Blockchain For Large-Scale Internet of Things Data Storage and Protection

Jinseong Park

Department of Computer Science & Engineering

Seoul National University of Science & Technology

April 16, 2019

Internet of Things (IoT)

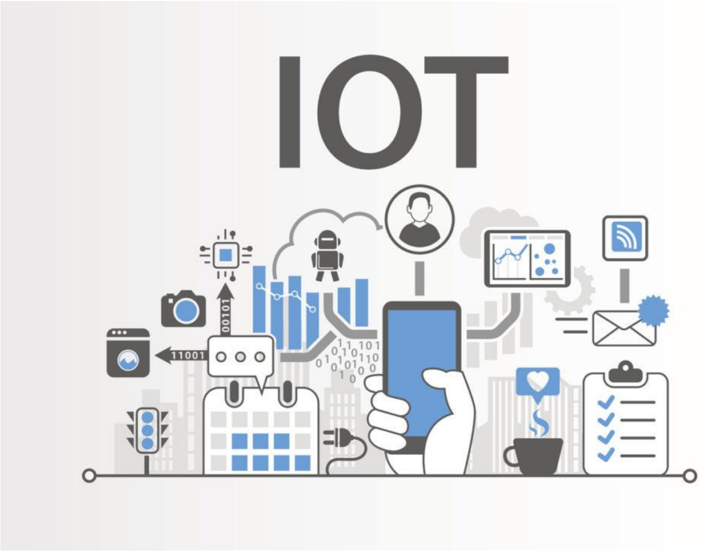


Table 1: IoT Units Installed Base by Category (Millions of Units)

Category	2016	2017	2018	2020
Consumer	3,963.0	5,244.3	7,036.3	12,863.0
Business: Cross-Industry	1,102.1	1,501.0	2,132.6	4,381.4
Business: Vertical-Specific	1,316.6	1,635.4	2,027.7	3,171.0
Grand Total	**6,381.8**	**8,380.6**	**11,196.6**	**20,415.4**

Source: Gartner (January 2017)

Ref: https://www.zdnet.com/article/iot-devices-willoutnumber-the-worlds-population-this-year-for-the-first-time/

- IoT is to describe the ubiquitous connection of everyday object.
- IoT device is dramatically increased and is estimated by Gartner that there will be over 20 billions by 2020
- IoT devices are used various field such as 'smart grid', 'smart factory', 'connected car' and 'medical system'.

Internet of Things (IoT) (Con't)

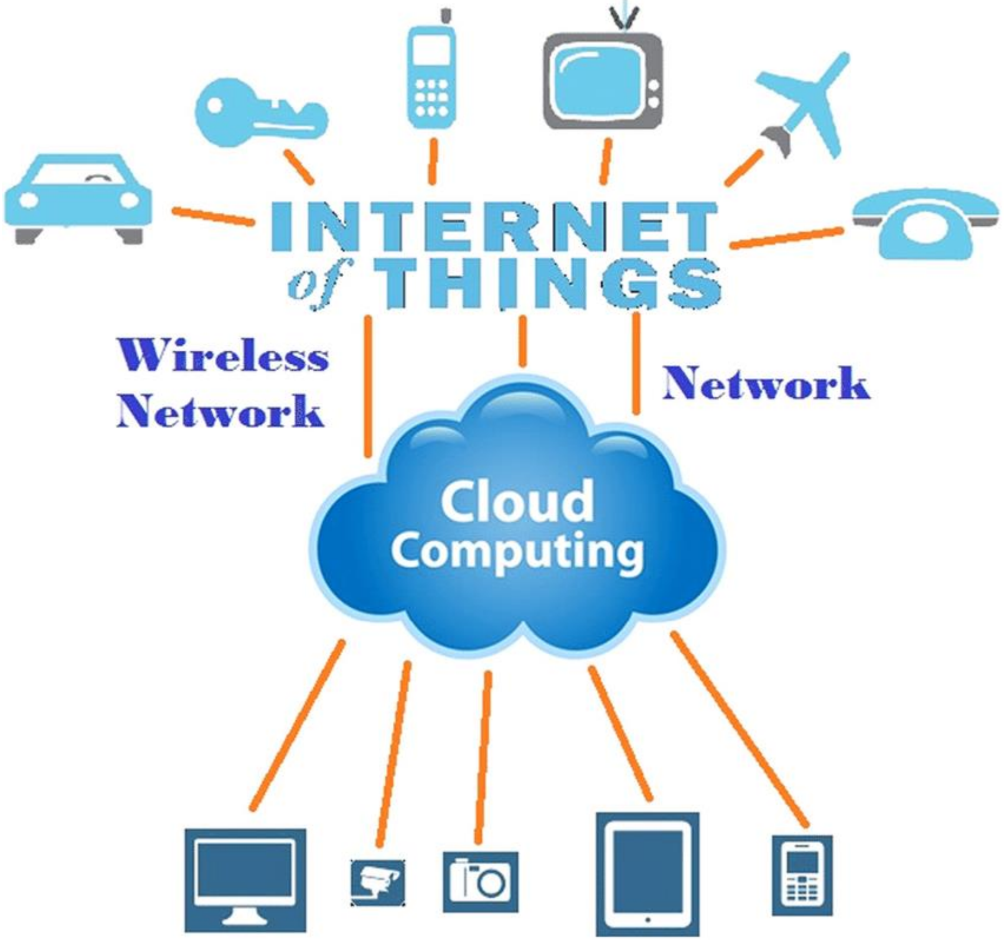
STCIS
Seoul National University of Science and Technology
Cryptography and Information Security Laboratory

- Traditional IoT application consist of cloud-based IoT structure. However this structure brings two drawbacks.

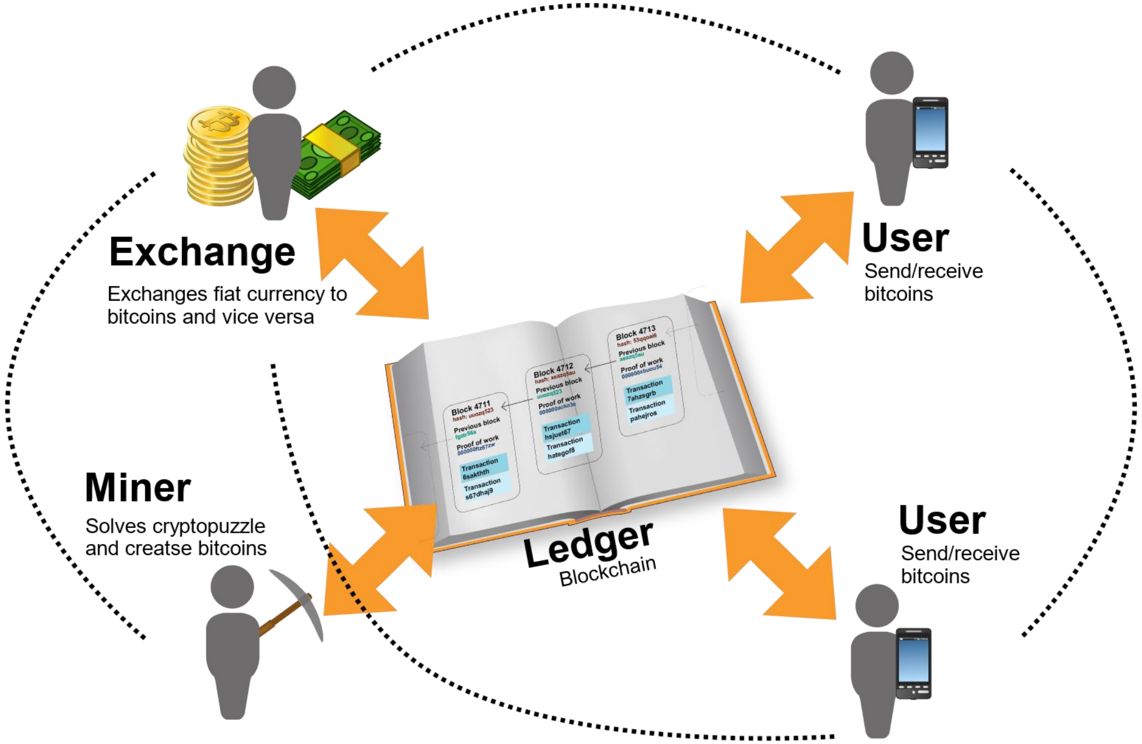
1. The cloud server needs very high storage capacity to store the IoT data.

2. Sensitive data can be easily leaked from the server.

→ This paper suggest a way using blockchain which represented decentralized structure to solve these drawbacks.



Blockchain



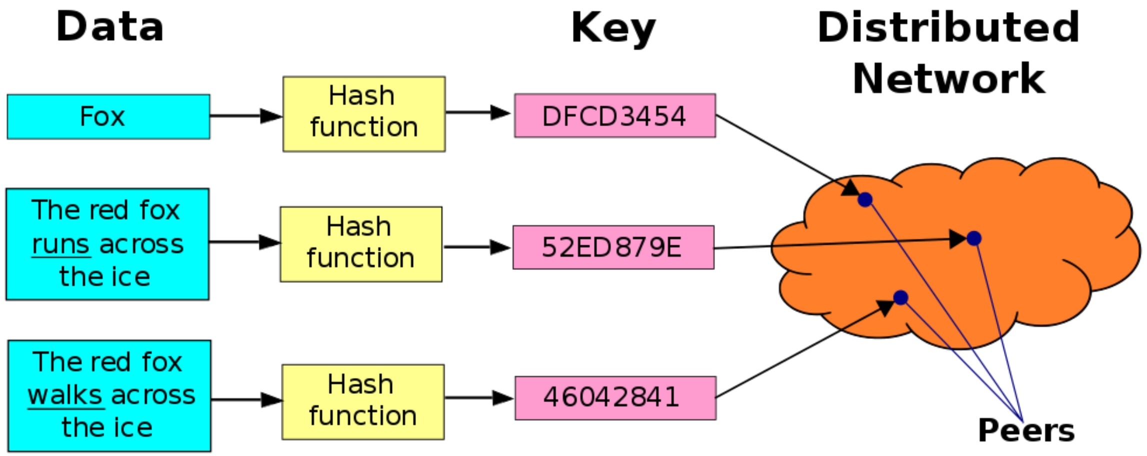
Ref: https://www.csg.uzh.ch/csg/en/research/blockchain.html

- Blockchain offers a convenient platform for distributed data storage and protection.
- In an IoT application on this paper, data can be stored in Distributed Hash Tables(DHTs).

Distributed Hash Table (DHT)

- DHT is data structure for fast & easy searching in distributed environment.

- DHT uses Hash Table that formed Key-Value for fast Data searching.
  - Key is hashed Data and it is also pointer of "Where data is saved" in distributed network.



Ref: https://en.wikipedia.org/wiki/Distributed_hash_table

- DHT isn't specific algorithm's name. There are some kinds of DHT algorithm, For examples, Chord, Kademlia, Voldemort, Apache Cassandra.

Distributed Hash Table (DHT) (Con't)

- Blockchain using DHT will decide whether the access can be granted of not. Therefore the authentication of the requester is handled by the distributed blockchain miners instead of a trusted centralized server. It brings following advantages

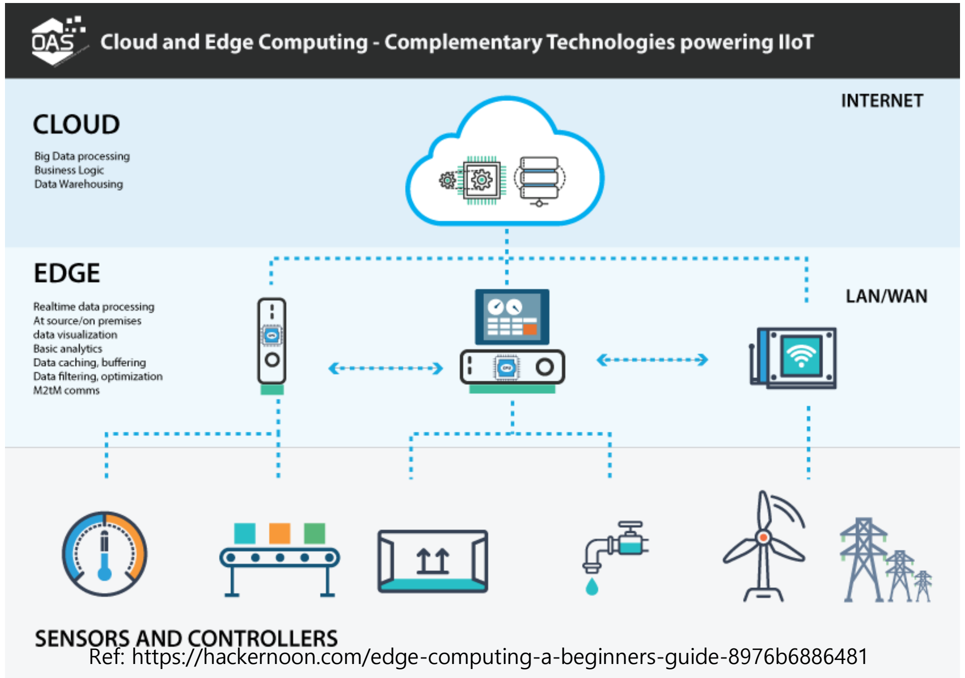
  1. Decentralized Storage : Blockchain using DHT makes be able to easily and fast searching and it can function even millions of nodes.

  2. No Centralized Trusted Server : The access to IoT data is controlled by the majority of the blockchain miners. Therefore Users don't need to worry about unauthorized access to User's data.

  3. Traceability and Accountability : Activities such as accessing and modifying the IoT data, can be recorded by the blockchain. No malicious attempts can be made undetected.

1. Introduction

Edge computing



Ref: https://hackernoon.com/edge-computing-a-beginners-guide-8976b6886481

- IoT devices have low computational power, therefore they are not capable of conducting complex computations.

- In contrast cloud computing, there are edge server between data source and cloud. It makes to be able to real-time computations and communications.

Certificateless Cryptography

- When Applying blockchain, the miners take charge of authentication when an entity requests to access the data. However, the miners should not have any knowledge of the credentials to perform authentation.

- Identity Based Encryption (IBE)
  - IBE is a type of public-key encryption. And IBE's public key is user's identity(e.g. user's name, user's email address)
  - IBE has key Escrow problem.
    - Key Escrow problem: Key Generation Center (KGC) is not trusted.

Certificateless Cryptography (Con't)

- Certificateless cryptography is different than IBE as a user's public key is generated by both the user's identity and some secret of which the KGC is not aware. Therefore, KGC has no knowledge of the user's private key, while a public key can be verified whether it belongs to certain user or not.

- The only drawback of certificateless cryptography compared to IBE is that the public key of a user, even though can be verified needs to be pre-broadcasted.
  - This drawback is solved by using blockchain.

Overview

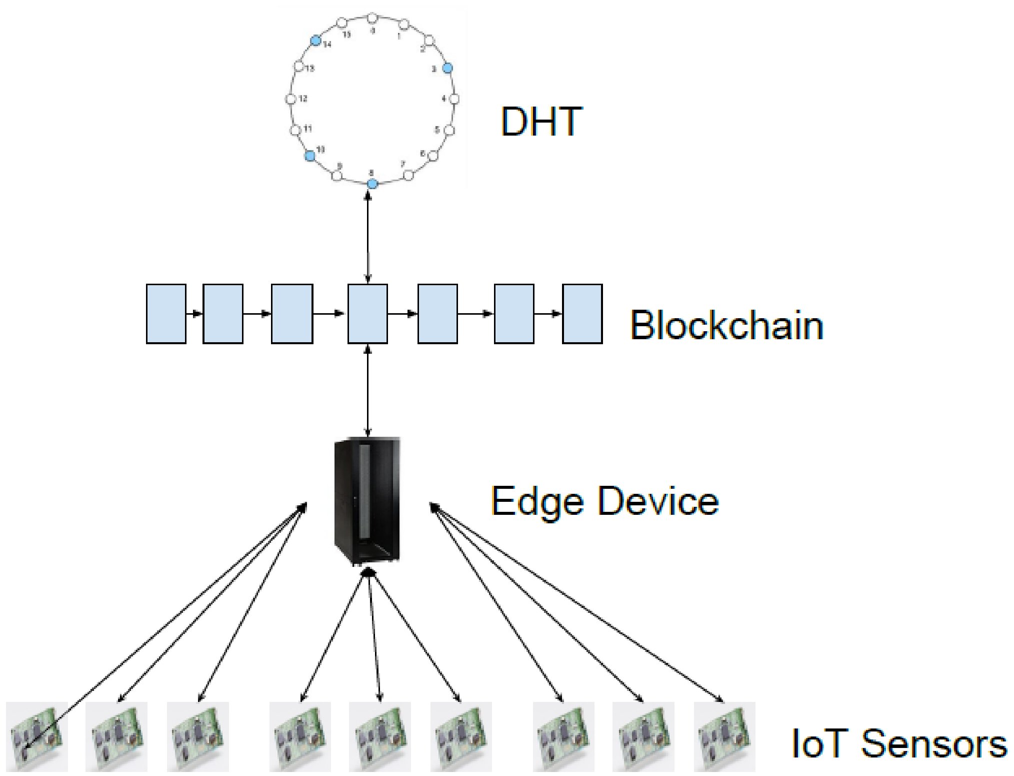


Fig. 1. The structure of data storage scheme with blockchain

- Edge device forwards data to DHT. And It also posts transaction to blockchain that means which IoT device data saved where in DHT.
- Blockchain verifies the transaction and records the identity of the IoT device and the storage address.

Overview

STCIS
Seoul National University of Science and Technology
Cryptography and Information Security Laboratory

- When an IoT device requests data from DHT, it posts a "transactions" to the blockchain.

- If the transaction is validated and written into a block, the DHT node storing the data will send data to the requester.

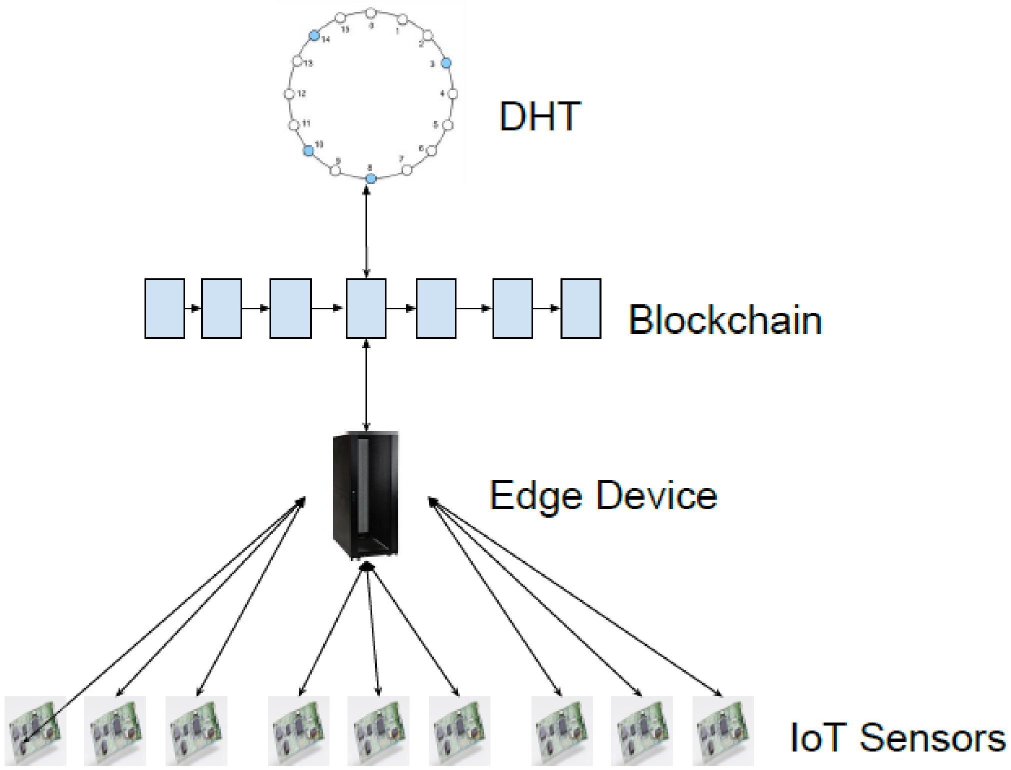


Fig. 1. The structure of data storage scheme with blockchain

Blockchain Description

- In this paper, using “Proof of Useful Work (PoUW)” mechanism.
- PoUW is achieved by adopting Intel’s Software Guard extensions “SGX”
- It is to let the miners compute useful work for Intel, and in return Intel provides workers with a proof of their work so that the workers can build a block.

Blockchain Transactions

STCIS
Seoul National University of Science and Technology
Cryptography and Information Security Laboratory

- In this paper, transactions are two cases.
  - Request service : $T = (ID_A, Timestamp, Action = store\ data\ in\ Addr.)$
  - Request data : $T = (ID_B, ID_A, Timestamp, Action = access\ data\ in\ Addr.)$
- Note that a DHT node in "$Addr.$" does not send data to a requester until the DHT node confirms that the transaction of the request has been verified and written into the blockchain.

Miner Awards

- The proposed system is built upon the blockchain run by a large group of miners.
- This system using following three aspects for reward to miners.
  - This system eliminates the centralized server, and the service fee from a traditional server should be transferred to the miners in the blockchain.
  - PoUW utilizes miners to compute useful work for large companies. In return, theses companies will pay back miners for their work.
  - The operations of blockchain will inevitability create block awards that can be split among the miners as their rewards.

Edge Computing

STCIS
Seoul National University of Science and Technology
Cryptography and Information Security Laboratory

- The roles of an edge device are as follow
  - Manage the identities of IoT devices. An edge server stores a copy of identities of all nearby IoT devices and helps each device build a pair of keys for authentication through a KGC.
  - Create transactions for IoT devices. Transaction signed IoT device's ID, and the signing process should be conducted by the edge server. Also data encryption and decryption should be conducted by the edge server.
  - Collect and forward data to DHT. The edge server continuously collects data from nearby devices. It determines the DHT address to store the data and sends the encrypted data to the designated address.

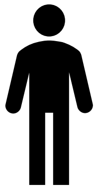
Security Model

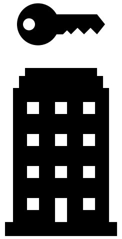
- In suggested design with certificateless cryptography, KGC is not able to obtain any user's private key.
- Data storage and protection are performed solely by the blockchain, without intervention of any other entity.
- Therefore, the security of our scheme is based on the security of the blockchain mechanism.

3. Authentication of Blockchain Transactions

Certificateless Cryptography

- Key generation





① $Setup(1^{\lambda}) \rightarrow (K, MSK)$

PSK_A ←

② $PSkeyGen(K, ID_A, MSK) \rightarrow (PSK_A)$

③ $SValGen(K, ID_A) \rightarrow (X_A)$

④ $SKeyGen(K, PSK_A, X_A) \rightarrow (SK_A)$

⑤ $PKeyGen(K, X_A) \rightarrow (PK_A)$ →

PK_A is broadcast

Only KGC know

Only User know

Certificateless Cryptography

- Functions
  - $Encrypt\ (K, M, ID_A, PK_A) \rightarrow C$
  - $Decrypt\ (C, SK_A) \rightarrow M$
  - $Sign\ (K, M, SK_A) \rightarrow Sig$
  - $Ver(M, Sig, ID_A, PK_A) \rightarrow Valid\ or\ Invalid$

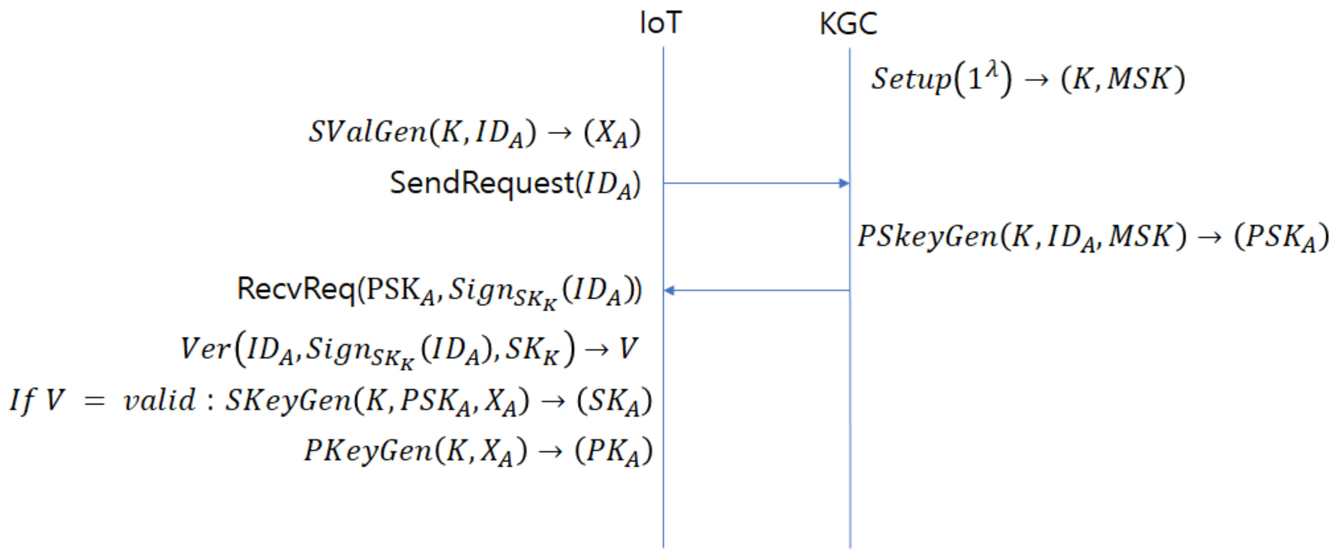
How Blockchain Transactions Work

- Registration

Algorithm 1 Device Registration

Input: ID_A
Output: PK_A, SK_A

```
procedure KeyGen(1^λ, ID_A)
    X_A ← SValGen(K, ID_A)
    SendRequest (ID_A)
    RecvReq(PSK_A, Sign_{SK_K}(ID_A))
    V ← Ver(ID_A, Sign_{SK_K}(ID_A), SK_K)
    if V = Valid then:
        SK_A ← SKeyGen(K, PSK_A, X_A)
        PK_A ← PKeyGen(K, X_A)
    end if
     return
end procedure
```



How Blockchain Transactions Work

- Transactions Description and Verification

Algorithm 2 Verify A Transaction

Input: T_A, σ_{T_A}

Output: a verified T_A

```
procedure VERTRANS( T_A, σ_{T_A} )
    s ← 0
    V_1 ← VerID(ID_A, PK_A, K)
    if V_1 = Valid then
        V_2 ← Ver(T_A, σ_{T_A}, ID_A, PK_A)
    elseAbort
        if V_2 = Valid then
            s ← 1
        else Abort
        end if
    end if
     return
end procedure
```

If the Public key PK_A is derived from the identity ID_A associated with it

If the signed transaction can be verified with the public key PK_A

How Blockchain Transactions Work

- IoT Data Storage and Protection

Algorithm 3 Store Data

Input: ID_A, ACL
Output: a verified T_A

procedure SETACL(ACL)
Create $T_A = (PK_A, ID_A, ACL, Addr)$
Broadcast (T_A, σ_A)
return
end procedure
procedure VERTRANS(T_A, σ_A) ▷ run by the miners
$s \leftarrow 0$
$V_1 \leftarrow VerID(ID_A, PK_A, K)$
if $V_1 = Valid$ **then**
$V_2 \leftarrow Ver(T_A, \sigma_A, ID_A, PK_A)$
elseAbort
if $V_2 = Valid$ **then**
$s \leftarrow 1$
elseAbort
end if
end if
return
end procedure

Algorithm 4 Access Data

Input: $ID_A||Addr, ID_B$
Output: a verified T_B

procedure REQUESTDATA($ID_B, ID_A||Addr$)
Create $T_B = (ID_B, ID_A||Addr)$
$\sigma_{T_B} \leftarrow Sign(K, T_B, Sk_B)$
Broadcast (T_B, σ_{T_B})
return
end procedure
procedure VERTRANS(T_B, σ_{T_B}) ▷ run by the miners
$s \leftarrow 0$
$V_1 \leftarrow VerID(ID_B, PK_B, K)$
if $V_1 = Valid$ **then**
$V_2 \leftarrow Ver(T_B, \sigma_{T_B}, ID_B, PK_B)$
elseAbort
if $V_2 = Valid$ **then**
if $ID_B \in ACL$ **then**
$s \leftarrow 1$
elseAbort
end if
end if
end if
return
end procedure

Access log

4. Extension to Data Trading

Data Trading

STCIS
Seoul National University of Science and Technology
Cryptography and Information Security Laboratory

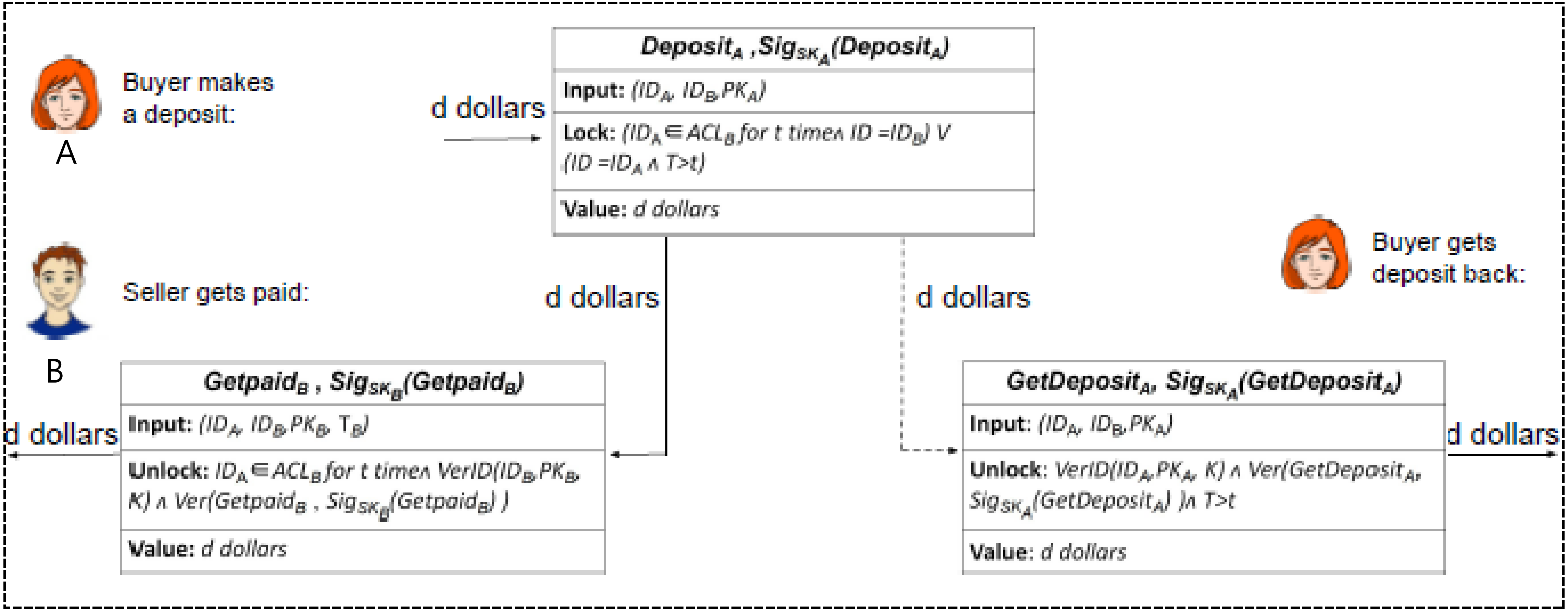


Fig. 3. Data trading with the blockchain

Protocol Security

- Assumption : Algorithm 1-4 are secure protocol in authentication, and certificateless cryptography algorithm is secure.

- If the security parameter K is sufficiently large, adversary is not able to guess the private key of a user.

Privacy

- This paper suggest using re-encryption for privacy.
  - Re-encryption is useful cryptography primitive that enables data encrypted under one public key to be transformed to data under another public key, without decrypting the message.
- Re-encryption has to be performed by the DHT node that holds that data.
  - Because if it is perfomed by user, ciphertext transformed under his own key.

Traceability and Accountability

- In the proposed scheme, all accessing data in a certain DHT address will be recorded in blockchain.

- It makes be not able to deny accessing attempt , and data owner can trace malicious attempts.

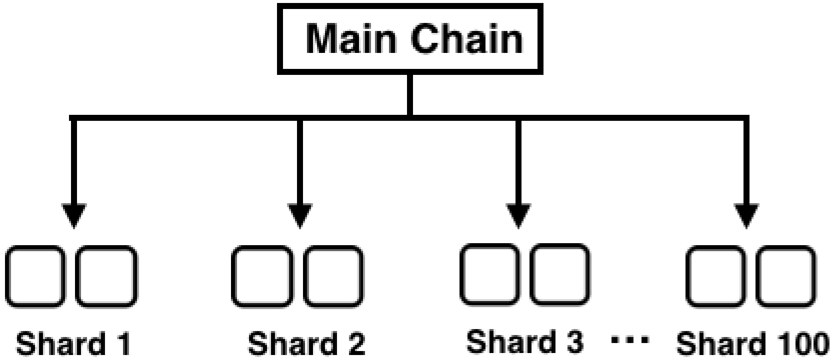
Blockchain Security

- The security of the proposed scheme is based on the security of blockchain and certificateless cryptography.
- Therefore the scheme's security is based on number of miners in the blockchain.

Scalability

- Scalability is a major problem in blockchain's design.

- Two most studied mechanisms to solve the problem are Sharding and sidechains



Ref: https://medium.com/decipher-media/%EB%B8%94%EB%A1%9D%EC%B2%B4%EC%9D%B8-%ED%99%95%EC%9E%A5%EC%84%B1-%EC%86%94%EB%A3%A8%EC%85%98-%EC%8B%9C%EB%A6%AC%EC%A6%88-4-1-sharding-%EC%83%A4%EB%94%A9-611a311c80e6

7. Conclusion

- The first paper tacking the problem of building a secure and accountable storage system for largescale IoT data.

- The first to combine edge computing, certificateless cryptography, and blockchain as a whole to serve IoT applications.

8. Opinion

STCIS
Seoul National University of Science and Technology
Cryptography and Information Security Laboratory

- Blockchain + Edge computing …?
  - Edge computing is used for reducing network transfer idle time.
  - But, if edge computing combined blockchain, edge computing's advantage disappear because blockchain's scalability problem.

Thank you