NETWORKS

Tsung-Ting (Tim) Kuo

Reference: Mankowitz S, editor. Clinical informatics board review and self assessment. Springer; 2018 Feb 8.

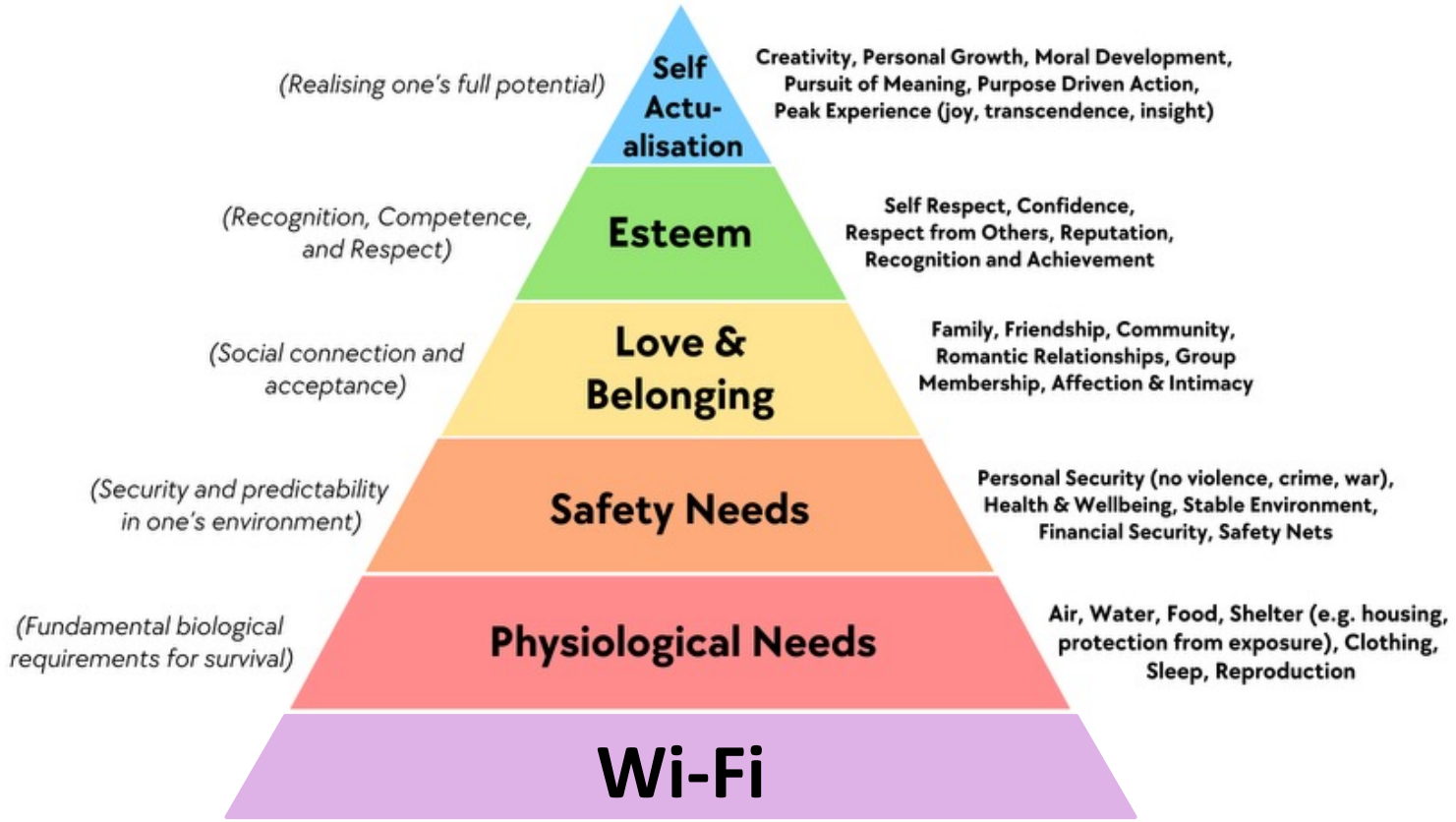
HOUSEKEEPING

- Assignments are due 11:59pm every Tuesday

PollEv.com/renawu484

MASLOW'S HIERARCHY OF NEEDS

Also important for establishing the interoperability of health informatic systems



Adapted from: https://en.wikipedia.org/wiki/Maslow's_hierarchy_of_needs#/media/File:Maslow's_Hierarchy_of_Needs_Diagram.png

TODAY

General Networks

- Layers
- Topologies
- Types
- Architectures

Blockchain Networks

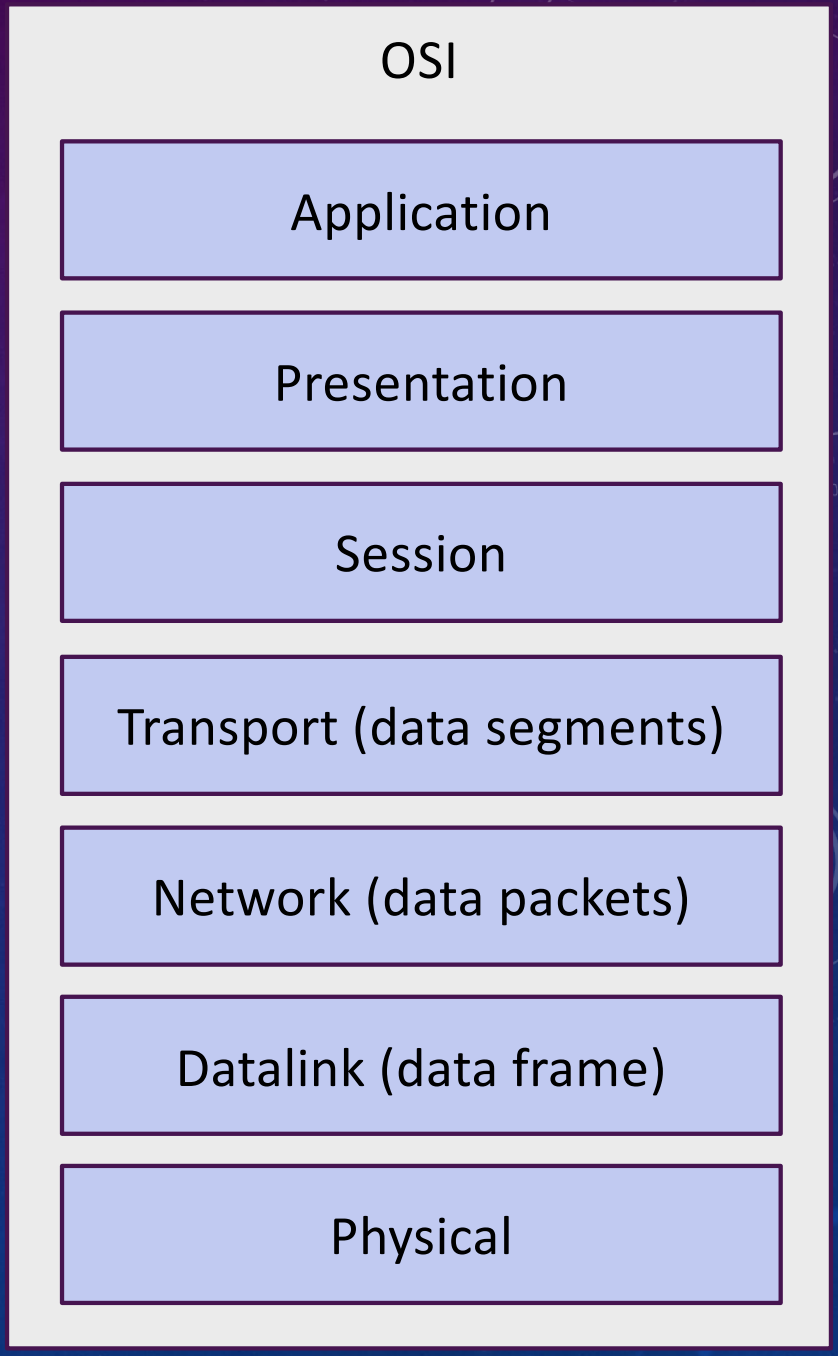
- Basics
- Healthcare

LAYERS

OSI
(OPEN SYSTEM INTERCONNECT)

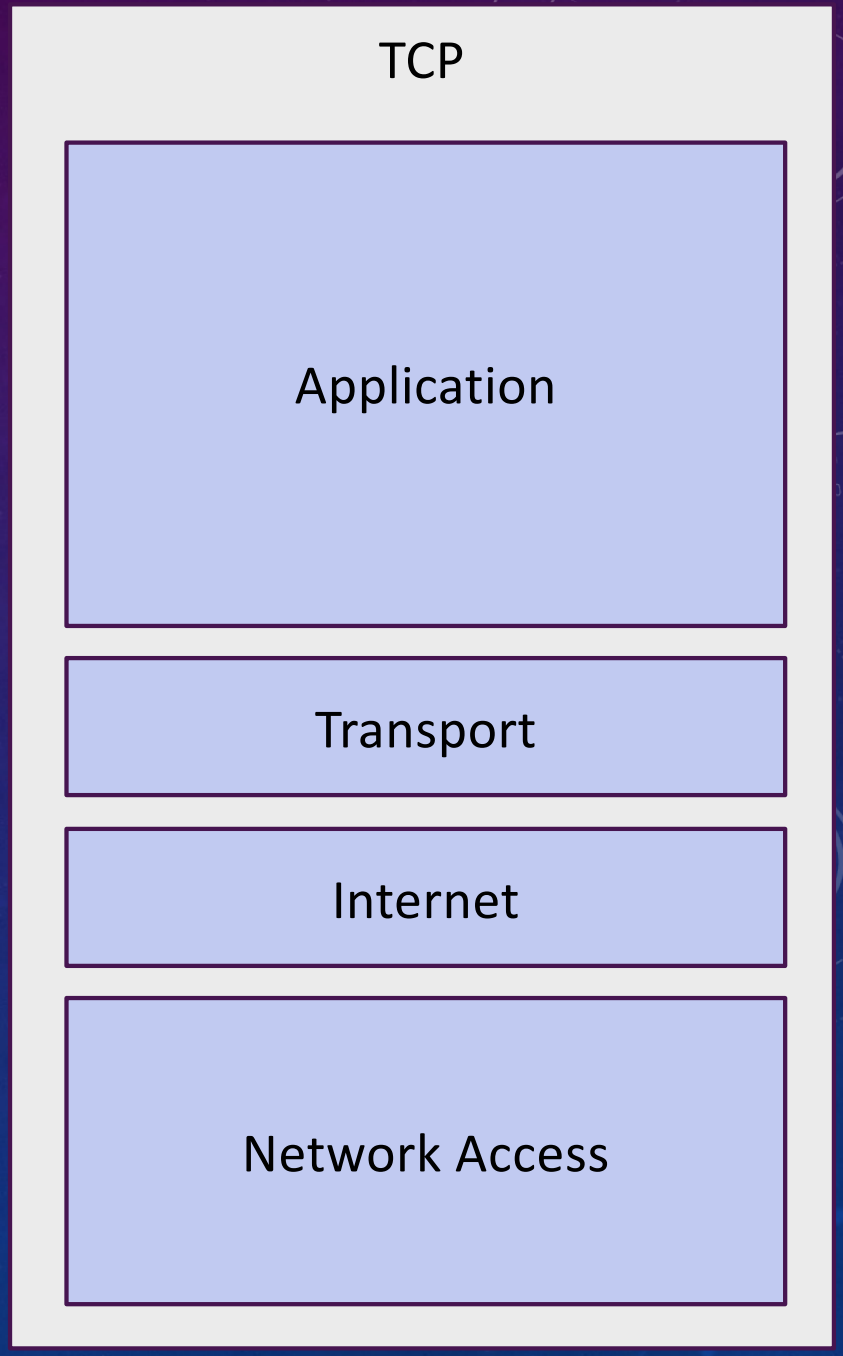
- Developed by the International Organization for Standardization (ISO) in 1970s
- Also called 7-layer model
- Example protocols
  - Layer 7: HL7
  - Layer 2: MAC (Medium Access Control)
  - Layer 1: RJ45

Reference: https://en.wikipedia.org/wiki/OSI_model
https://commons.wikimedia.org/wiki/File:Ethernet_RJ45_connector_p1160054.jpg



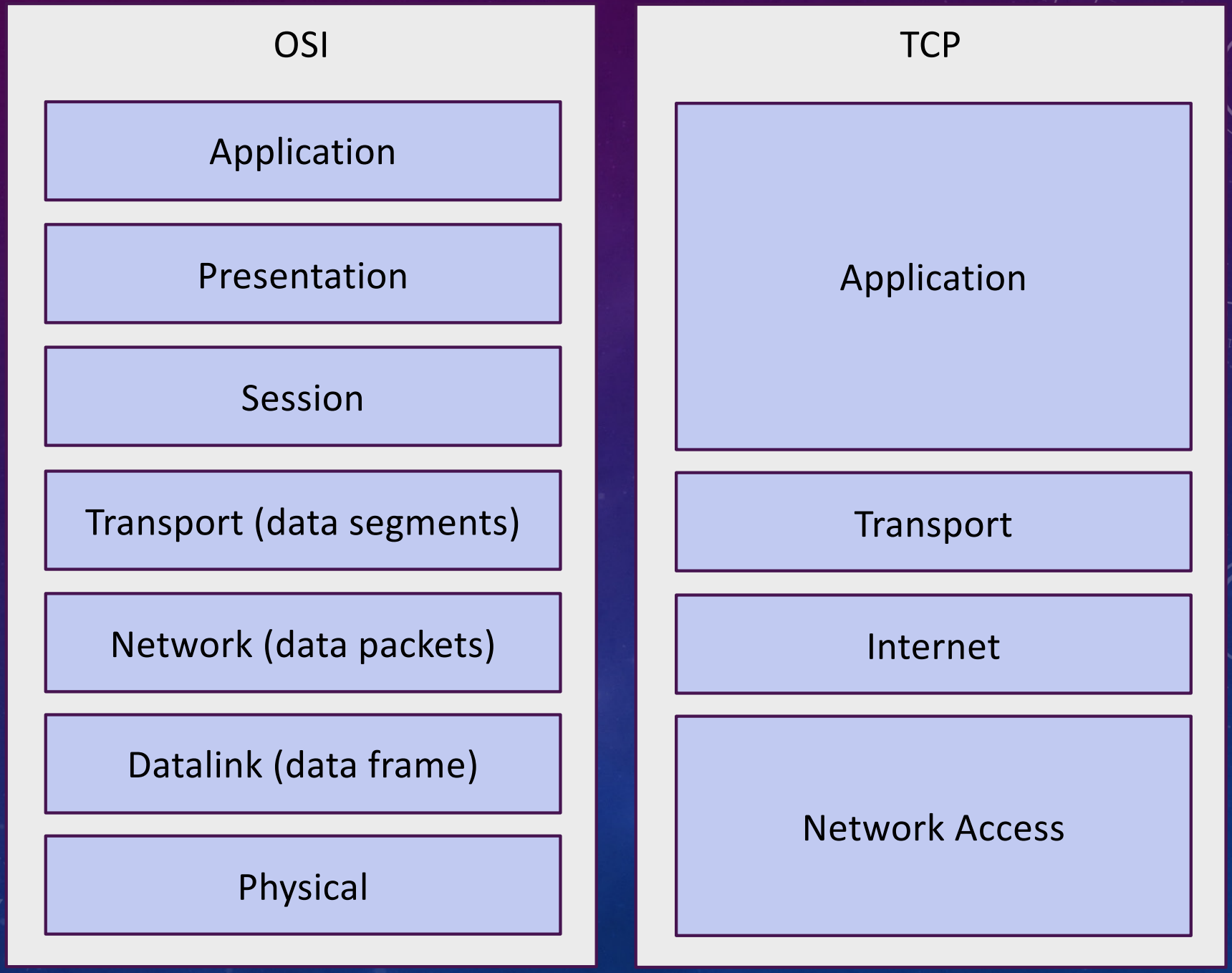
TCP/IP

- Development was supported by the U.S. Department of Defense (DoD) as ARPANET in 1960s
- Winner of "protocol wars" in 1970s - 1990s
- Example protocols
  - HL7
  - Transmission Control Protocol (TCP) or User Datagram Protocol (UDP)
  - Internet Protocol (IP)
  - MAC / RJ45



Reference: https://en.wikipedia.org/wiki/Internet_protocol_suite
https://en.wikipedia.org/wiki/Protocol_Wars

COMPARISON


OSI
Application
Presentation
Session
Transport (data segments)
Network (data packets)
Datalink (data frame)
Physical
TCP
Application
Transport
Internet
Network Access

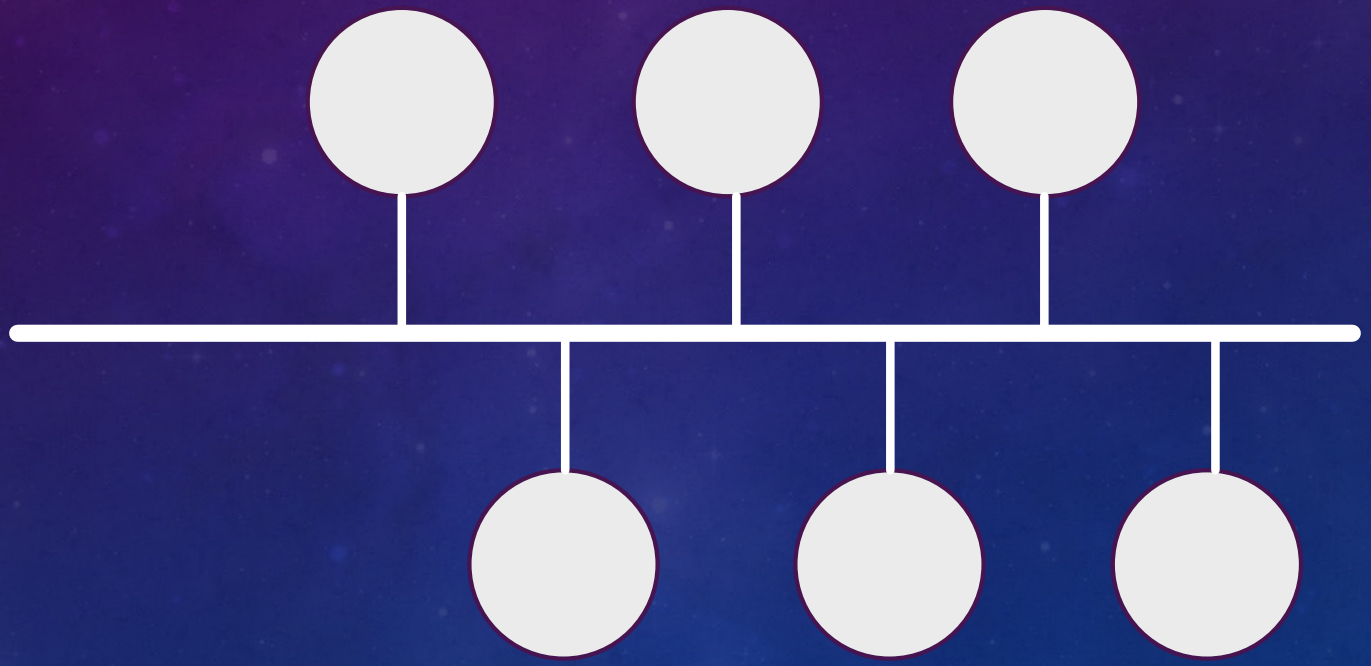
TOPOLOGIES

POINT-TO-POINT

- Simplest
- Limited use cases

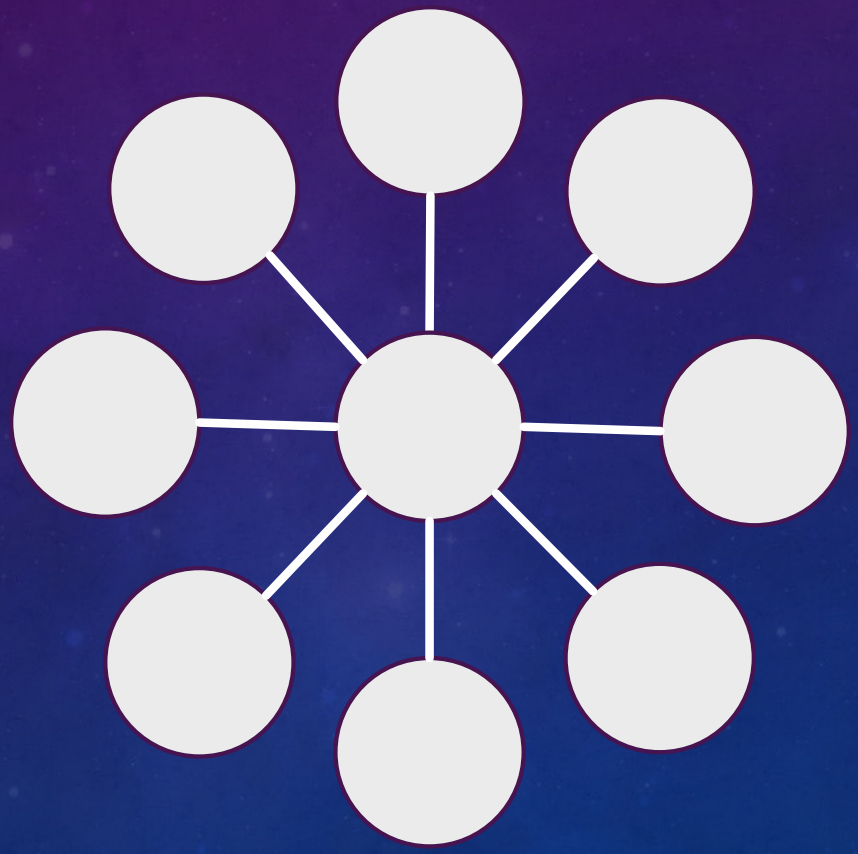
BUS

- Relatively cheap and easy to add nodes
- The bus (or backbone) presents a single-point-of-failure (SPOF)



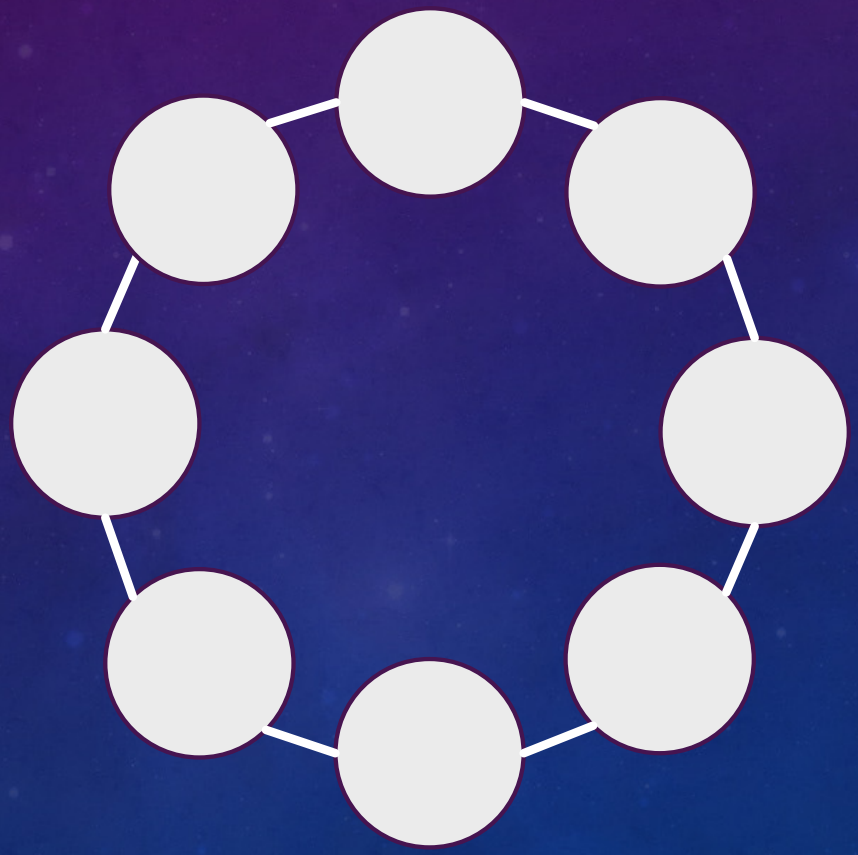
STAR

- Relatively easy to add nodes
- The hub also presents a SPOF



RING

- Uni-directional – less bandwidth usage yet every node is SPOF
- Bi-directional – no SPOF (becomes a daisy-chain) yet higher bandwidth usage

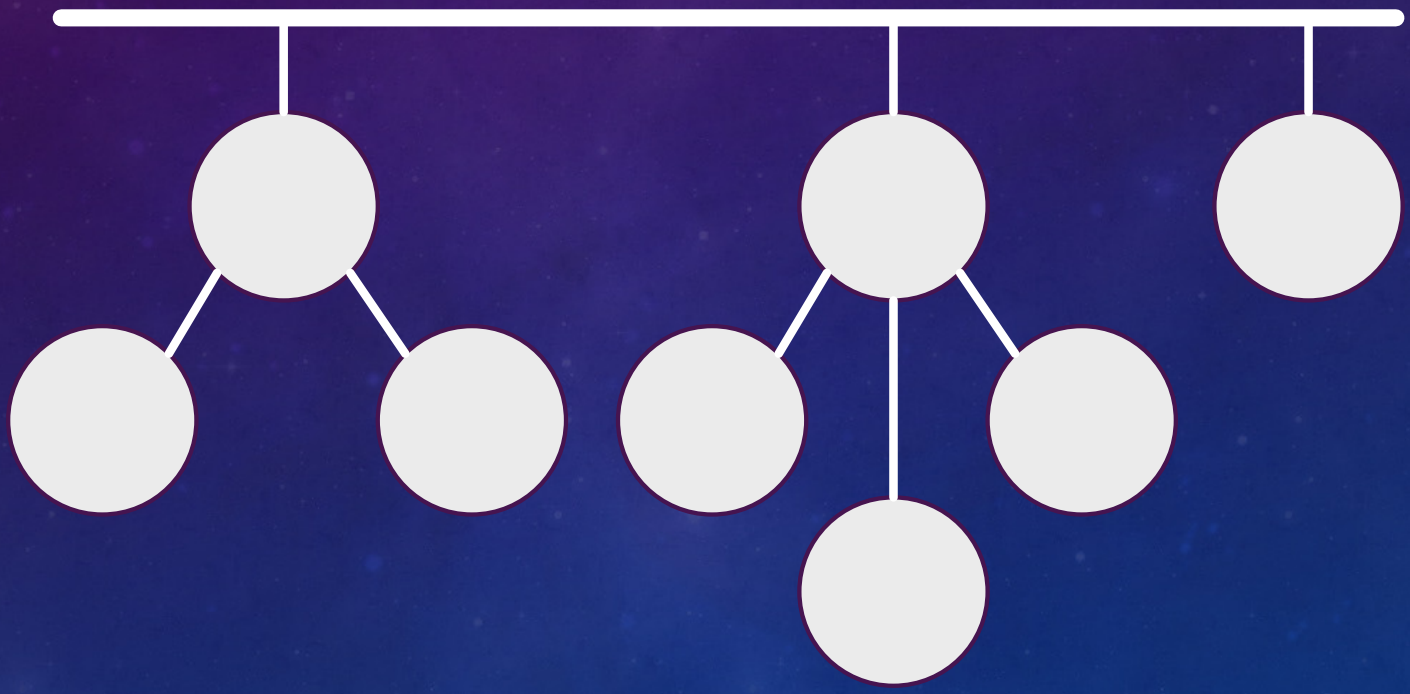


MESH

- Fully-connected – no SPOF with highest efficiency/security yet very high cost
- Partially-connected (at least 2 neighbors/node) – no SPOF with lower cost

HYBRID: TREE

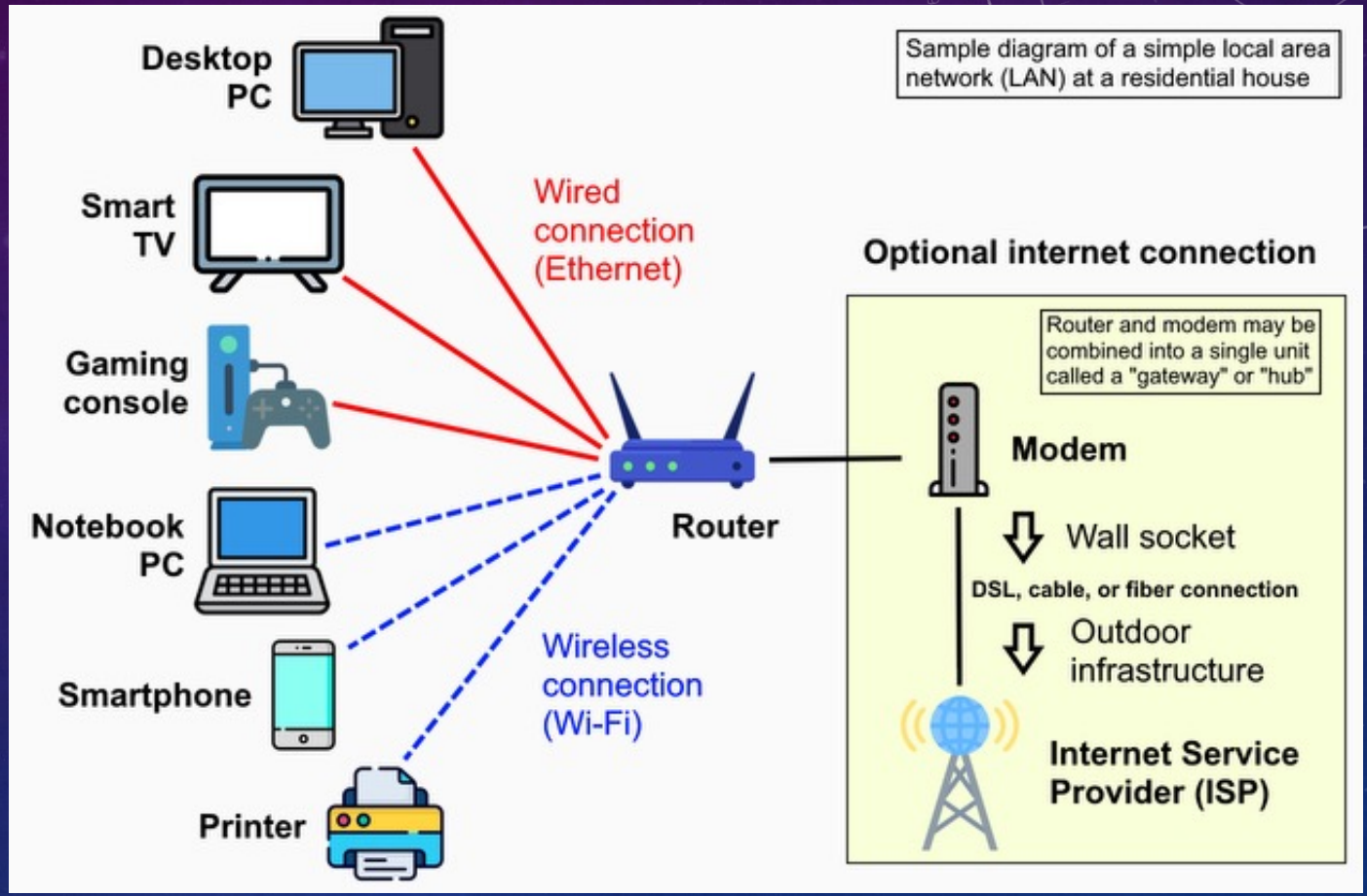
- Departments in large organizations



TYPES

LOCAL AREA NETWORK (LAN)

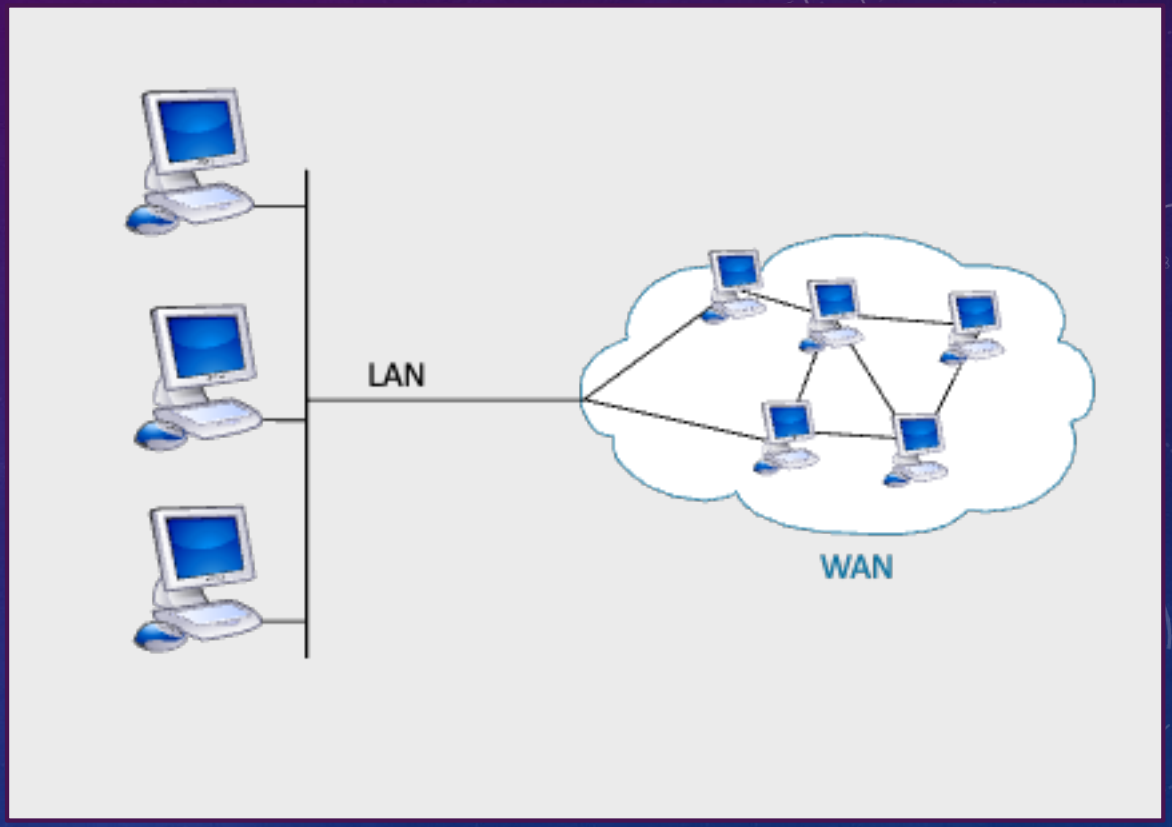
- Limited area
  - Residence
  - Campus
  - Building
- Example
  - Ethernet
  - Wi-Fi
- May use various topologies (e.g., star)



Reference: https://en.wikipedia.org/wiki/Network_topology
https://en.wikipedia.org/wiki/Local_area_network
https://en.wikipedia.org/wiki/Local_area_network#/media/File:Home_LAN_local_area_network_example_diagram.png

WIDE AREA NETWORK (WAN)

- Larger geographic area
  - May also use various topologies
  - Can be public (e.g., Internet)
  - Can also be private (e.g., many Yale and YNHHS websites)
- Private WANs can be bridged via a VPN (Virtual Private Network)

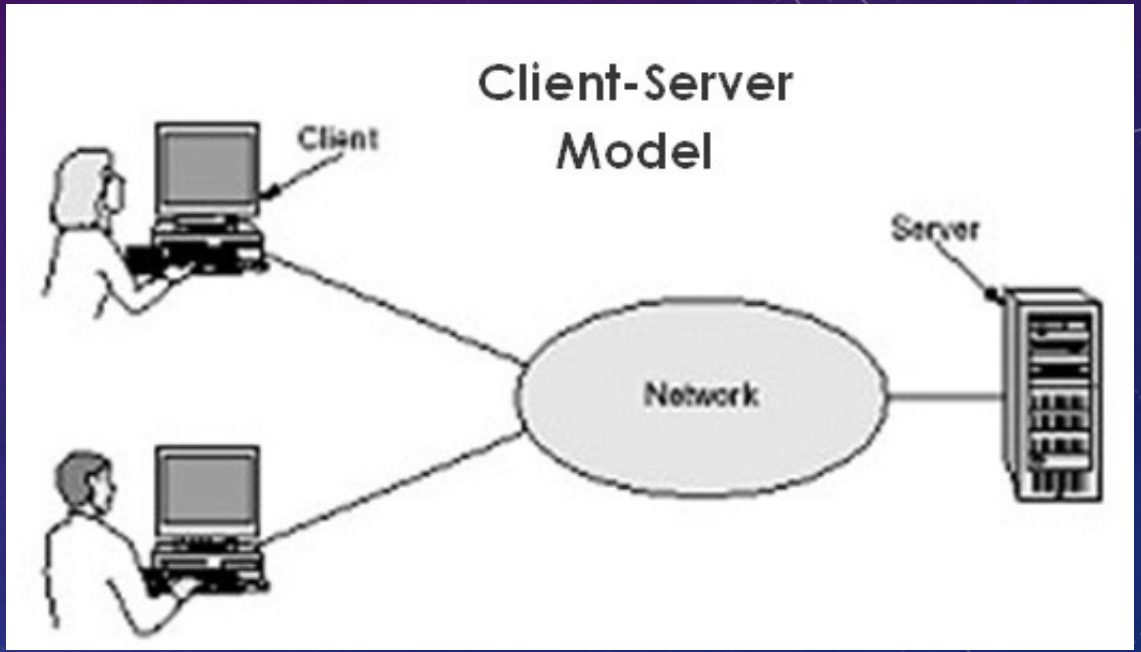


Reference: https://en.wikipedia.org/wiki/Network_topology
https://en.wikipedia.org/wiki/Wide_area_network
https://en.wikipedia.org/wiki/Wide_area_network#/media/File:LAN_WAN_scheme.svg

ARCHITECTURES

CLIENT-SERVER (TWO-TIER)

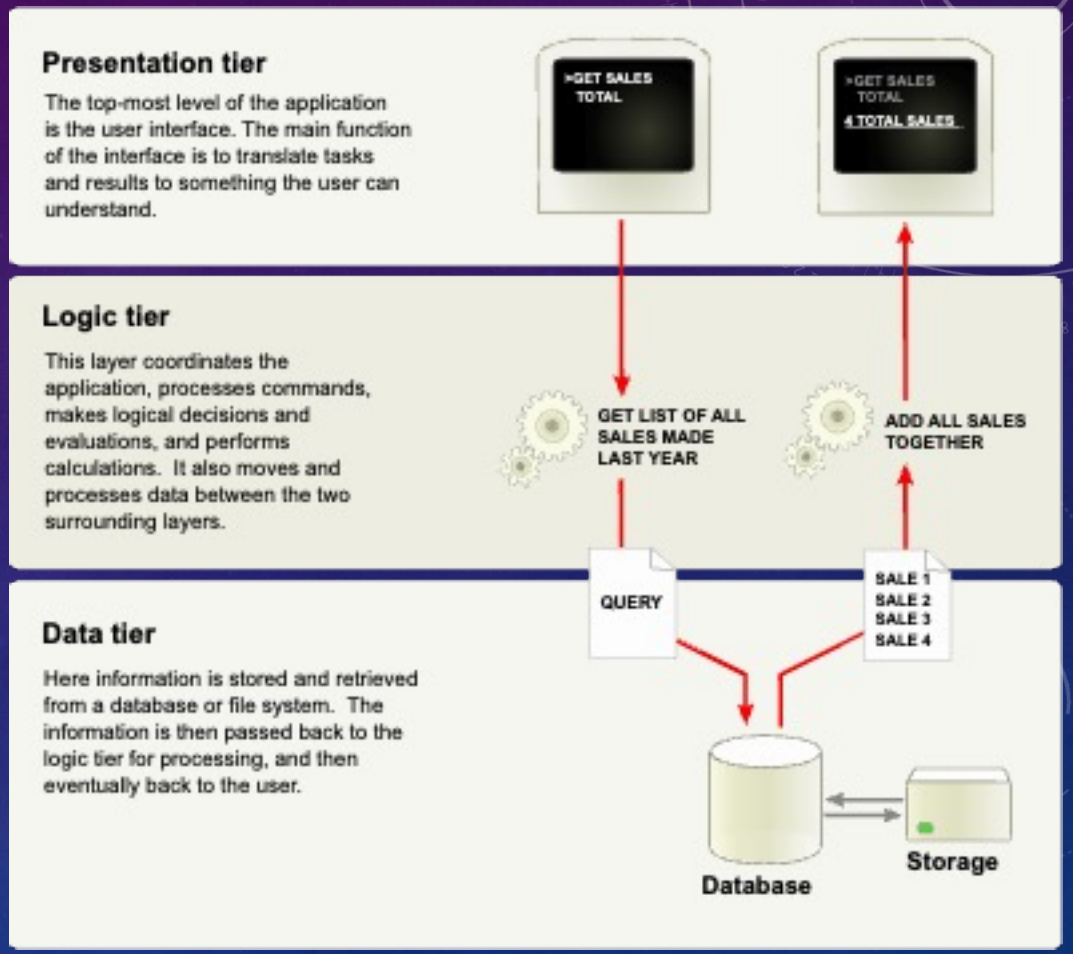
- Application Architecture
  - Underlying topology and type may vary
- One-to-many
  - Client/server can even be on the same device
  - There could be inter-server communications
- Client could be
  - Thin (e.g., cloud)
  - Thick (e.g., local computing)
  - Ultrathin (e.g., remote desktop)



Reference: https://en.wikipedia.org/wiki/Client–server_model
https://commons.wikimedia.org/wiki/File:Client-Server_Model.jpg

THREE-TIER

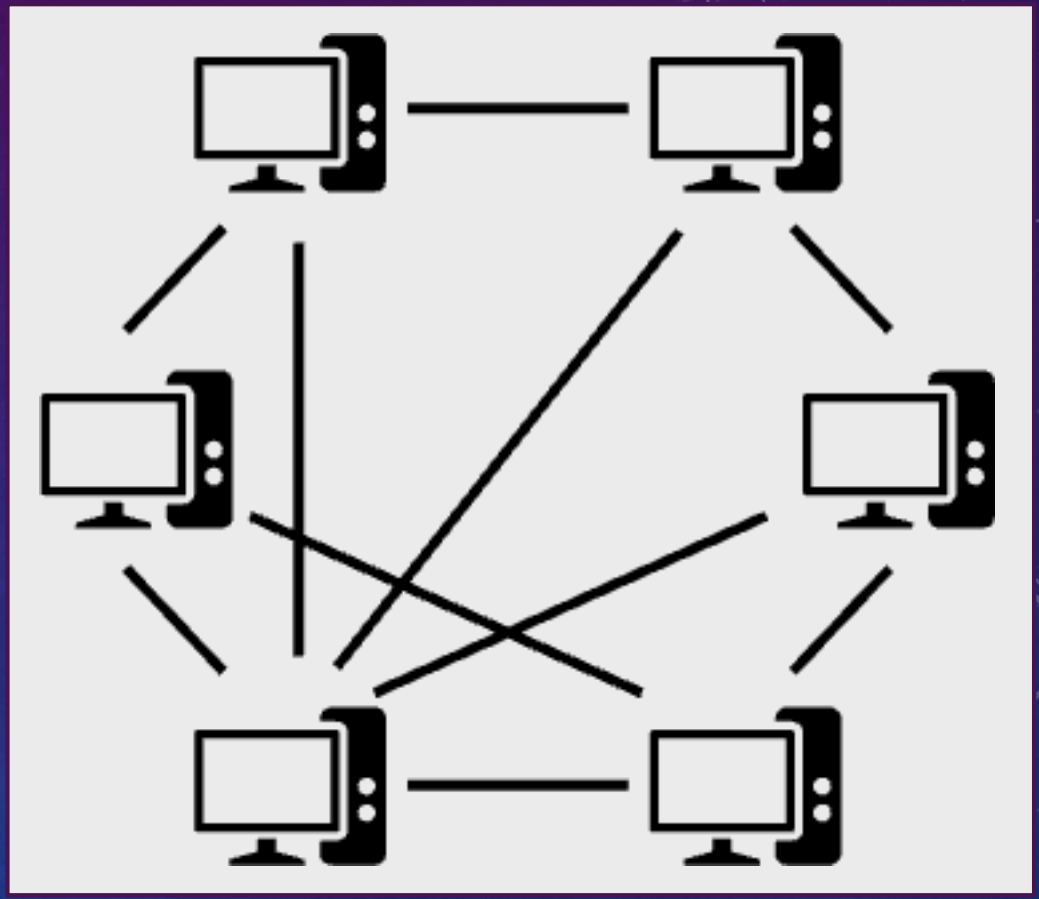
- Common for web and software design
  - Graphical User Interface (GUI) / View
  - Business logic / Control
  - Database / Model
- Each layer may be developed/tested independently
- Could be more tiers (i.e., n-tier or multi-tier)



Reference: https://en.wikipedia.org/wiki/Multitier_architecture#Three-tier_architecture
https://en.wikipedia.org/wiki/Multitier_architecture#/media/File:Overview_of_a_three-tier_application_vectorVersion.svg

PEER-TO-PEER

- No SPOF
  - Each node is an "equal" peer
  - Each node functions both as clients and servers to other nodes
- Examples
  - Napster for music sharing in 1999
  - BitTorrent for file sharing in 2001
  - Blockchain for crypto-currency in 2009

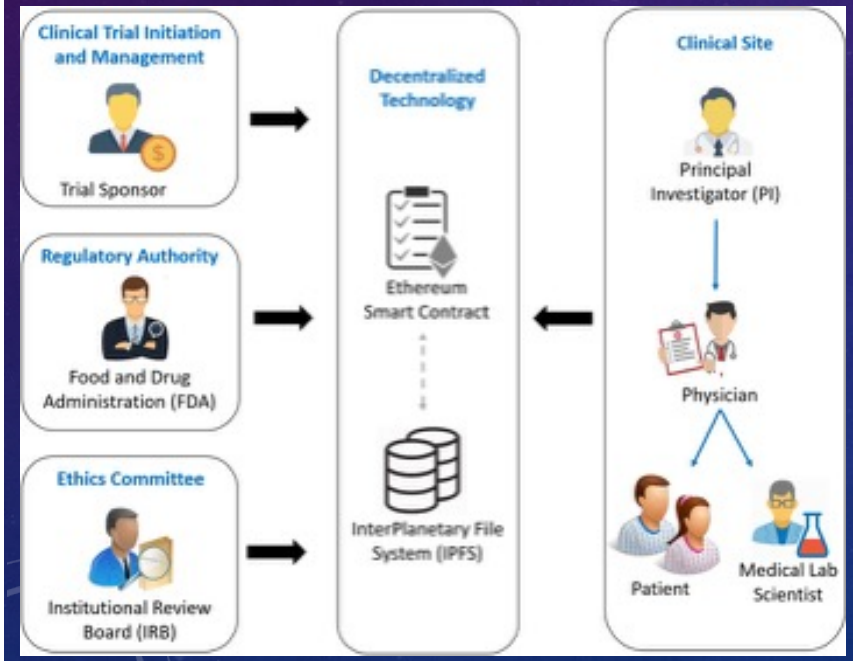


Reference: https://en.wikipedia.org/wiki/Peer-to-peer
https://en.wikipedia.org/wiki/Peer-to-peer#/media/File:P2P_network.svg

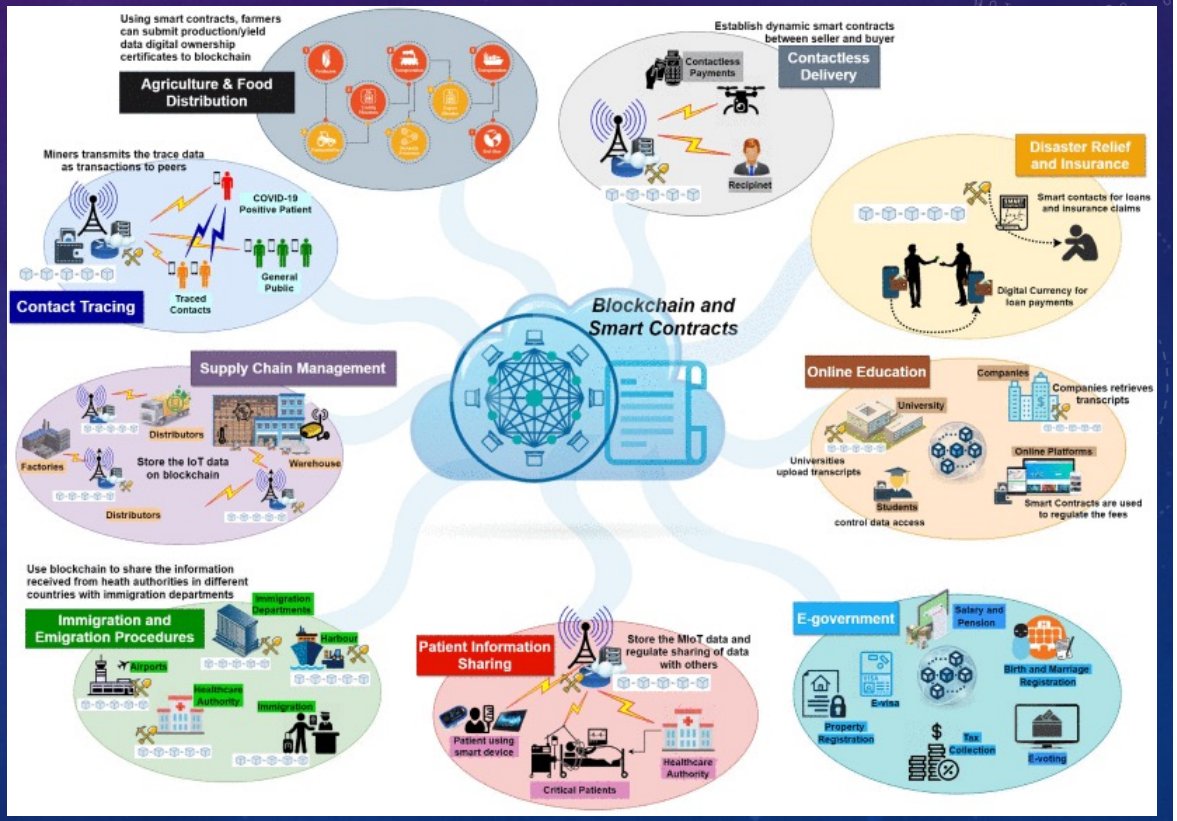
BLOCKCHAIN BASICS

BLOCKCHAIN

- The underlying technology of crypto-currencies
  - Different coins have different implementations (e.g., Bitcoin, Ethereum, …)
- A new way to share healthcare data
  - Clinical Trials
  - COVID-19
  - …



[Omar et al.]

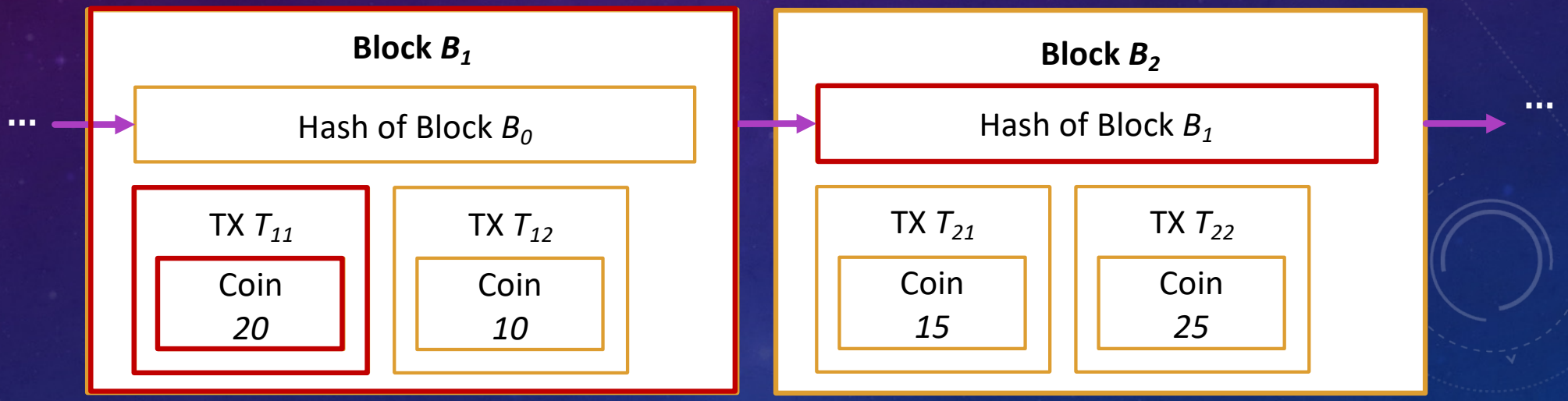


[Kalla et al.]

BLOCKCHAIN (CONT.)

Example Hash Functions
- **123 mod 10 = 3**
- **4256 mod 10 = 6**

- A new form of ledger to store data
  - Data are recorded in transactions (TXs)
  - TXs are enclosed in blocks
  - Blocks are chained using hashes to order TXs (and thus data)



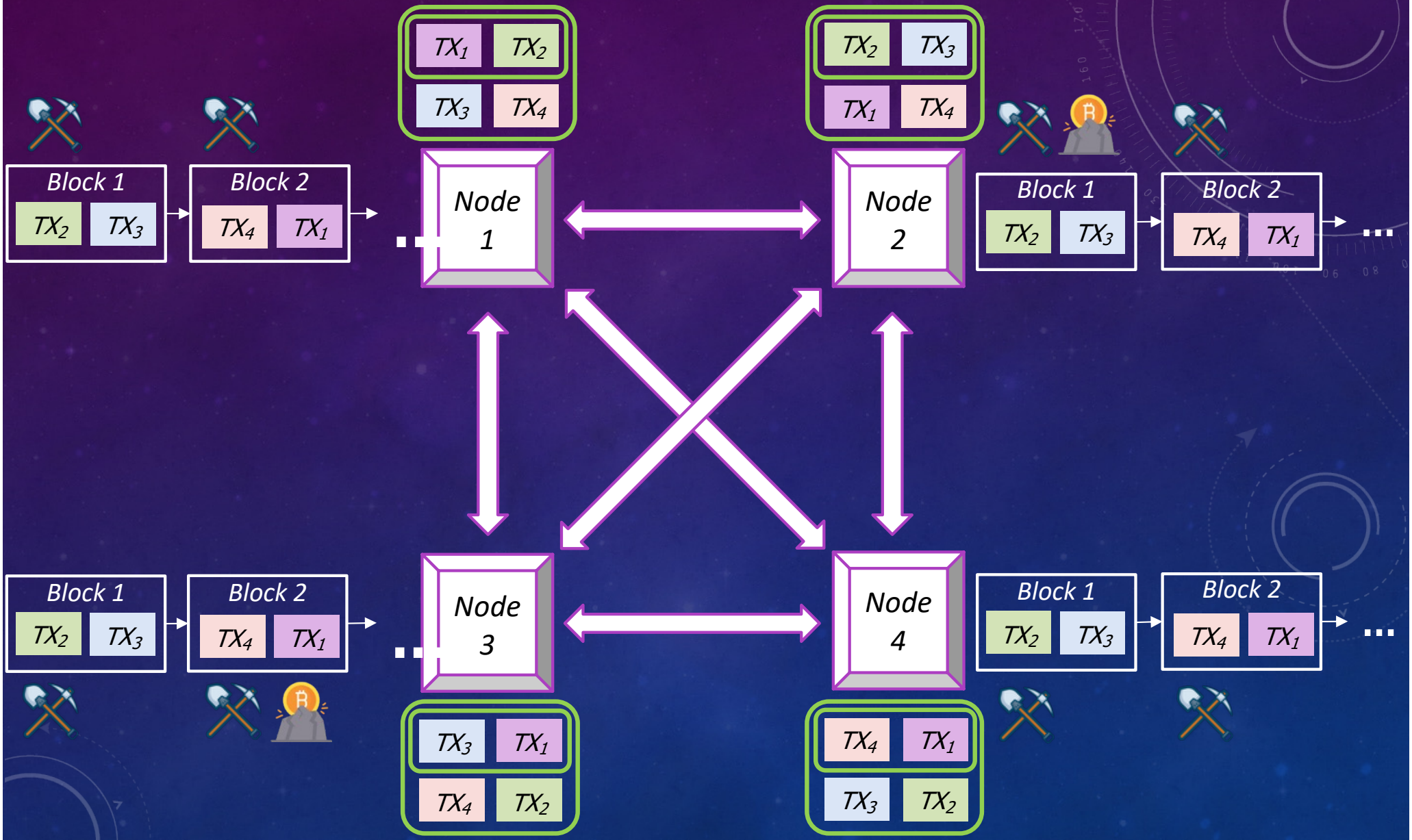
Ensuring the order of blocks and thus of TXs

Kuo T-T et al. ModelChain: Decentralized Privacy-Preserving Healthcare Predictive Modeling Framework on Private Blockchain Networks. Use of Blockchain for Healthcare and Research Workshop, 2016.

THE BITCOIN BLOCKCHAIN

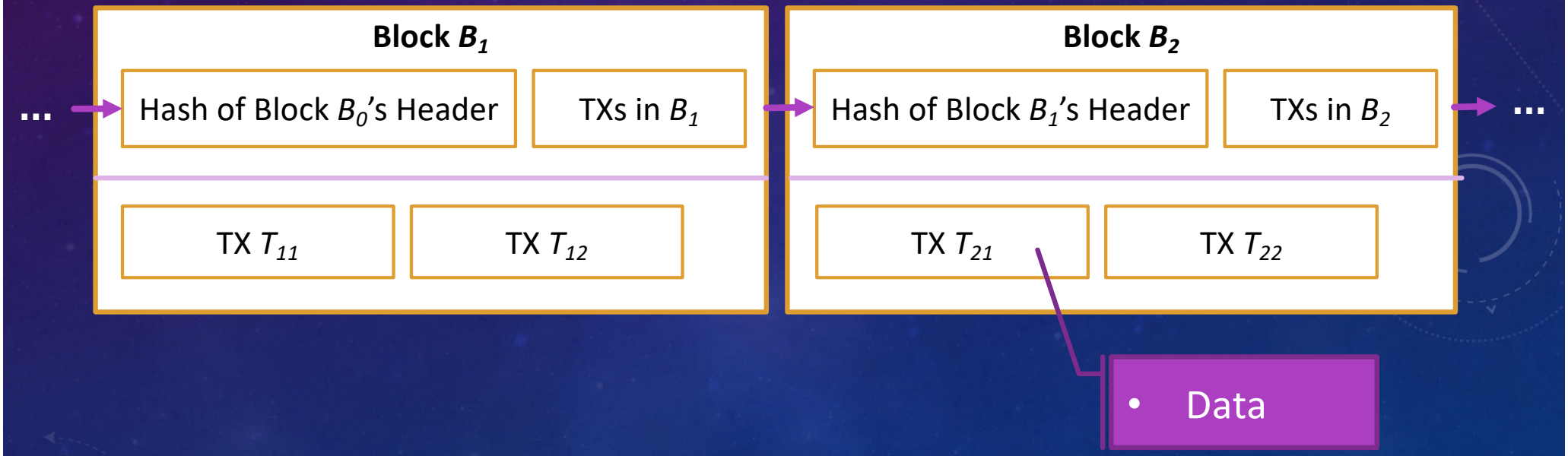
CB&B 7400

Simplified Overview [Nakamoto]



BEYOND CRYPTO-CURRENCY

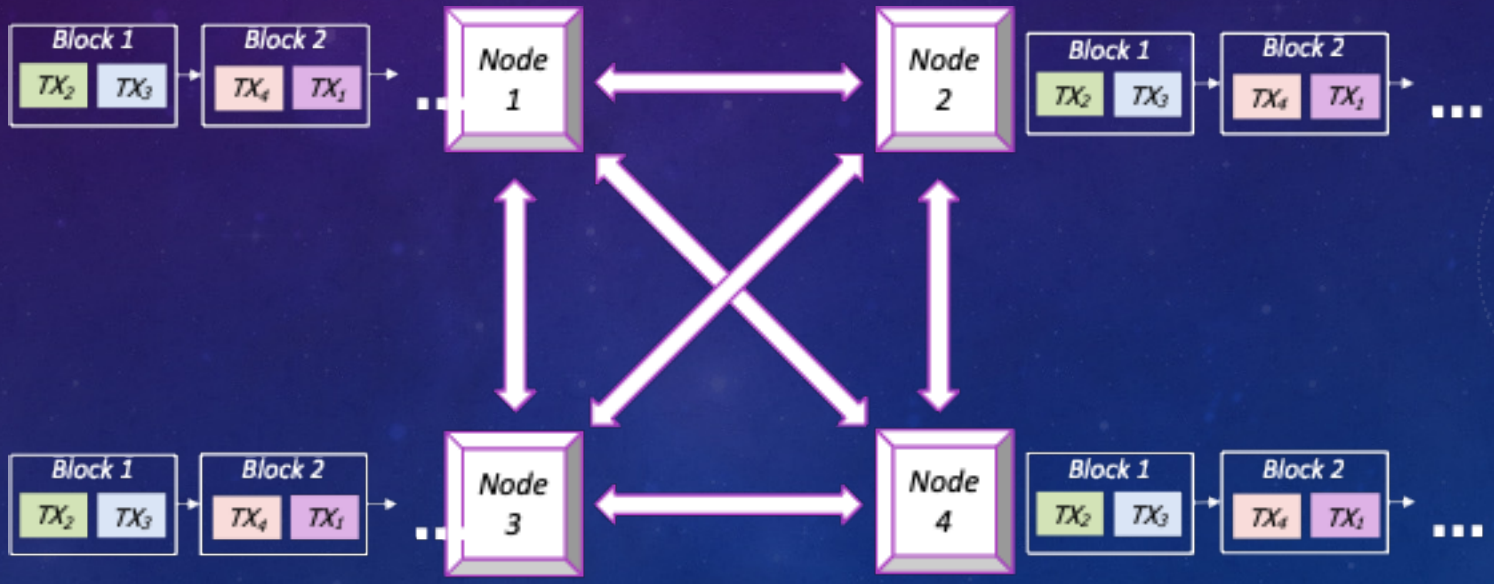
- Blockchain is also a new form of distributed ledger
  - Store data using transaction metadata
- Non-financial applications
  - Either permission-less or permissioned blockchain networks



Kuo T-T et al. Blockchain distributed ledger technologies for biomedical and health care applications. Journal of the American Medical Informatics Association (JAMIA). 2017

BLOCKCHAIN FOR HEALTHCARE

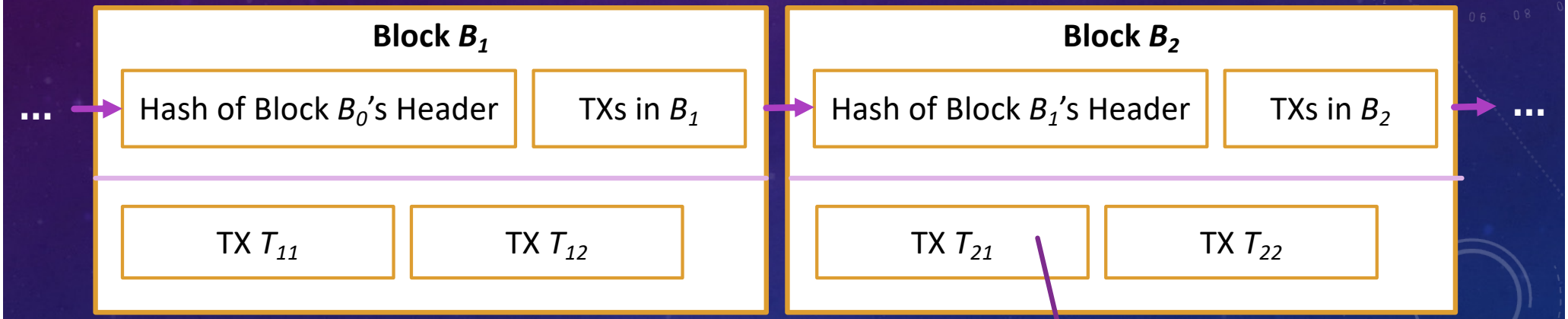
- Benefits compared to traditional distributed databases
  - Decentralized management
  - Immutable audit trail
  - Data provenance
  - Robustness/availability
  - Security/privacy



Kuo T-T et al. Blockchain distributed ledger technologies for biomedical and health care applications. Journal of the American Medical Informatics Association (JAMIA). 2017

SMART CONTRACT

- Store both data and program on chain
  - Code transparency
  - Code immutability
  - Code provenance



- Decentralized management
- Immutable audit trail
- Data provenance
- Robustness/availability
- Security/privacy

- Data
- Program

BLOCKCHAIN FOR HEALTHCARE

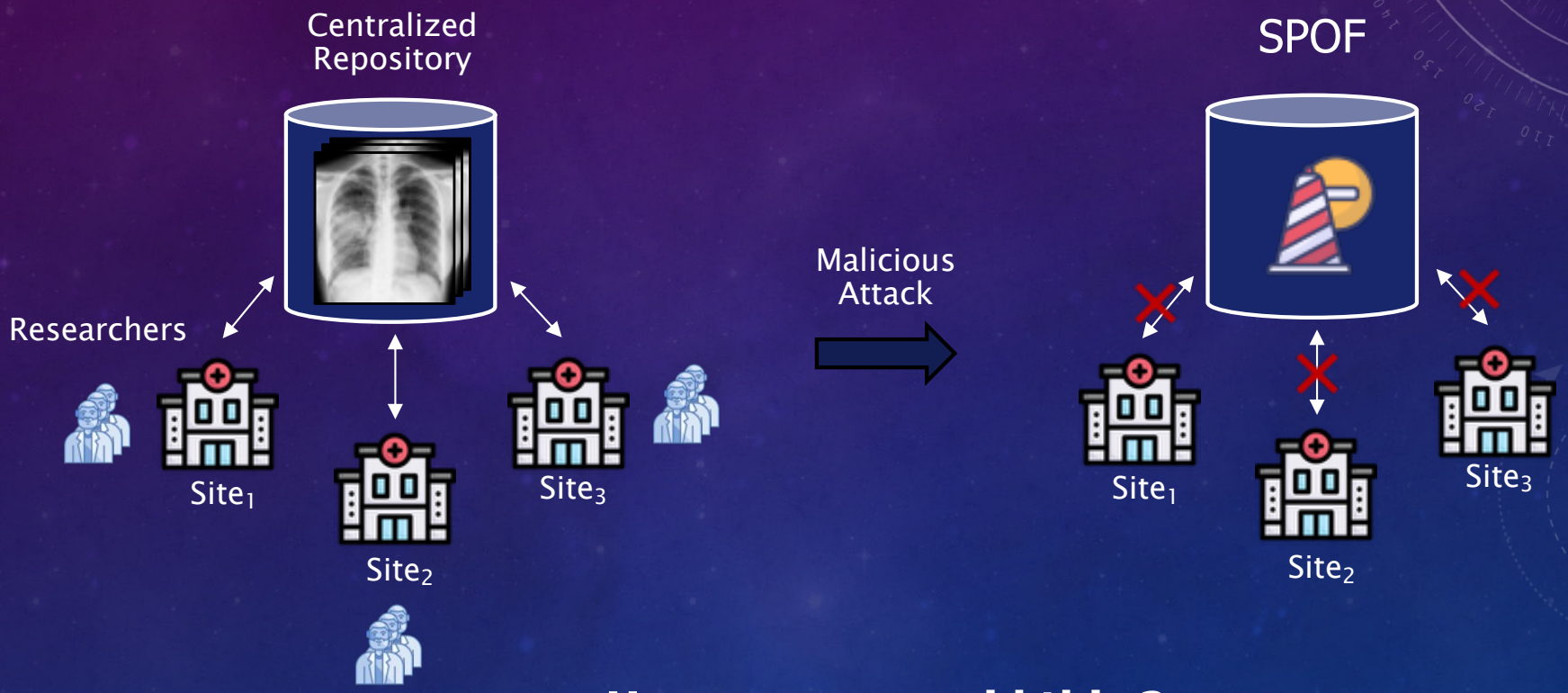
MEDICAL IMAGE SHARING

- COVID-19 data sharing between healthcare institutions
  - Aid downstream analyses that can find prevention methods and treatments
- COVID-19 chest x-ray image
  - Support imaging reviews to determine prognostic COVID-19 pneumonia features
  - Facilitate machine/deep learning for COVID-19 detection
- A need for medical image sharing between institutions
  - Require a trustworthy data exchanging framework

How about an intuitive solution ?

Image Sources:
Cohen, J.P., et al., Covid-19 image data collection: prospective predictions are the future, 2020
Cohen, J.P., P. Morrison, and L. Dao Covid-19 image data collection. 2020

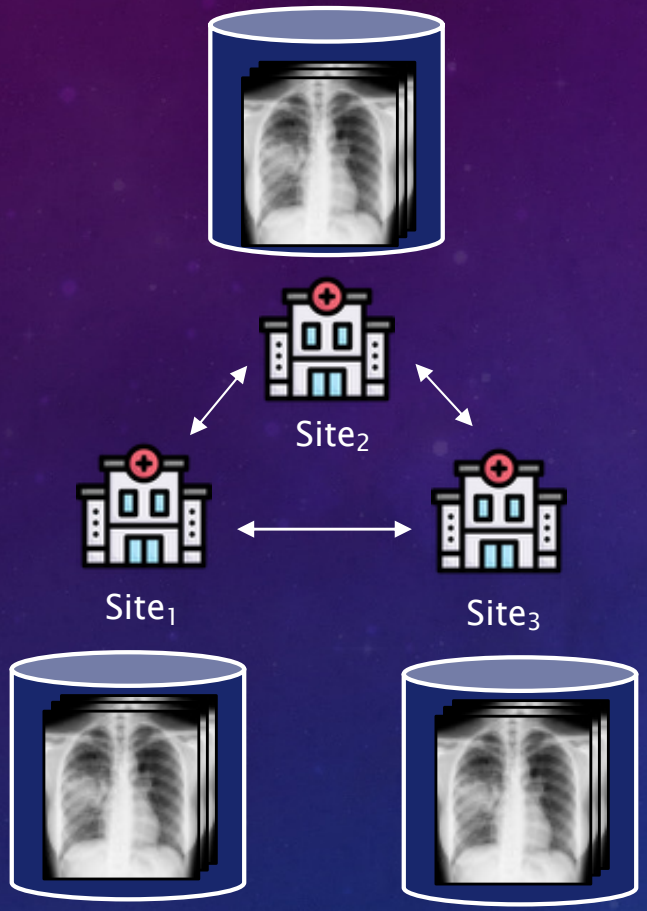
RISKS FOR CENTRALIZED SOLUTION



How can we avoid this ?

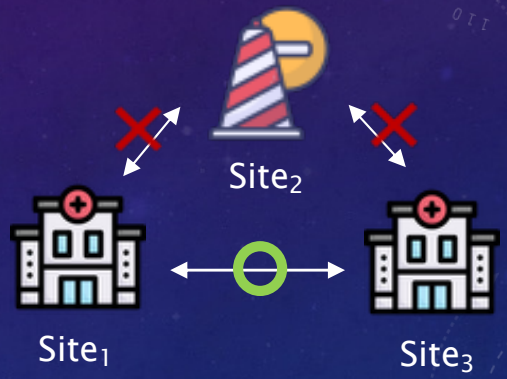
Adapted From: Mun Li M, **Kuo T-T**. Previewable Contract-Based On-Chain X-Ray Image Sharing Framework for Clinical Research. International Journal of Medical Informatics. 2021:104599. doi: 10.1016/j.ijmedinf.2021.104599. PubMed PMID: 34628257

DECENTRALIZED ARCHITECTURE



Malicious Attack

No SPOF

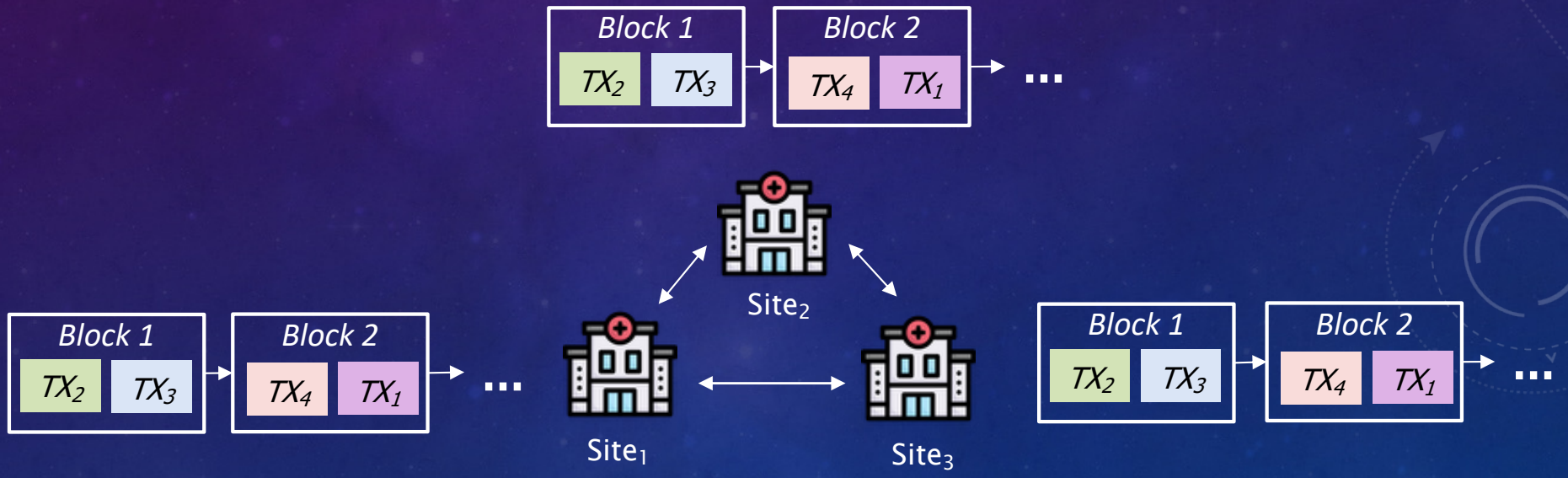


Which technology to adopt ?

Adapted From: Mun Li M, **Kuo T-T**. Previewable Contract-Based On-Chain X-Ray Image Sharing Framework for Clinical Research. International Journal of Medical Informatics. 2021:104599. doi: 10.1016/j.ijmedinf.2021.104599. PubMed PMID: 34628257

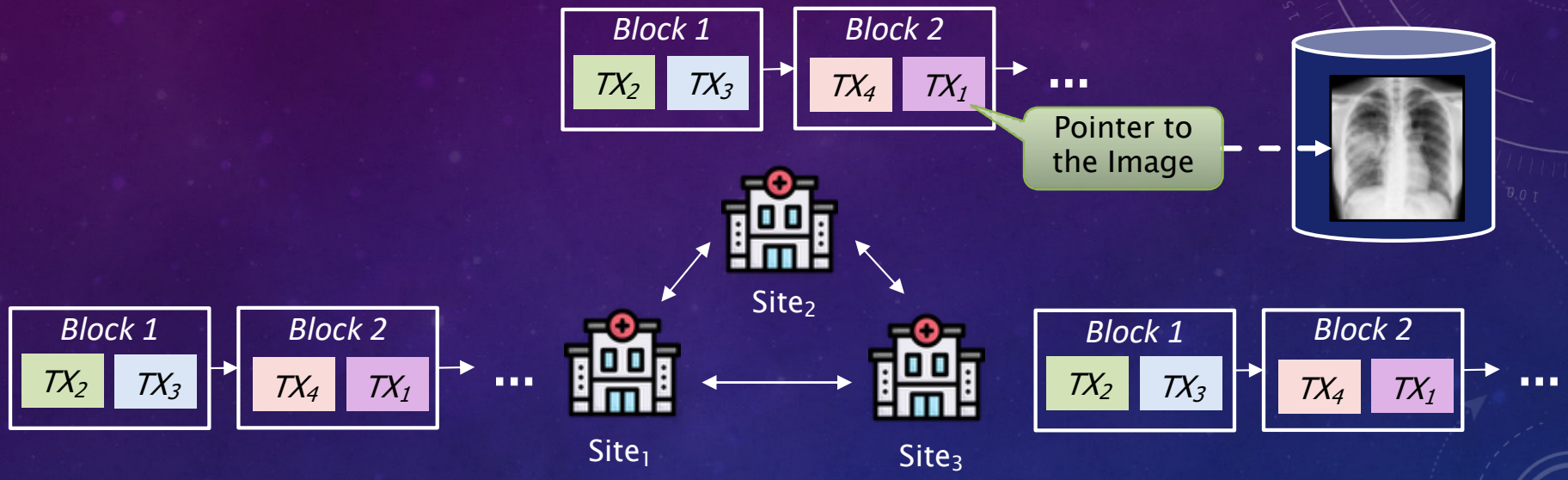
BLOCKCHAIN

- Benefits
  - Availability to avoid SPOF
  - Immutability to avoid changing of data on blockchain
  - Provenance to avoid denying the transactions



Adapted From: Mun Li M, **Kuo T-T**. Previewable Contract-Based On-Chain X-Ray Image Sharing Framework for Clinical Research. International Journal of Medical Informatics. 2021:104599. doi: 10.1016/j.ijmedinf.2021.104599. PubMed PMID: 34628257

EXISTING BLOCKCHAIN SOLUTIONS

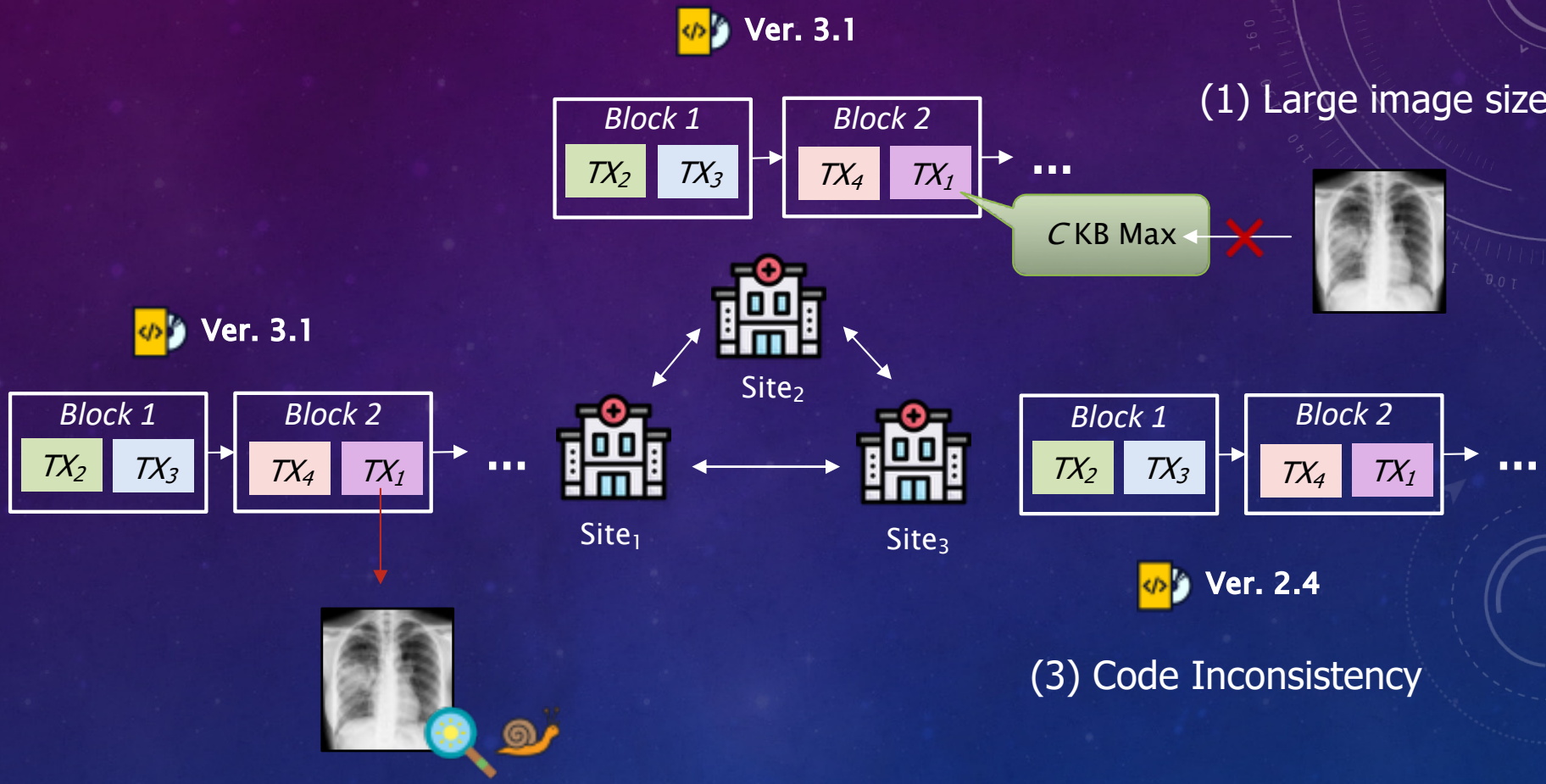


Only image pointers are available/immutable/provable, not images

Can we bring the same technical benefits to images (i.e., store images on-chain) ?

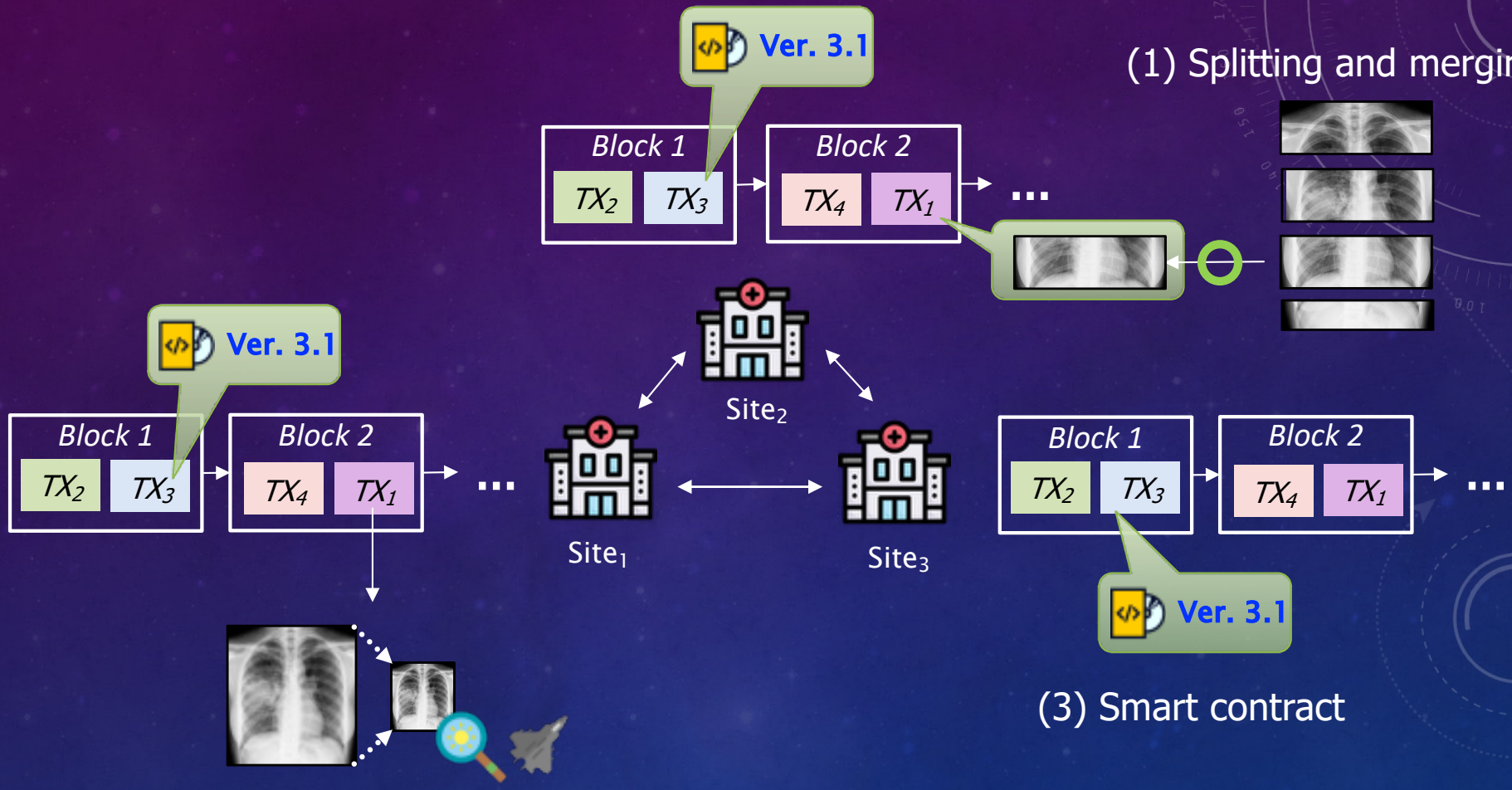
Adapted From: Mun Li M, **Kuo T-T**. Previewable Contract-Based On-Chain X-Ray Image Sharing Framework for Clinical Research. International Journal of Medical Informatics. 2021:104599. doi: 10.1016/j.ijmedinf.2021.104599. PubMed PMID: 34628257

CHALLENGES OF IMAGES-ON-CHAIN



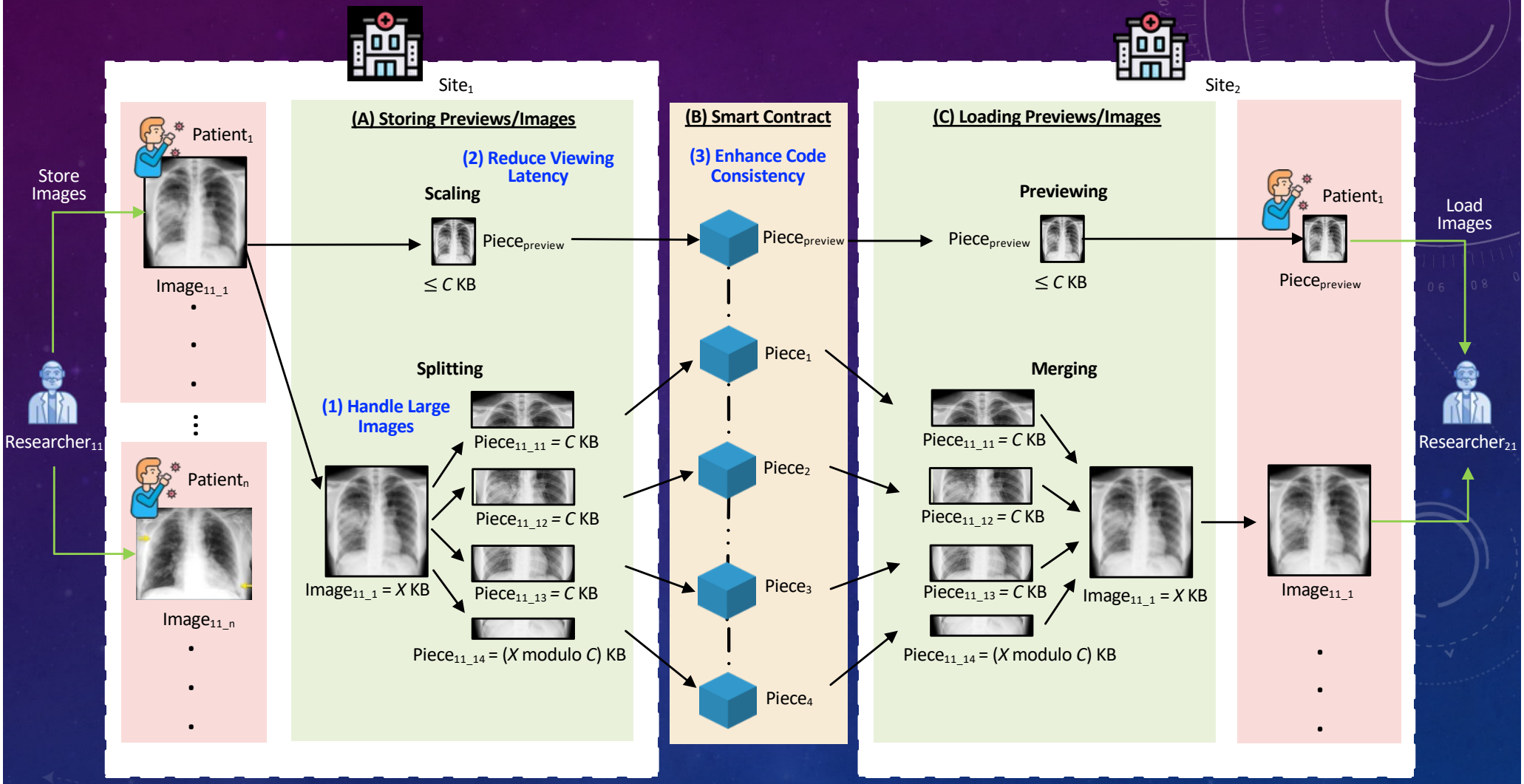
Adapted From: Mun Li M, **Kuo T-T**. Previewable Contract-Based On-Chain X-Ray Image Sharing Framework for Clinical Research. International Journal of Medical Informatics. 2021:104599. doi: 10.1016/j.ijmedinf.2021.104599. PubMed PMID: 34628257

OUR SOLUTION



Adapted From: Mun Li M, **Kuo T-T**. Previewable Contract-Based On-Chain X-Ray Image Sharing Framework for Clinical Research. International Journal of Medical Informatics. 2021:104599. doi: 10.1016/j.ijmedinf.2021.104599. PubMed PMID: 34628257

DATA SHARING WORKFLOW



Adapted From: Mun Li M, **Kuo T-T**. Previewable Contract-Based On-Chain X-Ray Image Sharing Framework for Clinical Research. International Journal of Medical Informatics. 2021:104599. doi: 10.1016/j.ijmedinf.2021.104599. PubMed PMID: 34628257

COVID-19 IMAGE DATASET

- Chest x-ray images
  - Patients positive/suspected of COVID-19 or other viral/bacteria pneumonias
  - Publicly available image repository from the University of Montreal

Category	Statistic	Value
Images	Total Number of Images	920
	JPG	711
	PNG	209
Image Size (KB)	Maximum	18,497
	Minimum	10
	Median	221
	Average	584
Patients	Total Number of Patients	450
Number of Images per Patient	Maximum	22
	Minimum	1
	Median	2
	Average	2
Total Image Size for a Patient (KB)	Maximum	57,043
	Minimum	10
	Median	401
	Average	1,195

Cohen, J.P., et al., Covid-19 image data collection: prospective predictions are the future, 2020
Cohen, J.P., P. Morrison, and L. Dao Covid-19 image data collection. 2020
Mun Li M, **Kuo T-T**. Previewable Contract-Based On-Chain X-Ray Image Sharing Framework for Clinical Research. International Journal of Medical Informatics. 2021:104599. doi: 10.1016/j.ijmedinf.2021.104599. PubMed PMID: 34628257

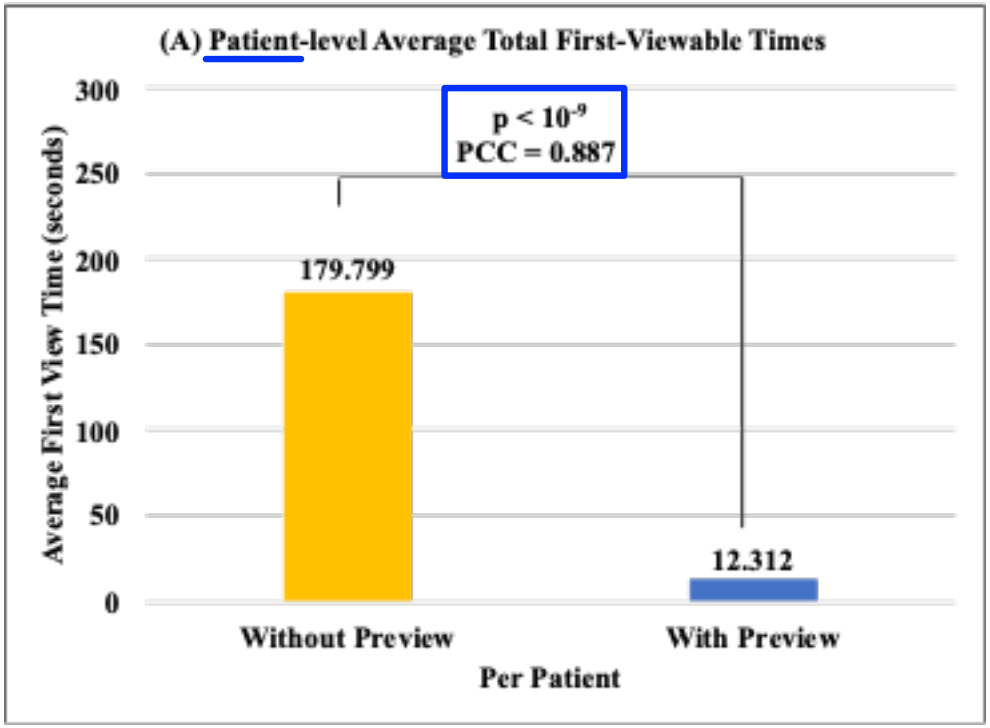
EXPERIMENT SETTING

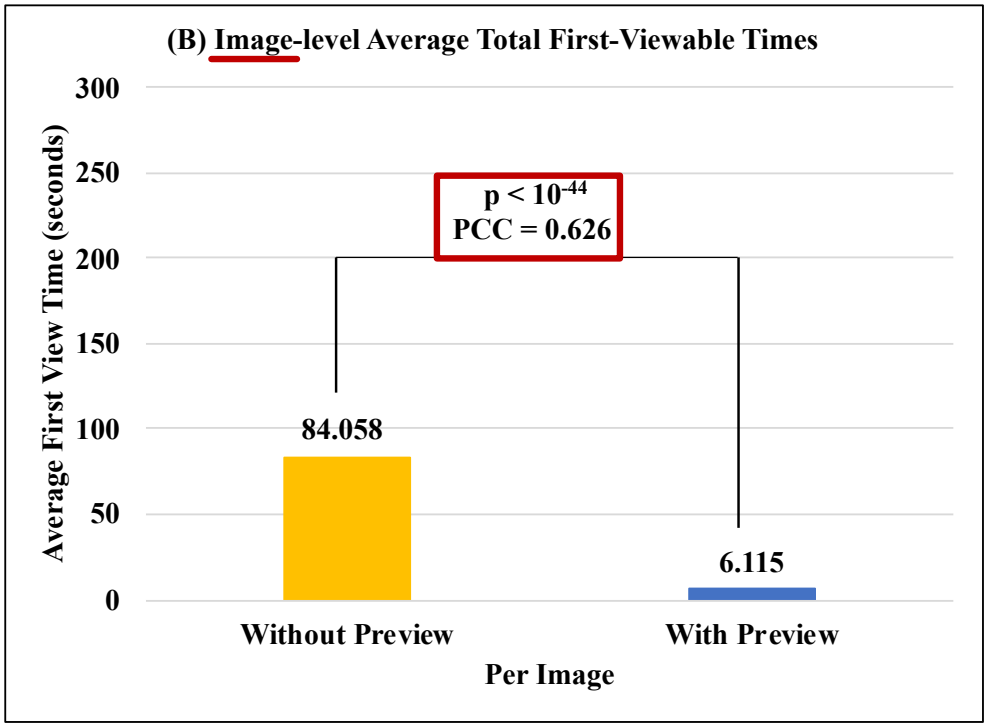
- Comparing methods
  - Patient-level with preview (our method, 450 patients total)
  - Patient-level without preview
  - Image-level with preview (assuming only 1 image per patient, 920 images total)
  - Image-level without preview
- Evaluation metrics
  - Storing time
  - First-viewable storing time (= storing time if no preview available)
  - Loading time
  - First-viewable loading time (= loading time if no preview available)
  - Total first-viewable time (= first-viewable storing + first-viewable loading time)

Mun Li M, **Kuo T-T**. Previewable Contract-Based On-Chain X-Ray Image Sharing Framework for Clinical Research. International Journal of Medical Informatics. 2021:104599. doi: 10.1016/j.ijmedinf.2021.104599. PubMed PMID: 34628257

RESULTS: TOTAL FIRST VIEWABLE TIME

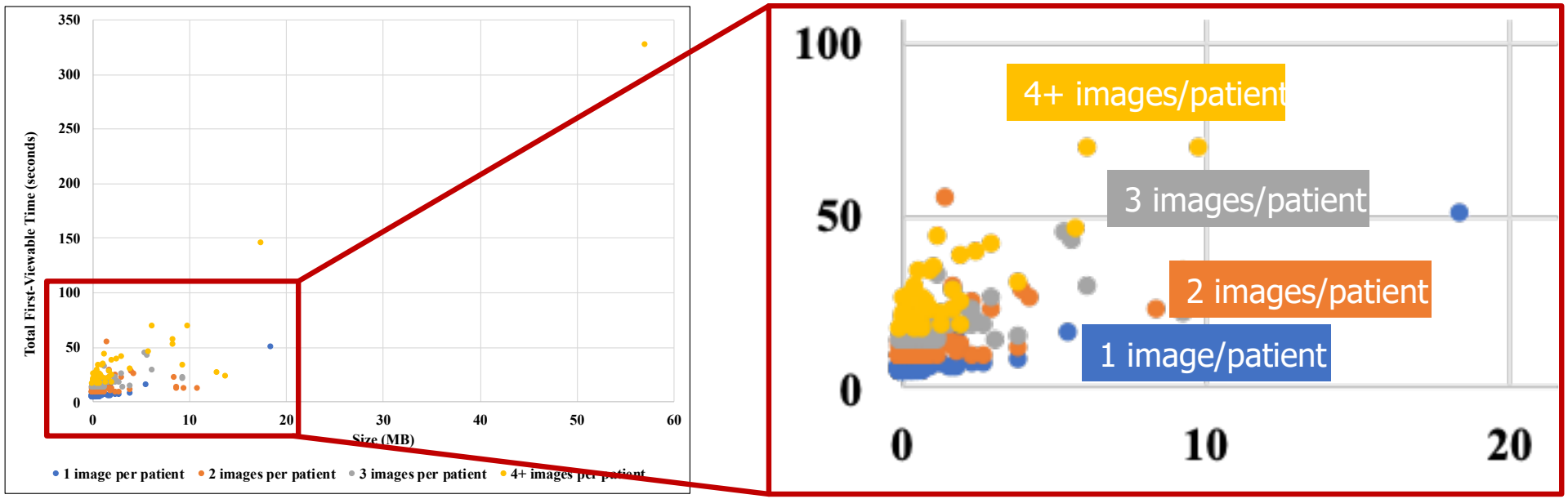
- Statistics
  - Paired two-sample t-test
  - Pearson Correlation Coefficient (PCC, between 1.0 and -1.0)





Mun Li M, **Kuo T-T**. Previewable Contract-Based On-Chain X-Ray Image Sharing Framework for Clinical Research. International Journal of Medical Informatics. 2021:104599. doi: 10.1016/j.ijmedinf.2021.104599. PubMed PMID: 34628257

RESULTS: BREAKDOWN



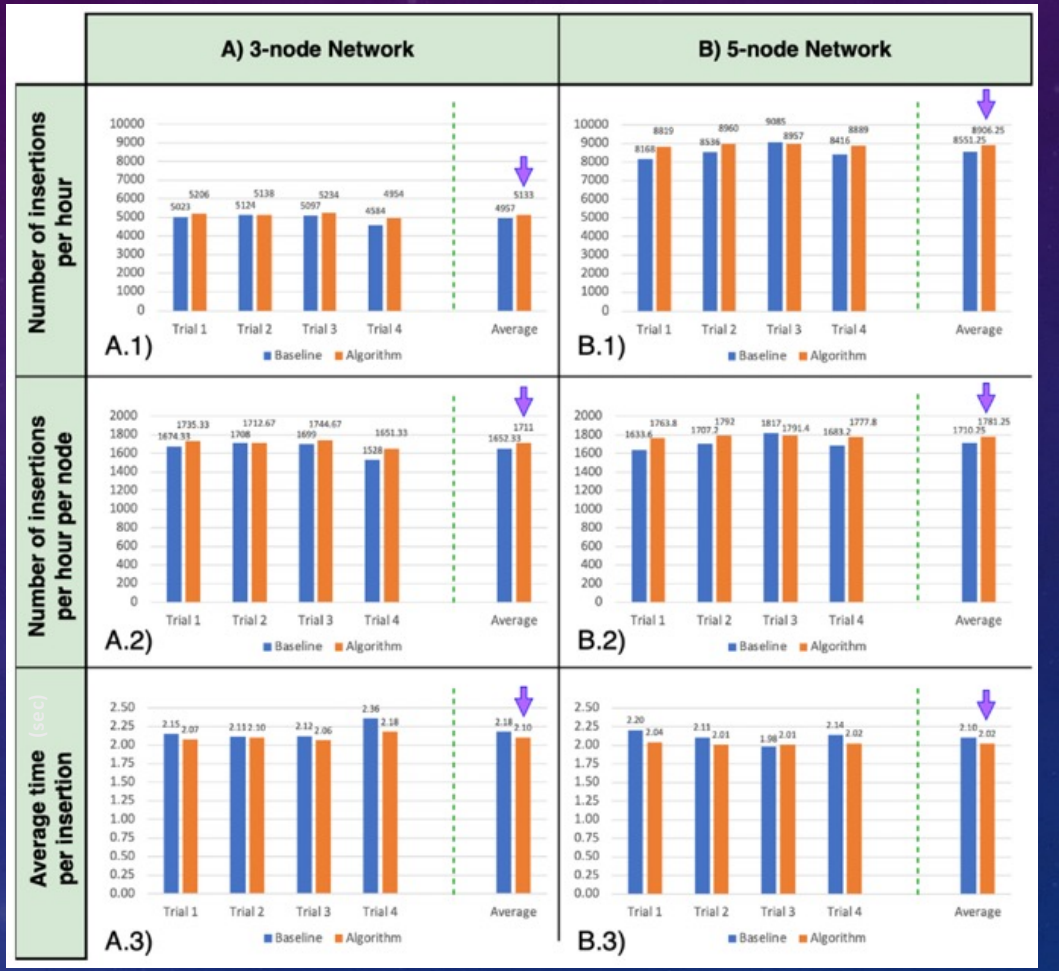
Need more time for patients with large # of images and large total image size

Mun Li M, **Kuo T-T**. Previewable Contract-Based On-Chain X-Ray Image Sharing Framework for Clinical Research. International Journal of Medical Informatics. 2021:104599. doi: 10.1016/j.ijmedinf.2021.104599. PubMed PMID: 34628257

PATIENT DATA SHARING CONSENT

- Related to iDASH Competition 2021
  - Storing 10K consent records with 127 queries
- Developed 2 mapping algorithms for
  - Insertion and Consented Patient
- Compared with a non-mapping baseline

Data Field	Data Description	Data Type	Sample value	Number of distinct values
Patient ID	The unique identification number assigned to a patient.	Integer	1675	4000
Study ID	The unique identification number assigned to a research study.	Integer	10	60
Timestamp	The timestamp as the patient made their data-sharing choices.	Unix format	1620315008	39991
Consent Record	The seven data categories that a patient may choose for or against sharing with researchers.	Boolean vector (size = 7)	[*true*, *false*, *false*, *false*, *true*, *true*, *false*]	128



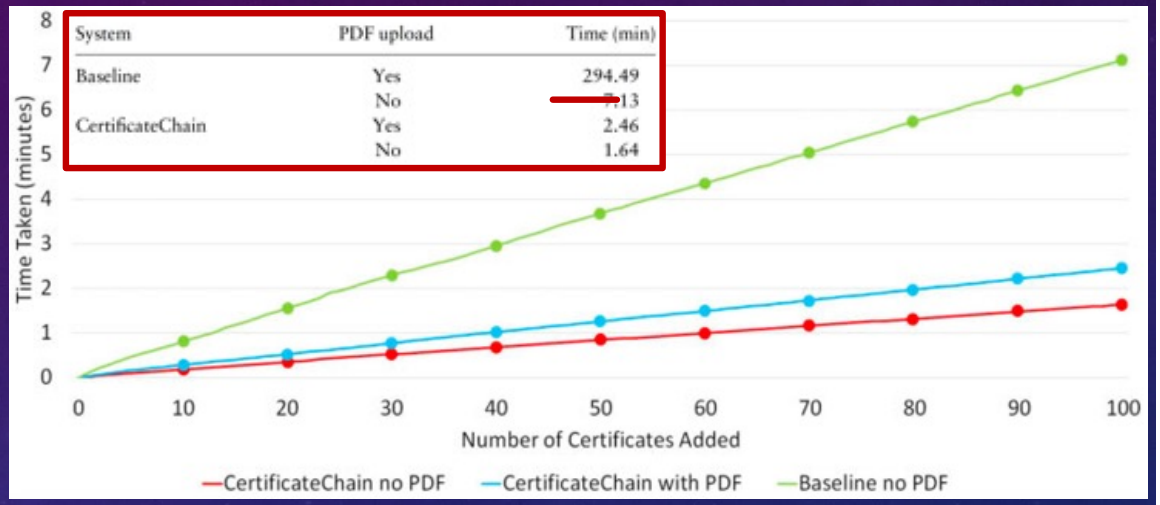
Using mapping technique is consistently better

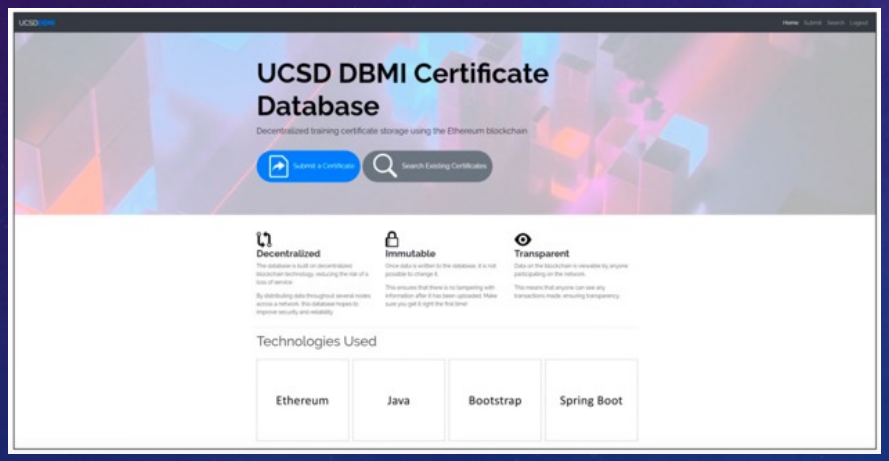
Reference: Pham A, Edelson M, Nouri A, **Kuo T-T**. Distributed management of patient data-sharing informed consents for clinical research. Computers in Biology and Medicine. 2024.

BIOMEDICAL TRAINING CERTIFICATE

- Related to iDASH Competition 2022
  - Storing 100 training certificates with 1 query
- Hash table indexing
  - Allow parallel (instead of sequential) TXs
- Compared with a non-indexing baseline
- Also developed GUI for real tests

System	PDF upload	Time (min)
Baseline	Yes	294.49
	No	7.13
CertificateChain	Yes	2.46
	No	1.64





Category	People	CITI	HIPAA	Certificates
Student	14	13	2	15
Staff	2	2	0	2
Faculty	1	2	1	3
Total	17	17	3	20

Using indexing technique can also improve speed

Reference: Tellew J, **Kuo T-T.** CertificateChain: decentralized healthcare training certificate management system using blockchain and smart contracts. Journal of the American Medical Informatics Association Open (JAMIA Open). 2022.

CLINICAL RESEARCH ACTIVITY

- Storing COVID-19 question/answer pairs
  - Collected 27 pairs, each from up to 14 institutions
  - Storing 9,166 logs (2,098 files) with 18,144 queries
- Smart contract to store and query logs/files
- Compared with a GitHub-based baseline
  - Either in log- or institution- level (all in seconds)
  - Evaluated cross-cloud (AWS, GCP, Azure)

Splitting	Measurements	Proposed	Baseline
Log-level	Total time	10 670.472 (4730.111)	
	Per-query time	0.588 (0.261)	~ 2.4
Institution-level	Total time	7537.861 (361.600)	
	Per-query time	0.415 (0.020)	~ 1.8

Using blockchain cross-cloud could be a feasible solution

Splitting	Measurements	Baseline	Proposed	*P* value
Log-level	Total time	28 811.467 (2571.143)	6405.482 (88.000)	$<10^{-28}$
	Average time over 9166 records	3.143 (0.281)	0.699 (0.010)	$<10^{-28}$
	Average time over the largest number of input logs	9.428 (0.841)	2.096 (0.029)	$<10^{-28}$
Institution-level	Total time	35 769.300 (7864.710)	13 401.609 (1882.121)	$<10^{-14}$
	Average time over 9166 records	3.902 (0.858)	1.462 (0.205)	$<10^{-14}$
	Average time over the largest number of input logs	5.713 (1.183)	2.130 (0.201)	$<10^{-15}$

Reference: **Kuo T-T**, et al. Blockchain-Enabled Immutable, Distributed, and Highly Available Clinical Research Activity Logging System for Federated COVID-19 Data Analysis from Multiple Institutions. Journal of the American Medical Informatics Association (JAMIA). 2023.

TODAY

GENERAL NETWORKS

BLOCKCHAIN NETWORKS