



# Secure Data Provenance in Cloud-centric Internet of Things via Blockchain Smart Contracts

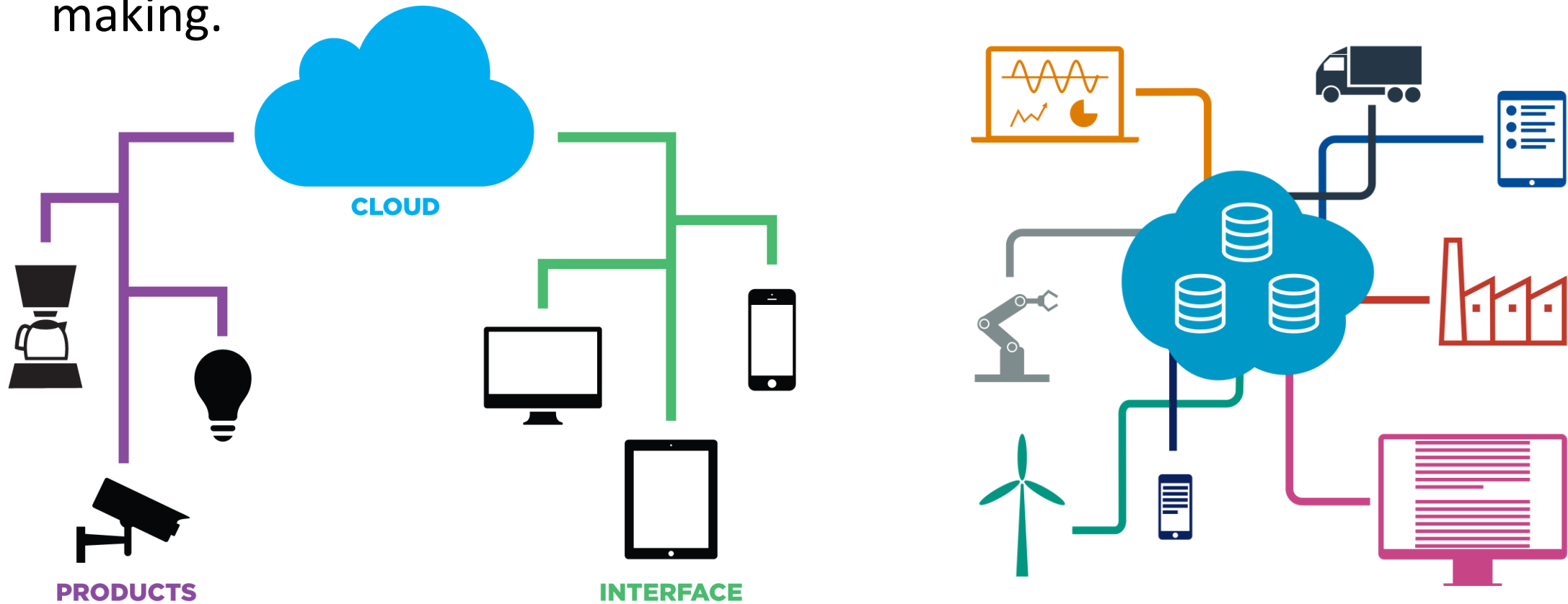
Saqib Ali (PhD)  
Post-doc fellow

**Supervisor:**

Dr. Guojun Wang, Pearl River Scholarship Distinguished Professor  
Director of Institute of Computer Networks,  
Vice Dean of School of Computer Science and Technology,  
Guangzhou University, Guangzhou.

# Use Case: IoT Devices

IoT devices harvest the data through sensors and transmit it over a wireless network towards a cloud storage for analysis and decision making.



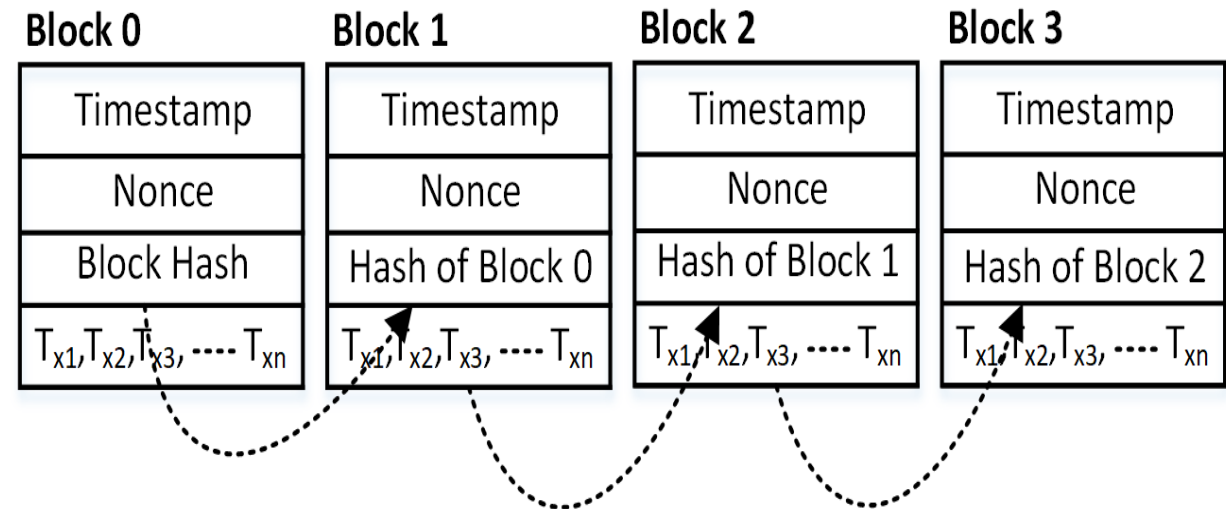
# Problems in the Collected Data

The collected data can be faulty due to:

- Environmental influence,
- Background noise,
- Faulty sensors,
- Dying battery,
- Device may generate biased or fake data due to security attacks (e.g., data modification, false information injection or fast sampling rate to drain the battery of the device).

# Proposed Solution:

- Blockchain & Smart Contracts
  - i. Highly Trusted
  - ii. No need for the trusted 3<sup>rd</sup> party
  - iii. Protected
  - iv. Tampered proof
  - v. Data transparency
  - vi. Visibility
  - vii. Auditability



# Distributed Ledger Technology (DLT)

A **distributed ledger** is a type of data structure which resides across multiple computer devices, generally spread across locations or regions.

## Key Components

P2P network

Distributed computation

Cryptography

Consensus algorithm

Smart Contracts

# Transactions

- The record of an event, cryptographically secured with a digital signature, that is verified, ordered, and bundled together into blocks, form the transactions in the blockchain.
- In the Bitcoin blockchain, transactions involve the transfer of bitcoins, while in other blockchains, transactions may involve the transfer of any asset or a record of some service being rendered.
- Furthermore, a smart contract within the blockchain may allow automatic execution of transactions upon meeting predefined criteria.

# What are Smart Contracts?

- i. Autonomous agents which execute automatically and independently on the top of the blockchain
- ii. Visible to everyone & Fully trusted and scrutinised
- iii. Can write or query the blockchain
- iv. Invoked using transactions (ID + Data)
- v. Logic in the smart contract is totally immutable
- vi. All transactions to and from the smart contracts are fully traceable

## Components:

- i. Unique address on the blockchain
- ii. Private storage or State
- iii. Associated code (where we can define our logic)

## Caution

- i. Once deployed, smart contracts cannot be reversed







# Contributions

1. We proposed a trustworthy **device registration and identity provenance smart contracts** via blockchain to ensure the device integrity and provenance in cloud-centric IoT network.
2. We stationed a comprehensive **data provenance smart contract** in the blockchain to guarantee the secure provenance towards the data stored in the cloud.
3. We deploy a **self-learning traffic profile provenance contract** in the network. The contract is used to certify the provenance to the device traffic profile while tracing the discrepancies in its traffic behavior.
4. Finally, we outline the security analysis & implementation details of the proposed framework using private blockchain i.e., Hyperledger Fabric

# Types of Contracts used in the Proposed Architecture

## **i. Device Registration Contract**

- This global contract maps device identification strings to their blockchain address identity (equivalent to a public key).
- Policies coded into the contract can regulate registering new identities or changing the mapping of existing ones.

## **ii. Device Provenance Contract**

- Using device signature & device metadata

## **iii. Data Provenance Contract**

- Using Data hash, timestamp & device Signatures

## **iv. Traffic Profile Provenance Contract**

- Using traffic profile

---

**Algorithm 1** Device Registration Contract

---

**Input:** transaction  $tx$

$tx \leftarrow tx\_type, reg\_type, device\ ID, publickey, gateway\ ID, \& \text{ other device metadata}$

$tx\_type \leftarrow device\_record$

$reg\_type \leftarrow \text{new or update}$

**Output:**  $tx_{id}$  &  $bc_{id}$  {successful transaction  $tx$  to the blockchain}

1: **if** ( $reg\_type = \text{new}$ ) **then**

2:  $deviceID \leftarrow bc_{id}$  {unique blockchain address (equivalent to a public key) is assigned by a Membership Service Provider}

3:  $tx \leftarrow set(tx \cup bc_{id})$  {create a new asset}

4: **else**

5: **if** ( $reg\_type = \text{update}$ ) **then**

6:  $query(tx)$  {query the blockchain to get the device registration information}

7:  $tx \leftarrow set(updated\_tx)$  {update an existing asset}

8: **end if**

9: **end if**

10:  $PutState(tx)$  {transaction  $tx$  is placed in a Write set as a data-write proposal. However, it does not affect the ledger}

11:  $WriteState(tx)$  {transaction  $tx$  is validated and successfully committed to the ledger}

12: **return**  $tx_{id}, bc_{id}$

---

---

**Algorithm 2** Device Provenance Contract

---

**Input:** transaction  $tx$

$tx \leftarrow tx\_type, publickey \text{ (device), \& device\_metadata}$

$tx\_type \leftarrow device\_proven$

**Output:** *True* or *False*

1:  $tx' \leftarrow query(publickey \text{ and } tx\_type)$  {query the blockchain to get the device metadata}

2: **if** ( $tx' = null$ ) **then**

3:    $throw$  {could not find the device}

4: **else**

5:   **if** ( $device'.sig = device.sig$ ) **then**

6:     **if** ( $tx'.metadata = tx.metadata$ ) **then**

7:       **return** *TRUE* {device identity is proven}

8:     **else**

9:       **return** *FALSE* {device identity is NOT proven}

10:    **end if**

11:   **else**

12:      $throw$  {device signature mismatch}

13:     **return** *FALSE*

14:   **end if**

15: **end if**

---

---

**Algorithm 3** Data Provenance Contract

---

**Input:** transaction  $tx$

$tx \leftarrow$  hash(data), tx\_type, publickey (device), & timestamp (data collected)

$tx\_type \leftarrow$  valid\_record

**Output:** *True* or *False*

- 1:  $tx' \leftarrow$  query(publickey and timestamp) {query the blockchain to get the transaction record}
  - 2: **if** ( $tx' = null$ ) **then**
  - 3:    *throw* {could not locate the record}
  - 4: **else**
  - 5:    **if** ( $device'.sig = device.sig$ ) **then**
  - 6:      **if** ( $tx'.hashvalue = hash(data)$ ) **then**
  - 7:        **return** *TRUE* {record is proven}
  - 8:      **else**
  - 9:        **return** *FALSE* {record is NOT proven}
  - 10:     **end if**
  - 11:    **else**
  - 12:      *throw* {signature mismatch}
  - 13:      **return** *FALSE*
  - 14:    **end if**
  - 15: **end if**
-

---

**Algorithm 4** Traffic Profile Provenance Contract

---

**Input:** transaction  $tx'$

$tx' \leftarrow$  publickey (device), tx\_type, timestamp, traffic\_profile with feature vector  $f$ , & threshold vector  $t$  for feature vector  $f$  in  $tx'$

$tx\_type \leftarrow$  traffic\_proven

**Output:**  $tx$ ,  $traffic\_proven$  (*True or False*)

{True: discrepancy detected & False: no discrepancy}

- 1: **for all**  $k : k \in$  set of recent traffic profiles  $tx_{n-1}$  **do**
  - 2:   generate correlation model  $C$  for the feature vector  $f$  in  $tx_{n-1}$  {learning from the previously stored data in the blockchain}
  - 3: **end for**
  - 4: **for all** features  $f$  in  $tx'$  **do**
  - 5:   **if**  $(\delta(tx'.f, C) > t)$  **then**
  - 6:      $discrepancy\_status \leftarrow TRUE$
  - 7:      $traffic\_proven \leftarrow FALSE$
  - 8:   **else**
  - 9:      $discrepancy\_status \leftarrow FALSE$
  - 10:     $traffic\_proven \leftarrow TRUE$
  - 11:   **end if**
  - 12: **end for**
  - 13:  $tx \leftarrow tx' \cup traffic\_proven$
  - 14:  $PutState(tx)$
  - 15:  $WriteState(tx)$  {transaction  $tx$  is validated and successfully committed to the ledger}
  - 16: **return**  $tx$ ,  $discrepancy\_status$
-

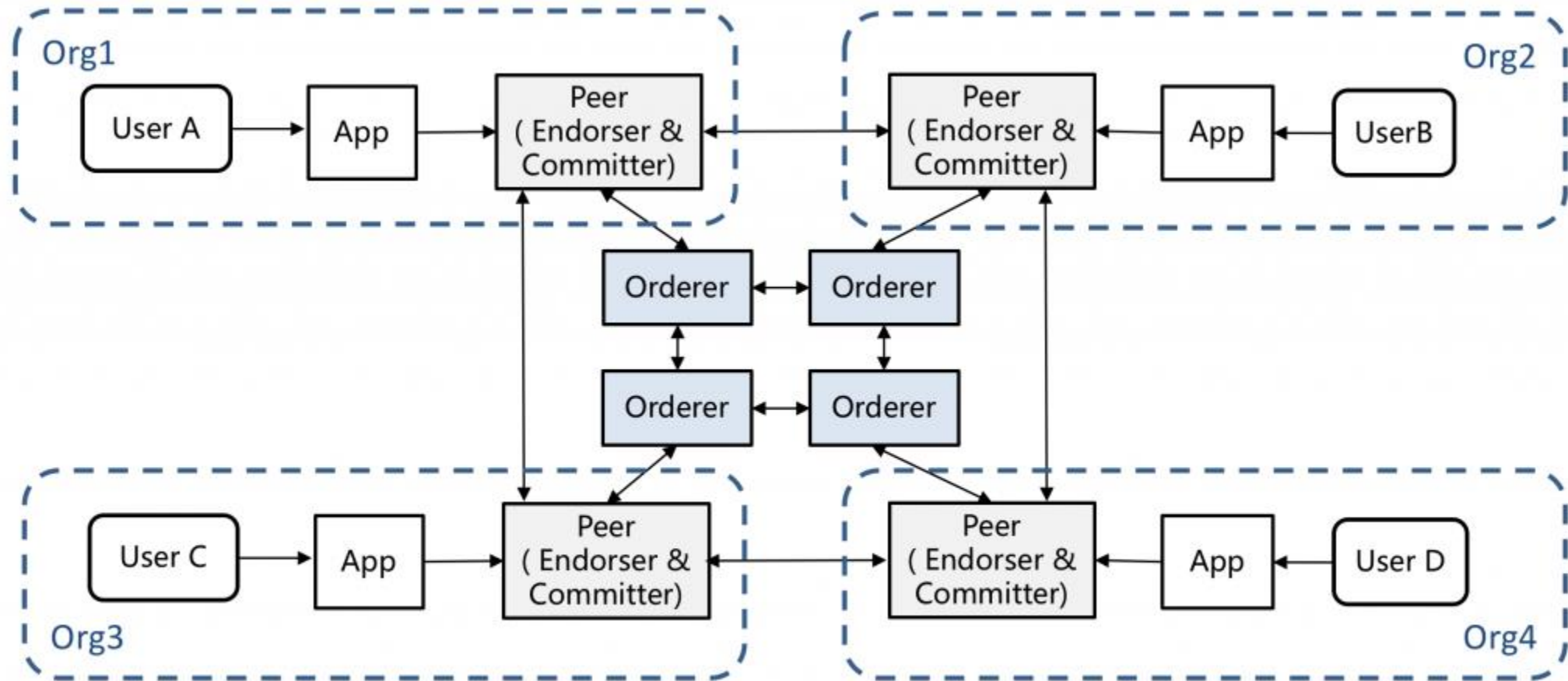
# Significance

- i. Trusted data collection
- ii. Protected
- iii. Tampered proof
- iv. Data transparency and auditability

# Implementation

- Hyperledger Fabric

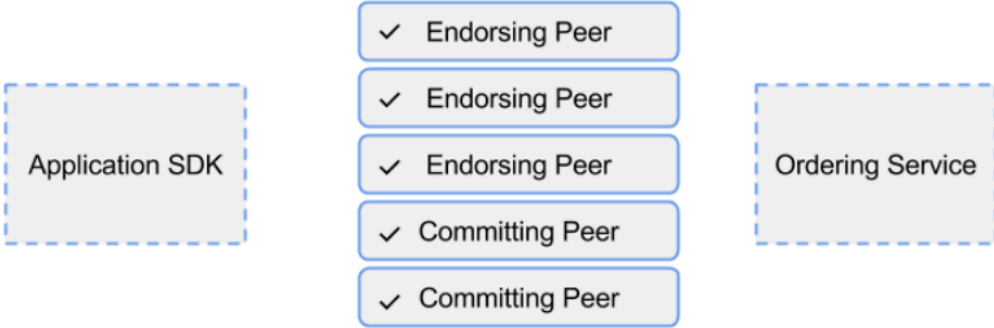
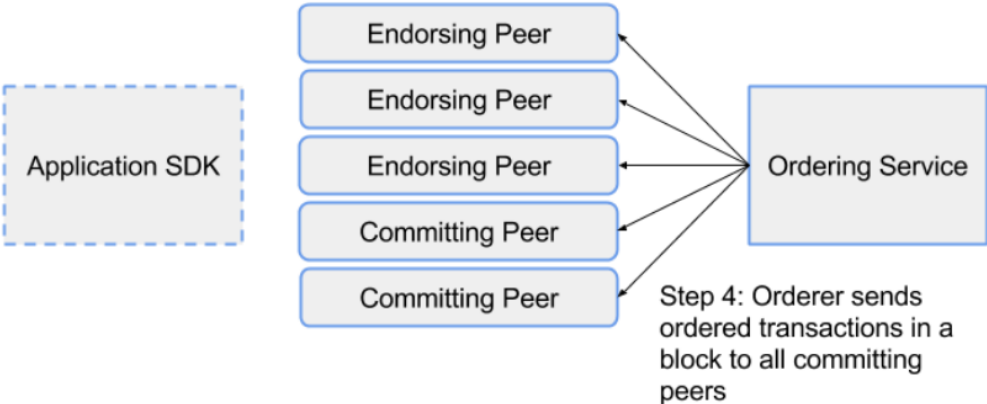
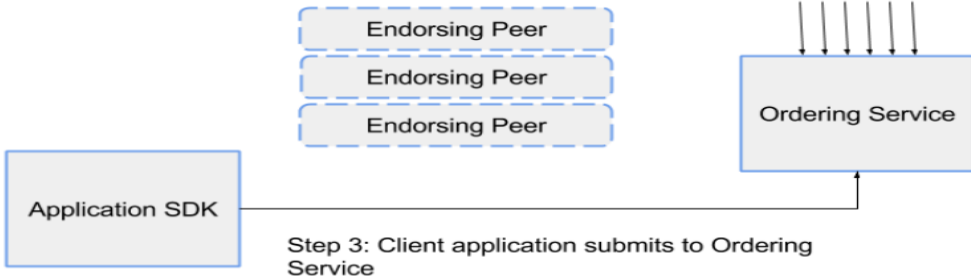
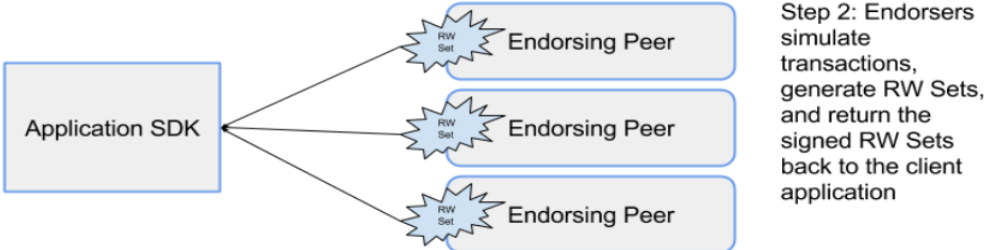
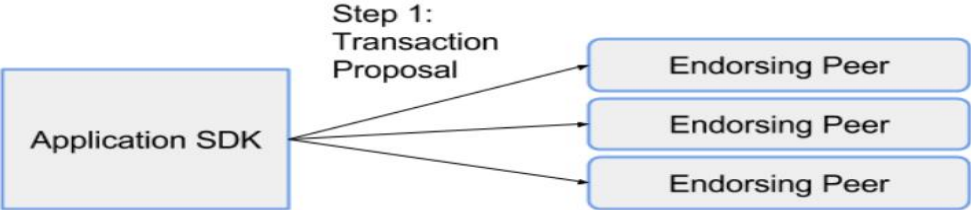
# Hyperledger Fabric



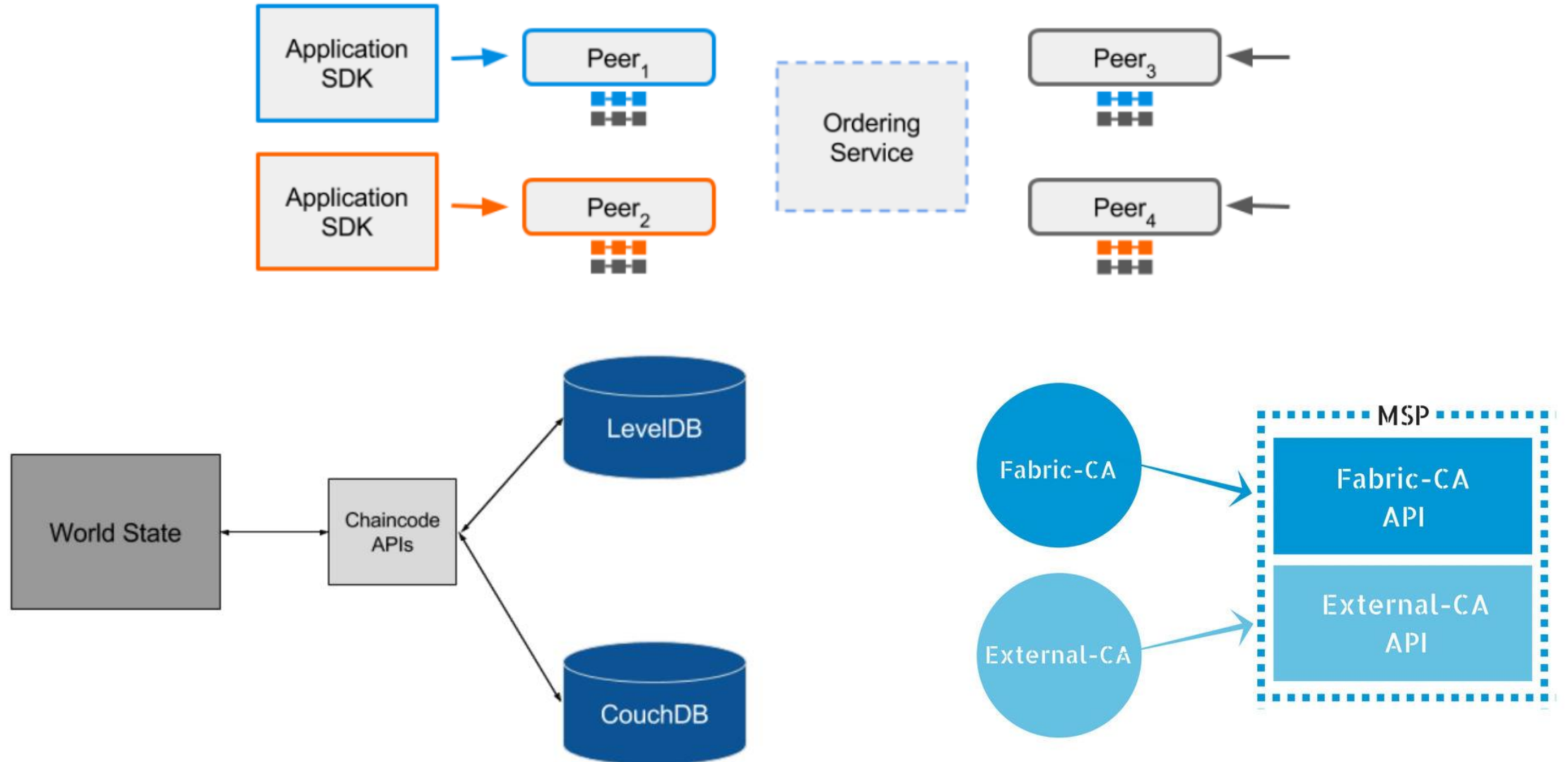
Fabric 1.0 deployment model. Some arrows are omitted for simplicity



# Transaction Flow



# Channels, World State Database, & MSP







```
Store path:/home/saqibali/.hfc-key-store
Successfully loaded admin from persistence
Assigned the admin user to the fabric client ::{"name":"admin","mspId":"Org1MSP","roles":null,"affiliation":"","enrollmentSecret":"","enrollment":{"signatureIdentity":"addc3b90966cbab5468e9a06a029e55a2a72845fc7359dbce60f40396c9b2987","identity":{"certificate":"-----BEGIN CERTIFICATE-----\nMIICAjCCAigAwIBAgIUCxv+G28gebCNbK00doLHo7XEjCAwCgYIKoZIzj0EAwIw\nczELMAKGA1UEBhMCVVMxEzARBgNVBAGTCkNhbgGmb3JuaWEwXjAUBgNVBACQDQgAE68nxLGBALHrPrgSohZhWx3GsbDspu/+09a9Dc/rn\ndoChIIJNLAORxAX4bbc/BWatHVWVswdzTAcqRXduzt69fKNsMGowDgYDVR0PAQH/\nBAQDAgeAMAwGA1UdEwEB/wQCAAAwHQYDVR0OBBYEF0+HgGBWA8jSXOZ10ur3R6Ds\n2NCdMCsGA1UdIwQkMCKAIEI5qg3NdtruuLoM2nAYUdFFBNMarRst3dusalc2Xkl8\nnMAoGCCqGSM49BAMCA0gAMEUCIQDPRIEr89N4VxS6mRLXChKbVo1nULdlvZXJaGwP\nnSM8VHgIgatIi7bQLazxC+E/x+OUVgCJ8qYtIaXAb6cCEBzuTDry=\n-----END CERTIFICATE-----\n"}}}
```

```
Store path:/home/saqibali/.hfc-key-store
Successfully loaded user1 from persistence
Query has completed, checking results
Response is [{"Key":"1", "Record":{"holder":"Miriam","location":"67.0006, -70.5476","timestamp":"1504054225","vessel":"923F"}},{ "Key":"10", "Record":{"holder":"Fatima","location":"51.9435, 8.2735","timestamp":"1487745091","vessel":"49W4"}},{ "Key":"2", "Record":{"holder":"Dave","location":"91.2395, -49.4594","timestamp":"1504057825","vessel":"M83T"}},{ "Key":"3", "Record":{"holder":"Igor","location":"58.0148, 59.01391","timestamp":"1493517025","vessel":"T012"}},{ "Key":"4", "Record":{"holder":"Amalea","location":"-45.0945, 0.7949","timestamp":"1496105425","vessel":"P490"}},{ "Key":"5", "Record":{"holder":"Rafa","location":"-107.6043, 19.5003","timestamp":"1493512301","vessel":"S439"}},{ "Key":"6", "Record":{"holder":"Shen","location":"-155.2304, -15.8723","timestamp":"1494117101","vessel":"J205"}},{ "Key":"7", "Record":{"holder":"Leila","location":"103.8842, 22.1277","timestamp":"1496104301","vessel":"S22L"}},{ "Key":"8", "Record":{"holder":"Yuan","location":"-132.3207, -34.0983","timestamp":"1485066691","vessel":"EI89"}},{ "Key":"9", "Record":{"holder":"Carlo","location":"153.0054, 12.6429","timestamp":"1485153091","vessel":"129R"}]}
```

```
submit recording of a tuna catch:
[ '11', '28.012, 150.425', '498256368', 'Johon', '0239L' ]
Store path:/home/saqibali/.hfc-key-store
Successfully loaded user1 from persistence
Assigning transaction_id: bad37fa295c4e77b00ee9754a21c7d3326fe73328ab56340d3f2ef6f2a1c8425
Transaction proposal was good
Successfully sent Proposal and received ProposalResponse: Status - 200, message - ""
info: [EventHub.js]: _connect - options {}
The transaction has been committed on peer localhost:7053
Send transaction promise and event listener promise have completed
Successfully sent transaction to the orderer.
Successfully committed the change to the ledger by the peer
```

# Future Research Directions

# Key Issues in Blockchain & Smart Contracts

1. Privacy
2. Scalability
3. Smart Contracts are not Intelligent

Parno, B., Howell, J., Gentry, C., & Raykova, M. (2013). Pinocchio: Nearly practical verifiable computation. *Proceedings - IEEE Symposium on Security and Privacy*, 238–252. <https://doi.org/10.1109/SP.2013.47>

Ames, S., Hazay, C., Ishai, Y., & Venkatasubramanian, M. (2017). Ligerio : Lightweight Sublinear Arguments Without a Trusted Setup. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security - CCS '17*, 2087–2104. <https://doi.org/10.1145/3133956.3134104>

# 1. Zero knowledge Proofs

**Def.** Zero-knowledge proofs are encryption schemes used to prove that you know something without revealing what it is.

**For example,** you can show without a doubt that you know the answer to a puzzle without actually disclosing the solution.

## **Types:**

1. Interactive zero knowledge proofs
2. Non interactive zero knowledge proofs

Oded Goldreich and Yair Oren. Definitions and Properties of Zero-Knowledge Proof Systems. Journal of Cryptology. Vol 7(1). 1–32. 1994 ([PS](#))

# 1. Interactive Zero Knowledge Proofs

**Def.** Interactive zero-knowledge proofs require interaction between the individual (or computer system) proving their knowledge and the individual validating the proof.

**Actors in the system:**

1. **Proof** - who claim some knowledge
2. **Verifier** - who verify the claim of the prover

**Key requirement: Interaction between the Prover & Verifier** - in the form of the challenges such that the responses from the prover will convince the verifier if and only if the statement is true.

Goldwasser, S.; Micali, S.; Rackoff, C. (1989), "[The knowledge complexity of interactive proof systems](#)" (PDF), *SIAM Journal on Computing*, Philadelphia: [Society for Industrial and Applied Mathematics](#), **18** (1): 186–208, [doi:10.1137/0218012](#), [ISSN 1095-7111](#)



# 1. How Interactive Zero Knowledge Proofs work?

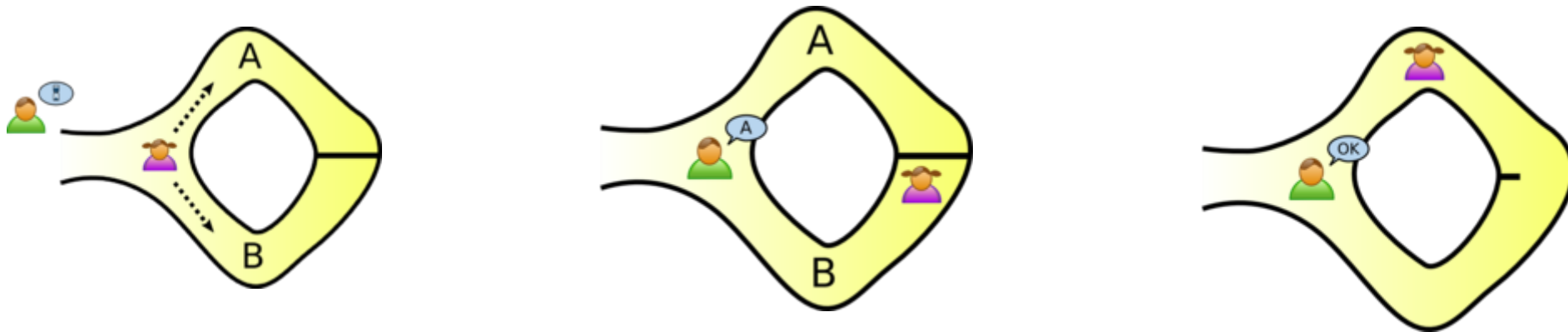
**Example:** Ali Baba Cave

Peggy (the **prover** of the statement) and Victor (the **verifier** of the statement).

Peggy has uncovered the secret word used to open a magic door in a cave.

Victor wants to know whether Peggy knows the secret word.

Peggy does not want to reveal her knowledge (the secret word) to Victor



Quisquater, Jean-Jacques; Guillou, Louis C.; Berson, Thomas A. (1990). ["How to Explain Zero-Knowledge Protocols to Your Children"](#) (PDF). *Advances in Cryptology – CRYPTO '89: Proceedings*. **435**: 628–631.

# 1. Issue with Interactive Zero Knowledge Proofs?

**Advantage:** An interactive zero-knowledge proof has the advantage that only the verifier can be absolutely convinced that the prover has the knowledge.

**Disadvantage:** If bystanders and observers can't verify the claim, the prover then has to interact with every verifier independently—which takes time and is resource intensive.

## 2. Non-interactive zero-knowledge proofs

The reason for non-interactive zero-knowledge proofs is to allow a large number of observers to verify the proof efficiently.

In Blockchains: every block need to be computed and verified by every node in the network which raise the issue of scalability

Example: **Sudoku Puzzle**

*Blum, Manuel; Feldman, Paul; Micali, Silvio (1988).  
["Non-Interactive Zero-Knowledge and Its Applications"](#). Proceedings of the twentieth annual ACM symposium on Theory of computing (STOC 1988): 103–112. [doi:10.1145/62212.62222](#).*

	COLUMN	COLUMN	COLUMN	
ROW	3	4	5	7
	SECTOR	SECTOR	SECTOR	
ROW	5	2	7	1
		3	2	5
ROW	9		1	8
			4	6
ROW	1	2	9	7
				5

# Practical Applications of Zero Knowledge Proofs

	<b>prover scalability (quasilinear time)</b>	<b>verifier scalability (polylogarithmic time)</b>	<b>Transparency (public randomness)</b>	<b>Post-quantum security</b>
hPKC	Yes	Only repeated computation	No	No
DLP	Yes	No	Yes	No
IP	Yes	No	Yes	No
MPC	Yes	No	Yes	Yes
IVC+hPKC	Yes	Yes	No	No
ZK-STARK	Yes	Yes	Yes	Yes

Figure 2: Theoretical comparison of universal (NP complete) realized ZK systems.

# References

- [1] M. Díaz, C. Martín, and B. Rubio, “State-of-the-art, challenges, and open issues in the integration of Internet of things and cloud computing,” *J. Netw. Comput. Appl.*, vol. 67, pp. 99–117, 2016.
- [2] A. Botta, W. De Donato, V. Persico, and A. Pescapé, “Integration of Cloud computing and Internet of Things: A survey,” *Futur. Gener. Comput. Syst.*, vol. 56, pp. 684–700, 2016.
- [3] K. Lee, S. Kim, J. Jeong, S. Lee, H. Kim, and J. S. Park, “A framework for DNS naming services for Internet-of-Things devices,” *Future Generation Computer Systems*, North-Holland, 31-Jan-2018.
- [4] D. He and S. Zeadally, “An Analysis of RFID Authentication Schemes for Internet of Things in Healthcare Environment Using Elliptic Curve Cryptography,” vol. 2, no. 1, pp. 72–83, 2015.
- [5] O. Novo, “Blockchain Meets IoT: an Architecture for Scalable Access Management in IoT,” *IEEE Internet Things J.*, vol. 14, no. 8, pp. 1–1, 2018.
- [6] S. Wilkinson and J. Lowry, “MetaDisk: Blockchain-Based Decentralized File Storage Application,” pp. 1–11, 2014.
- [7] D. T. T. Anh, M. Zhang, B. C. Ooi, and G. Chen, “Untangling Blockchain: A Data Processing View of Blockchain Systems,” *IEEE Trans. Knowl. Data Eng.*, vol. 4347, no. c, pp. 1–1, 2018.
- [8] J. R. Douceur, “The Sybil Attack,” pp. 251–260, 2002.
- [9] M. Castro and B. Liskov, “Practical byzantine fault tolerance and proactive recovery,” *ACM Trans. Comput. Syst.*, vol. 20, no. 4, pp. 398–461, 2002.
- [10] M. Castro, M. Castro, B. Liskov, and B. Liskov, “Practical Byzantine fault tolerance,” *OSDI {'}99 Proc. third Symp. Oper. Syst. Des. Implement.*, no. February, pp. 173–186,

Q/A