incentive from tokenomics + partial coverage by token deposit + fairness guarantees liveliness

Casper FFG Evaluation

- Anti-attacking

Attacks	Anti-attacking
Selfish-mining	Probabilistic: during a checkpoint interval depends on the computing power of attackers
Eclipse attacks	Network connection initiation Enable an adversary to carry out 51% attacks with less than 51% computing power
Double-spending	Probabilistic: during a checkpoint interval depends on the computing power of attackers
Nothing-at-Stake	Absolute: accountability If a validator violates a rule, consensus can detect the violation and know which validator violated and penalize the validator with the entire deposit

Tutorial Outline

- Data Asset: What and Why
- Data Asset Core Components
- Data Asset Governance for Decentralized Collaborative Intelligence
  - Governance principles
  - "Trust" for data asset governance for decentralized collaborative intelligence
    - Agreement
    - Accounting
    - Auditing
    - Privacy
  - "Incentive" for data asset governance for decentralized collaborative intelligence
- Data Asset Ecosystems
- Challenges and Future Directions

Data Asset for Collaborative Intelligence

Data Auditing (DA)

Xin Mu[1], Feida Zhu[2]

1. Peng Cheng National Laboratory, Shenzhen, China, mux@pcl.ac.cn
2. Singapore Management University, fdzhu@smu.edu.sg,

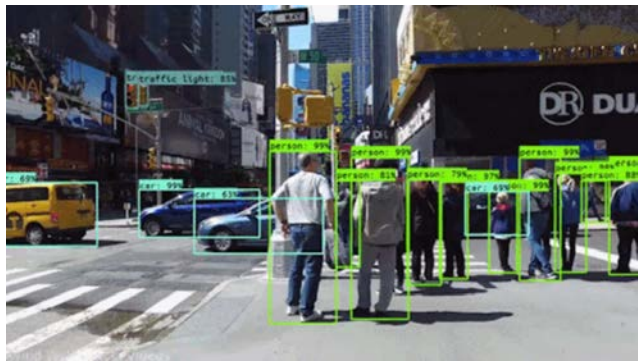
Outline

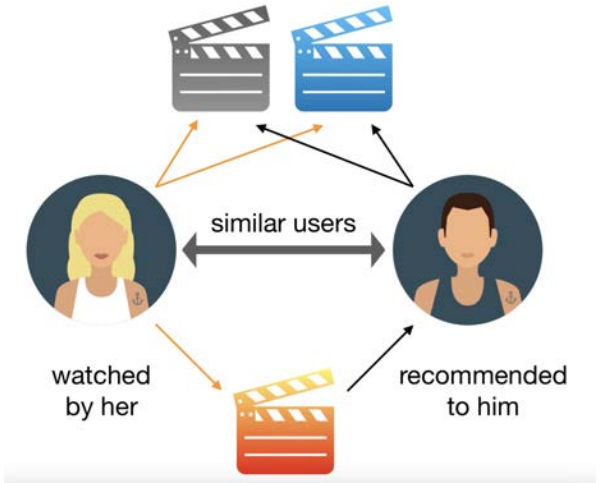
- The definition of DA
- The importance of DA
- The three research directions of DA
- Some research papers of DA
- Discussion
- Conclusion

KDD2021

Machine learning

- Machine learning becomes a very useful technique in many real applications.

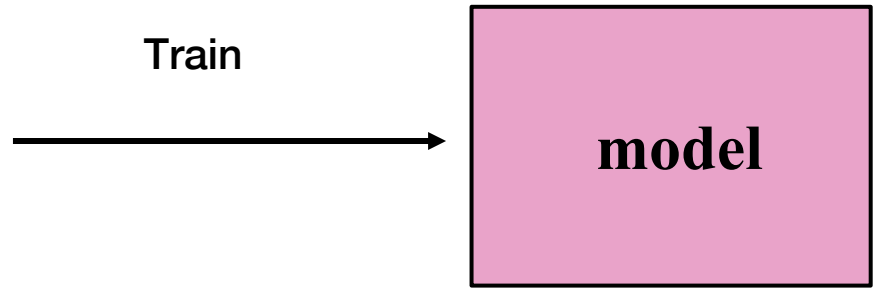


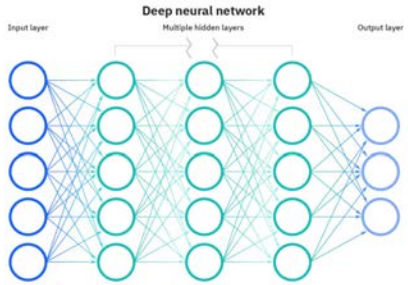


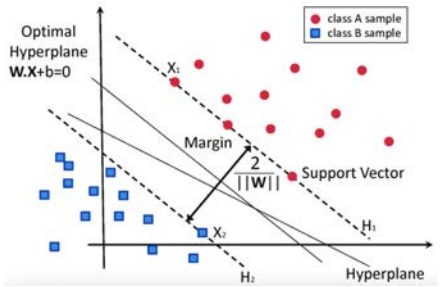
KDD2021

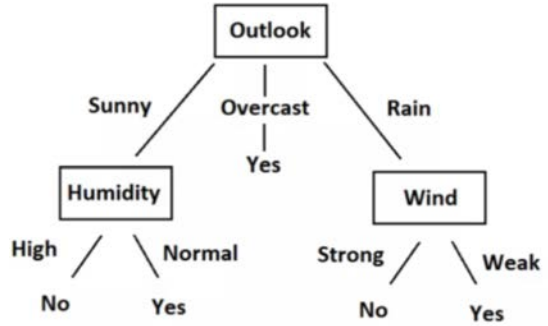
Machine learning model

- A training dataset $D_{train.}$
- The machine learning model is trained on $D_{train.}$



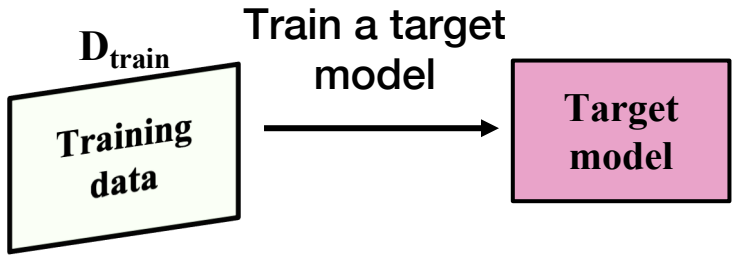




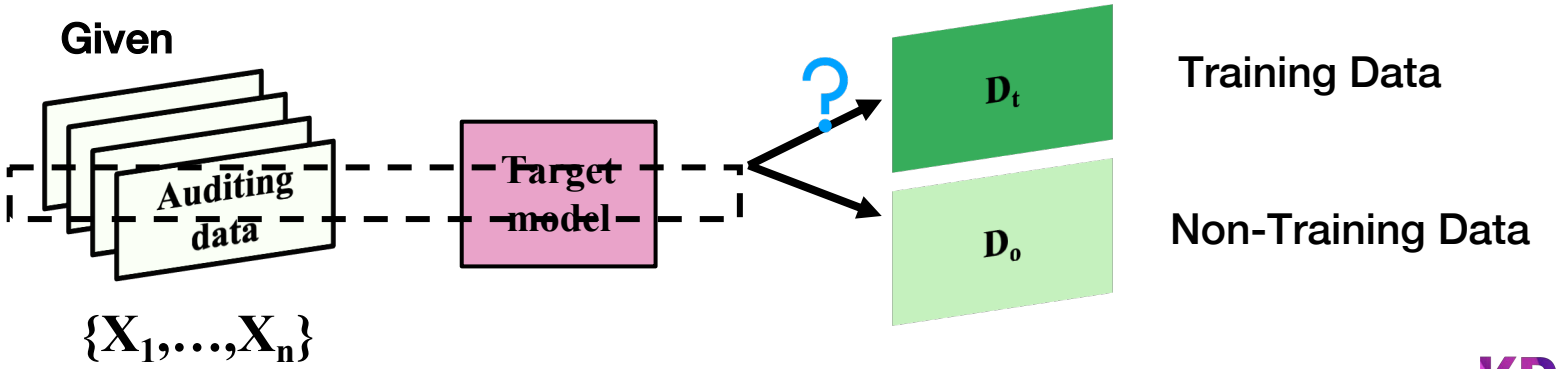


The definition of DA

- A training dataset D_{train}.
- The target model is trained on D_{train}.



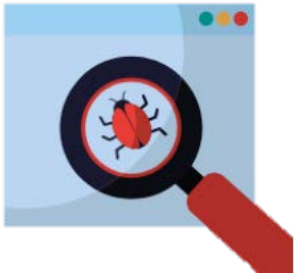
- **The Problem**: Given a data point *x*, an auditor is to determine if the data *x* was in the target model's training dataset.



The importance of DA

For data privacy

Data auditing enables users and businesses to safeguard their data.



For incentive governance

Data auditing offers a trustworthy basis for fair incentive allocation.

The common assumption of DA

Auditor		Target model
Model knowledge	Access to the model	White-box
	Obtain the model' s prediction results	Black-box
Data knowledge	Know the format of data	Full outputs results or Partial outputs results
Machine learning background		Limited number of query

Three research directions

Audit-training techniques

Audit the target model‘s training process.

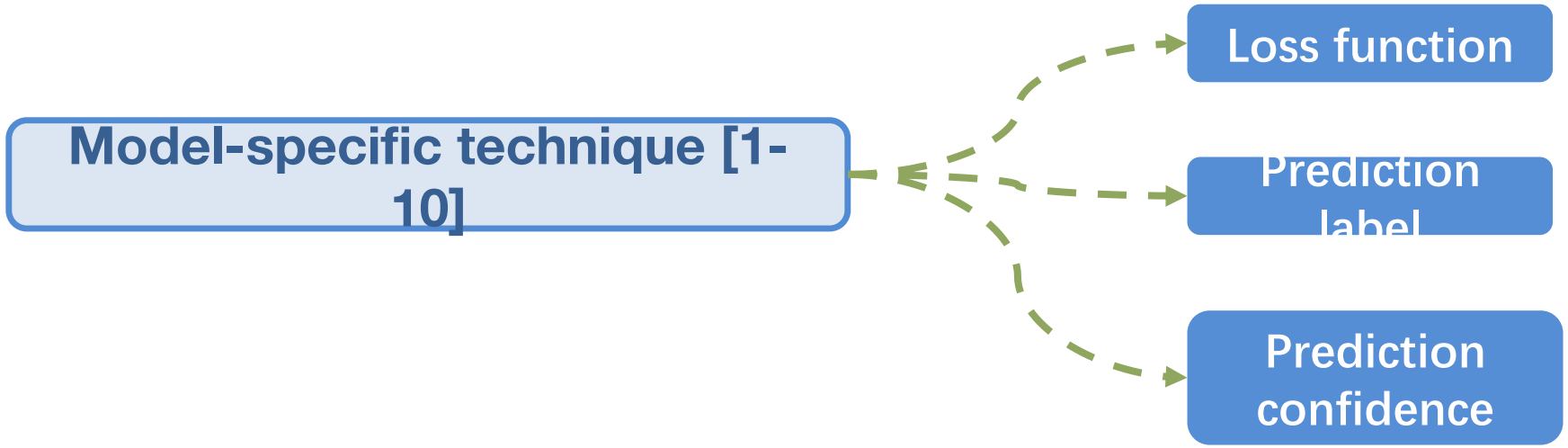
Model-specific technique

Design a criterion on the model output to compare training data and non-training data, e.g., the prediction loss or the prediction confidence.

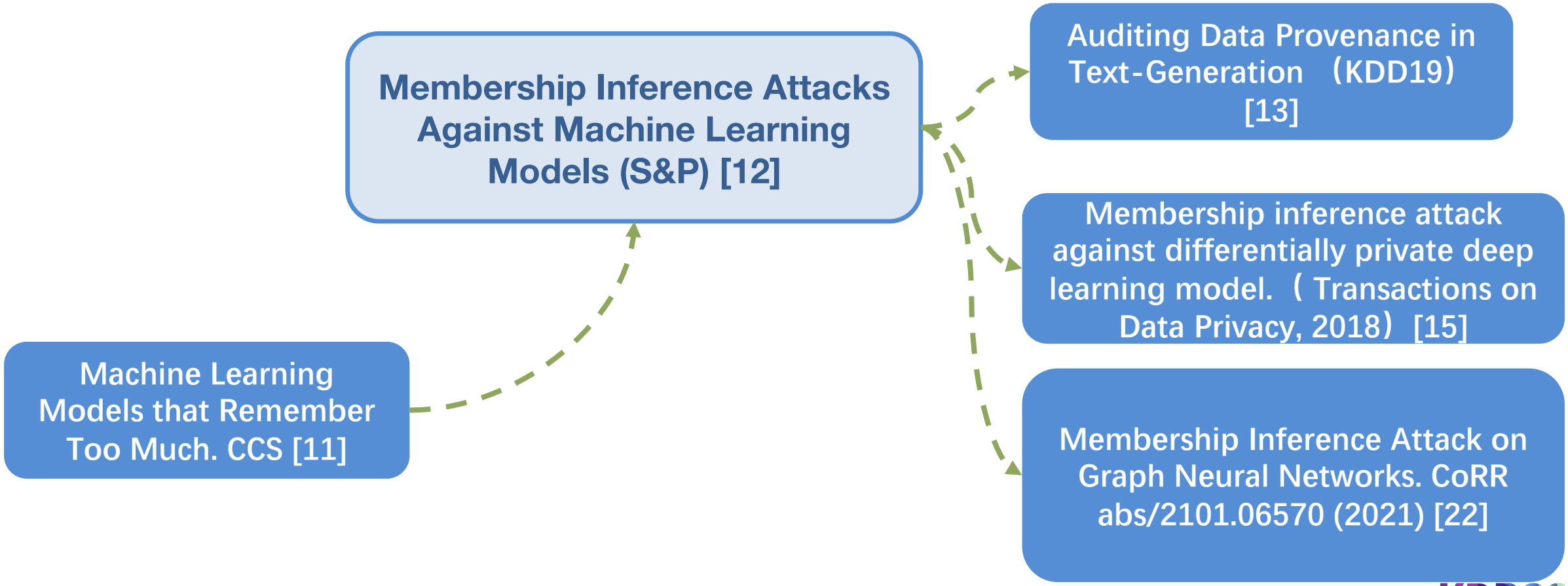
Shadow-training technique

Use multiple ``shadow models'' to imitate the behavior of the target model, and train an attack model to classify training and non-training.

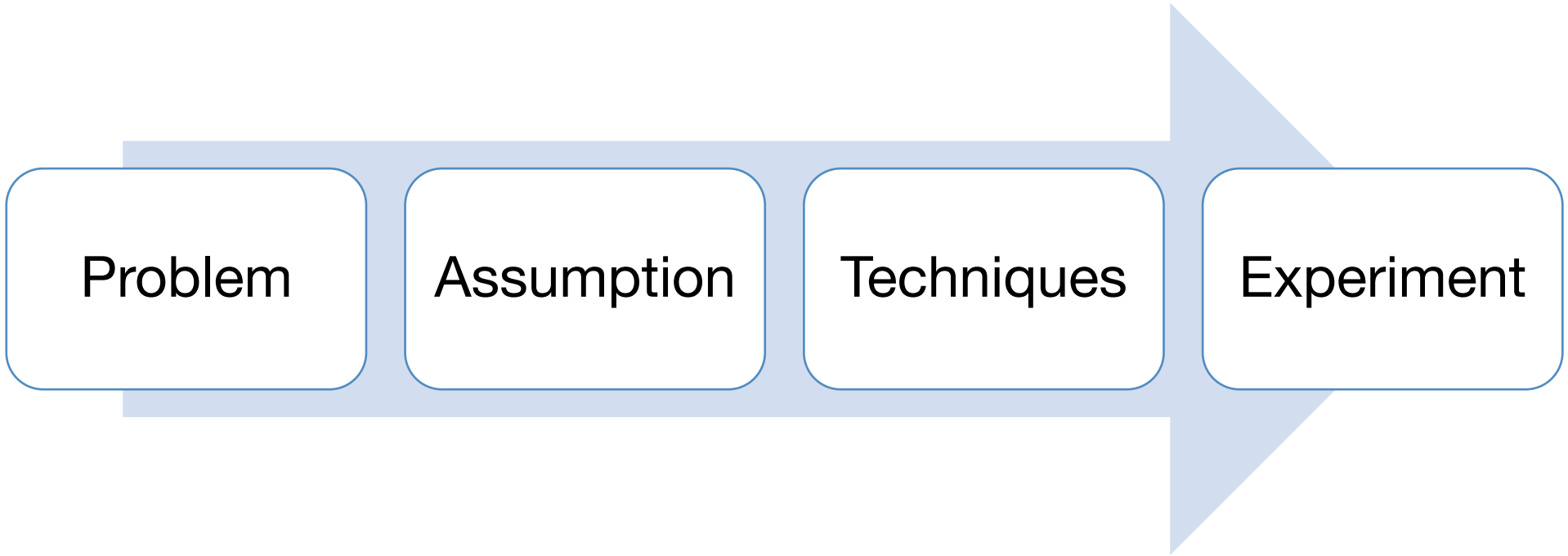
The map of model-specific technique



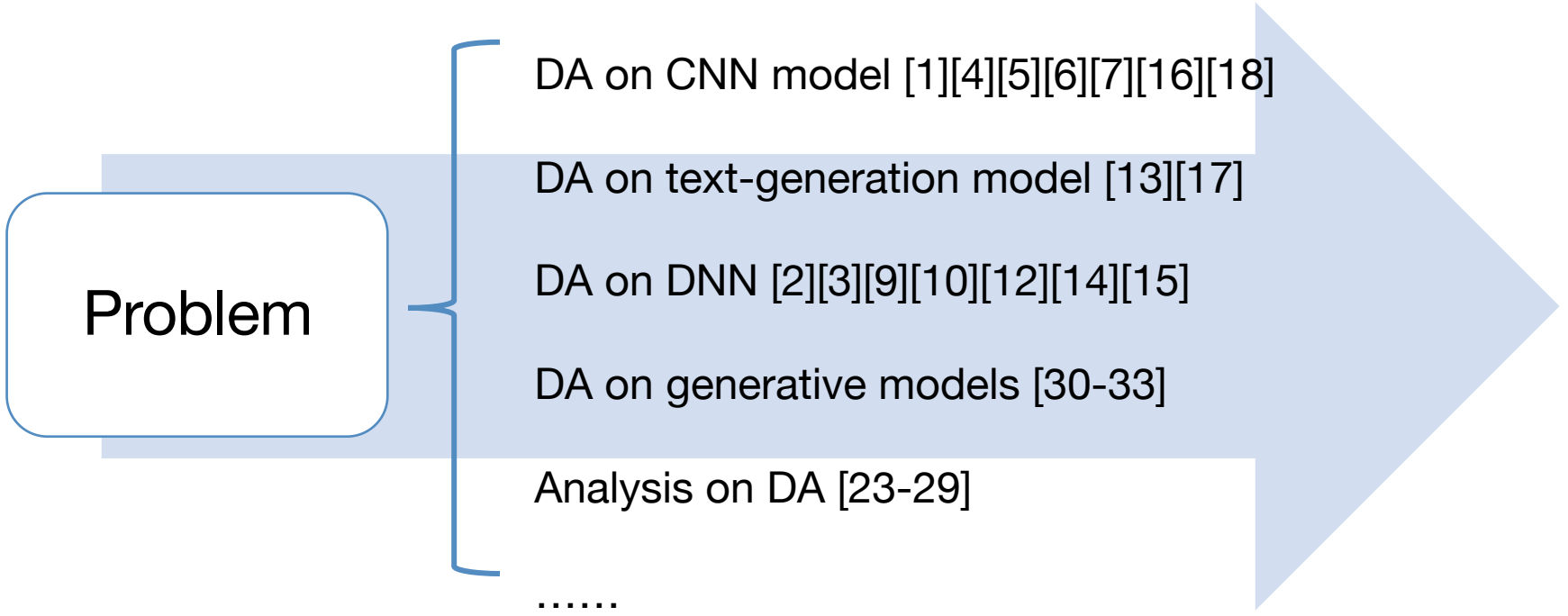
The map of shadow-training technique [11-23]



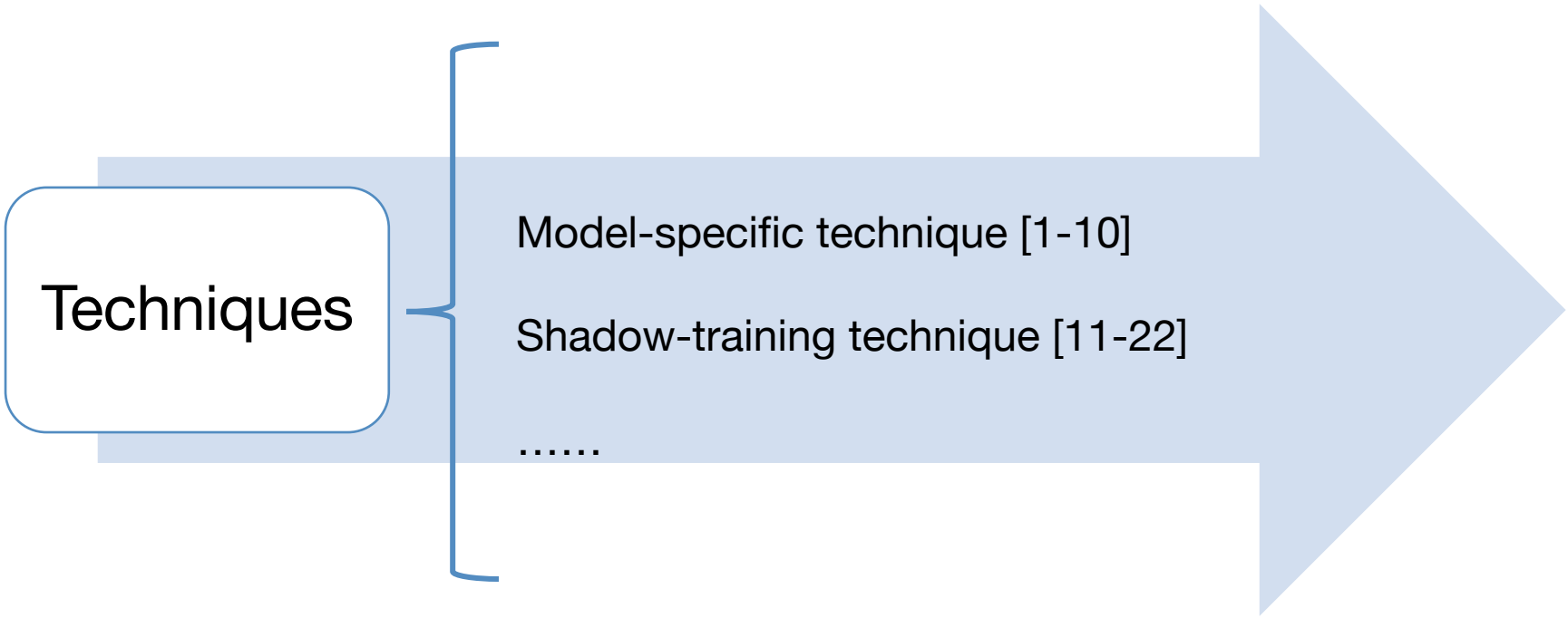
Some state-of-the-art papers



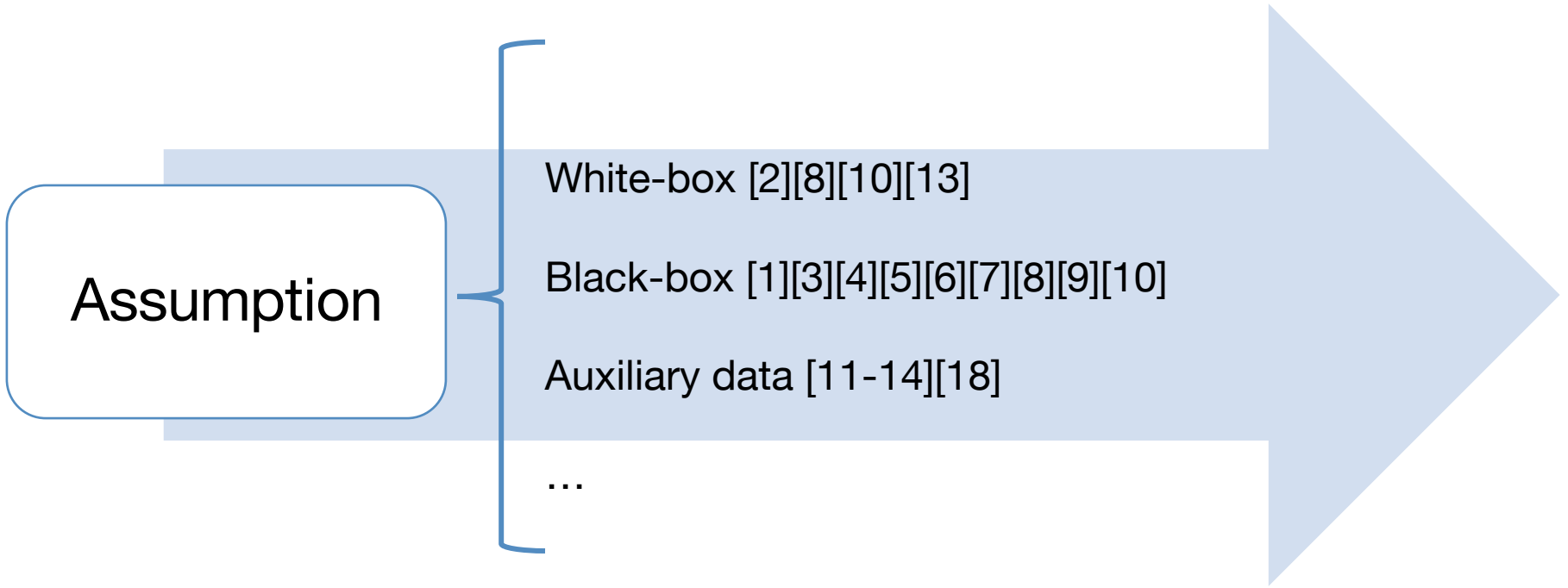
Some state-of-the-art papers



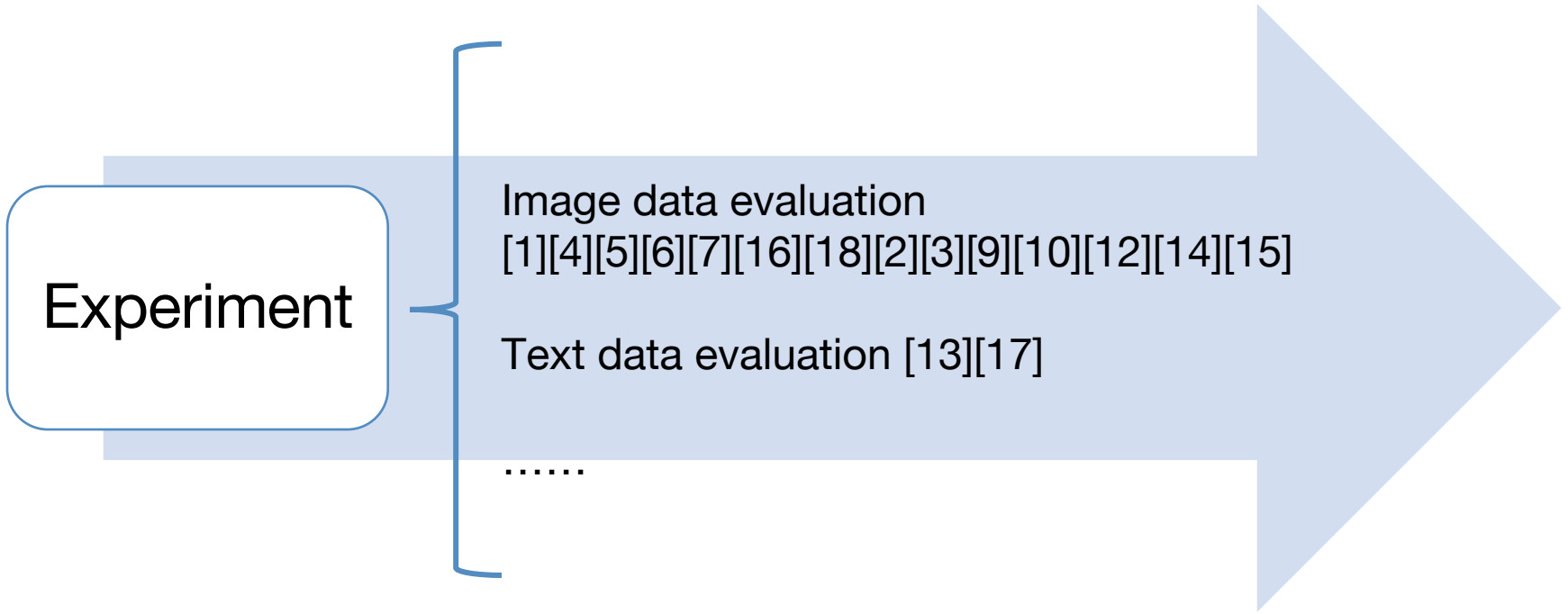
Some state-of-the-art papers



Some state-of-the-art papers



Some state-of-the-art papers



Shokri et al., "Membership Inference Attacks Against Machine Learning Models", S&P'17

Problem Setting

Problem

Given a machine learning model and a record, determine whether this record was used as part of the model's training dataset or not.

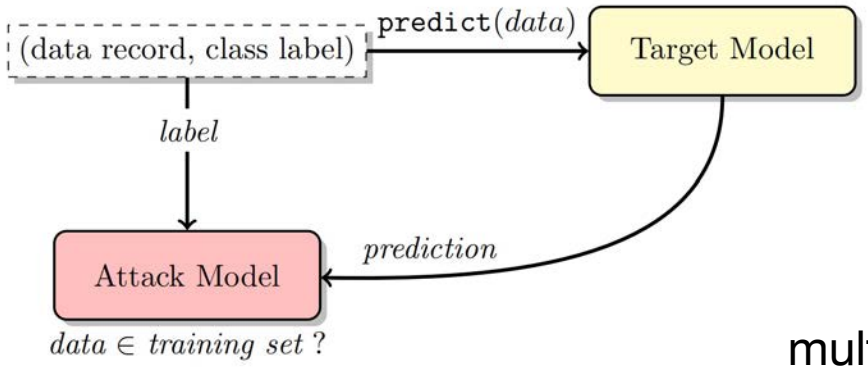
Assumption

Output results: (1) access to the model and can obtain the model's prediction vector on any data record. (2) know the format of the inputs and outputs of the model.

Background knowledge: background knowledge about the population from which the target model's training dataset was drawn.

Shokri et al., "Membership Inference Attacks Against Machine Learning Models", S&P'17

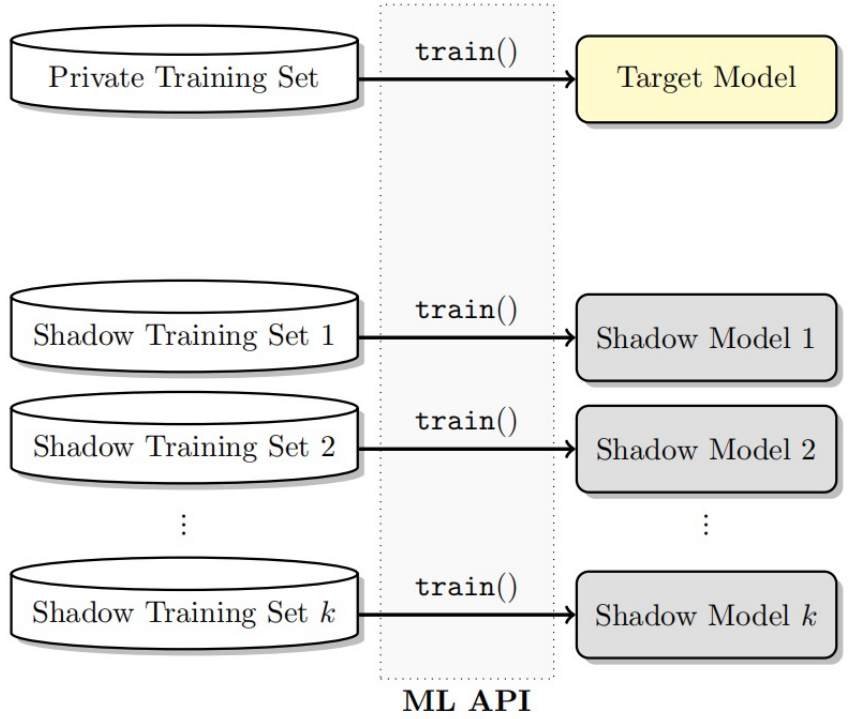
Technology



The main challenge is how to train the attack model to distinguish members from non-members of the target model's training dataset

Shadow models

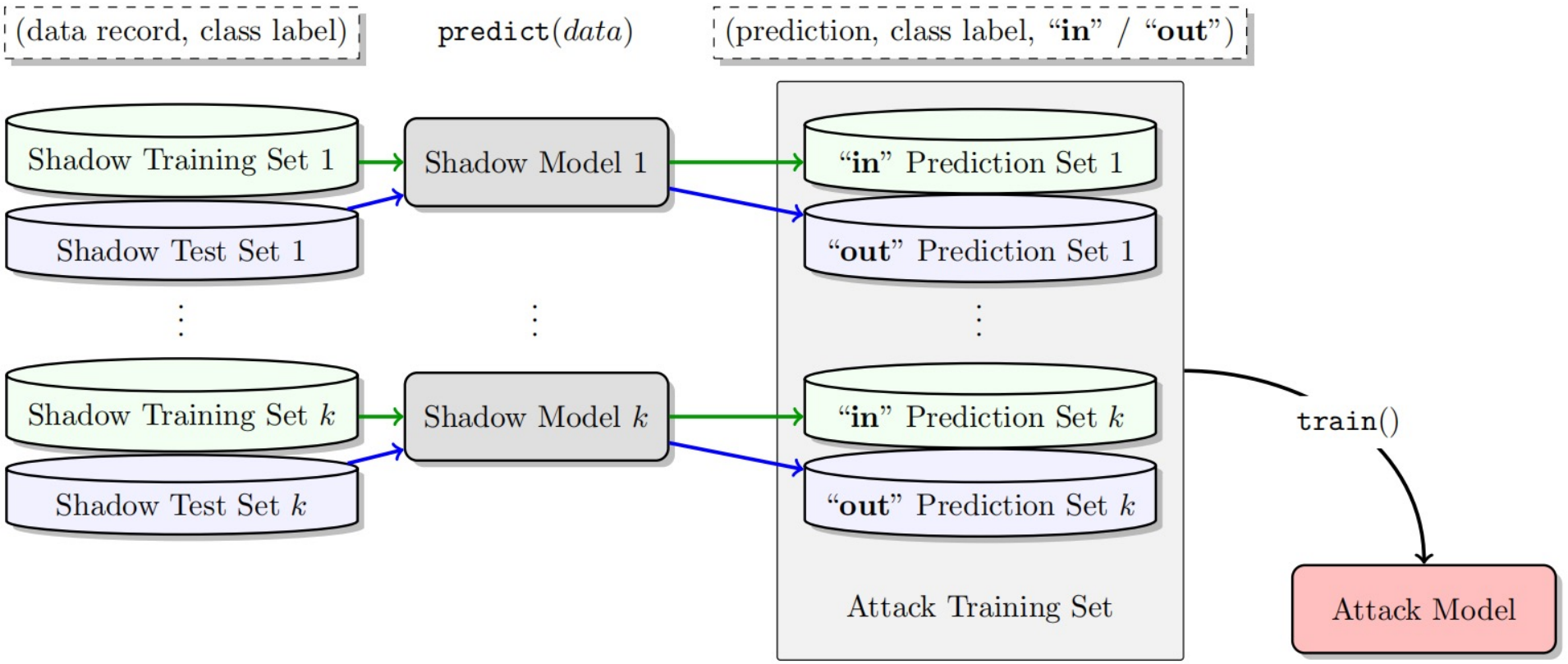
multiple "shadow" models intend to behave similarly to the target model.



Shokri et al., "Membership Inference Attacks Against Machine Learning Models", S&P'17

Technology

Training the attack model on the inputs and outputs of the shadow models.



Shokri et al., "Membership Inference Attacks Against Machine Learning Models", S&P'17

Experiment

Metric

Precision: what fraction of records inferred as members are indeed members of the training dataset.

Recall: what fraction of the training dataset's members are correctly inferred as members by the attacker

Data

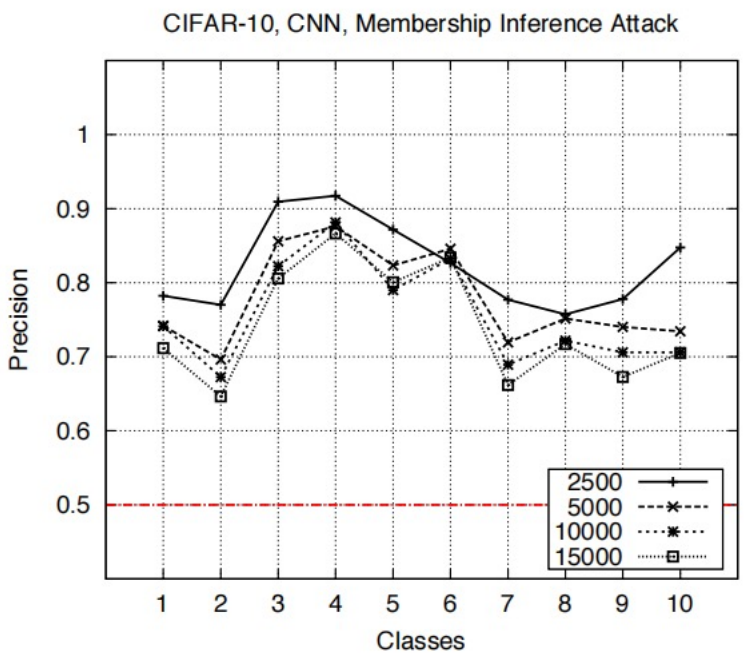
CIFAR, MNIST, Purchases, Locations, Texas hospital stays, UCI Adult

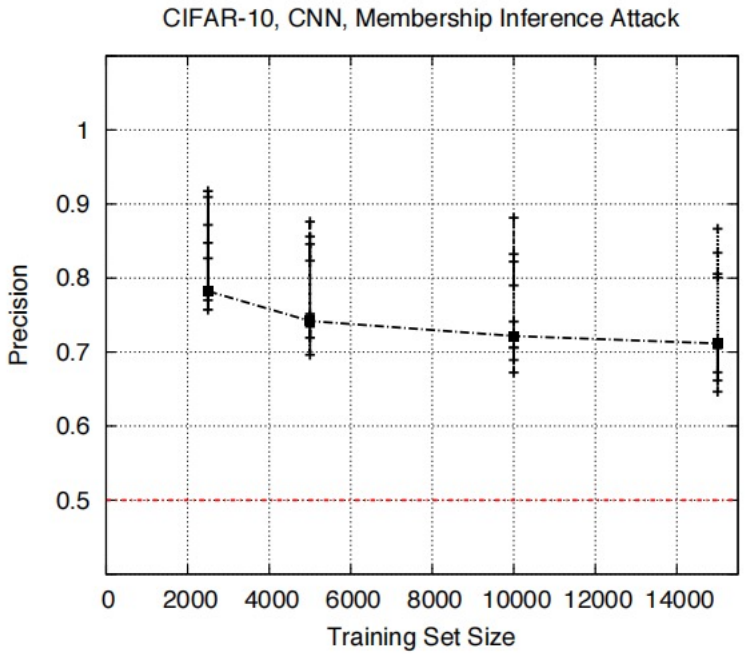
Target model

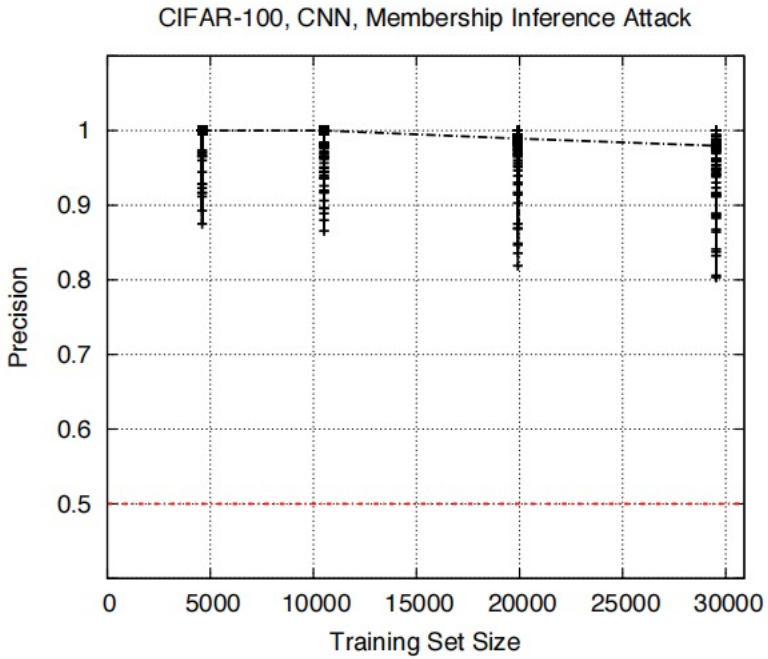
NN (neural network with one hidden layer of size 64 with ReLU activation functions and a SoftMax layer.

Shokri et al., "Membership Inference Attacks Against Machine Learning Models", S&P'17

Experiment







Precision of the membership inference attack against neural networks trained on CIFAR datasets

Song et al., "Auditing Data Provenance in Text-Generation Models", KDD'19

Problem Setting

Problem

Audit ML models to determine if their data was used to train these models. We focus specifically on auditing that generate natural-language text.

Assumption

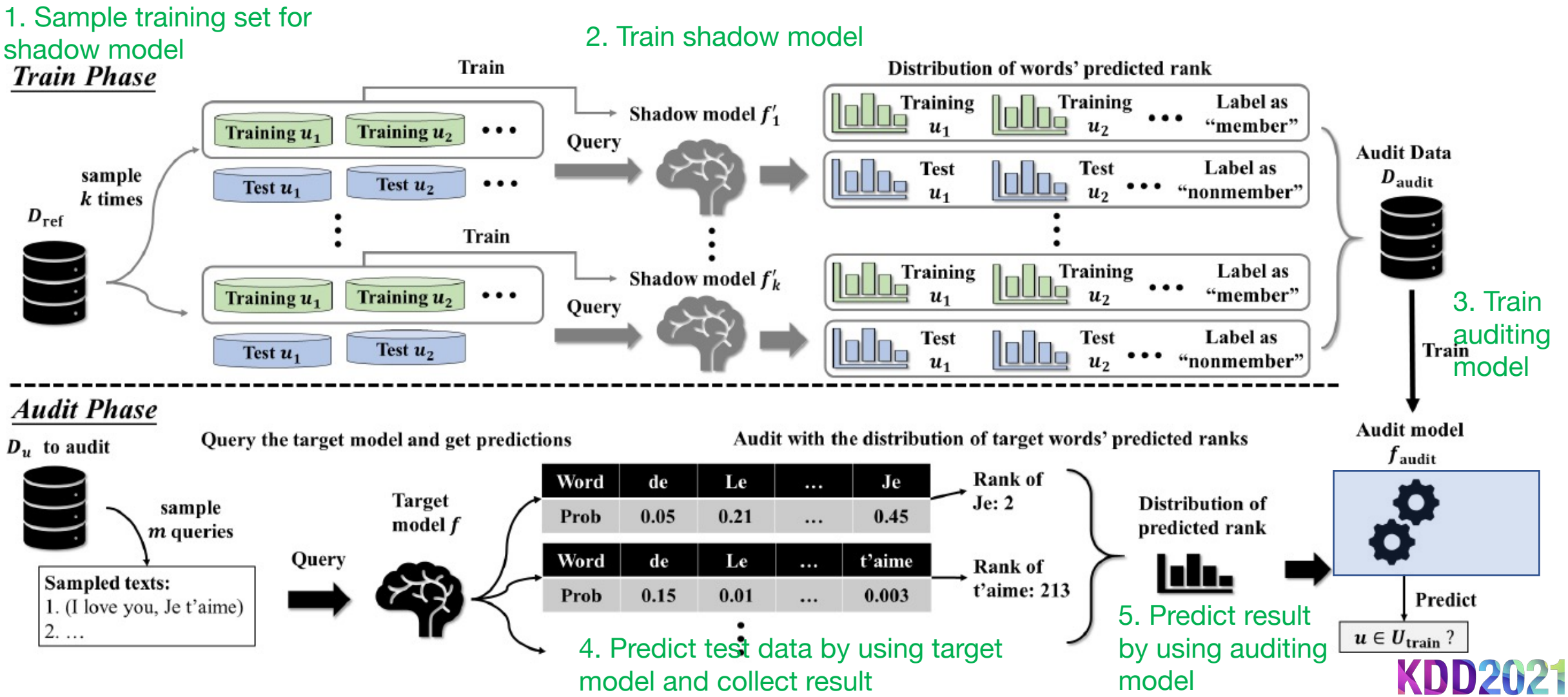
(1) Auditor knows the learning algorithm used to create model but he may or may not know the training hyper-parameters.

(2) The auditor also needs an auxiliary dataset to train shadow models that perform the same task as target model.

(3) We assume that the tokens in the model's output space are ranked.

Song et al., "Auditing Data Provenance in Text-Generation Models", KDD'19

Technology



KDD2021

Song et al., "Auditing Data Provenance in Text-Generation Models", KDD'19

Experiment

Metric

Precision: what fraction of records inferred as members are indeed members of the training dataset.

Recall: what fraction of the training dataset's members are correctly inferred as members by the attacker

Accuracy: the percentage of all users who are classified correctly

AUC: the area under the ROC curve that shows the gap between the scores

Data

Reddit, SATED, Dialogs, Locations, Texas hospital stays, UCI Adult

Target model

both the target and shadow models are one-layer LSTMs or GRUs.

Song et al., "Auditing Data Provenance in Text-Generation Models", KDD'19

Experiment

Effect of different hyper-parameters.

Effect of the number of users.

Effect of the number and selection of audit queries.

Effect of the size of the model's output.

Effect of noise and errors in the queries.

Effect of training shadow models with different hyper-parameters.

Dataset	Accuracy	AUC	Precision	Recall
Reddit	0.990	0.993	0.983	0.996
SATED	0.965	0.981	0.937	0.996
Dialogs	0.978	0.998	0.958	1.000

M. A. Rahman et al. Membership inference attack against differentially private deep learning model. Transactions on Data Privacy, 11(1):61–79, 2018.

Problem Setting

Problem

The membership inference attack against a state-of-the-art differentially private deep model

Assumption

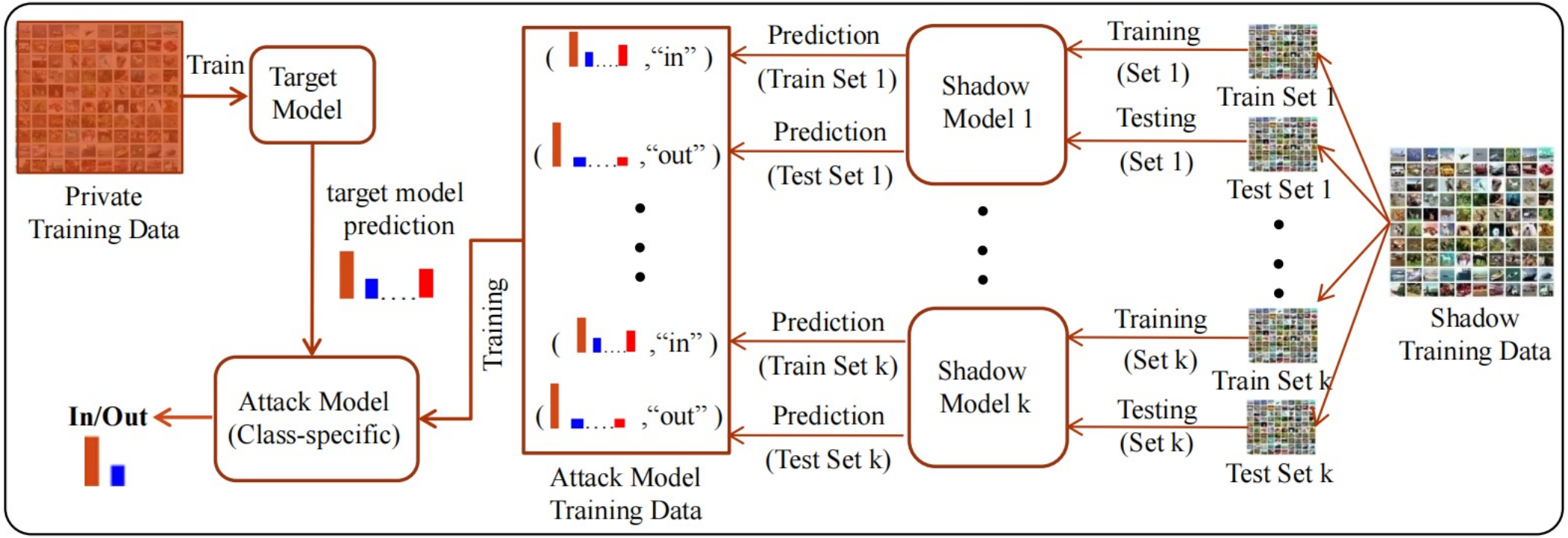
(1) White-box deep models

(2) The target model is a classification model

(3) Some background knowledge about the population the target models' training dataset

M. A. Rahman et al. Membership inference attack against differentially private deep learning model. Transactions on Data Privacy, 11(1):61–79, 2018.

Technology



Overview of the membership inference attack.

M. A. Rahman et al. Membership inference attack against differentially private deep learning model. Transactions on Data Privacy, 11(1):61–79, 2018.

Experiment

Metric

Accuracy: accuracy measure simply reports the percentage of examples that are correctly predicted to be members of the target model's training dataset.

F1-score: combines the precision and recall measures into a single value.

Data

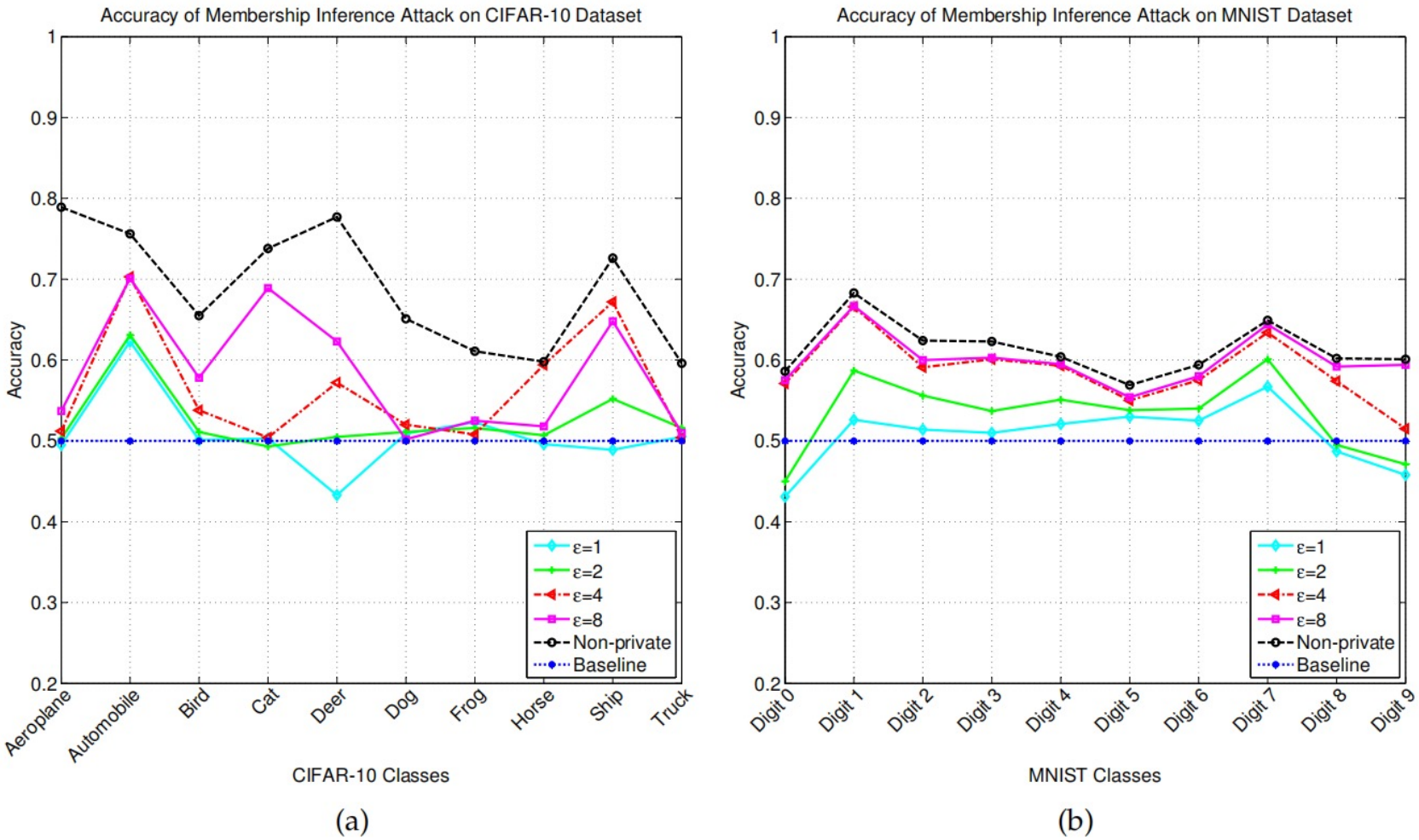
CIFAR-10, MNIST

Target model

CNN for CIFAR, NN For MNIST

M. A. Rahman et al. Membership inference attack against differentially private deep learning model. Transactions on Data Privacy, 11(1):61–79, 2018.

Experiment



J. Hayes et al. LOGAN: Membership inference attacks against **generative models**. In PETS, 2019.

Problem Setting

Problem

Membership Inference Attacks Against **Generative Models**

Assumption

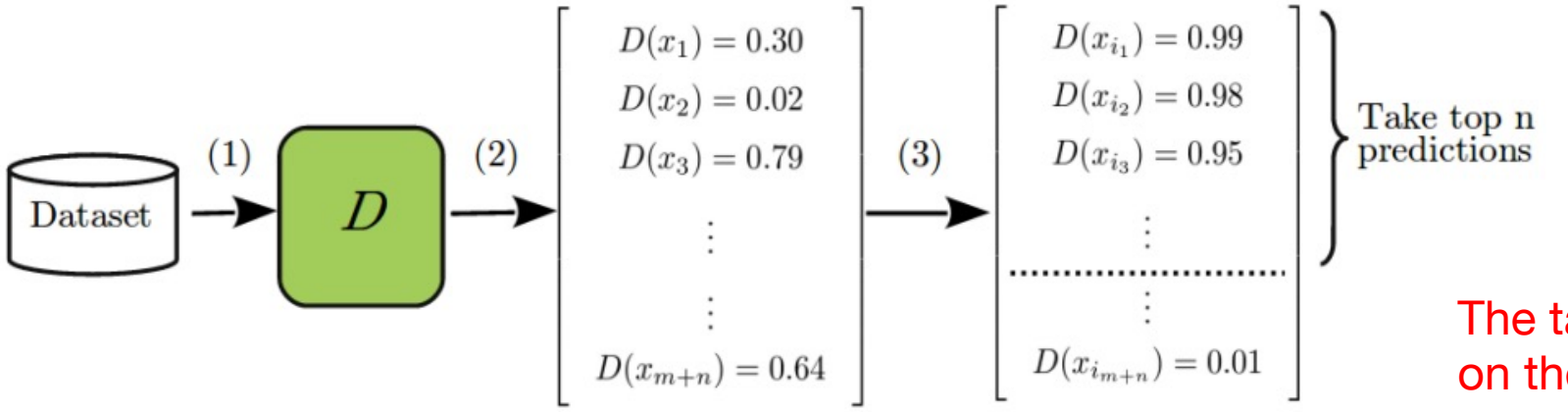
(1) The size of the training set.

(2) In white-box, the adversary only needs access to the discriminator of a target GAN model.

(3) In black-box, we assume the attacker does not have prior or side information about training records or the target model.

J. Hayes et al. LOGAN: Membership inference attacks against generative models. In PETS, 2019.

White-Box Prediction Method:

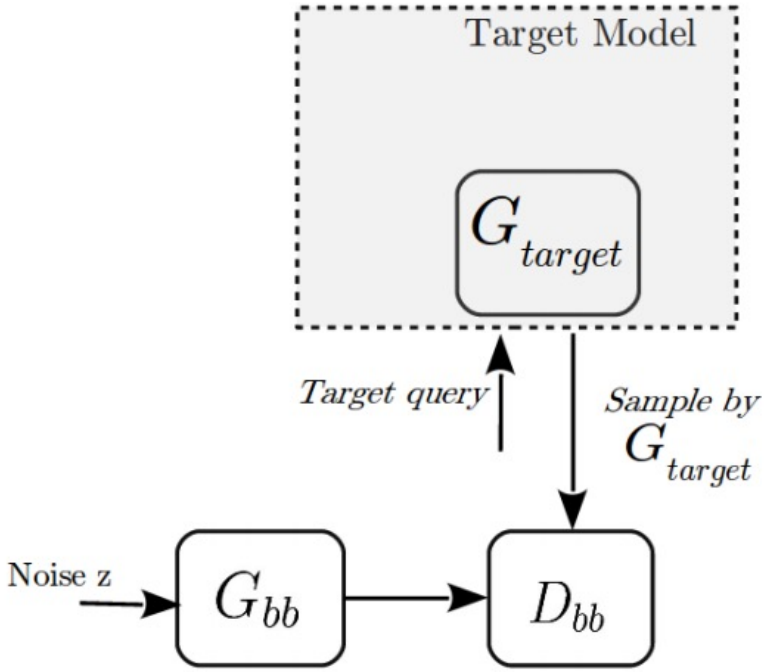


The target model overfitted on the training data

- inputs data-points to the Discriminator *D* (1),
- extracts the output probabilities (2),
- and sorts them (3).

J. Hayes et al. LOGAN: Membership inference attacks against generative models. In PETS, 2019.

Black-Box Assumption



Idea: trains a new GAN in order to imitate the target model

J. Hayes et al. LOGAN: Membership inference attacks against generative models. In PETS, 2019.

Experiment

Metric

Accuracy: accuracy measure simply reports the percentage of examples that are correctly predicted to be members of the target model's training dataset.

Data

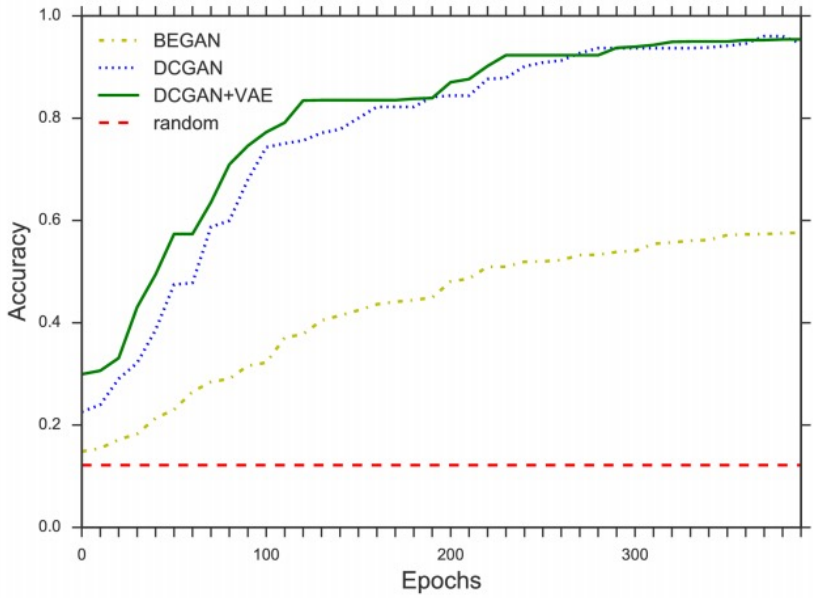
CIFAR-10, Labeled Faces in the Wild (LFW) , Diabetic Retinopathy (DR)

Target model

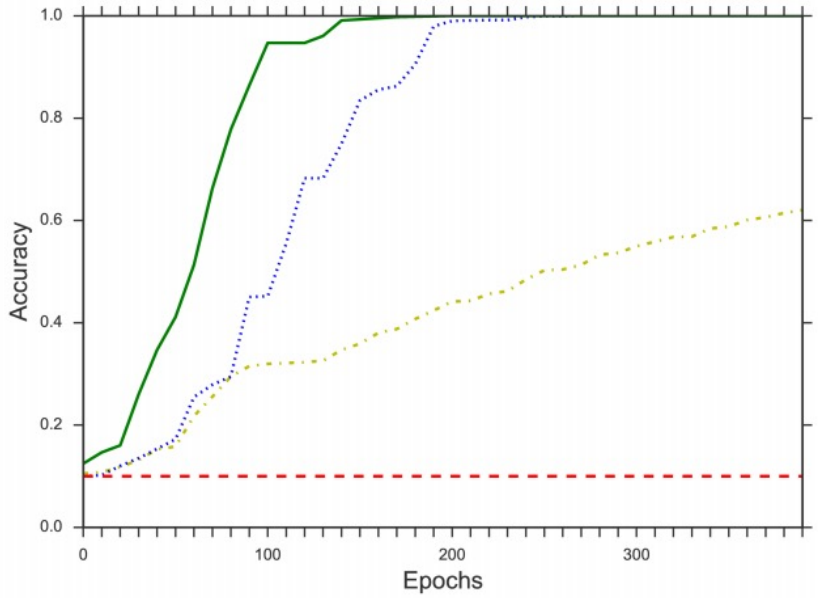
1. DCGAN
2. DCGAN+VAE
3. BEGAN

M. A. Rahman et al. Membership inference attack against differentially private deep learning model. Transactions on Data Privacy, 11(1):61–79, 2018.

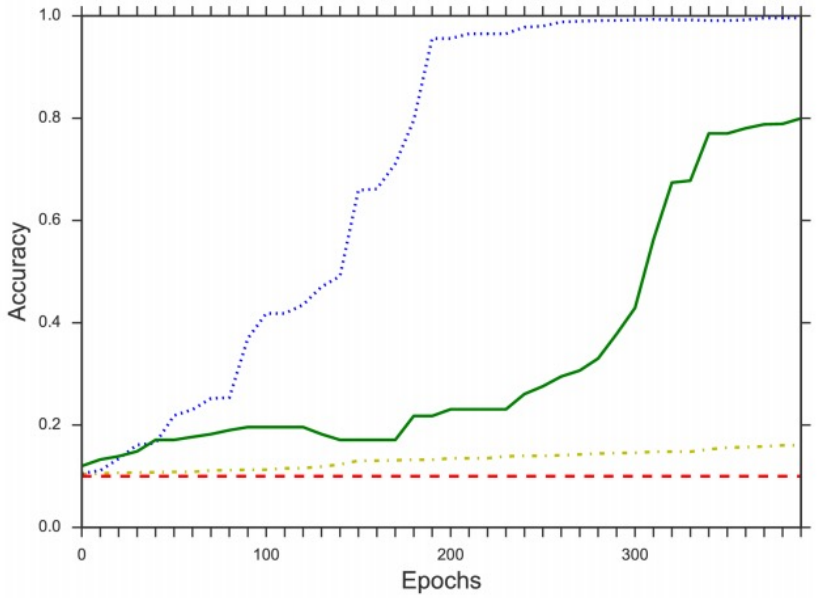
Experiment



(a) LFW, top ten classes



(b) LFW, random 10% subset

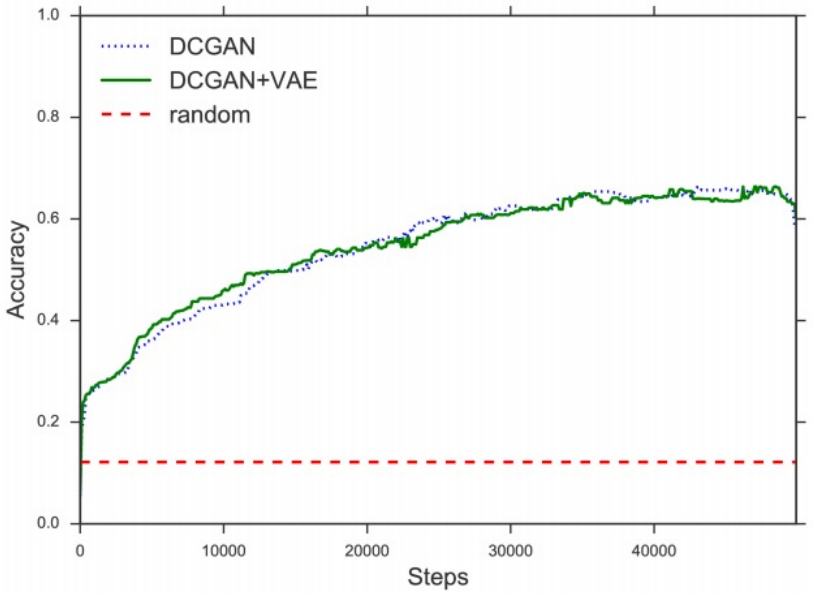


(c) CIFAR-10, random 10% subset

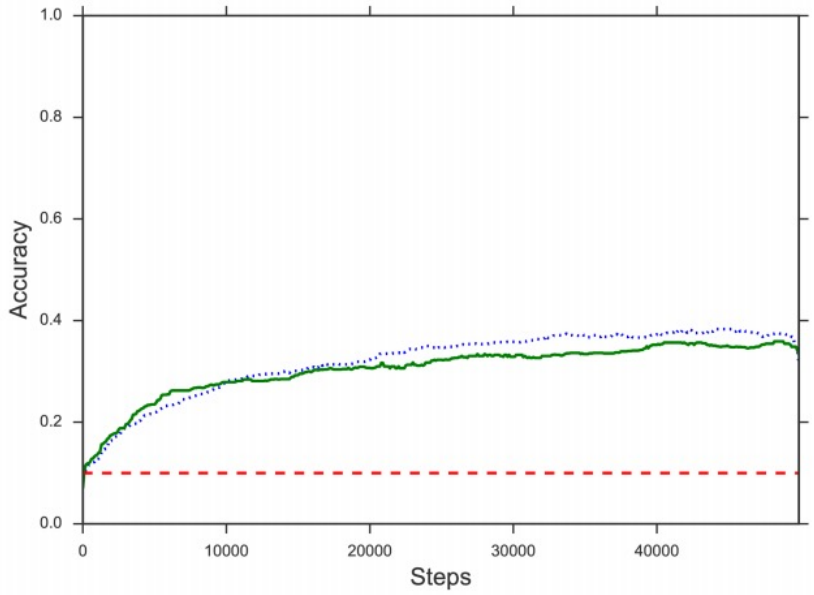
Accuracy of white-box attack with different datasets and training sets.

M. A. Rahman et al. Membership inference attack against differentially private deep learning model. Transactions on Data Privacy, 11(1):61–79, 2018.

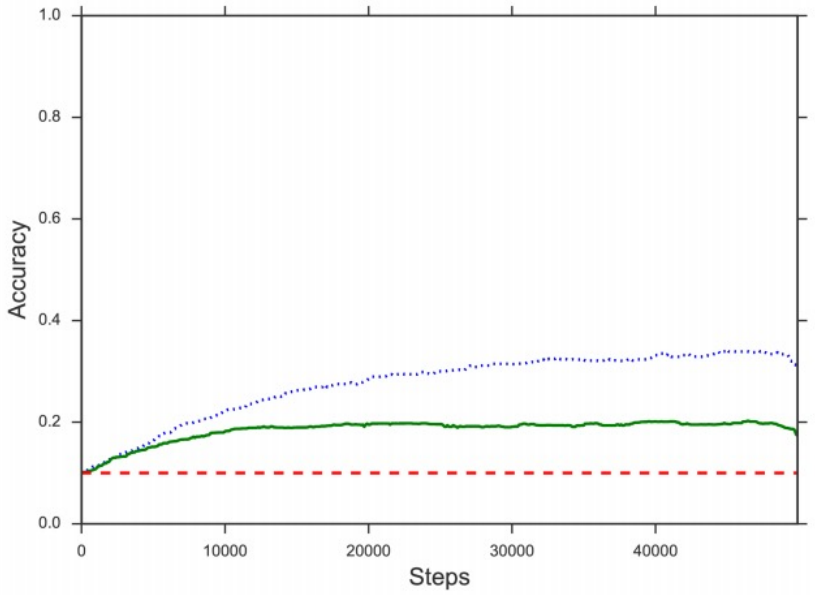
Experiment



(a) LFW, top ten classes



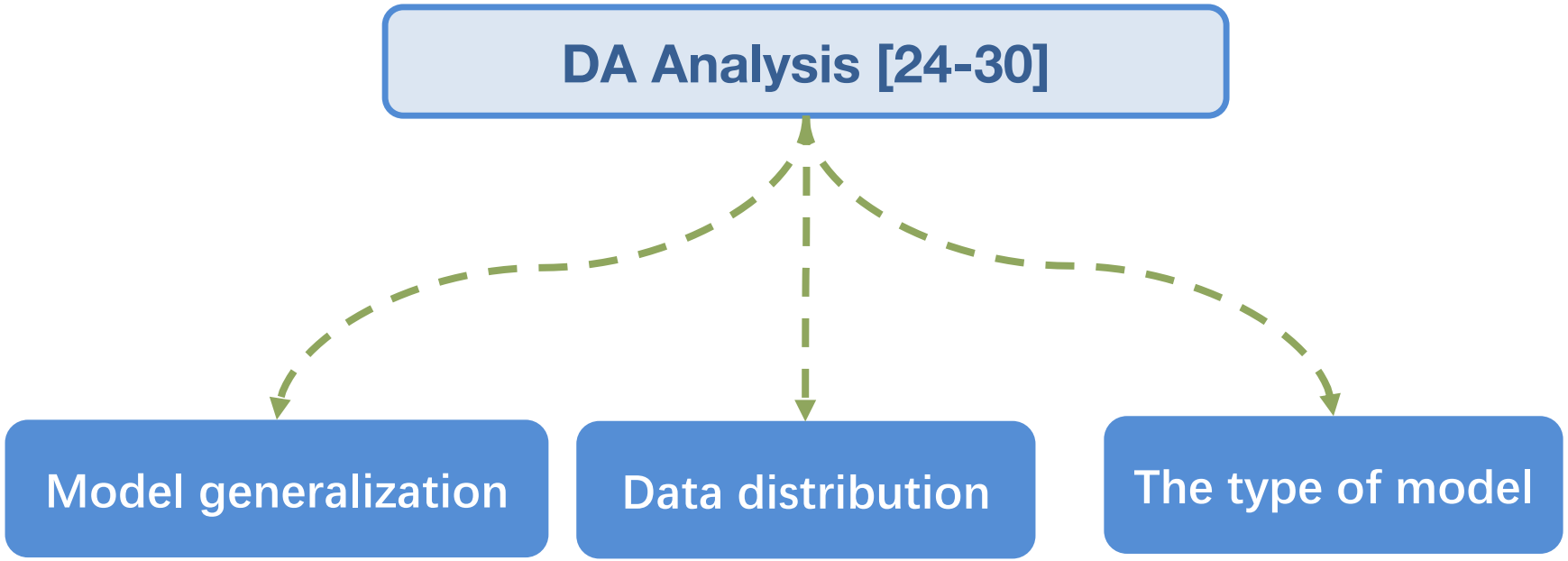
(b) LFW, random 10% subset



(c) CIFAR-10, random 10% subset

Accuracy of black-box attack on different datasets and training sets.

The map of DA Analysis



Y. Long et al., Understanding membership inferences on well-generalized learning models. arXiv:1802.04889, 2018.

Problem Setting

(1) A study that discovers overfitting to be a sufficient but not a necessary condition for an DA to succeed.

(2) The unique influence of a target record is the key for a successful DA.

Related Areas

- Understanding machine model (From model owner)[30][34]
  - Understand the effect of training points on a model's predictions.
- Machine unlearning (From model owner)[31][32]
  - Given a trained model, unlearning assures the user that the model is no longer trained using the data which the user elected to erase.
- Data Valuation[33]
  - how to fairly allocate the reward generated by a ML model to the data contributors
- Defense DA [35][36][38]

Conclusion

- Auditing Data is an important problem in data provenance.
- From a taxonomy point of view, there are three directions for this problem.
- Some important works are reported.
- Related areas

Tutorial Outline

- Data Asset: What and Why
- Data Asset Core Components
- Data Asset Governance for Decentralized Collaborative Intelligence
  - Governance principles
  - "Trust" for data asset governance for decentralized collaborative intelligence
    - Agreement
    - Accounting
    - Auditing
    - Privacy
  - "Incentive" for data asset governance for decentralized collaborative intelligence
- Data Asset Ecosystems
- Challenges and Future Directions

Data Asset for Collaborative Intelligence

"Trust" for data asset governance - Privacy

Huiwen Liu[1], Feida Zhu[2]

1. Singapore Management University, hwliu.2018@phdcs.smu.edu.sg
2. Singapore Management University, fdzhu@smu.edu.sg,

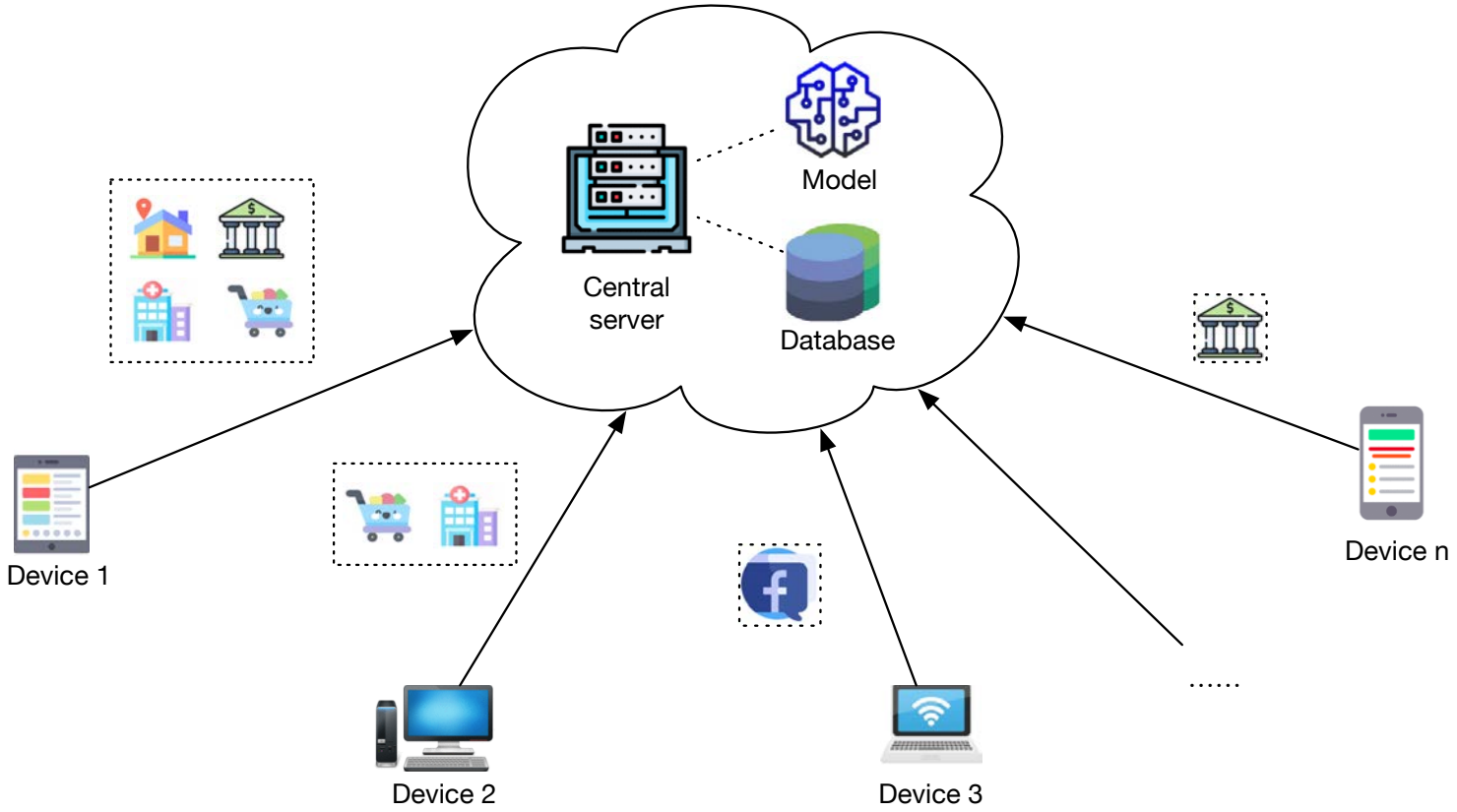
Outline

- Motivation & Background
- Privacy issue in federated learning
- Threat models
- Defensive techniques

Motivation & Background

Motivation: separation of computing and data

- Privacy issue in traditional collaborative machine learning (ML)

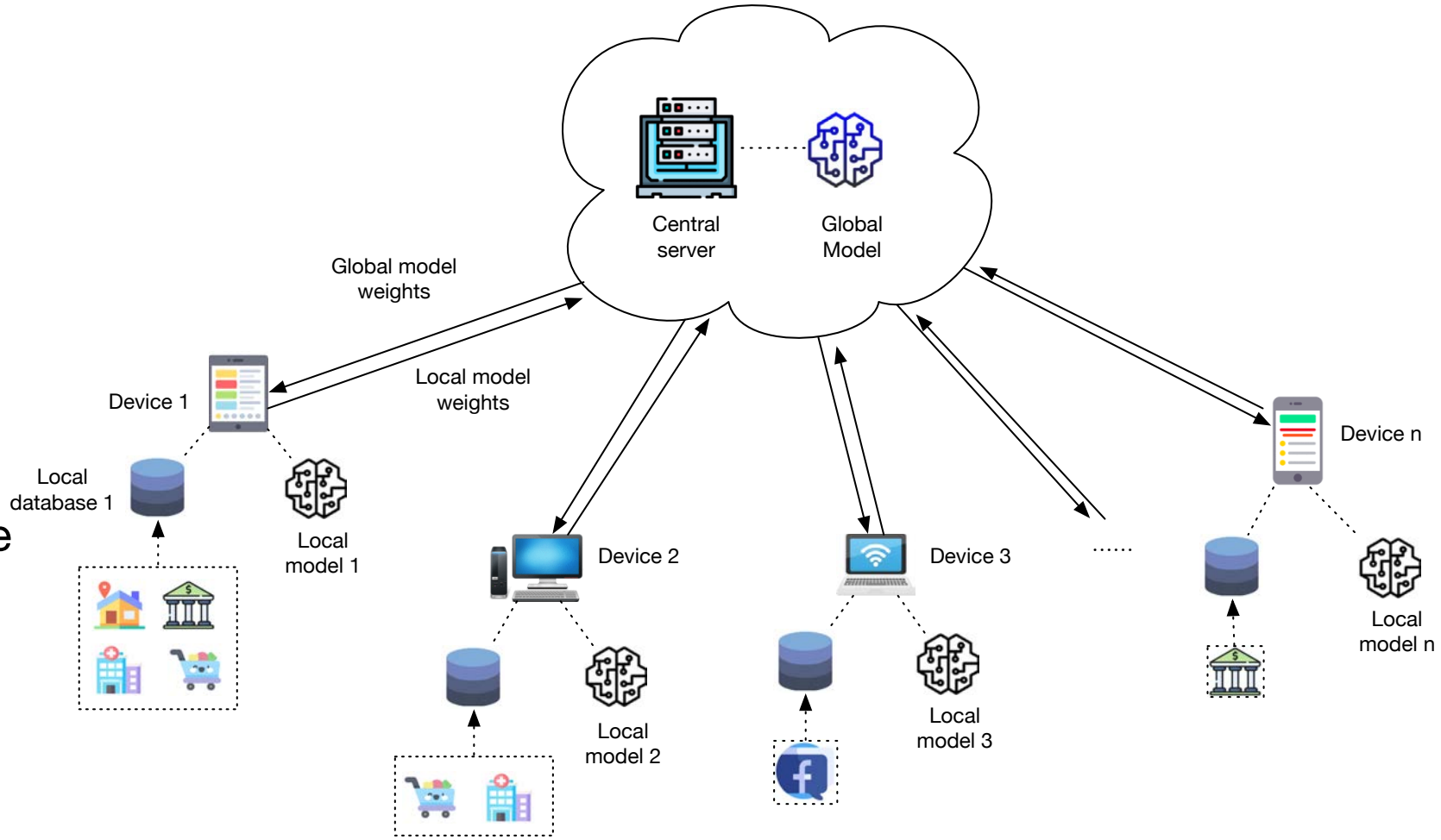


Reveal sensitive information
-> address information
-> health information
-> shopping hobbies
.....

Solution:
Federated Learning (FL)
-> function of traditional ML
-> privacy protection:
separation of computation and data

Motivation: separation of computation and data

- Federated learning (FL)
  - Store data locally and push network computation to the edge devices
  - Central server aggregates the local model weights and generates the global model weights with some specific protocols



Background: definition of FL

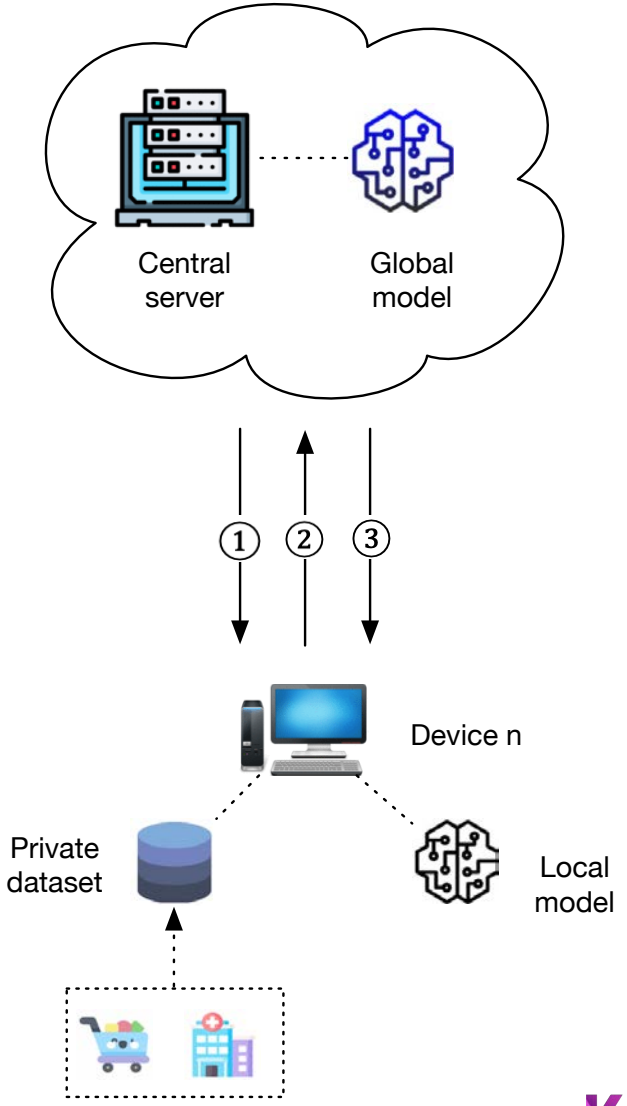
- Learn a single, global statistical model from data stored locally on multiple remote devices
- Goal:

$$\min_{w} F(w), \text{ where } F(w) := \sum_{k=1}^{m} p_k F_k(w)$$

  - m is the total number of devices;
  - F_k is the local objective function for the kth device;
  - p_k specifies the relative impact of each device with $p_k \geq 0$ and $\sum_{k=1}^{m} p_k = 1$.

Background: process flow of FL

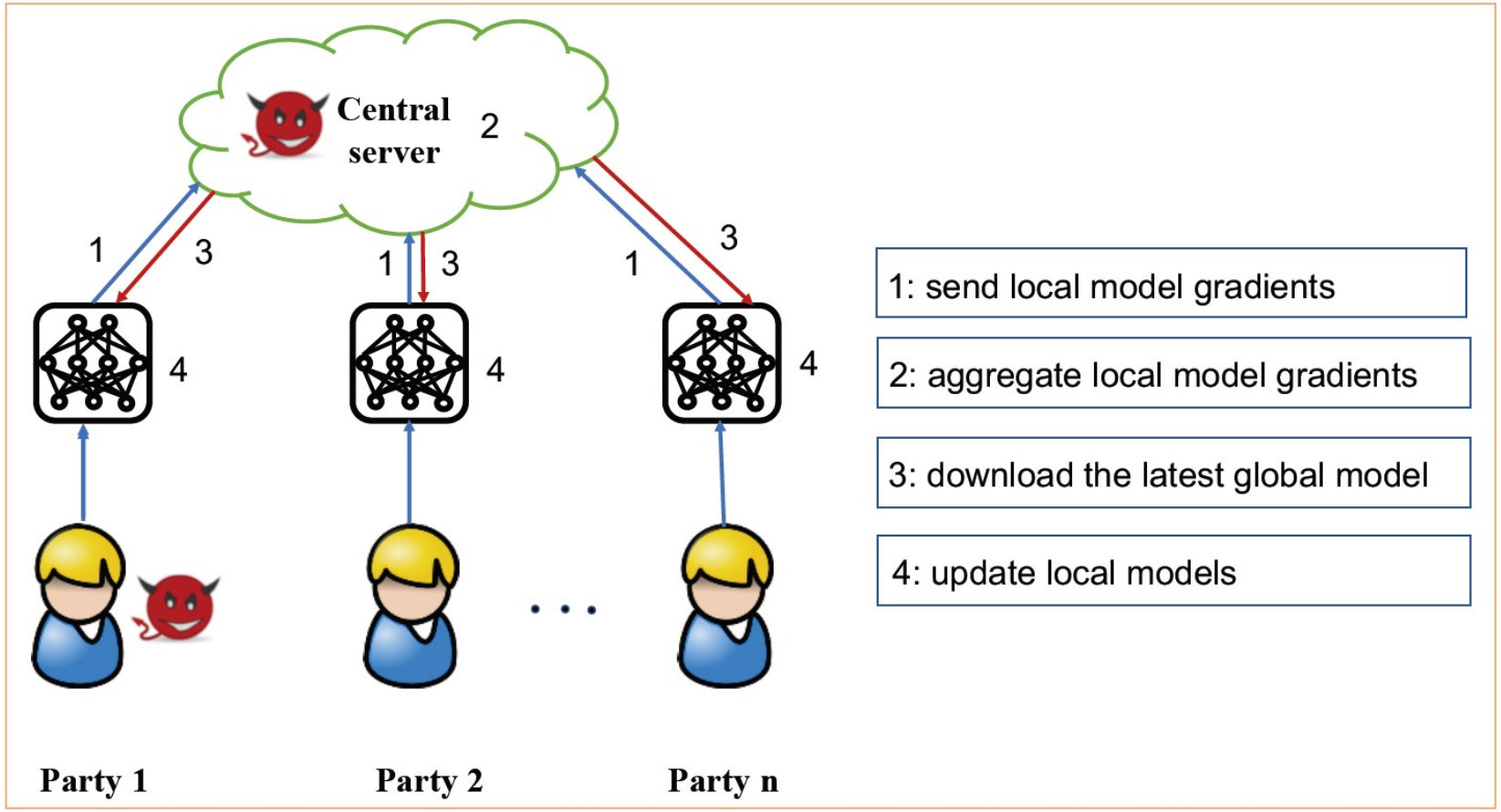
- General work flow of FL
  - ① Model selection: central server shares initial model parameters with all the edge devices
  - ② Local model training: edge devices train local model with initial parameters and share local model parameters with central server
  - ③ Aggregation of local models: central server aggregates the local model parameters and shares result global model parameter with edge devices



Privacy issue in federated learning

Privacy issue in federated learning

- Both malicious FL server and participants may compromise the FL system



[1] Lyu L, Yu H, Yang Q. Threats to federated learning: A survey[J]. arXiv preprint arXiv:2003.02133, 2020.

Threat models

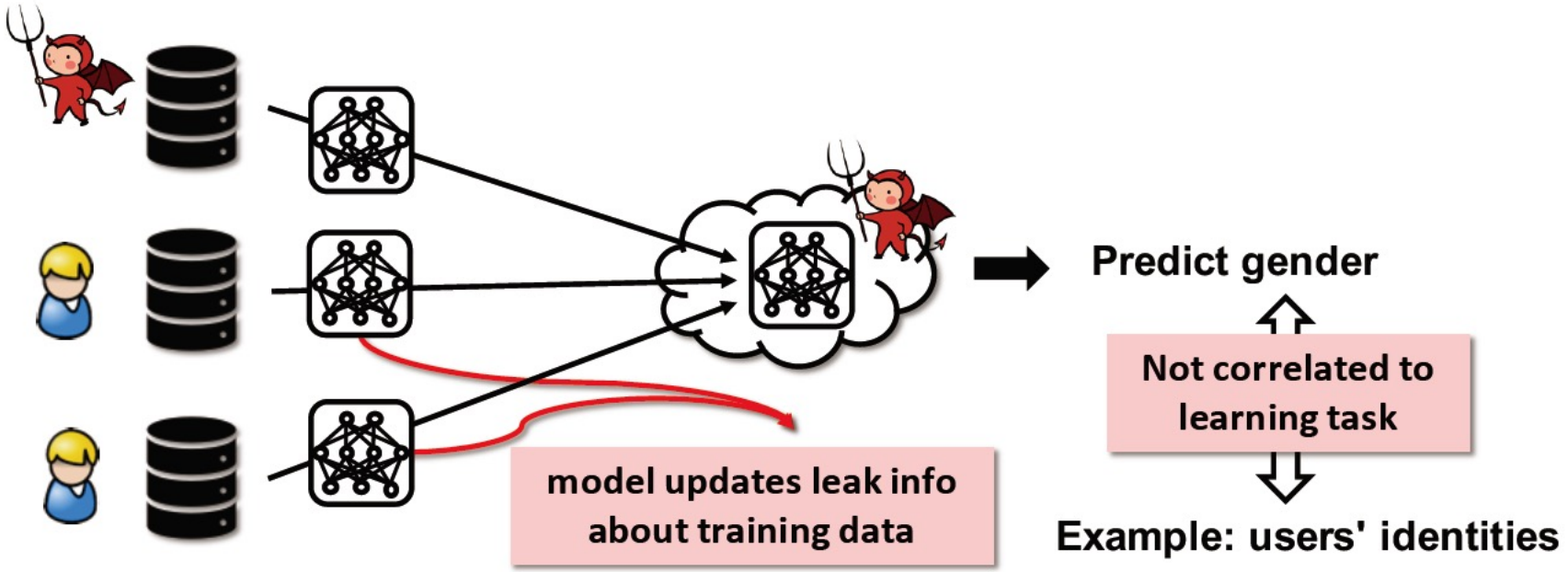
Threat models

- Two main goals of FL: global model & privacy protection
- Attacks: destroy one of these two goals, or even both at once

Attacker	Attacking model		Description	Objective
Honest-but-curious devices	Inference/evasion attacks		Target participant privacy	Privacy protection
Byzantine devices	Poisoning attacks	Random attacks	Attempt to prevent a model from being learned at all	Global model
		Targeted attacks	Attempt to bias the model to produce inferences that are preferable to the adversary	Global model & Privacy protection

Inference attacks

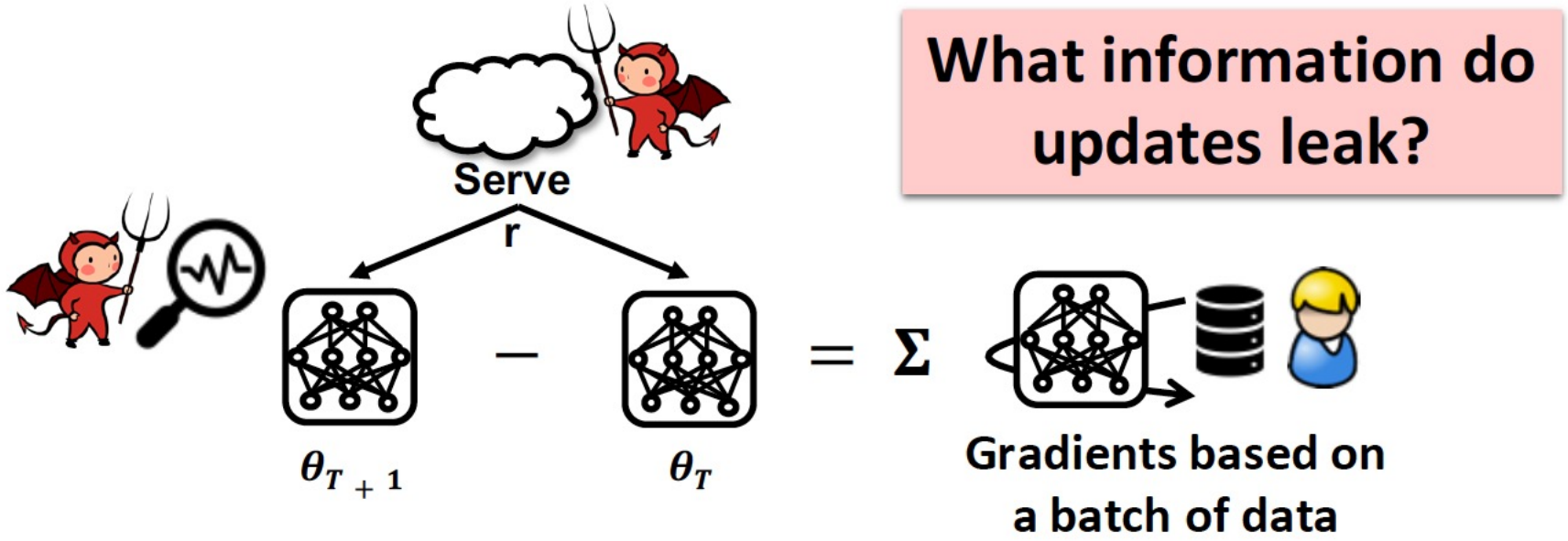
- Exchanging gradients during FL training can result in serious privacy leakage
  - Attacker infers information unrelated to the learning task



[1] Lyu L, Yu H, Yang Q. Threats to federated learning: A survey[J]. arXiv preprint arXiv:2003.02133, 2020

Inference attacks

- Exchanging gradients during FL training can result in serious privacy leakage
  - Attacker infers gradients from a batch of training data



[1] Lyu L, Yu H, Yang Q. Threats to federated learning: A survey[J]. arXiv preprint arXiv:2003.02133, 2020

Inference attacks

- Inferring class representatives
- Inferring membership
- Inferring properties
- Inferring training inputs and labels

Inferring class representatives

- Attacking mode: Generative Adversarial Networks (GAN) [1] attack
  - Exploits the real-time nature of the FL learning process that allows the adversarial party to train a GAN that generates prototypical samples of the targeted training data which were meant to be private
- Representative work
  - Work [2] proposes the mGAN-AI framework for exploring GAN-based attacks on FL
    - mGAN-AI attacks are experimented on a malicious central server of the FL environment
    - It explores user level privacy leakage against the federated learning by the attack from a malicious server
    - The inference attack gains the highest accuracy with mGAN-AI framework because it does not interfere with the training process.

[1] Briland Hitaj, Giuseppe Ateniese, and Fernando P´erez-Cruz. Deep models under the gan: information leakage from collaborative deep learning. In CSS, pages 603–618, 2017.
[2] Wang Z, Song M, Zhang Z, et al. Beyond inferring class representatives: User-level privacy leakage from federated learning[C]//IEEE INFOCOM 2019-IEEE Conference on Computer Communications. IEEE, 2019: 2512-2520.

Inferring membership

- Attacking mode [1]
  - Aim to get information by checking if the data exists on a training set
  - The attacker misuses the global model to get information on the training data of the other users.
  - The information on the training data set is inferred through guesswork and training the predictive model to predict original training data
- Representative work
  - Work [2] explores the vulnerability of the neural network (NN) to memorize their training data which is prone to passive and active inference attacks.

[1] Truex S, Liu L, Gursoy M E, et al. Demystifying membership inference attacks in machine learning as a service[J]. IEEE Transactions on Services Computing, 2019.
[2] Nasr M, Shokri R, Houmansadr A. Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning[C]//2019 IEEE symposium on security and privacy (SP). IEEE, 2019: 739-753.

Inferring properties

- Attacking mode [1]
  - An adversary can launch both passive and active property inference attacks to infer properties of other participants' training data that are independent of the features that characterize the classes of the FL model
- Representative work
  - Work [2] uses multi-task learning to trick the FL model into learning a better separation for data with and without the property

[1] Lyu L, Yu H, Yang Q. Threats to federated learning: A survey[J]. arXiv preprint arXiv:2003.02133, 2020.
[2] Luca Melis, Congzheng Song, Emiliano De Cristofaro, and Vitaly Shmatikov. Exploiting unintended feature leakage in collaborative learning. In SP, pages 691–706, 2019

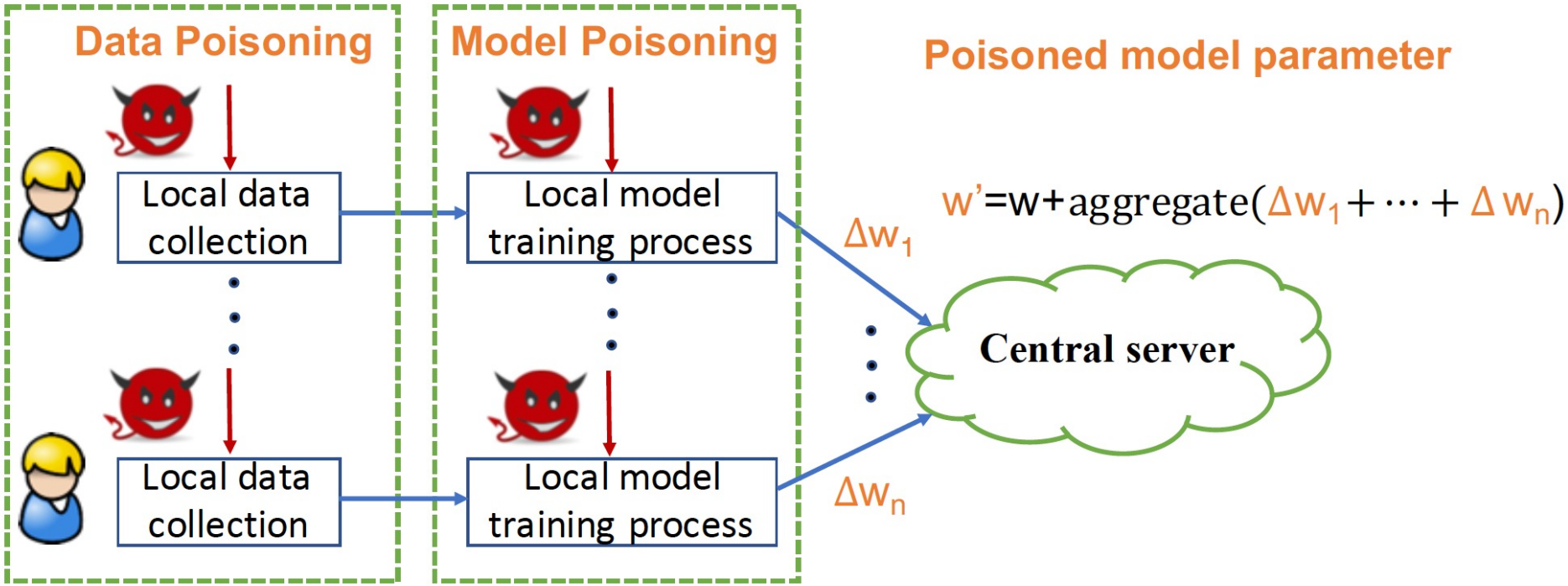
Inferring training inputs and labels

- Attacking mode
  - Infers labels from the shared gradients and recover the original training samples without requiring any prior knowledge about the training set
- Representative works
  - Work [1] proposes Deep Leakage from Gradient (DLG), which is an optimization algorithm that can obtain both the training inputs and the labels in just a few iterations
  - Work [2] presents an analytical approach called Improved Deep Leakage from Gradient (iDLG), which can certainly extract labels from the shared gradients by exploiting the relationship between the labels and the signs of corresponding gradients

[1] Ligeng Zhu, Zhijian Liu, and Song Han. Deep leakage from gradients. In NeurIPS, pages 14747–14756, 2019.
[2] Bo Zhao, Konda Reddy Mopuri, and Hakan Bilen. idlg: Improved deep leakage from gradients. CoRR, arXiv:2001.02610, 2020.

Poisoning attacks

- Objective: random attacks & targeted attacks
- Data v.s. model poisoning attacks in FL



[1] Lyu L, Yu H, Yang Q. Threats to federated learning: A survey[J]. arXiv preprint arXiv:2003.02133, 2020

Defensive techniques

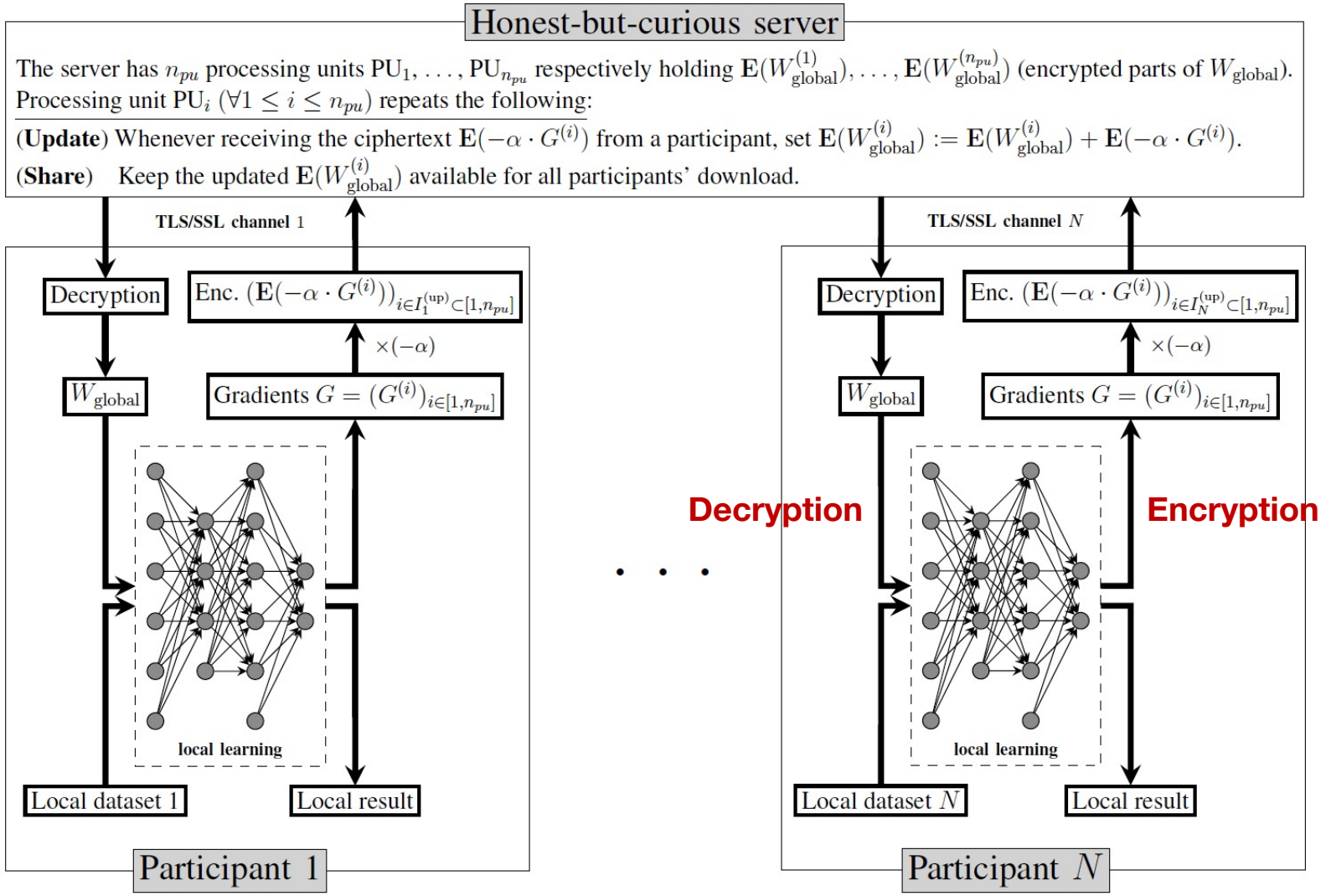
- Secure Multi-party Computation (SMC)
- Differential Privacy (DP)
- Hybrid: SMC + DP
- VerifyNet
- Adversarial training

Solution 1: Secure multi-party computation

- Traditional SMC [1]
  - Utilize cryptographic methods to secure the inputs of multi-participant while they jointly compute a model or a function
- Application in FL
  - Key idea: encrypt uploaded parameters
  - Representative work
    - Work [2] combines encryption with asynchronous stochastic gradient descent (SGD) which efficiently prevents data leakage of clients at the central server.

[1] Canetti R, Feige U, Goldreich O, et al. Adaptively secure multi-party computation[C]//Proceedings of the twenty-eighth annual ACM symposium on Theory of computing. 1996: 639-648.
[2] Aono Y, Hayashi T, Wang L, et al. Privacy-preserving deep learning via additively homomorphic encryption[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(5): 1333-1345.

Representative work of SMC



Key technology:
Gradient-encrypted (i.e., additively homomorphic encryption
) Asynchronous Stochastic Gradient Descent (SGD)

[1] Aono Y, Hayashi T, Wang L, et al. Privacy-preserving deep learning via additively homomorphic encryption[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(5): 1333-1345.

Representative work of SMC

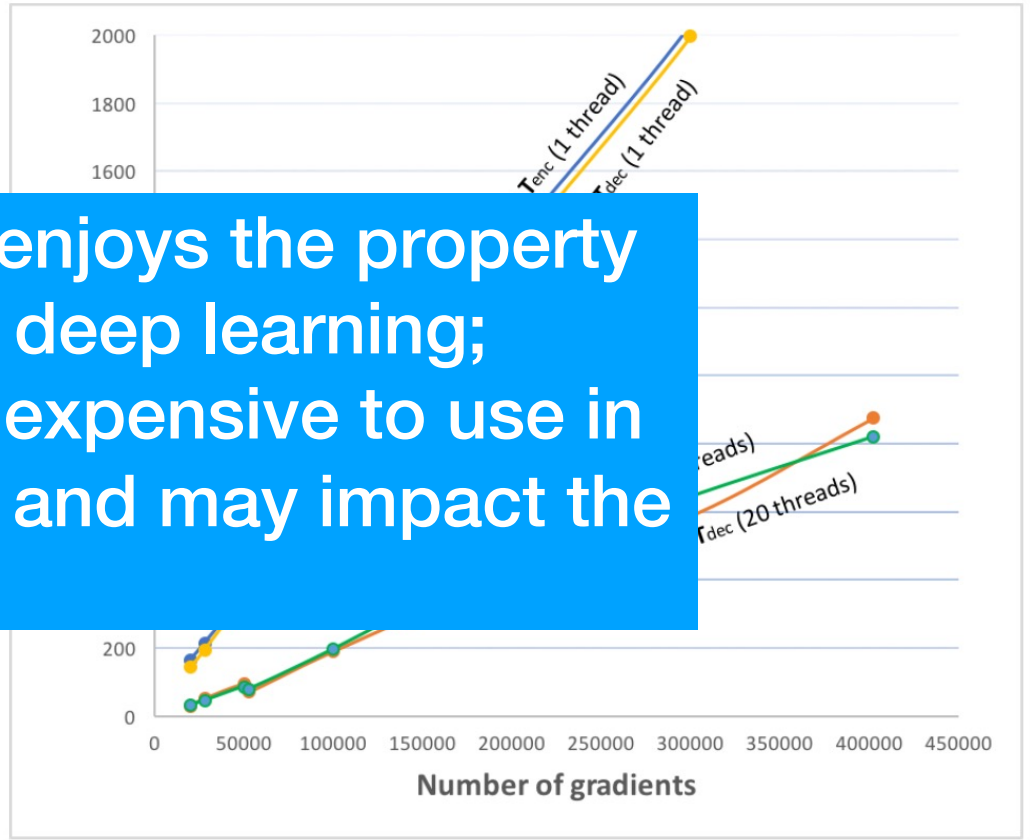
Experiment

- Accuracy: Tensorflow code achieves around 97% accuracy over the testing set
- Computational costs estimation

Enc. via (10)	
Dec. of (10)	
Add of ciphertexts via (	

Enc. via (10)	
Dec. of (10)	
Add of ciphertexts via (	

For privacy-preserving, system enjoys the property of not declining the accuracy of deep learning; But the encryption technique is expensive to use in a larger landscape environment and may impact the efficiency of the ML model.



[1] Aono Y, Hayashi T, Wang L, et al. Privacy-preserving deep learning via additively homomorphic encryption[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(5): 1333-1345.

Solution 1: Analysis

- SMC mainly protects against attacks by encrypting uploaded parameters
- Key challenge: efficiency loss
  - SMC based solutions have a higher time complexity than typical FL frameworks which may negatively affect the model training.

Solution 2: Differential privacy (DP)

- Traditional DP [1,2]
  - Preserve privacy by adding noise to local sensitive data
  - Statistic data quality loss caused by the added noise of each user is relatively low compared with the increased privacy protection
- Application in FL
  - Key idea: add random noise to uploaded parameters
  - Representative work:
    - Work [3] adds artificial noises to the parameters at the clients side before aggregating to effectively prevent information leakage.

[1] Dwork C. Differential privacy[C]//International Colloquium on Automata, Languages, and Programming. Springer, Berlin, Heidelberg, 2006: 1-12.
[2] Xie L, Lin K, Wang S, et al. Differentially private generative adversarial network[J]. arXiv preprint arXiv:1802.06739, 2018
[3] Wei K, Li J, Ding M, et al. Federated learning with differential privacy: Algorithms and performance analysis[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3454-3469.

Representative work of DP

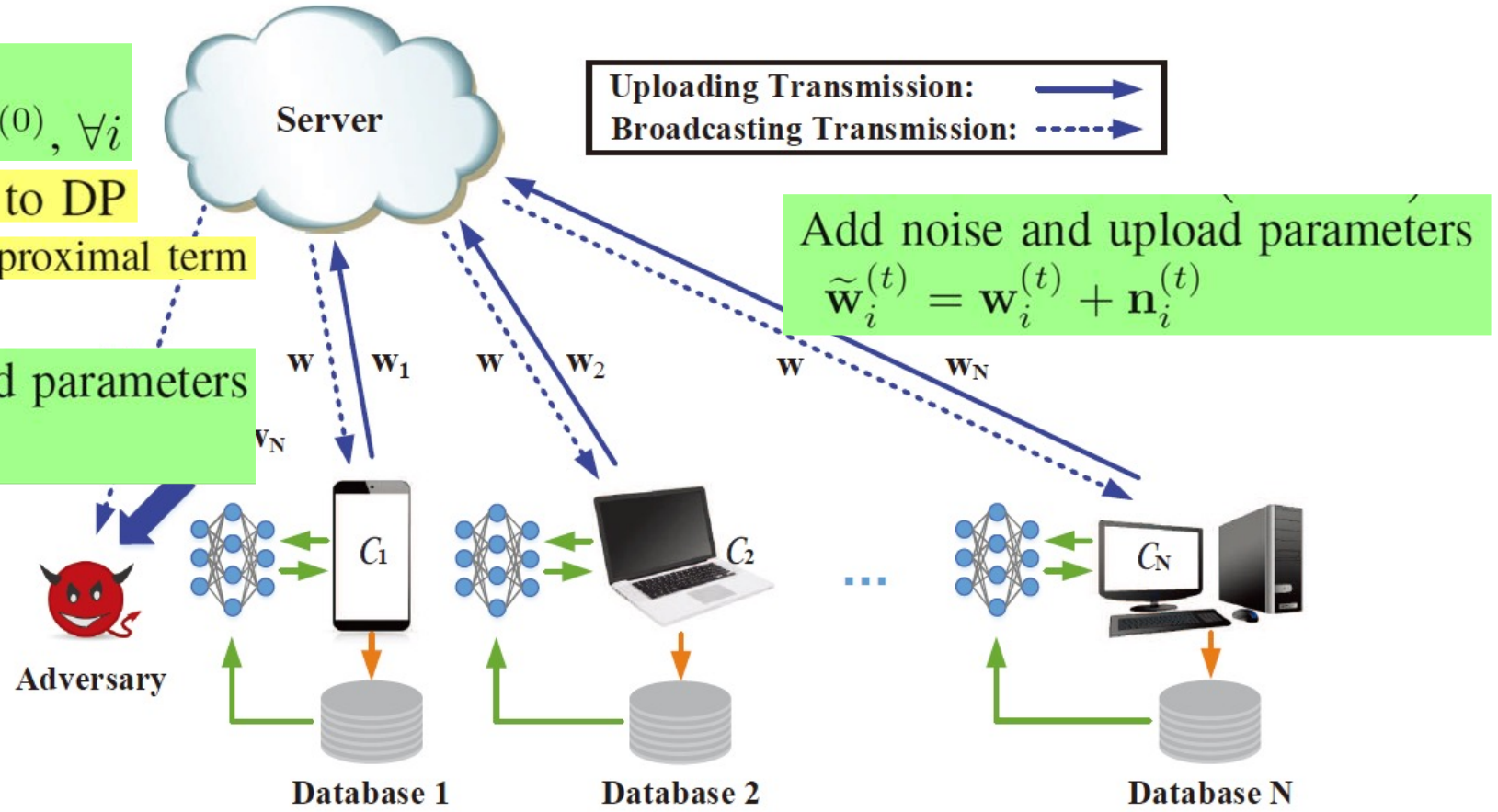
- NbAFL: noising before model aggregation FL (key technology)

Data: T, $\mathbf{w}^{(0)}$, μ, ϵ and δ
Initialization: $t = 1$ and $\mathbf{w}_i^{(0)} = \mathbf{w}^{(0)}, \forall i$

ϵ, δ The parameters related to DP

μ A presetting constant of the proximal term

The server broadcasts global noised parameters
$\widetilde{\mathbf{w}}^{(t)} = \mathbf{w}^{(t)} + \mathbf{n}_{\mathrm{D}}^{(t)}$

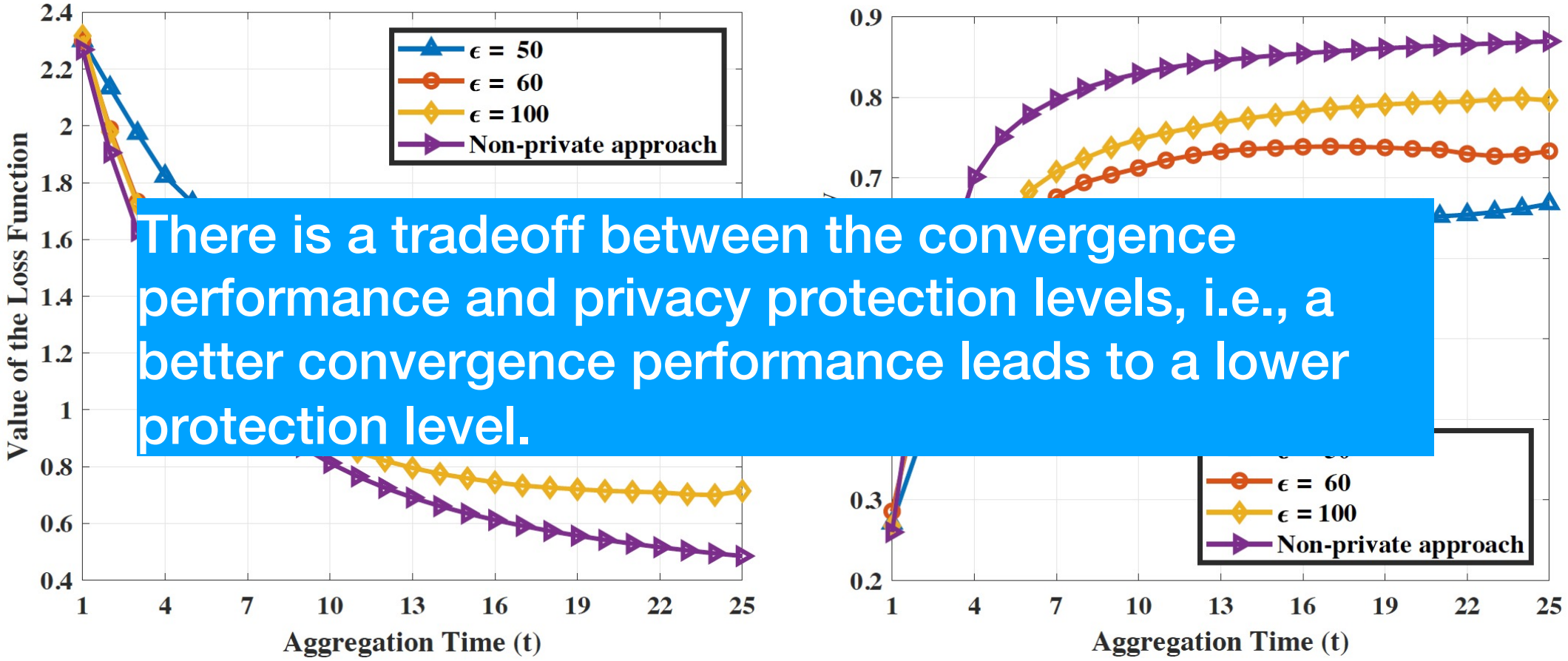


[1] Wei K, Li J, Ding M, et al. Federated learning with differential privacy: Algorithms and performance analysis[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3454-3469.

Representative work of DP

Experiment

- Performance evaluation on protection levels ϵ



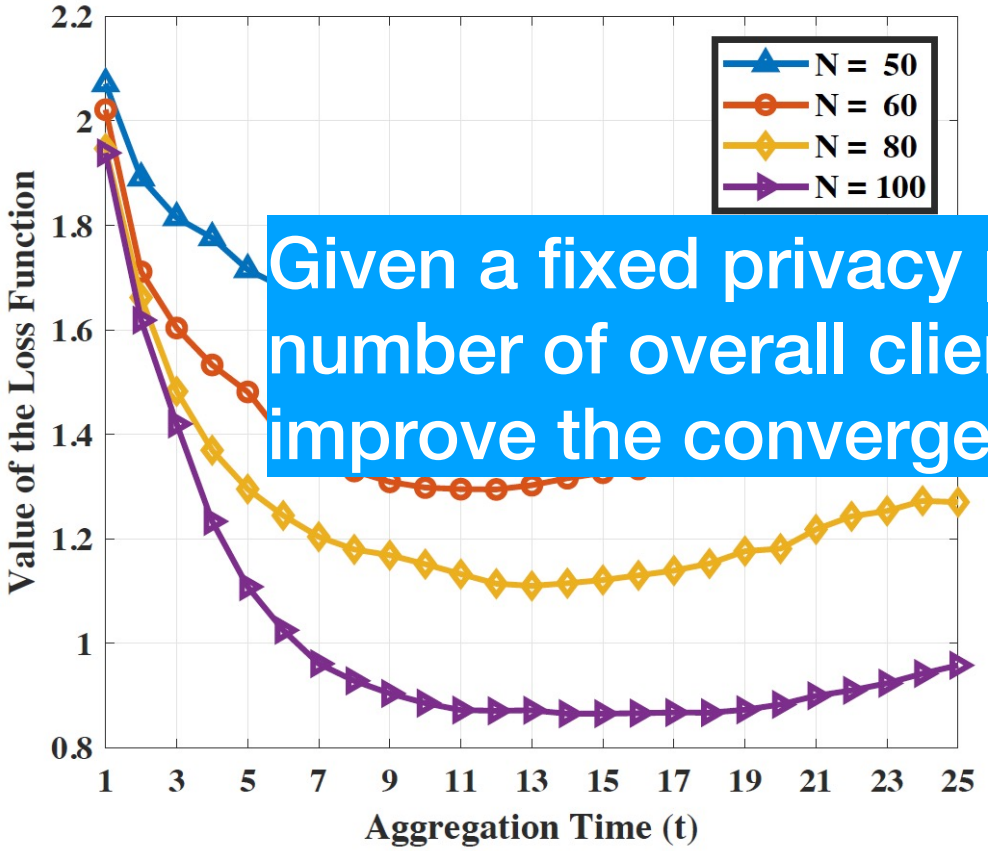
There is a tradeoff between the convergence performance and privacy protection levels, i.e., a better convergence performance leads to a lower protection level.

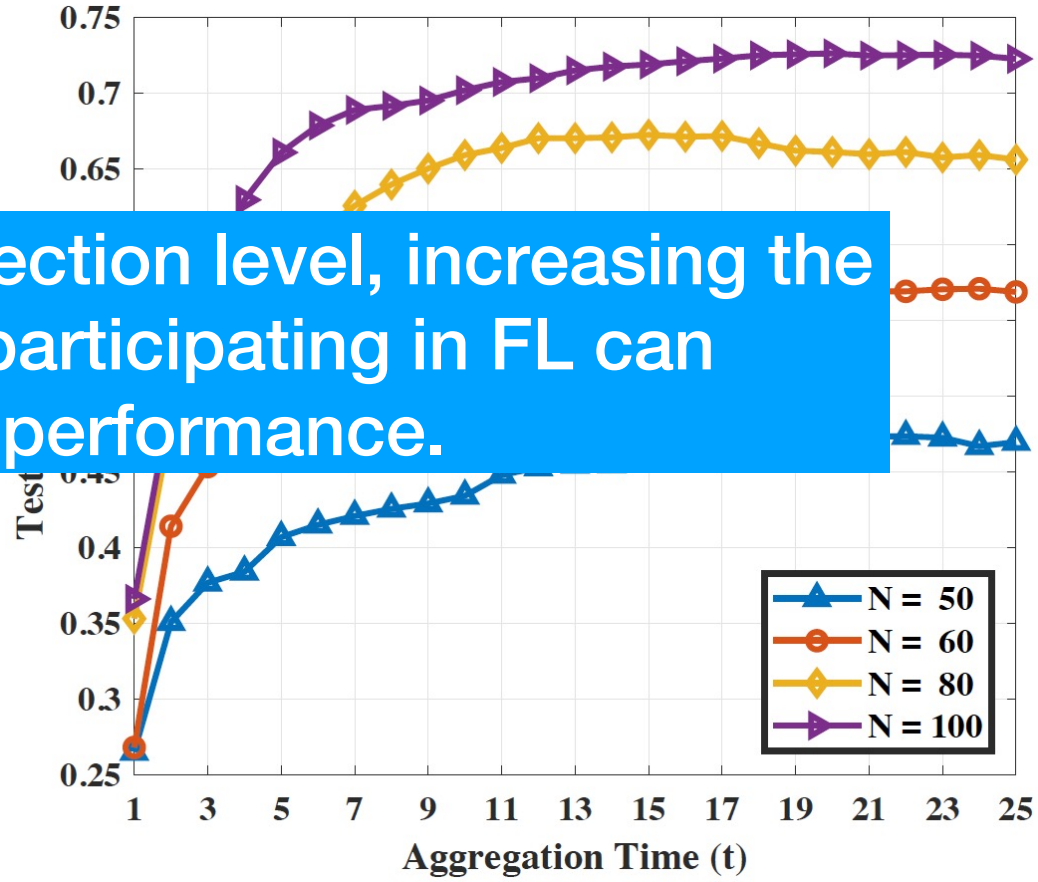
[1] Wei K, Li J, Ding M, et al. Federated learning with differential privacy: Algorithms and performance analysis[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3454-3469.

Representative work of DP

Experiment

- Impact of the number of clients N





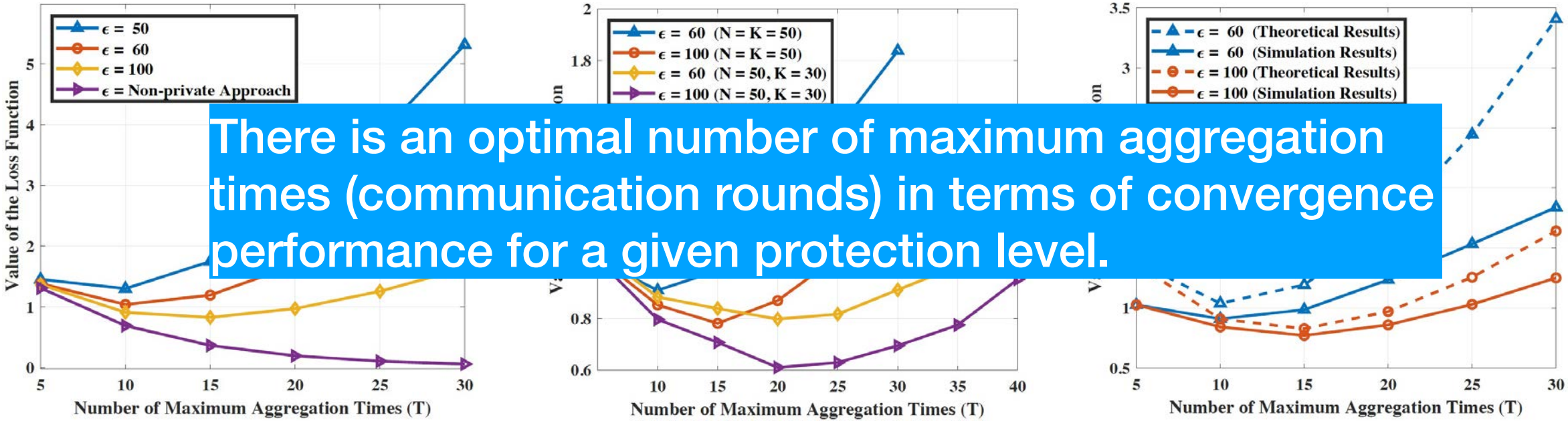
Given a fixed privacy protection level, increasing the number of overall clients participating in FL can improve the convergence performance.

[1] Wei K, Li J, Ding M, et al. Federated learning with differential privacy: Algorithms and performance analysis[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3454-3469.

Representative work of DP

Experiment

- Impact of the number of maximum aggregation times T



There is an optimal number of maximum aggregation times (communication rounds) in terms of convergence performance for a given protection level.

[1] Wei K, Li J, Ding M, et al. Federated learning with differential privacy: Algorithms and performance analysis[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3454-3469.

Solution 2: Analysis

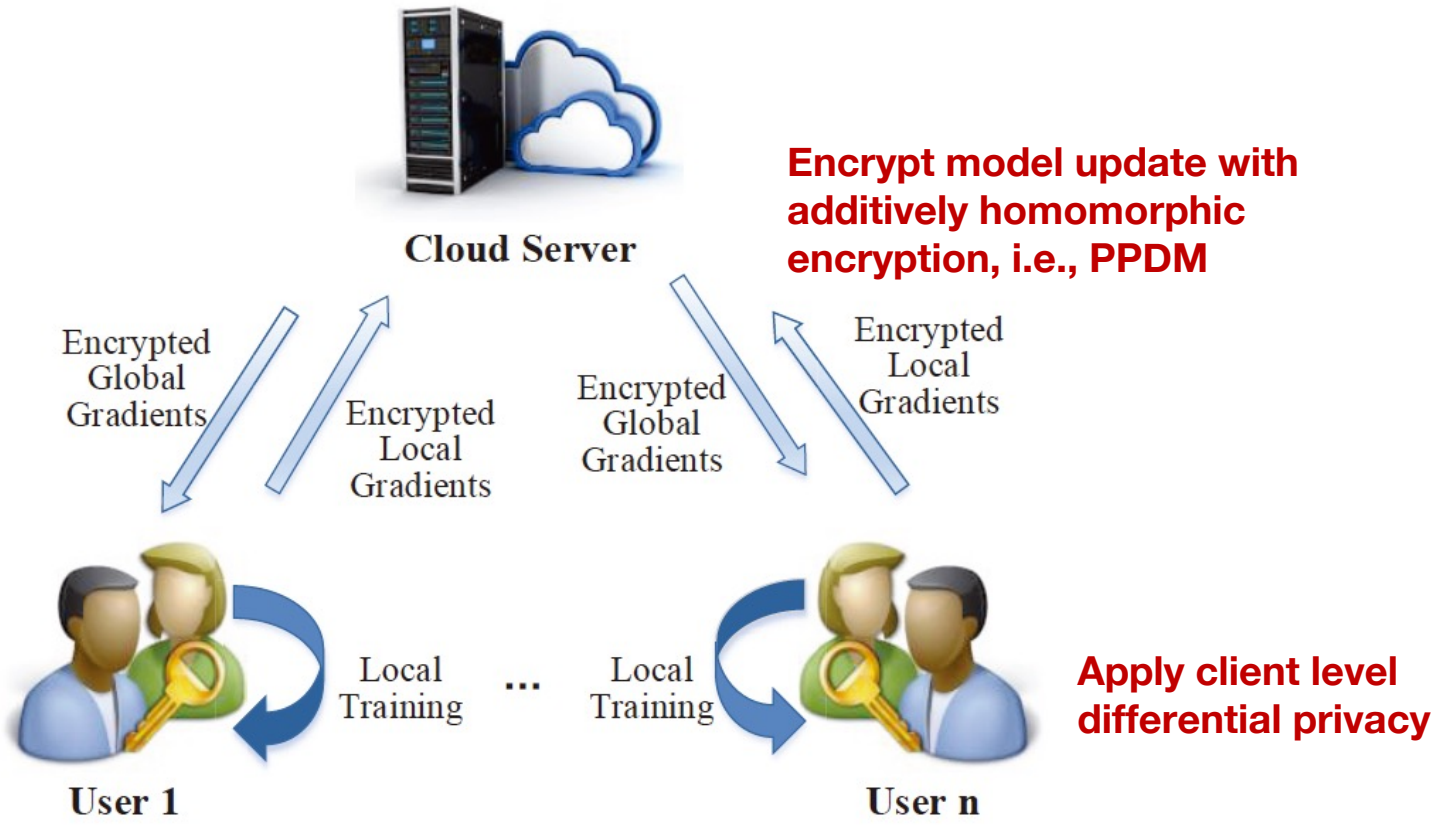
- DP add random noise to uploaded parameters to protect against privacy
- Key challenge: accuracy loss
  - Bring uncertainty into the upload parameters and may harm the training performance
  - Make the FL server more difficult to evaluate the client's behavior to calculate payoff

Solution 3: Hybrid = SMC + DP

- Key idea
  - Encrypt the manipulated parameter
  - Aim to achieve a secured federated learning model with high efficiency and accuracy simultaneously
- Representative works
  - Work [1] combines homomorphic encryption and differential privacy
  - Work [2] combines differential privacy with secure multiparty computation

[1] Hao M, Li H, Xu G, et al. Towards efficient and privacy-preserving federated deep learning[C]//ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE, 2019: 1-6.
[2] Truex S, Baracaldo N, Anwar A, et al. A hybrid approach to privacy-preserving federated learning[C]//Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security. 2019: 1-11

Representative work I of Hybrid Technology



Encrypt model update with additively homomorphic encryption, i.e., PPDM

Apply client level differential privacy

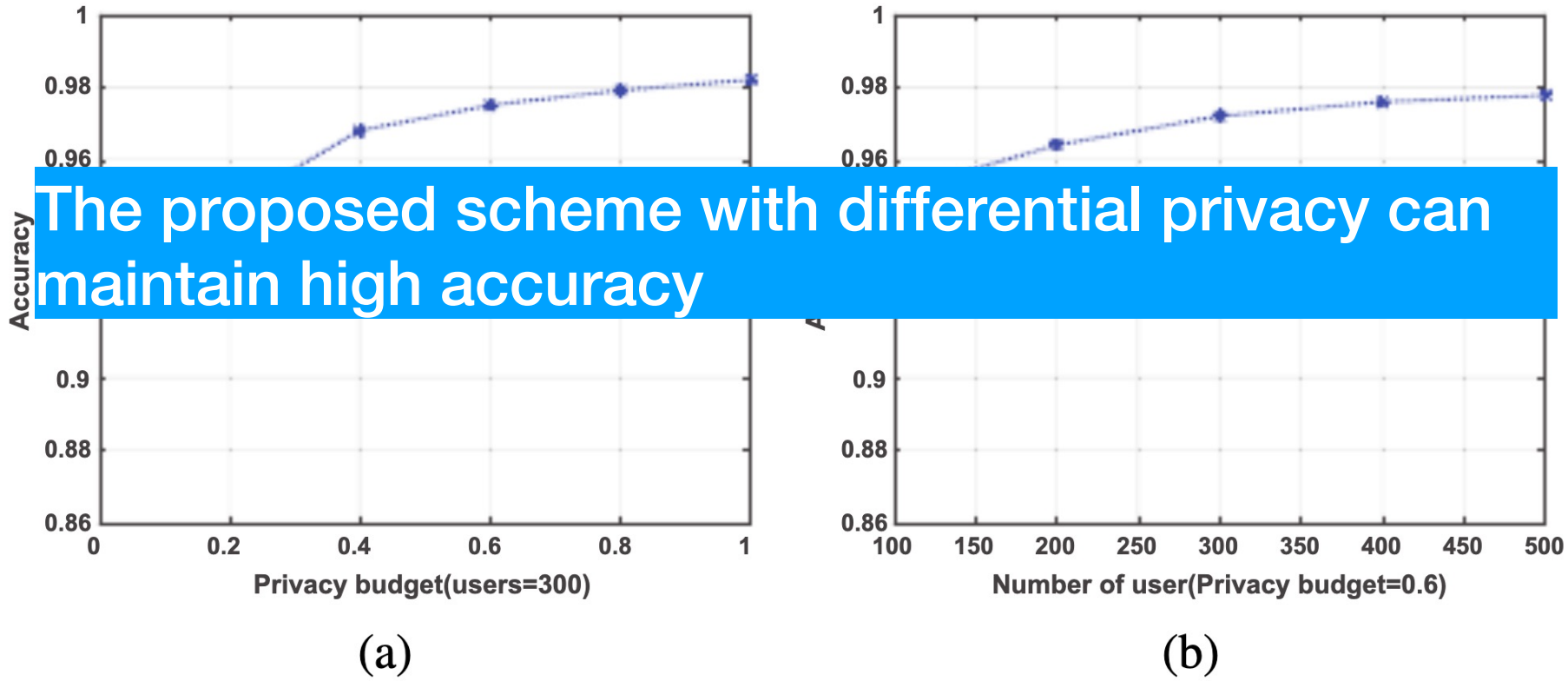
Key technology:
Integration of homomorphic encryption and differential privacy

[1] Hao M, Li H, Xu G, et al. Towards efficient and privacy-preserving federated deep learning[C]//ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE, 2019: 1-6.

Representative work I of Hybrid

Experiment

- Accuracy



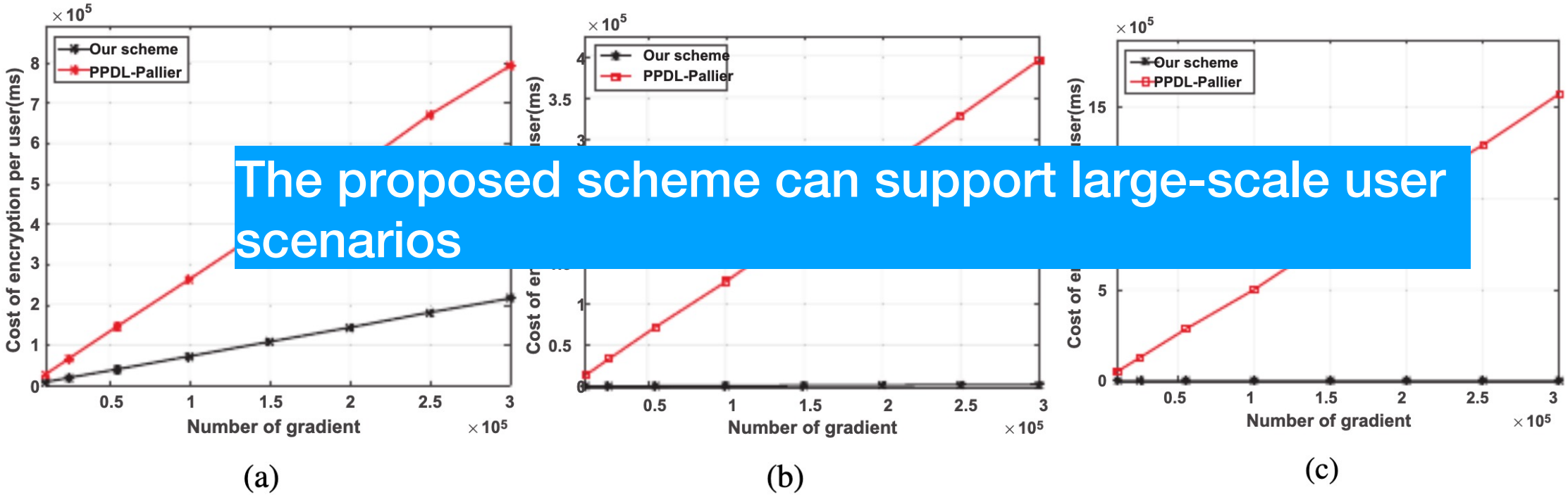
The proposed scheme with differential privacy can maintain high accuracy

[1] Hao M, Li H, Xu G, et al. Towards efficient and privacy-preserving federated deep learning[C]//ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE, 2019: 1-6.

Representative work I of Hybrid

Experiment

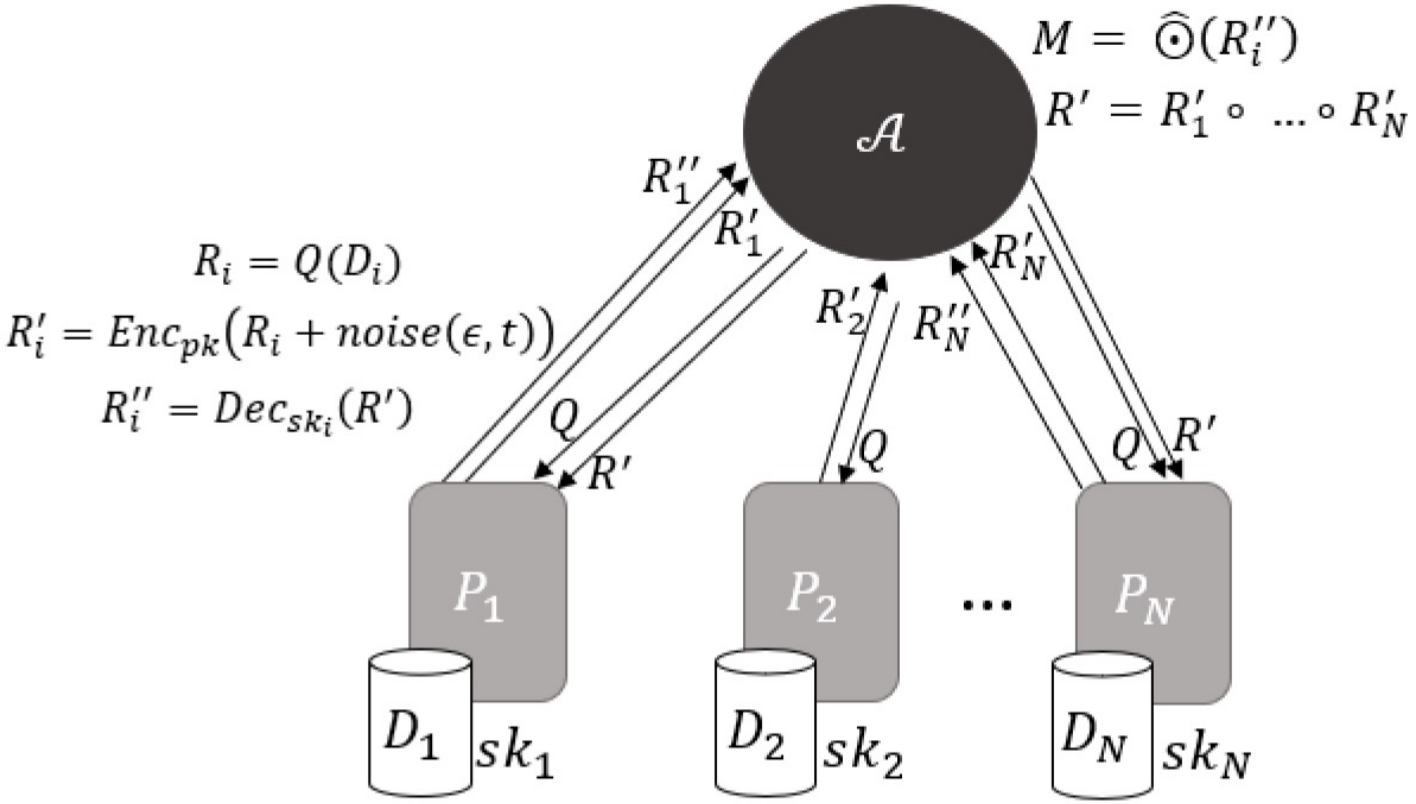
- Communication cost



The proposed scheme can support large-scale user scenarios

[1] Hao M, Li H, Xu G, et al. Towards efficient and privacy-preserving federated deep learning[C]//ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE, 2019: 1-6.

Representative work II of Hybrid Technology



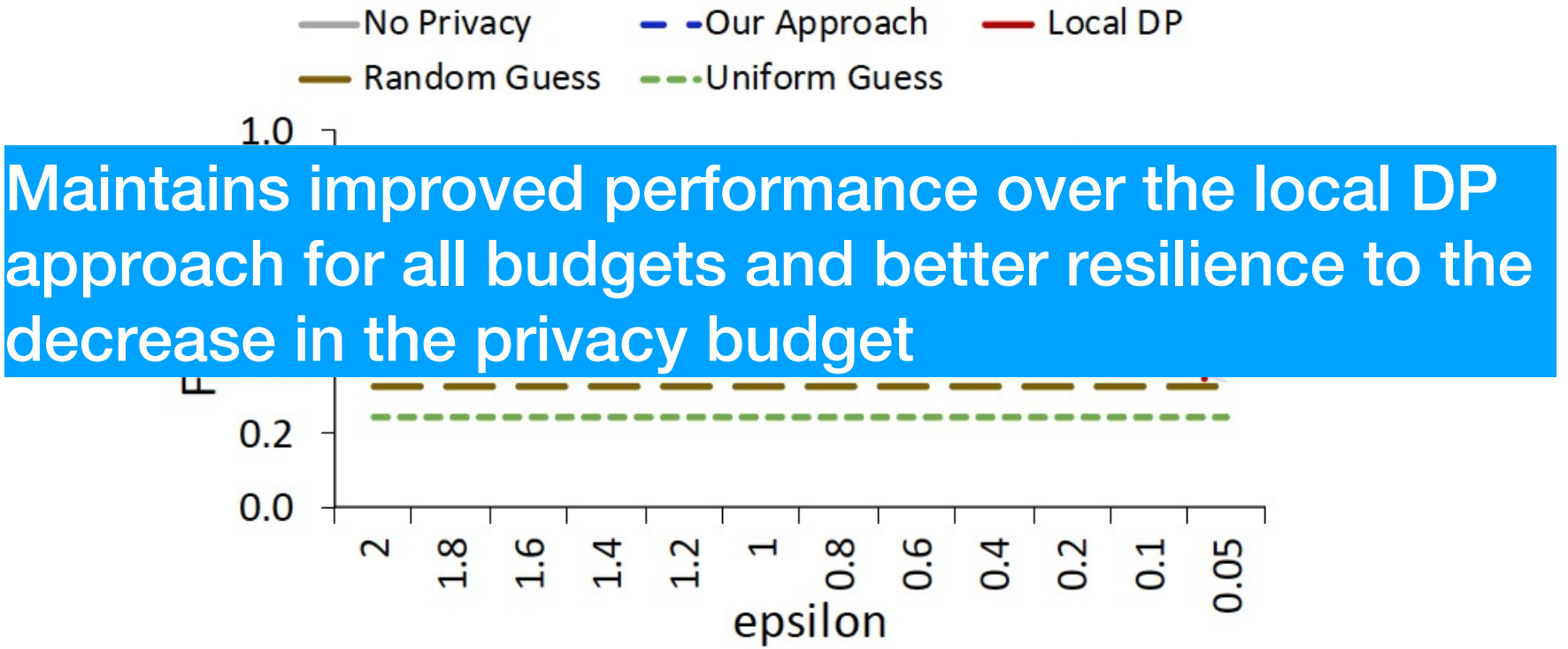
Key technology:
Integration of secure multiparty computation (SMC) and differential privacy (DP)

[1] Truex S, Baracaldo N, Anwar A, et al. A hybrid approach to privacy-preserving federated learning[C]//Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security. 2019: 1-11.

Representative work II of Hybrid

Experiment

- Impact of privacy budget



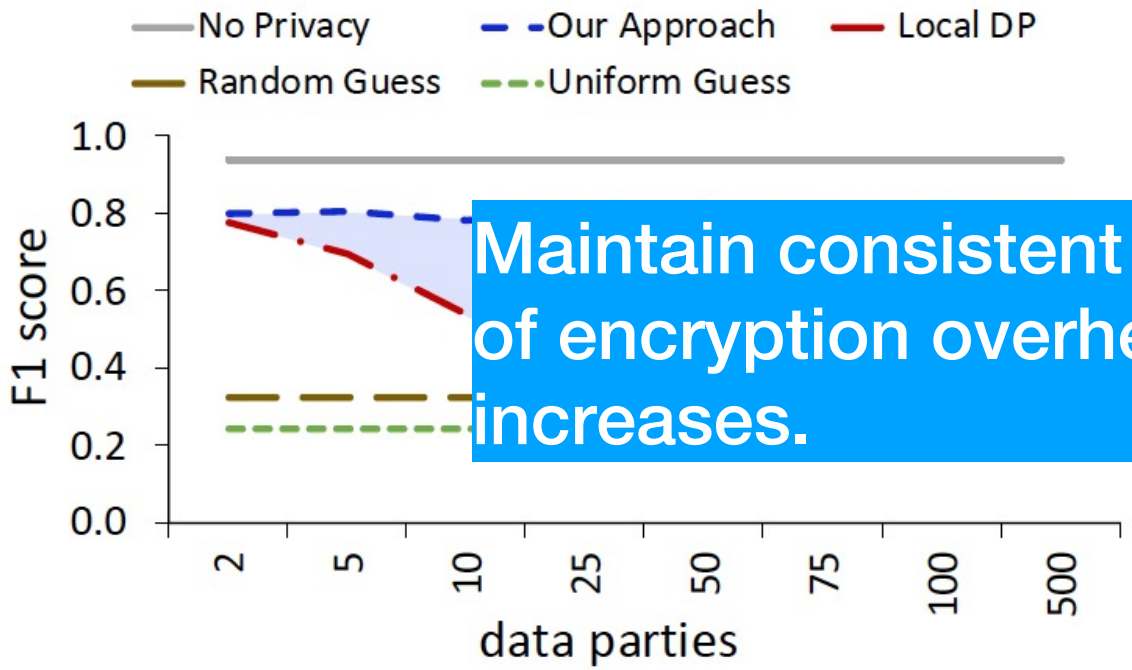
Maintains improved performance over the local DP approach for all budgets and better resilience to the decrease in the privacy budget

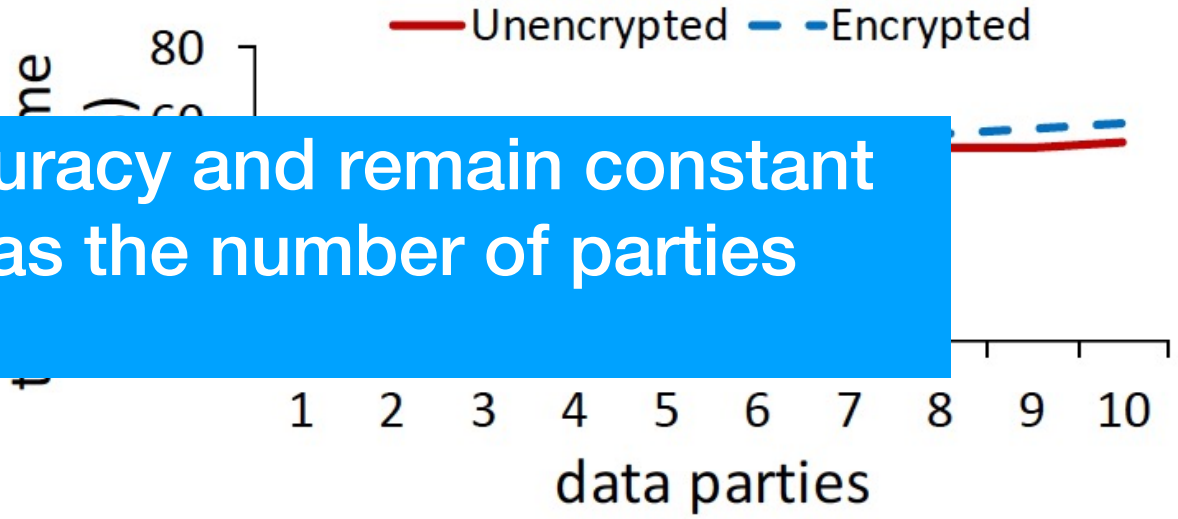
[1] Truex S, Baracaldo N, Anwar A, et al. A hybrid approach to privacy-preserving federated learning[C]//Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security. 2019: 1-11.

Representative work II of Hybrid

Experiment

- Impact of number of parties





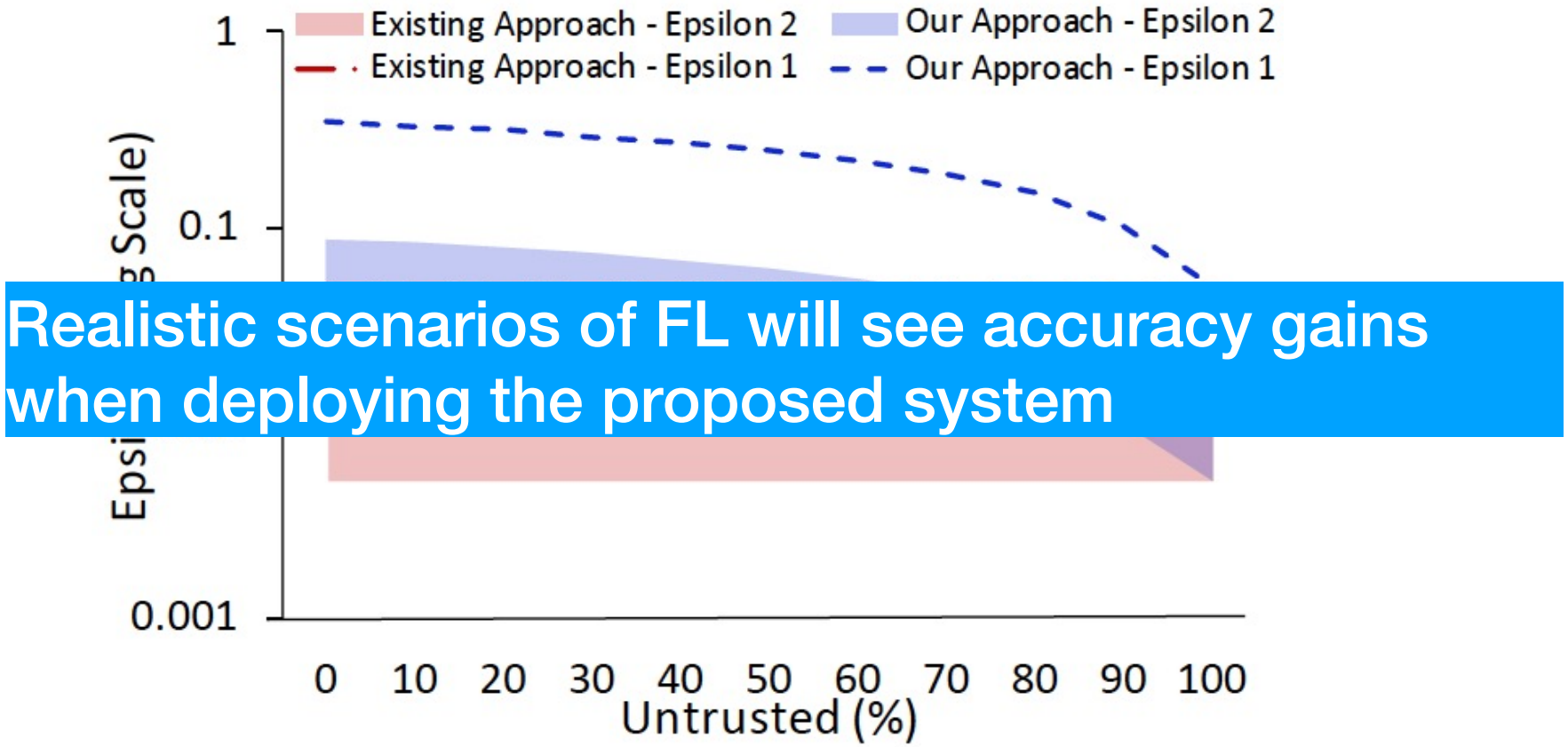
Maintain consistent accuracy and remain constant of encryption overhead as the number of parties increases.

[1] Truex S, Baracaldo N, Anwar A, et al. A hybrid approach to privacy-preserving federated learning[C]//Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security. 2019: 1-11.

Representative work II of Hybrid

Experiment

- Impact of number of trust



Realistic scenarios of FL will see accuracy gains when deploying the proposed system

[1] Truex S, Baracaldo N, Anwar A, et al. A hybrid approach to privacy-preserving federated learning[C]//Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security. 2019: 1-11.

Solution 3: Analysis

- Hybrid solution mainly combines DP and SMC to improve model accuracy while preserving provable privacy guarantees and protecting against extraction attacks and collusion threats.
- Key challenge
  - Subdued cost on both efficiency and accuracy

Solution 4: VerifyNet [1]

- Motivation: address three problems existing in federated training process
  - Protect the privacy of the user's local gradients in the workflow
  - Prevent malicious spoofing by the central server
  - Users' offline during training process

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

Solution 4: VerifyNet [1]

- A privacy preserving and verifiable FL framework
  - Privacy preserve: provides double-masking protocol which makes it difficult for attackers to infer training data
  - Reliability guarantee: provides a way for clients to verify central server results which ensures the reliability of central server
- Robust to handle multiple dropouts

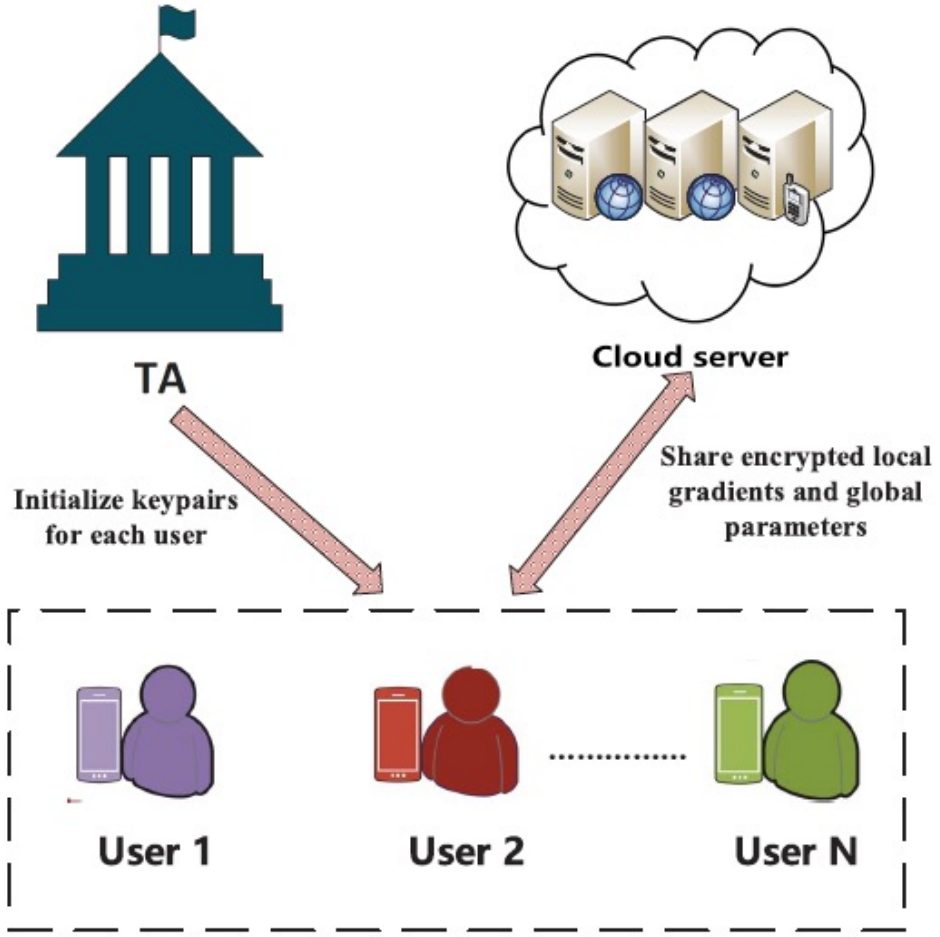
[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

VerifyNet

Technology

1. Trusted Authority (TA):
The main job of TA is to initialize the entire system, generate public parameters, and assign public and private keys to each participant. Afterwards, it will go offline unless a dispute arises.

2. User:
Each user needs to send his/her encrypted local gradients to the cloud server during each iteration. Besides, the cloud server will also receive some other encrypted information to prepare for generating $Proof$ of its calculated results
.



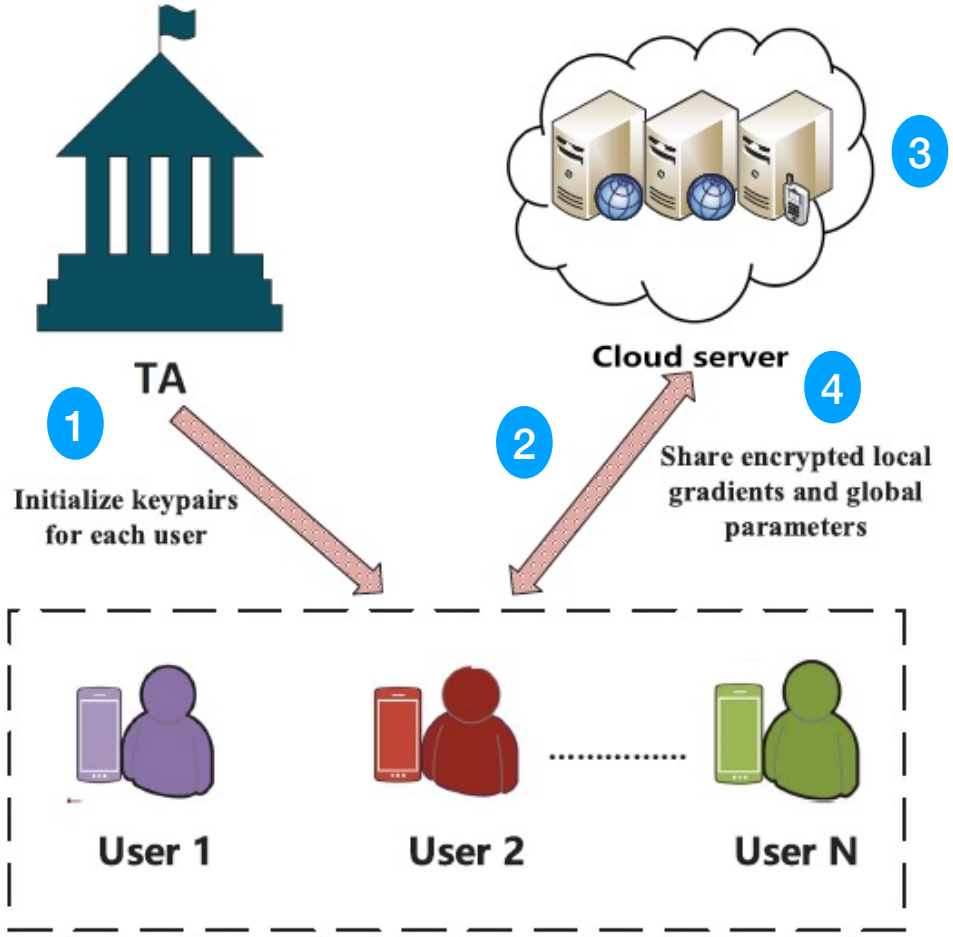
3. Cloud server:
The cloud server aggregates the gradients uploaded by all online users and sends the results along with the $Proof$ to each user.

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

VerifyNet

Technology

1. TA initializes the entire system and generates all the public and private keys



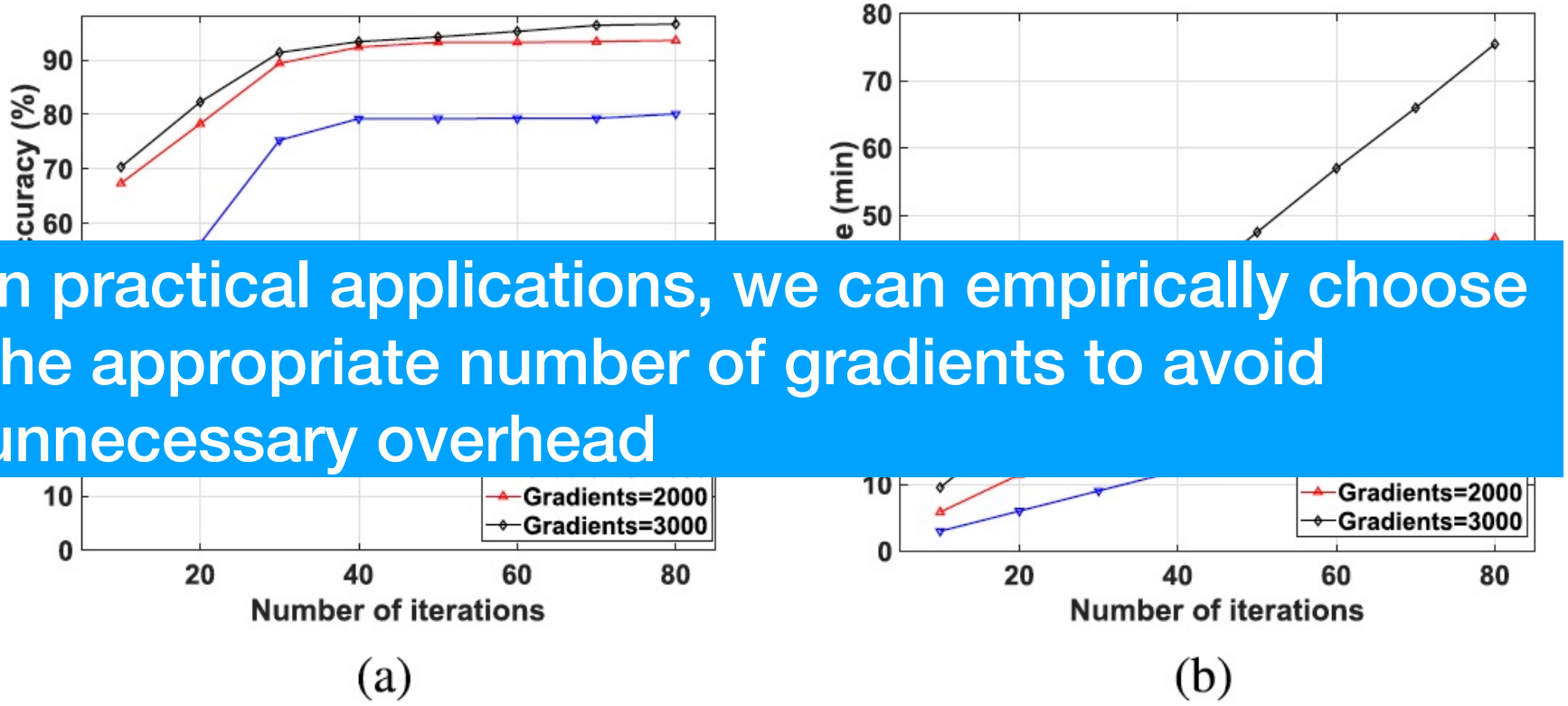
2. Each user encrypts its local gradient and submits it to the cloud server

3. After receiving enough message from all online users, the cloud server aggregates the gradients of all online users and returns the results along with $Proof$ to each user

4. Every user decides to accept or reject the calculation results by verifying the $Proof$ and returns to the round 1 to start a new iteration.

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

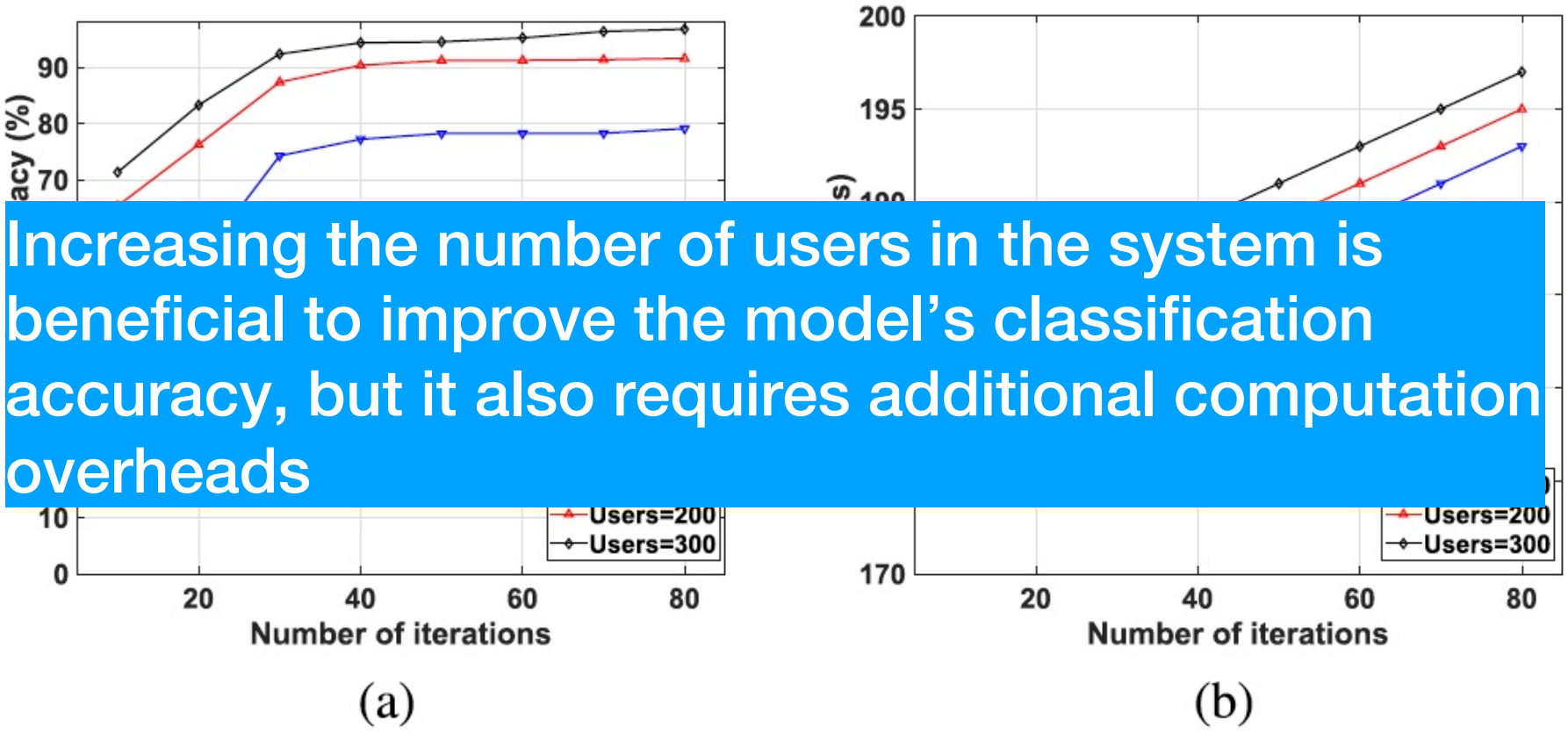
- Classification accuracy



In practical applications, we can empirically choose the appropriate number of gradients to avoid unnecessary overhead

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

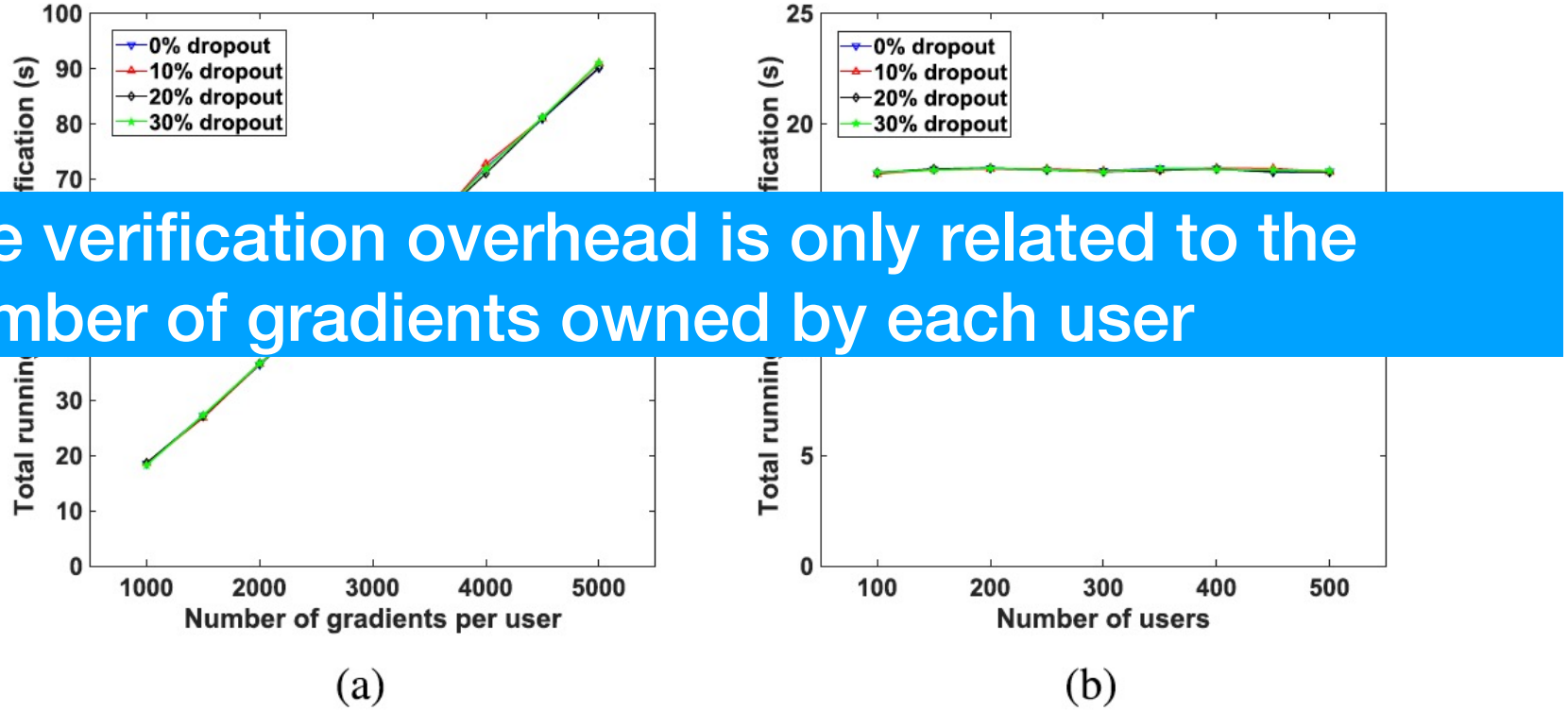
- Classification accuracy



Increasing the number of users in the system is beneficial to improve the model's classification accuracy, but it also requires additional computation overheads

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

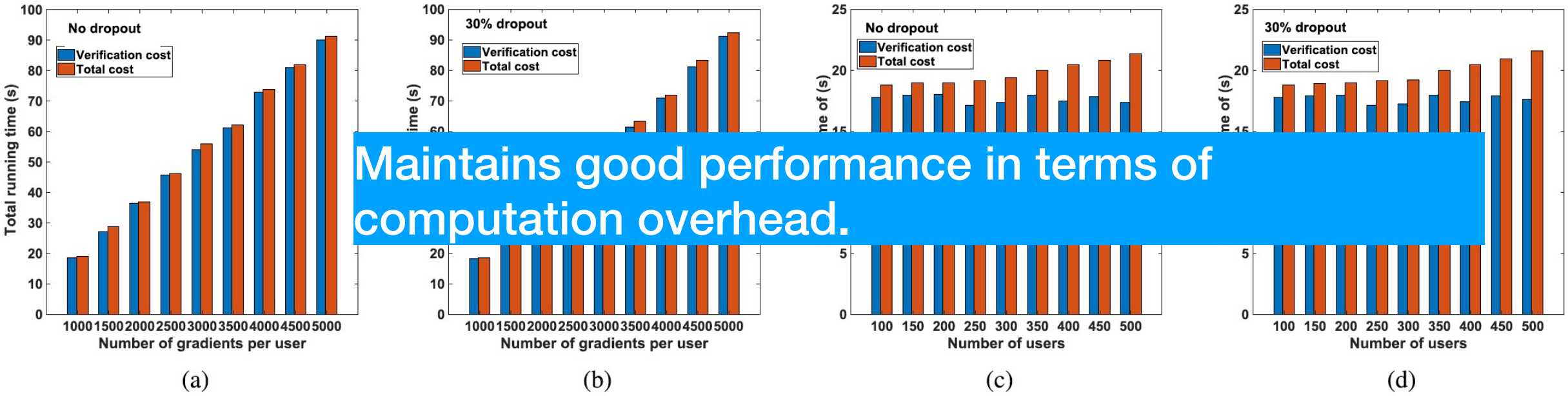
- Performance of client
  - Computation overhead



The verification overhead is only related to the number of gradients owned by each user

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

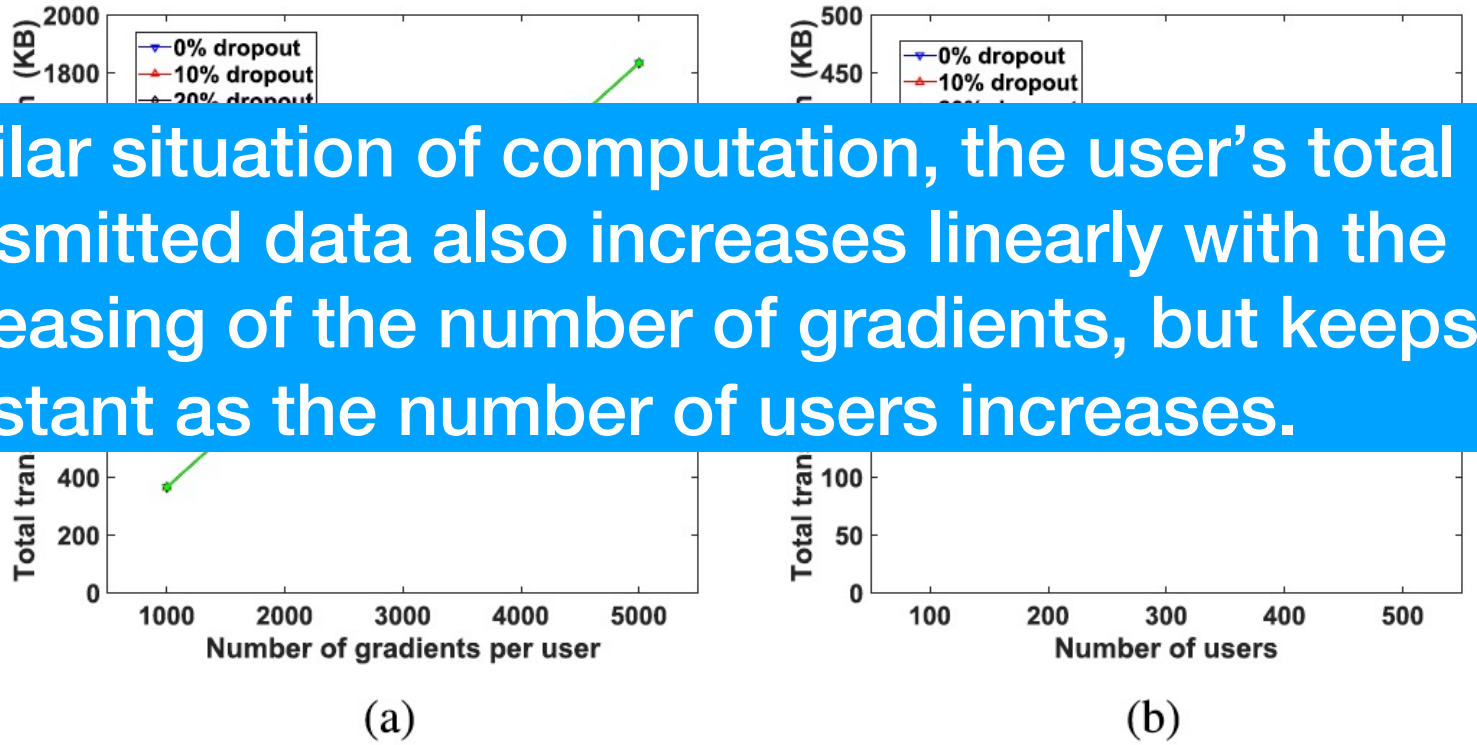
- Performance of client
  - Computation overhead



Maintains good performance in terms of computation overhead.

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

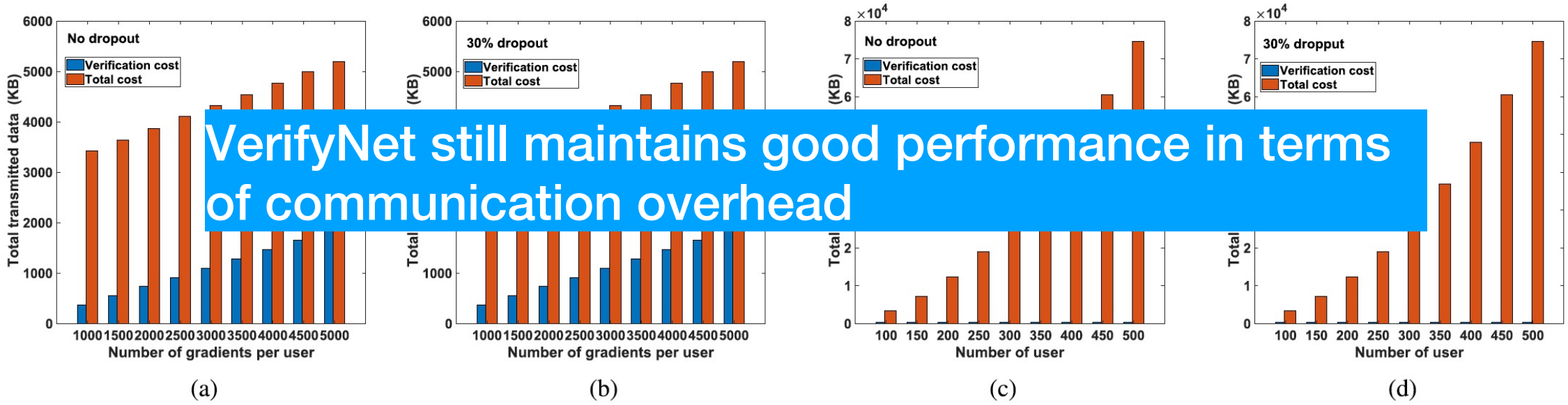
- Performance of client
  - Communication overhead



Similar situation of computation, the user's total transmitted data also increases linearly with the increasing of the number of gradients, but keeps a constant as the number of users increases.

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

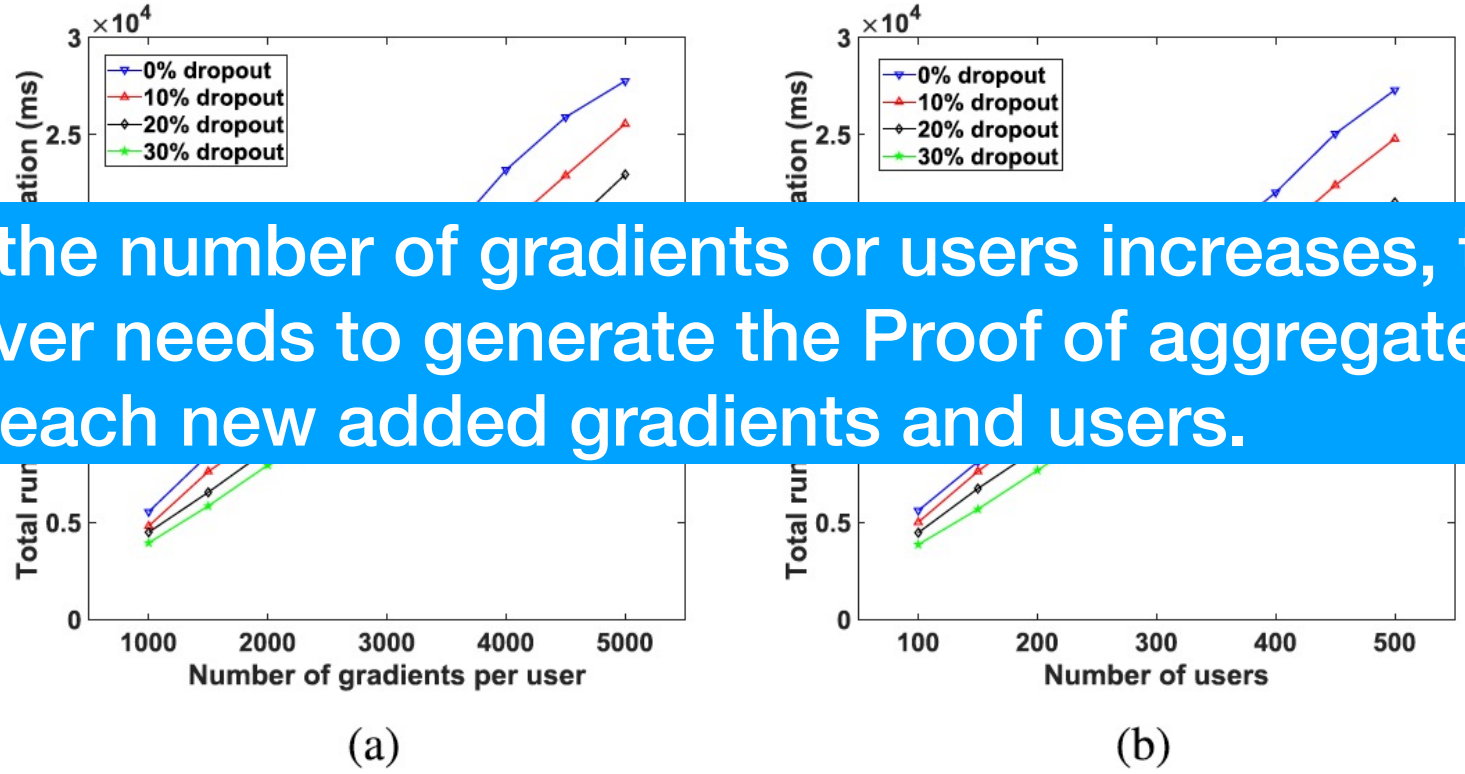
- Performance of client
  - Communication overhead



VerifyNet still maintains good performance in terms of communication overhead

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

- Performance of server
  - Computation overhead



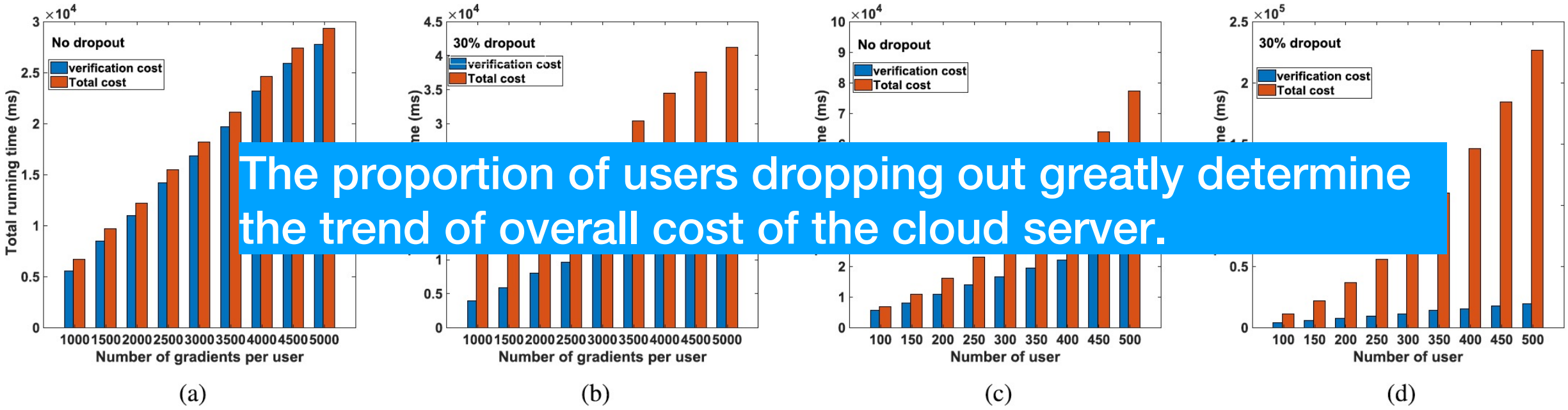
As the number of gradients or users increases, the cloud server needs to generate the Proof of aggregated result for each new added gradients and users.

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

VerifyNet

Experiment

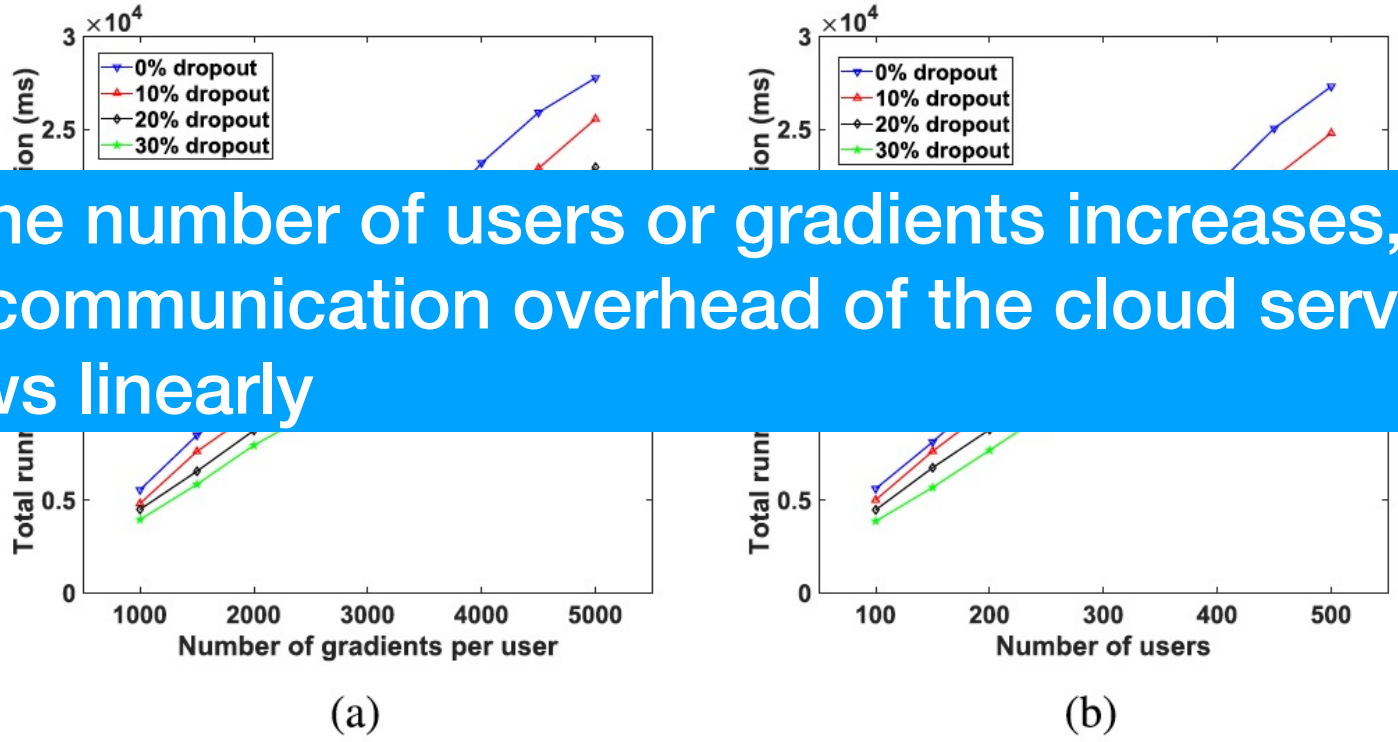
- Performance of server
  - Computation overhead



The proportion of users dropping out greatly determine the trend of overall cost of the cloud server.

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

- Performance of server
  - Communication overhead



As the number of users or gradients increases, the communication overhead of the cloud server also grows linearly

[1] Xu G, Li H, Liu S, et al. Verifynet: Secure and verifiable federated learning[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 911-926.

Solution 4: Analysis

- VerifyNet
  - Provides double-masking protocol to protect privacy
  - Uses verifiable aggregation results to ensure the reliability of central server
- Key challenge: computation overhead

Solution 5: Adversarial training

- Adversarial training [1]
  - A proactive defensive technique, tries all permutations of an attack from the beginning of the training phase to make the FL global model robust to known adversarial attacks.
- Application in FL
  - Key idea: the inclusion of adversarial examples obscures the raw data
  - Representative work
    - Work [2] proposes FedDynAT to study the feasibility of using adversarial training (AT) in the communication constrained federated learning

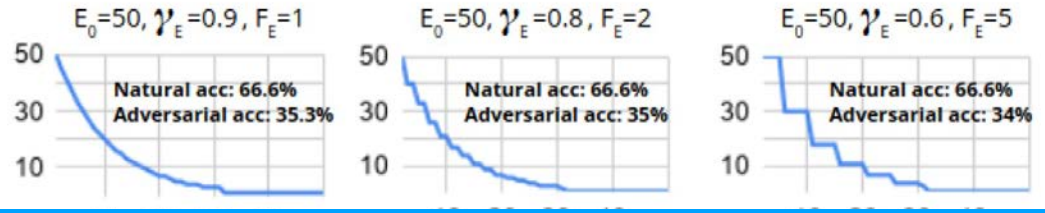
[1] Tramèr F, Boneh D, Kurakin A, et al. Ensemble adversarial training: Attacks and defenses[C]//6th International Conference on Learning Representations, ICLR 2018-Conference Track Proceedings. 2018.
[2] Shah D, Dube P, Chakraborty S, et al. Adversarial training in communication constrained federated learning[J]. arXiv preprint arXiv:2103.01319, 2021.

FedDynAT: Federated Dynamic AT

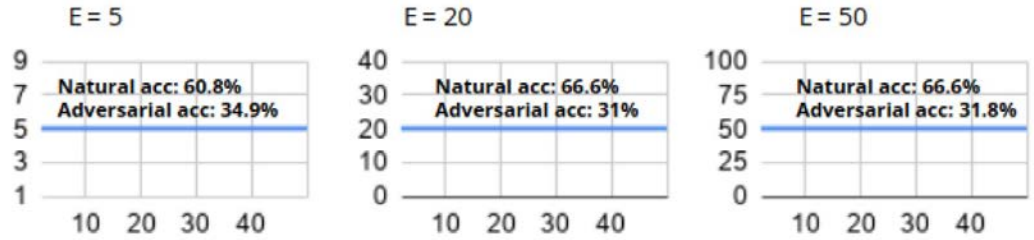
- Motivation: challenges in adopting AT to federated learning
  - Increased drop in natural and adversarial accuracy with federated AT and non-iid data
  - Increased communication overhead
- Key idea
  - Follow a dynamic E-schedule for the number of local adversarial training epochs at each round
  - Use FedCurve as fusion algorithm

[1] Shah D, Dube P, Chakraborty S, et al. Adversarial training in communication constrained federated learning[J]. arXiv preprint arXiv:2103.01319, 2021.

- Evaluation of varying E-schedule
  - Initial value E_0, decay rate γ_{E} and decay frequency F_{E}



Improved performance in both natural and adversarial accuracy with FedDynAT with a smooth drop in E compared to FedCurvAT with fixed E

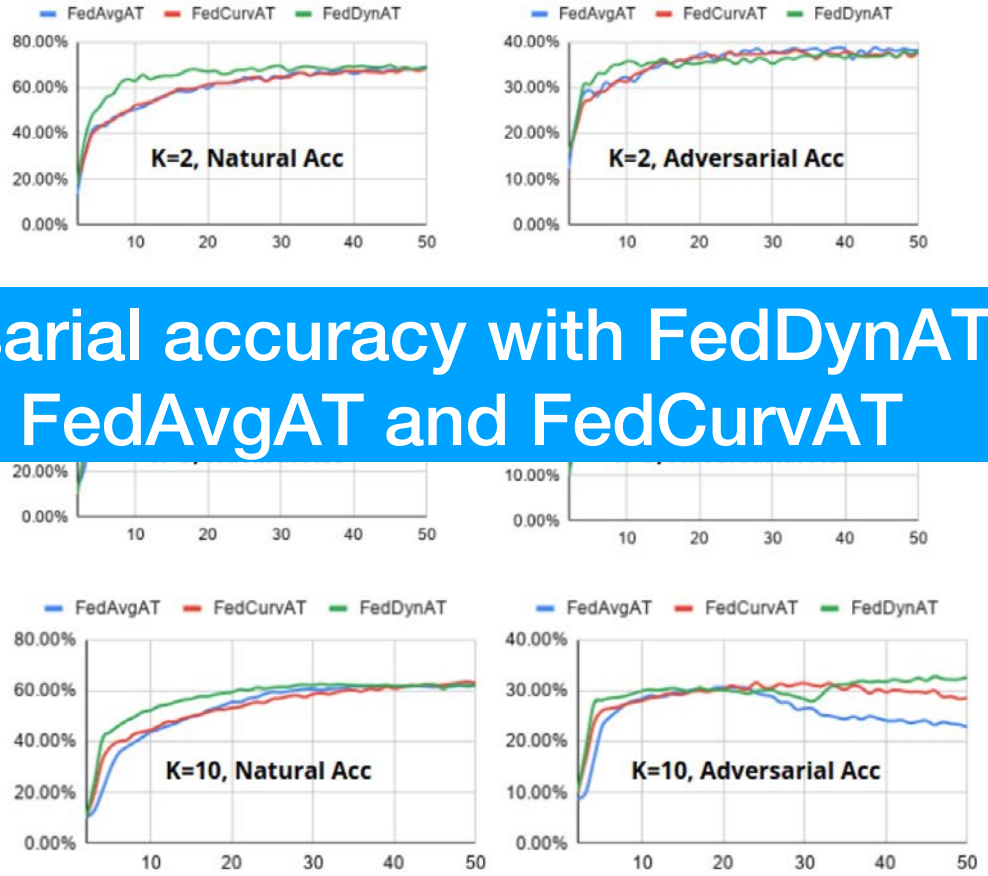


[1] Shah D, Dube P, Chakraborty S, et al. Adversarial training in communication constrained federated learning[J]. arXiv preprint arXiv:2103.01319, 2021.

FedDynAT

Experiment

- Accuracy with different communication budgets



The adversarial accuracy with FedDynAT is significantly better than FedAvgAT and FedCurvAT

[1] Shah D, Dube P, Chakraborty S, et al. Adversarial training in communication constrained federated learning[J]. arXiv preprint arXiv:2103.01319, 2021.

Solution 5: Analysis

- Adversarial training adds adversarial samples to actual training data to improve privacy
- Key challenge
  - Computation power and extra training time for adversarial samples

Comparison

- Approaches and associated cost to enhance privacy preservation in FL

Approach	Methodology	Cost
Secure Multi-party Computation	Encrypt uploaded parameters	Efficiency loss due to encryption
Differential Privacy	Add random noise to uploaded parameters	Accuracy loss due to added noise in clients' models
Hybrid	Encrypt the manipulated parameter	Subdued cost on both efficiency and accuracy
VerifyNet	Double-masking protocol; Verifiable aggregation results	Communication overhead
Adversarial Training	Include adversarial samples in training data	Computation power, training time for adversarial samples

[1] Mothukuri V, Parizi R M, Pouriyeh S, et al. A survey on security and privacy of federated learning[J]. Future Generation Computer Systems, 2021, 115: 619-640.

Tutorial Outline

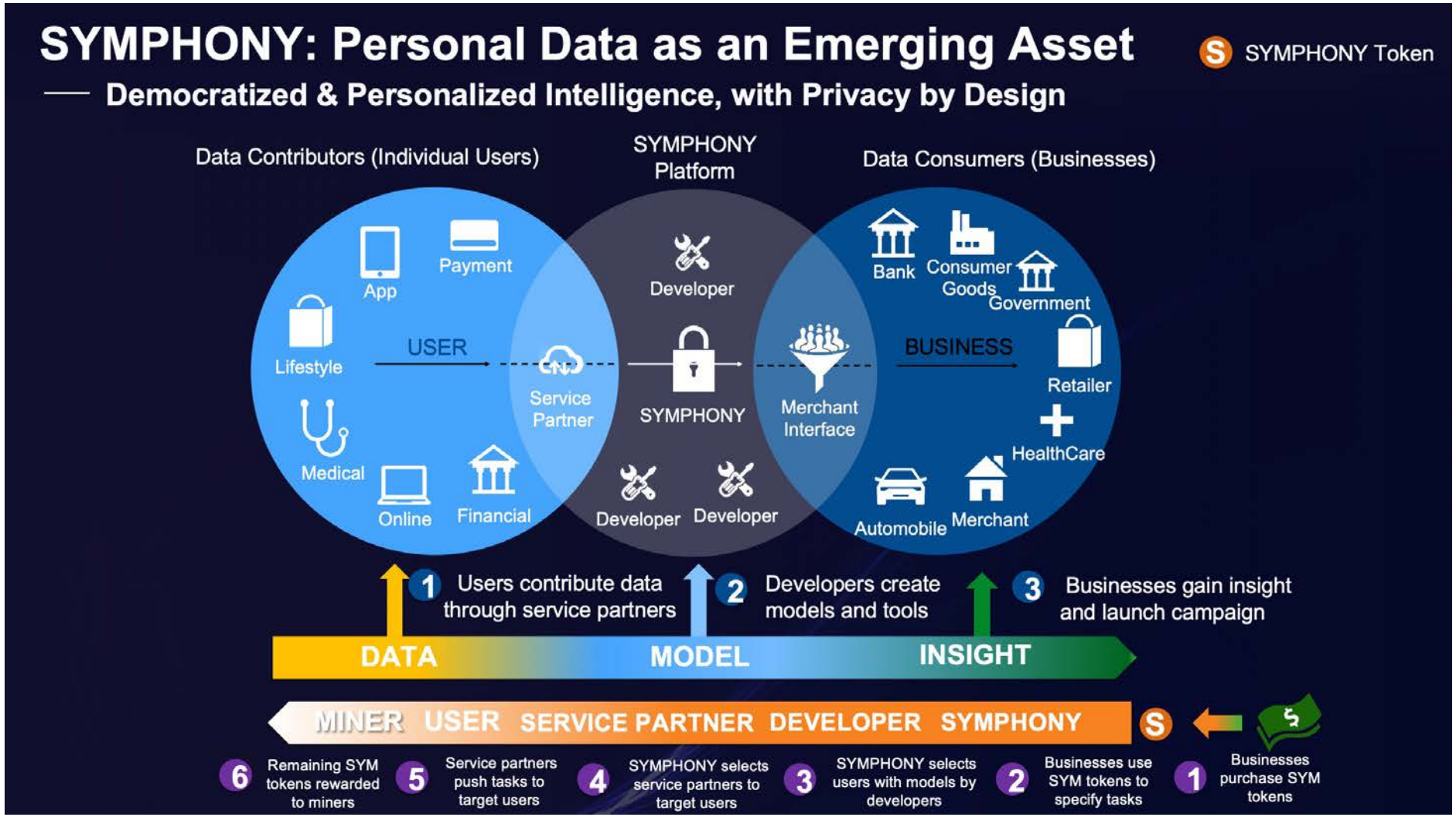
- Data Asset: What and Why
- Data Asset Core Components
- Data Asset Governance for Decentralized Collaborative Intelligence
  - Governance principles
  - "Trust" for data asset governance for decentralized collaborative intelligence
  - "Incentive" for data asset governance for decentralized collaborative intelligence
    - Data pricing (covered in Part A)
    - Value allocation model (not covered in this tutorial)
    - Tokenomics design (not covered in this tutorial)
- Data Asset Ecosystems
- Challenges and Future Directions

Tutorial Outline

- Data Asset: What and Why
- Data Asset Core Components
- Data asset governance for decentralized collaborative intelligence
- Data Asset Ecosystems
  - Case Study: Personal Data as Emerging Asset Class
  - Case Study: B-to-B Data Sharing and Exchange
- Challenges and Future Directions

Data Asset Ecosystem

- Case Study: Personal Data as Emerging Asset Class



Data Asset Ecosystem

- Case Study: B-to-B Data Sharing and Exchange

Study on data sharing between companies in Europe

The study focuses on sharing and/or re-using of machine-generated data as a key priority of data sharing. Such type of data, "created without the direct intervention of a human by computer processes, applications and services or by sensors"

- Data generated by the Internet-of-Things (IoT) and physical devices, including sensors or mobile phones
- Data generated by internal IT business systems, mainly containing information about products, services, sales, logistics and customers, partners or suppliers (CRM24, ERP25, etc.)
- Data generated through users' interaction with websites (i.e. cookies, web tracking, logs), which contain information about a user's behaviour on a particular website or when surfing the web, about his/her interests and preferences, etc.
- Data generated through crowdsourcing or web collaboration.

https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

Data Asset Ecosystem

- Case Study: B-to-B Data Sharing and Exchange

Study on data sharing between companies in Europe

Data monetisation

- ✓ Unilateral approach to share data
- ✓ Generate additional revenues
- ✓ Add value to services provided

https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

Data Asset Ecosystem

- Case Study: B-to-B Data Sharing and Exchange

Study on data sharing between companies in Europe

Data marketplaces

- ✓ Trusted intermediary between data suppliers and data users
- ✓ Data suppliers sell their data to interested data users
- ✓ Revenue is generated from each data transaction

DAWEX
SELL BUY SHARE DATA

Figure 50. Main characteristics and example of a data marketplace

https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

Data Asset Ecosystem

- Case Study: B-to-B Data Sharing and Exchange

Study on data sharing between companies in Europe

Industrial Data Platform

- ✓ Strategic and collaborative partnerships
- ✓ Mutual benefits for all parties
- ✓ Data shared (for free) in a closed, exclusive and secure environment
- ✓ Develop new or improved products and/or services
- ✓ Enhance internal performance

AIRBUS MAN

Figure 51. Main characteristics and examples of industrial data platforms

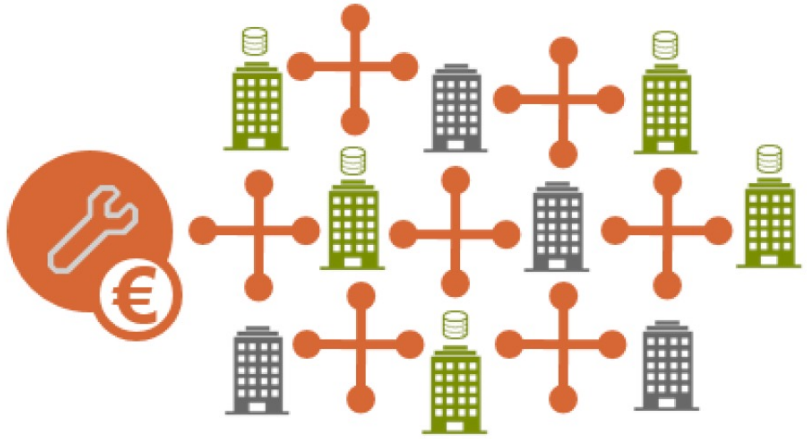
https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

Data Asset Ecosystem

- Case Study: B-to-B Data Sharing and Exchange

Study on data sharing between companies in Europe

Technical Enabler



- ✓ Businesses specialised in and specifically dedicated to enabling data sharing through a technical solution
- ✓ Fee is requested for setting up, using and/or maintaining the solution
- ✓ Revenue is not generated from data shared

SENSATIVE API-AGRO nallian DKE DATA

Figure 52. Main characteristics and examples of technical enablers

https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

Data Asset Ecosystem

- Case Study: B-to-B Data Sharing and Exchange

Study on data sharing between companies in Europe

Open Data Policy

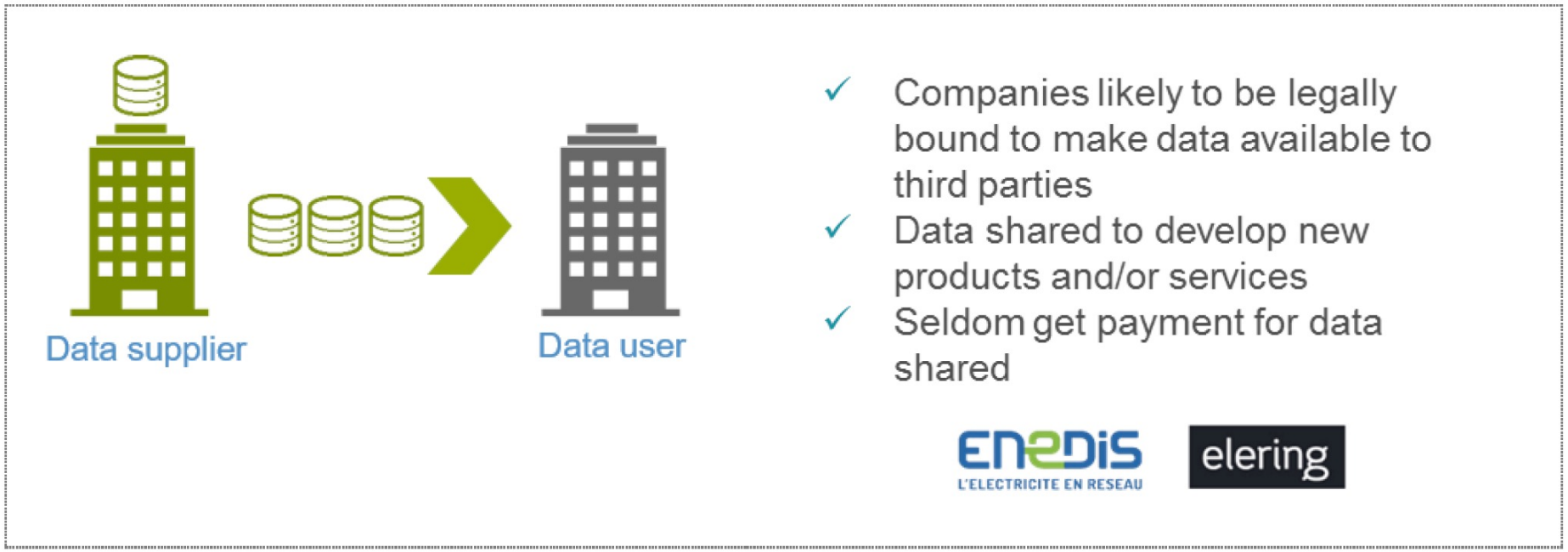


Figure 53. Main characteristics and examples of companies that follow an open data policy

https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

Tutorial Outline

- Data Asset: What and Why
- Data Asset Core Components
- Data asset governance for decentralized collaborative intelligence
- Data Asset Ecosystems
- Challenges and Future Directions

Challenges and Future Directions

- Data asset ready awareness
- Data control
- Data security
- Data asset technology
- Data asset law and regulation
- Trends in Data Asset Governance
  - Regulation---More control on data privacy and security
  - Usufruct---From corporate monopoly to individual ownership
  - Distribution---From centralized platforms to distributed devices
  - Heterogeneity --- Unstructured data and standards-based framework

The model-specific

1. Ahmed Salem, Yang Zhang, Mathias Humbert, Pascal Berrang, Mario Fritz, Michael Backes. **ML-Leaks: Model and Data Independent Membership Inference Attacks and Defenses on Machine Learning Models**. NDSS 2019.
2. Milad Nasr, Reza Shokri, Amir Houmansadr. **Comprehensive Privacy Analysis of Deep Learning: Passive and Active White-box Inference Attacks against Centralized and Federated Learning**. IEEE Symposium on Security and Privacy 2019: 739-753
3. Samuel Yeom, Irene Giacomelli, Matt Fredrikson, Somesh Jha. **Privacy Risk in Machine Learning: Analyzing the Connection to Overfitting**. CSF 2018: 268-282
4. Liwei Song, Prateek Mittal. **Systematic Evaluation of Privacy Risks of Machine Learning Models**. CoRR abs/2003.10595 (2020)
5. Choquette-Choo C A, Tramer F, Carlini N, et al**. Label-only membership inference attacks**. International Conference on Machine Learning. PMLR, 2021: 1964-1974.
6. Zheng Li, Yang Zhang. **Label-Leaks: Membership Inference Attack with Label**. CoRR abs/2007.15528 (2020)
7. Yunhui Long, Lei Wang, Diyue Bu, Vincent Bindschaedler, XiaoFeng Wang, Haixu Tang, Carl A. Gunter, Kai Chen: **A Pragmatic Approach to Membership Inferences on Machine Learning Models**. EuroS&P 2020: 521-534.
8. Alexandre Sablayrolles, Matthijs Douze, Cordelia Schmid, Yann Ollivier, Hervé Jégou. **White-box vs Black-box: Bayes Optimal Strategies for Membership Inference**. ICML 2019: 5558-5567
9. Bo Hui, Yuchen Yang, Haolin Yuan, Philippe Burlina, Neil Zhenqiang Gong, Yinzhi Cao. **Practical Blind Membership Inference Attack via Differential Comparisons**. NDSS 2021.
10. Bargav Jayaraman, Lingxiao Wang, Katherine Knipmeyer, Quanquan Gu, David Evans. **Revisiting Membership Inference Under Realistic Assumptions**. Proc. Priv. Enhancing Technol. 2021(2): 348-368 (2021)

The shadow-training

11. C. Song, T. Ristenpart, and V. Shmatikov**. Machine learning models that remember too much**. In CCS, 2017.
12. R. Shokri, M. Stronati, C. Song, and V. Shmatikov. **Membership inference attacks against machine learning models**. In S&P, 2017.
13. C. Song, V. Shmatikov. **Auditing Data Provenance in Text-Generation Models**. KDD 2019: 196-206
14. M. A. Rahman et al. **Membership inference attack against differentially private deep learning model**. Transactions on Data Privacy, 11(1):61–79, 2018.
15. Truex S, Liu L, Gursoy M E, et al. **Demystifying membership inference attacks in machine learning as a service**. IEEE Transactions on Services Computing, 2019.
16. Shadi Rahimian, Tribhuvanesh Orekondy, Mario Fritz. **Sampling Attacks: Amplification of Membership Inference Attacks by Repeated Queries**. CoRR abs/2009.00395 (2020)
17. Hisamoto S, Post M, Duh K. **Membership inference attacks on sequence-to-sequence models: Is my data in your machine translation system?**. Transactions of the Association for Computational Linguistics, 2020, 8: 49-63.
18. Yang He, Shadi Rahimian, Bernt Schiele, Mario Fritz. Segmentations-Leak. **Membership Inference Attacks and Defenses in Semantic Image Segmentation**. ECCV (23) 2020: 519-535
19. Yang Zou, Zhikun Zhang, Michael Backes, Yang Zhang. Privacy Analysis of Deep Learning in the Wild. **Membership Inference Attacks against Transfer Learning**. CoRR abs/2009.04872 (2020)
20. Seng Pei Liew, Tsubasa Takahashi**. FaceLeaks: Inference Attacks against Transfer Learning Models via Black-box Queries**. CoRR abs/2010.14023 (2020)
21. Iyiola E. Olatunji, Wolfgang Nejdl, Megha Khosla. **Membership Inference Attack on Graph Neural Networks**. CoRR abs/2101.06570 (2021)
22. Xinlei He, Rui Wen, Yixin Wu, Michael Backes, Yun Shen, Yang Zhang. **Node-Level Membership Inference Attacks Against Graph Neural Networks**. CoRR abs/2102.05429 (2021)

The DA analysis

23. Klas Leino, Matt Fredrikson:Stolen Memories. **Leveraging Model Memorization for Calibrated White-Box Membership Inference**. USENIX Security Symposium 2020: 1605-1622.
24. Stacey Truex, Ling Liu, Mehmet Emre Gursoy, Wenqi Wei, Lei Yu. **Effects of Differential Privacy and Data Skewness on Membership Inference Vulnerability**. TPS-ISA 2019: 82-91
25. Shahbaz Rezaei and Xin Liu. **On the Difficulty of Membership Inference Attacks**. arXiv:2005.13702, 2020.
26. Yugeng Liu, Rui Wen, Xinlei He, Ahmed Salem, Zhikun Zhang, Michael Backes, Emiliano De Cristofaro, Mario Fritz, Yang Zhang. **ML-Doctor: Holistic Risk Assessment of Inference Attacks Against Machine Learning Models**. CoRR abs/2102.02551 (2021)
27. Liwei Song, Reza Shokri, Prateek Mittal. **Membership Inference Attacks Against Adversarially Robust Deep Learning Models**. IEEE Symposium on Security and Privacy Workshops 2019: 50-56
28. Avital Shafran, Shmuel Peleg, Yedid Hoshen. **Reconstruction-Based Membership Inference Attacks are Easier on Difficult Problems**. CoRR abs/2102.07762 (2021)
29. Reza Shokri, Martin Strobel, and Yair Zick. 2020. **Exploiting Transparency Measures for Membership Inference: a Cautionary Tale**. In AAAI Workshop on Privacy-Preserving Artificial Intelligence (PPAI).

Generative models-based

30. J. Hayes, L. Melis, G. Danezis, and E. De Cristofaro. LOGAN: Membership inference attacks against generative models. In PETS, 2019.
31. Benjamin Hilprecht, Martin Härterich, Daniel Bernau. Monte Carlo and Reconstruction Membership Inference Attacks against Generative Models. Proc. Priv. Enhancing Technol. 2019(4): 232-249 (2019)
32. Kin Sum Liu, Chaowei Xiao, Bo Li, Jie Gao. Performing Co-membership Attacks Against Deep Generative Models. ICDM 2019: 459-467
33. Dingfan Chen, Ning Yu, Yang Zhang, Mario Fritz. GAN-Leaks: A Taxonomy of Membership Inference Attacks against Generative Models. CCS 2020: 343-362

Other references

30. Pang Wei Koh, Percy Liang: Understanding Black-box Predictions via Influence Functions. ICML 2017: 1885-1894
31. Bourtoule L, Chandrasekaran V, Choquette-Choo C, et al. Machine unlearning. arXiv preprint arXiv:1912.03817, 2019.
32. Baumhauer T, Schöttle P, Zeppelzauer M. Machine unlearning: Linear filtration for logit-based classifiers. arXiv preprint arXiv:2002.02730, 2020.
33. Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nick Hynes, Nezihe Merve Gürel, Bo Li, Ce Zhang, Dawn Song, Costas J. Spanos: Towards Efficient Data Valuation Based on the Shapley Value. AISTATS 2019: 1167-1176.
34. Zhu, Ligeng, et al. Deep Leakage from Gradients, NeurIPS 2019.
35. J. Jia, A. Salem, M. Backes, Y. Zhang, and N. Z. Gong, “MemGuard: Defending against Black-Box Membership Inference Attacks via Adversarial Examples,” in Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security - CCS ’19
36. Yang Z, Shao B, Xuan B, et al. Defending model inversion and membership inference attacks via prediction purification[J]. arXiv preprint arXiv:2005.03915, 2020.
37. Jiacheng Li, Ninghui Li, Bruno Ribeiro: Membership Inference Attacks and Defenses in Classification Models. CODASPY 2021: 5-16

Reference - Privacy

- [1] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial intelligence and statistics. PMLR, 2017: 1273-1282.
- [2] Aledhari M, Razzak R, Parizi R M, et al. Federated learning: A survey on enabling technologies, protocols, and applications[J]. IEEE Access, 2020, 8: 140699-140725.
- [3] Li T, Sahu A K, Talwalkar A, et al. Federated learning: Challenges, methods, and future directions[J]. IEEE Signal Processing Magazine, 2020, 37(3): 50-60.
- [4] Wu Y, Cai S, Xiao X, et al. Privacy Preserving Vertical Federated Learning for Tree-based Models[J]. Proceedings of the VLDB Endowment, 13(11).
- [5] Mothukuri V, Parizi R M, Pouriyeh S, et al. A survey on security and privacy of federated learning[J]. Future Generation Computer Systems, 2021, 115: 619-640.