Blockchain-based Data Privacy Management with Nudge Theory in Open Banking

Hao Wang[b], Shenglan Ma[a], Hong-Ning Dai[c,*], Muhammad Imran[d], Tongsen Wang[a]

[a] Department of Computer Science, Norwegian Uni. of Sci. & Tech., Norway

[b] Division of Science and Technology, Fujian Rural Credit Union, Fuzhou, China

[c] Faculty of Information Tech., Macau Univ. of Sci. and Tech., Macau

[d] College of Computer and Information Sciences, King Saud University, Saudi Arabia

Corresponding authors: Hao Wang (hawa@ntnu.no) and Hong-Ning Dai (hndai@ieee.org)

Blockchain-based Data Privacy Management with Nudge Theory in Open Banking

Hao Wang[a,*], Shenglan Ma[b], Hong-Ning Dai[c,*], Muhammad Imran[d], Tongsen Wang[a]

[a] Department of Computer Science, Norwegian Uni. of Sci. & Tech., Norway

[b] Division of Science and Technology, Fujian Rural Credit Union, Fuzhou, China

[c] Faculty of Information Tech., Macau Univ. of Sci. and Tech., Macau

[d] College of Computer and Information Sciences, King Saud University, Saudi Arabia

Abstract

Open banking brings both opportunities and challenges to banks all over the world especially in data management. A blockchain as a continuously growing list of records managed by a peer-to-peer network is widely used in various application scenarios; and it is commonly agreed that the blockchain technology can improve the protection of financial data privacy. However, current blockchain technology still poses some challenges in fully meeting the needs of financial data privacy protection. In order to address the existing problems, this paper proposes a new data privacy management framework based on the blockchain technology for the financial sector. The framework consists of three components: 1) a data privacy classification method according to the characteristics of financial data; 2) a new collaborative-filtering-based model; and 3) a data disclosure confirmation scheme for customer strategies based on the *Nudge Theory*. We implement a prototype and propose a set of algorithms for this framework. The framework is validated through field experiments and laboratory experiments.

Keywords: Blockchain, Open Banking, Nudge Theory, Data Privacy Management

[*] Corresponding authors: Hao Wang (hawa@ntnu.no) and Hong-Ning Dai (hndai@ieee.org)

1. Introduction

Entering the big data era, banks are now implementing comprehensive digital transformations to meet evolving needs and provide faster and better customer experiences for digital services. The financial and banking sector is moving towards "*Open Banking*", which can promote the maximized benefits of customers through data sharing and deeper cooperation between financial institutions [1]. Because the shared core financial data are connected with the interests of different stakeholders, open banking faces many challenges and difficulties. Privacy concerns prevent the owners of data from sharing and exchanging data outside their institutions for fear of fraud and abuse [2][3]. In addition, the issue of personal data ownership and privacy has significant impact on open banking, especially regulations such as General Data Protection Regulation (GDPR) by European Union (EU) have come into force in May 2018 [4].

In view of these challenges, banks and financial technology companies are exploring new technologies to transform existing systems, products and services to cope with different data privacy requirements and regulations [5]. A blockchain as a continuously growing list of records managed by a peer-to-peer (P2P) network is widely used in various application scenarios, and often combined with artificial intelligence [6], cloud computing [7], big data [8], Internet of Things (IoT) [9] and other technologies [10][11]. A blockchain can run secure computations while no one but the data owner has access to the raw data [12]. Very recently we witness a quick surge of interest and efforts in the application of blockchain and smart contract technologies [14] for many areas, such as government governance [15], insurance [13], medical treatment [16][17], electricity [18], digital storage [19], and education [20].

Although the blockchain technology is commonly agreed to be an enabler for the data privacy and data security, there are still some deficiencies when it comes to the financial data: (1) the granularity of privacy-preserving [22] for customer data are not suitable for the existed applications in banks [23], and customers have limited information on how their data are being used; (2) the complexity and variety in data and sub-systems require banks to have dynamic and convenient methods to deal with customers' data authorization to avoid tedious operations; (3) due to the existing laws and regulations as well as guidelines on banks' operation and customer management, data should be hierarchically managed in blockchains [24].

In order to address these issues, we propose in this paper a new blockchain-based data privacy management framework in combination with the Nudge theory. The framework dissolves the default data disclosure schemes of new customers with the customer-strategy collaborative filtering model, and then the schemes are confirmed in the application scenario of banks. A concrete blockchain-based financial customer data privacy management prototype is implemented. This paper is divided into seven sections. Section 2 introduces the related work in open banking and highlights the shortcomings in existence data privacy management of blockchain. Section 3 presents a new data privacy management framework with a data privacy classification method according to the characteristics of financial data. Section 4 describes a new collaborative-filtering-based model and a confirmation data disclosure scheme for customer strategies based on the Nudge Theory. Section 5 presents the smart contracts and data on-chain and off-chain algorithms. Section 6 evaluates the framework. Finally, we conclude this paper in Section 7.

2. Related Work

Pushed by the need of data sharing imposed by Open Banking, three different management approaches about global data privacy protection emerged:

Firstly, the more restrictive regulations found in the US w.r.t. the development of American technology-driven financial organizations. By functional supervision, the US regulations demand that financial technology should bring the essence of financial business into existing financial supervision system. Today the US's regulations are being loosened so as to enhance the financial data sharing.

Secondly, the passive regulations found in China w.r.t. the development of China's market-driven financial organizations. Chinese financial technologies are driven by market and business models. At present China's regulations on financial technologies are being gradually adjusted in order to solve some problems and risks resulting from the earlier rapid development period. The amendment of criminal law in 2017 has explained the infringement of citizens' personal information behavior standards, and clearly defined that it is not only the company, but also the company leaders should bear corresponding responsibility. China Banking Regulatory Commission (CBRC) also requires banks to carry out data privacy management to enhance the consciousness of the data security, and prevent misuse excessive data, aiming to protect customer data privacy.

Thirdly, the active regulations found in the UK, EU countries, Singapore, Australia, Hong Kong, Japan and South Korea. In order to develop financial technologies, supervision has stepped up to become the main guiding and driving force, promoting "Regulatory Sandbox" and "Open Bank Project". At the same time, GDPR has come into force in May 2018. It will replace the obsolete Data Protection Act that was used for more than 20 years, and provide a good regulation platform for open banking.

Currently, existing solutions to financial data privacy management are mainly based on the *centralized* systems that essentially enforce the operation permissions or data permissions across different financial sectors. If one financial sector initiates a retrieval request for the data contained in other sectors, it must obtain the permission from the centralized party first. The throughput and information security can be guaranteed by the enterprise firewall when the financial data retrieval operations occur within the enterprise intranet. However, the centralized system cannot support the extensive tasks as most of current financial transactions happen throughout the entire Internet. In addition, the centralized design also results in the single-point-of-failure (SPF) or deny-of-service (DoS) attack.

Blockchain addresses two important problems that the electronic coin system has long faced: the dual payment problem and the Byzantine general problem [25]. It can solve the double payment problem of decentralized systems in the absence of local agencies through the verification of distributed nodes and consensus mechanism, and accomplish the value transfer in the process of information transmission [26]. One important application of the blockchain is the smart contract. Smart contracts are self-executing and autonomous computing protocols that facilitate the performance and execution of agreements between two or more parties. The advantages of smart contracts are numerous. They can provide better security performance than traditional contract law and reduce transaction costs associated with the negotiation, verification and enforcement of agreements [27].

There are two types of blockchains, one is public or permissionless blockchain (such as Bitcoin) and the other is permissioned blockchain [28][29] (such as Hyperledger Fabric). The main differences between them are the privacy and consensus algorithms.

In a permissionless blockchain anyone can participate without a specific identity in the process of block verification to create consensus and also create smart contracts. Permissionless blockchains typically involve a native cryptocurrency. One major issue of permissionless blockchain is the high latency of block generation. Permissioned blockchain restricts the actors who can contribute to the consensus of the system state. So, it provides a way to secure the interactions among a group of entities that have a common goal but which do not fully trust each other, such as businesses that exchange funds (finance), goods (supply chain), or information (public service). By relying on the identities of the peers, a permissioned blockchain can use traditional Byzantine-fault tolerant (BFT) [30], RAFT, or Paxos consensus.

With the blockchain technology, it brings multiple solutions for data privacy [31] and open banking can enable multiple application scenarios [32]. The first application scenario is the provision of peer-to-peer (P2P) transactions, such as cross-border payments and remittances based on P2P transactions. The second scenario is that a blockchain can be used as a reliable database to record all kinds of information with the characteristics of credibility and traceability, such as registration of the anti-money laundering information. The third scenario is the confirmation of rights, such as land ownership, equity and other contracts or property authenticity verification and transfer. The fourth scenario is intelligent management, where smart contracts can be used to automatically detect the trigger conditions and the contract will be automatically processed in automatic payment, participation in profit, etc.

Even though the data stored in the blockchain allows only the original user to read and that other users would need authorization to access, data privacy management still has many shortcomings:

1. Through the big data analysis technology, it is easily to locate those public keys of big transactions, and then locate the owner of the public key by the transaction time, counterparties and the open bank information, etc.

2. According to the GDPR, the company must completely delete their personal data after obtaining the citizens' requirements for data erasing. For banks and financial companies that use blockchain technology, the actual removal of this information is divided into different situations and conditions, and there are no clear rules currently.

3. In the data protection acts, banks need to redesign their own system, so as to adapt to collaborative work between blockchains and the huge applications of the banks, rather than simply moving bank data on or off blockchain. Moreover, the data disclosure schemes need to be confirmed by each customer, which is costly for the banks.

4. In some cases, such as the GDPR act, the definition of personal data is very broad. In principle, it covers any information related to identifiable, living individuals, or any data type that can be identified directly or indirectly. However, in a complete banking application scenario, the needs of customer's business are different. Moreover, when a customer requires to cancel his/her account, the bank needs to remove all the copies of relevant data shared across the whole banking system. The smart ledger for customer data privacy management strategies should also be complied with by different banks.

3. Data Privacy Management Framework

This section proposes the data privacy management framework and a technical prototype.

3.1 Characteristics of Financial Data and Framework

Banking customers are divided into individual and corporate customers. Banks gather data from these customers while conducting the finance business, and then use them for marketing purposes, product recommendation, and anti-fraud control and so on. The financial data, better known as financial customer data (reflecting the financial attributes of customer) includes capital transactions and customer portrait information. For example, the data entity of the customer in our bank, the Fujian Rural Credit Union (FRCU), is shown in Fig. 1.

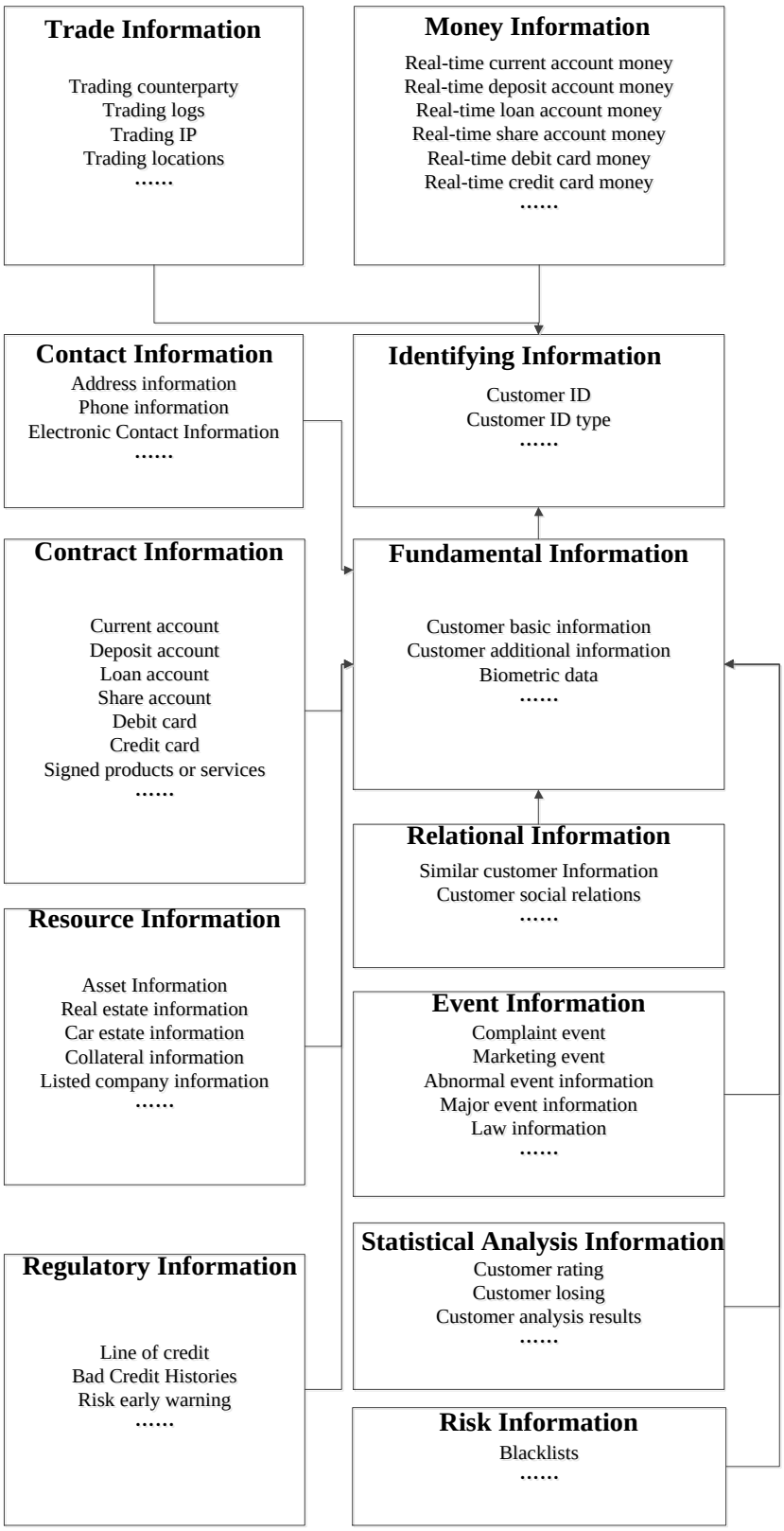

Fig.1 Customer Portrait Data Entity

Financial data privacy should be given priority including the identity of the basic information (such as name, address and ID number, etc.), network data (such as location, IP address, cookie data and RFID), biometric data (such as fingerprint, iris, etc.) and ethnic data. Data privacy management also needs to extend the access rights of the data subject, to clarify which banking institutions are used, and where and how to deal with the data. Therefore, the key points of data privacy management in financial block chain mainly involve three aspects:

Firstly, use the 'Privacy by Design' to implement the data privacy management. Banks should keep the handling of the minimal necessary financial business data and access to personal data to relevant personnel. It is necessary to implement the classification management of different dimensions of customers' information and provide the customers and regulators the facility for a quick inquiry and monitoring through the data retention characteristics of the blockchain.

Secondly, apply the effective technologies and algorithms to quickly build customer data disclosure schemes among the huge customer groups and reduce artificial contract signing.

Thirdly, realize the dynamic procedures of data to be put on chain and off chain with blockchain, enabling the regular update, additions, and removal of customer information.

Therefore, the designed framework is shown in Fig. 2.

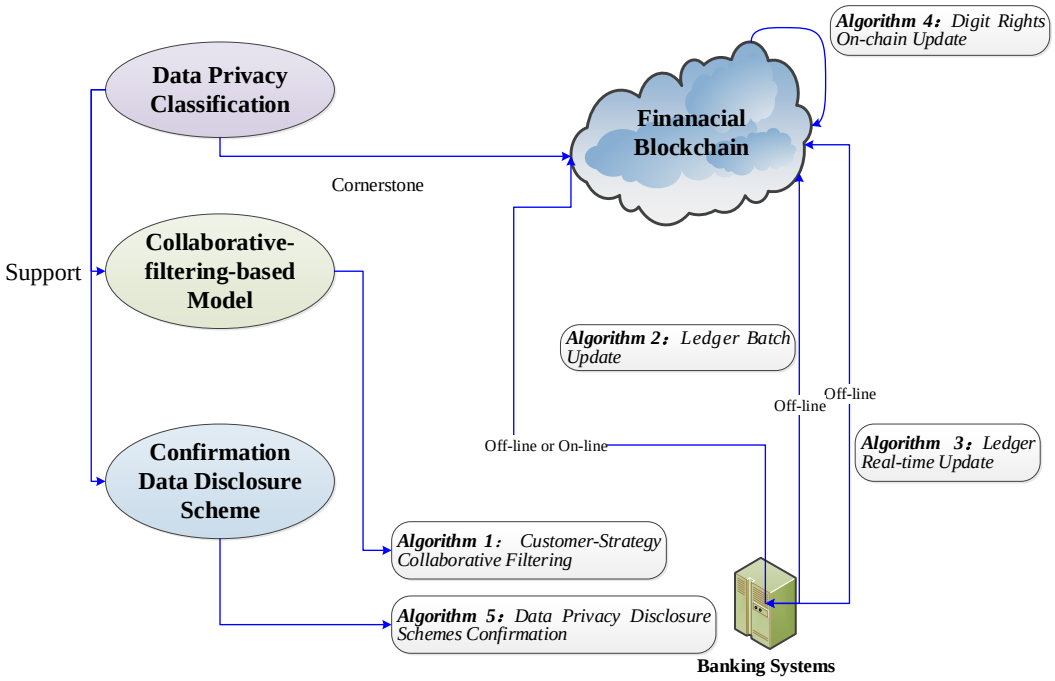

Fig. 2 Data Privacy Management Framework

The framework consists of three components: a data privacy classification method according to the characteristics of financial data, a new collaborative-filtering-based model and a confirmation data disclosure scheme for customer strategies based on the Nudge Theory. The privacy classification method is the cornerstone of the framework. Using the classification, the customer data disclosure schemes are confirmed by the collaborative-filtering-based model and nudging prompts. Five key algorithms are used in different data privacy protection scenarios.

3.2 Data Privacy Classification Method

Customer data, as shown in Fig 1, can be divided into basic information, identifying information, trade information, contact information, contract information, relational information, statistical analysis information, event information, resource information, money information, risk information and regulatory information. Therefore, the data privacy classification structures are:

1. Most stringent protection information (a.k.a. MSPI) -this type information is default not put on chain. Once it is put on the chain, it can only be accessed by the supervision parties, trading central bank and the original customer on the verified federation chain.

2. Customer sensitive information (a.k.a. CSI) - this type of information is selected by the customers to put on the chain, but must be allowed access by supervision parities.

3. Banking sensitive information (a.k.a. BSI) - as the private information required for the operation of the bank, this is determined by the bank itself to decide to put on the chain, sometimes is not on the chain.

4. Important information (a.k.a. II) – this refers to which type of customer information should be selected by the customer or authorized to the bank to put on the chain.

5. Public information (a.k.a. PI) - this must be disclosed on the chain.

Table 1 shows the common data classification strategies.

Table 1 Common financial data classification

CSI	**BSI**	**II**	**MSPI**	**PI**
Trading	Statistical analysis	Basic	Resource	Regulatory
Contact	Event	Identifying	Money	Risk
Contract				
Relational				

3.3 Prototype based on Data Privacy Management.

As banks need to deal with blockchain oriented data flow and interbank liquidation network (such as SWIFT) oriented capital flow when accessing to the financial chain.

Therefore, in the banking network, banks should take advantage of liquidation network to convert material money to digital rights, and at the same time, they would put the customer data on and off the chain through real-time or batch methods to share the data among banks. In the financial blockchain, there are also government, third-party service providers and regulatory agencies. Therefore, we did not choose the complex and resource-intensive consensus mechanism of permissionless blockchain. On the contrary, we selected the permissioned blockchain Hyperledger Fabric [30]. The enterprise customer information facility system (ECIF) is used to save the bank's customer unified information, to store the customer information off the chain, to meet the daily business needs of the bank. The on-chain data are organized from ECIF and bank core systems. The financial blockchain based data privacy framework is shown in Fig. 3.

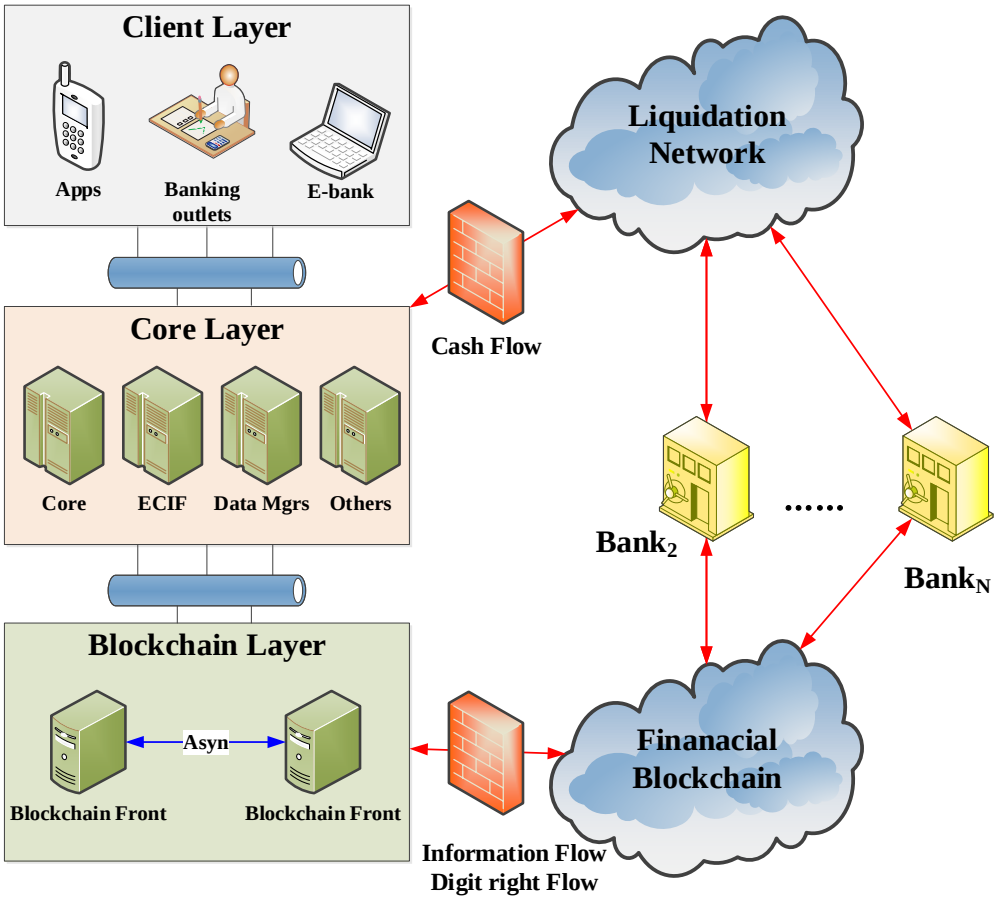

Fig. 3 Financial Blockchain based Framework

In order to elaborately facilitate the prototype, this paper assumes the following symbols:

$\bigcup Cus$: the customer dataset owned by each bank.

$Cus_i \in \bigcup Cus$: each customer belonging to the datasets.

$E = \{e_1, e_2, \ldots, e_n\}$: customer information, which e_{ij} represents the j-th subset in the i-th data type;

$R = \{r_1, r_2, \ldots, r_n\}$: digital rights.

$P = \{p_1, p_2, \ldots, p_n\}$: privacy policies.

C : condition attributes of customer-strategy collaborative filtering model.

$D = \{d_1, d_2, \ldots, d_n\}, d_i = \{p_t \mid e_{ij}\}$: decision attributes of customer-strategy collaborative filtering model, each element represents the e_{ij} chooses the p_t, and the d_i is not the default scheme.

M : the computed customer-strategy collaborative filtering model.

$S = \{s_{e_1d_1}, s_{e_1d_2}, \ldots, s_{e_nd_m}\}$: the computed disclosure schemes, which $s_{e_id_j}$ represents the i-th data type using the j-th disclosure scheme. The S includes the default schemes, so that $D \subseteq S$.

$T = \{t_{e_1}, t_{e_2}, \ldots, t_{e_n}\}$: nudging confirmation operations corresponding to e_i.

$U = \{u_{e_1}, u_{e_2}, \ldots, u_{e_n}\}$: customer information usage situations among banks.

$L = \{E、R、S、U\}$: smart ledger of blockchain.

Tr : transaction triggered by the banking business flow.

4. Nudging Collaborative Filtering Model and Promote Schemes

In this section, the data disclosure schemes are constructed using the Nudge theory, which reduces the manual operation and the massive transformation of the system. A *nudge* is "any aspect of the overall choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives." [33] And is a concept in behavioral science, political theory and economics. It gears not only for the customer, but also for the bank operators to know when and where to do relative data protection jobs [34]. This paper proposes to employ default data disclosure schemes of new customers with the customer-strategy collaborative filtering model. The disclosure schemes to be confirmed by customers are confirmed in steps embedded in the banking processes.

4.1 Customer-Strategy Collaborative Filtering Model

Based on our research findings, the habits of bank customers in the business implicit for the tendency of data privacy, this paper uses the collaborative filtering algorithm [35] to calculate the data disclosure schemes that are default to be set for a new customer and nudge the customers to make an initial decision. C is retrieved by the minimum set of reductions [36] : customer age, education, industry, position, income attribute, age, cross-sell score, financial term preferences, debit card consumer preferences, credit card spending preferences, the potential loan customers, credit card potential customers, debit card potential customers, forex potential customers, credit card high frequency transactions, customer loyalty, investment preferences.

Combined with the data privacy classification method (as shown in Table 1), the designed P is shown as follow:

p_1 : Confirmation prompts when the customers do business on phone bank.

p_2: Confirmation prompts by the staff service.

p_3: Confirmation prompts by customer manager.

p_4: Confirmation prompts by lobby manager.

p_5: Confirmation prompts by the bank teller.

p_6: Authorize the bank to disclosure.

p_7: Non-disclosure.

p_8: Disclosure.

Each $e_{i,j}$ can choose one policies p_j as D. Algorithm 1 describes customer-strategy collaborative filtering model in detail.

Algorithm 1: Customer-Strategy Collaborative Filtering

Input: Cus_{new}, $\cup Cus\_chg$

Output: M, S_{new}

1. Check M, if M ==null, go to step 2.

2. Start calculating the scores of each customer in collaborative filtering algorithm.

2.1 Calculate the cosine similarity set by C:

$SIM = \cup Sim(Cus_i, Cus_j)$.

2.2 Calculate the nearest neighbor set $NEIGHBOR$:

$Neighbor(Cus_i) = \{Cus_j, Sim(Cus_i, Cus_j) isTop5\}$.

2.3 Build the schemes D set for each Cus_i.

2.4 Combine the C and D to M.

3. Towards $\cup Cus\_chg$, reconstruct D, go to step2.

---

4. Towards Cus_{new} , start calculating the recommend schemes in collaborative filtering algorithm.

4.1 Calculate the top 5 nearest neighbor set by C.

$$Neighbor(Cus_{new}) = \{Cus_{new\,j}, Sim(Cus_{new}, Cus_j) isTop5\}$$

4.2 Calculate D of Cus_{new}:

$$d_i = p_t\left(\frac{\sum_{Neighbor(Cus_{new})} ((p_t \mid e_{ij} \in E_{Cus_{new\,j}}) * Sim(Cus_{new}, Cus_{newj}))}{\sum_{Neighbor(Cus_{new})} Sim(Cus_{new}, Cus_{newj})}\right)$$

4.3 Calculate S_{new} by combining the D and the default p_j of e_{ij}.

4.2 Nudging Customers to Determine Data Disclosure Schemes in Business Scenarios

In P, the first six schemes need to be confirmed. This section will further design suitable operations for customers and bank staffs embedded in the different business scenarios and system processes, but not add additional tasks for them, namely using the Nudge theory to subtly confirm Data disclosure solution.

Different information with different P should use different T. For example, trading information is mainly used for the internal risk monitoring of banks, and does not require disclosure except for the third-party applications access for verification and reconciliation. In the meanwhile, money information is usually used only when it is accessed by third-party applications, and for this reason, it can be used as a non-disclosure in most cases, and it is authorized by the customer to disclose to the fixed authorized institutions. Table 2 shows the different processing methods under different customer data disclosure schemes and different prompts.

Table 2 Different prompts of different data privacy management policies

Info.	Nudging Schemes T in P
Trading	p_1: When customers deal with the business, it is necessary to provide a reconciliation check whether to disclosure the trading record to the third party so as to ensure the correctness of the client's funds. If not permitted, the customer should accept the money criterion is based on the third-party. p_6: Customers authorize to the bank.
Money	p_6: Customers authorize to the bank.
Basic	$p_1 \sim p_6$: When opening an account after the contract signing or products marketing, it will have tips for confirmation that the basic information of the customer can be disclosed.
Relation	p_3: Customer managers conduct customer information surveys and inquiries and then ask customers if they are willing to disclose relational information to support the supply chain information and get better secured loan services. p_5: When custombe consult on some products that are bought by his relations, the bank tellers inquire the confirmation. p_6: When a company handles the payroll service, it applies to the bank to disclose information on its own employees.
Contact	p_1: When activating mobile phone banking and other E-bank applications, fill in contact information and prompt for confirmation. p_2: The customers are asked through the customer service hotline. p_3: When handling the loan, the customer manager fills in contact information and inquires on the confirmation

	p_4: The business process triggers the SMS notification procedure and pops up a prompt p_5: When opening the account, the contact information is confirmed together with the basic customer information.
Resource	p_3: Customer managers inquire when marketing. Or the business process may trigger a special application to obtain the disclosure of the resource information. p_6: Customers authorize to the bank.
Contract	p_1: When opening the corresponding service on the E-bank, the systems start to prompt. p_2: Consulting product information prompt p_3: Marketing products prompt. p_4: When the customer enters the branch, the lobby manager will be prompted to conduct marketing when he obtains the customer's purchase information. p_5: Trigger when a business is started.
Risk	p_6: Choices made by the banks
Event	$p_2 \sim p_3$: When a customer manager or an online staff does a customer marketing survey, he asks the customer if information can be disclosed for a particular event record.

5. Blockchain based Data Privacy Management

This section describes the implementation of a blockchain-based data privacy management prototype including the smart ledger, smart contract, and data on-chain and off-chain switching algorithms.

5.1 Customer Smart Ledger and Smart Contract

A customer's smart ledger includes customer information, digital rights, customer data disclosure schemes, and customer information usage situations [37]:

$L = \{E, R, S, U\}$.

The digital right R is the blockchain currency, released in the specific application scenarios, and customer information usage situations U filled by each bank explain how the disclosure data is to be used by different application systems and scenarios, and let the customer know how his/her data is used. Fig. 4 shows the smart ledger of customers.

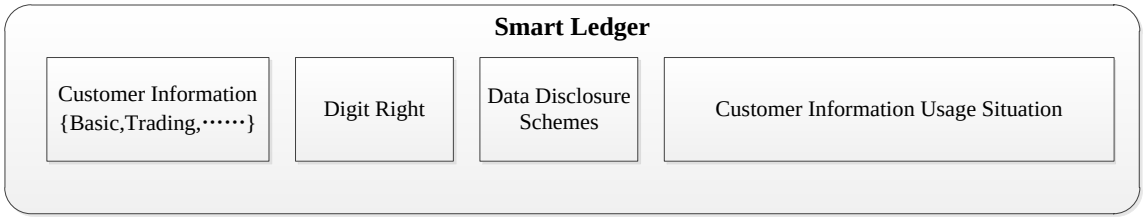

Fig. 4 Smart Ledger of Customers

When the $S_{e_i d_j}$ is changed to be non-disclosure, the smart contract [38] should be triggered to lock all the e_{ij} to be not visible by blockchain proprietary key.

Smart Contract: Encrypting the Non-disclosure Customer Information

Input: $S_{e_i d_j}$

Output: encrypted e_{ij}

1. Check $S_{e_i d_j}$, if $S_{e_i d_j}$ equals non-disclosure, then go to step 2, otherwise go to step 4.

2. Search the nearest L in blockchain, if $S_{e_i d_j}$ is changed then go to step 3,

otherwise go to step 4.

3. Using the blockchain proprietary key to encrypt the e_{ij}.

4. Update e_{ij} to the chain for consensus confirmation.

5.2 Data on-chain and off-chain Switching Algorithms

The encrypted smart ledgers are often updated by batch and real time [36], and Algorithm 2 and 3 show these procedures.

Algorithm 2: Ledger Batch Update

Input: Batch task, $L_{\text{on}-chain}$

Output: L_{new}

1. Banks schedule the batch task to synchronize the smart ledger, download $L_{\text{on}-chain}$.

2. For $1 \in L_{\text{on}-chain}$, check 1 in the ECIF.

3. If it is non-existence of ECIF, synchronize E to ECIF.

4. If it is existence of ECIF, update:

4.1 For S in $L_{\text{on}-chain}$, compared with the S strategy in the ECIF off the chain, and the data is updated by the $S_{e_{ij}d_k}$:

If $S_{e_{ij}d_k}$ is non-disclosure, set e_{ij} in the internal data analysis system is not available, and at the same time prevent the e_{ij} corresponding system to transmit;

If $S_{e_{ij}d_k}$ is disclosure, then put the local e_{ij} in the L_{new}, and set e_{ij} in the internal data analysis system is available;

If $S_{e_{ij}d_k}$ is authorized to banks to disclosure, then the decryption data is extracted by using its private key.

4.2 Put L_{new} on the chain for consensus confirmation.

Algorithm 3: Ledger Real-time Update

Input: Tr, $L_{on-chain}$

Output: L_{new}

1. Check the Tr.

2. If Tr is using or changing the customer data Cus_i, reads the $L_{on-chain}$:

2.1 For $S_{e_{ij}}$ is disclosure, then update e_{ij} to ECIF;

2.2 For $S_{e_{ij}}$ is authorized to banks to disclosure, then the decryption data is extracted by using its private key;

2.3 If $S_{e_{ij}}$ is non-disclosed, the business operation uses the data in the local ECIF.

3. If Tr is the account opening, for $l \in L_{on-chain}$, check l in the ECIF. If it does exit, then synchronizes, if not, then check $L_{on-chain}$ whether the customer information exists:

3.1 If exists, then download $L_{on-chain}$ directly;

3.2 If does not exist, banks store the information in the ECIF, calculate the S based on Algorithm 1, and construct the L_{new}.

3.3 Put L_{new} on the chain for consensus confirmation.

Algorithm 4 shows the digit rights of customer privacy data on-chain method.

Algorithm 4: Digit Rights On-chain Update

Input: Tr, $L_{on-chain}$

Output: L_{new}

1. Check the traction type.

2. If Tr is registration.

2.1 Initiative bank establishes the digital rights D_i;

2.2 Transfers money through liquidation network to the D_i.

2.3 Initiative bank construct the L_{new} and put L_{new} on the chain for consensus confirmation.

3. If Tr is recharging.

3.1 Banks launch money through the liquidation network, the initiative bank updates the value of D_i to consensus on confirmation

3.2 Refreshes the corresponding value of electronic account.

3.3 Refreshes the corresponding value of banking account.

4. If Tr is cashing.

4.1 The initiative bank deducts the value of D_i to do consensus confirmation.

4.2 Unfreeze the electronic account.

5. If Tr is cancellation.

5.1 Remove the D_i in L_{new}, and set S_{d_i} =non-disclosure.

5.2 Put L_{new} on the chain for consensus confirmation.

The last algorithm shows the data disclosure schemes confirmation processes.

Algorithm 5: Data Disclosure Schemes Confirmation

Input: Tr, $L_{\text{on}-chain}$

Output: L_{new}

1. The banking core system charged the account.

2. Using the common ECIF trading operations ECP which extracts the modified and

changed information to the X set as XML:

2.1 If $S_{e_{i,j}d_k}$ is disclosure, then stores the data $e_{i,j}$ in the X;

2.2 If $S_{e_{i,j}d_k}$ is authorized to banks to disclosure, then encrypt the data with the bank's public key and save it to X;

2.3 If $S_{e_{i,j}d_k}$ is between p_1 and p_6, then use the T_p.

2.3.1 If confirmation is agree, the data is saved to X， and encrypts the operation footprints and saves it to the ECIF, changes $S_{e_{i,j}d_k}$ to disclosure;

2.3.2 If confirmation is disagree, update S and save it to $\cup Cus\_chg$;

2.4 If $S_{e_{i,j}d_k}$ is non-disclosure, then skips.

3. Combine X and S to L_{new}, put L_{new} on the chain for consensus confirmation, and trigger the Smart Contract.

4. If customers want to change $S_{e_{i,j}d_k}$ that is disclosure, they use the E-bank or teller service to directly change $S_{e_{i,j}d_k}$, and then trigger Algorithm 2.

6. Experiments

This section presents the field experiments and laboratory experiments to validate our proposed framework. The field experiments are sampling statistical analysis to validate that a customer's financial characteristics implies its tendency to data privacy protection, thereby verifying the usability of the customer-strategy collaborative filtering model. The laboratory experiments include the evaluation of proposed model's recommending scheme capabilities and the ability of blockchain security and throughput.

6.1 Field Experiments

1. The general conclusion is that customers with loans are more willing to disclosure information.

This experiment conducted a telephone interview with 1000 loan customers randomly selected to communicate and understand their willingness to disclosure the personal information to achieve better banking services. In the meanwhile, 1000 customers that have not handled the loan business are selected. The entire interview gets valid responses. The results are shown in Fig. 5 (Y denotes “Agree to disclosure information”, N denotes “Disagree with disclosure information”).

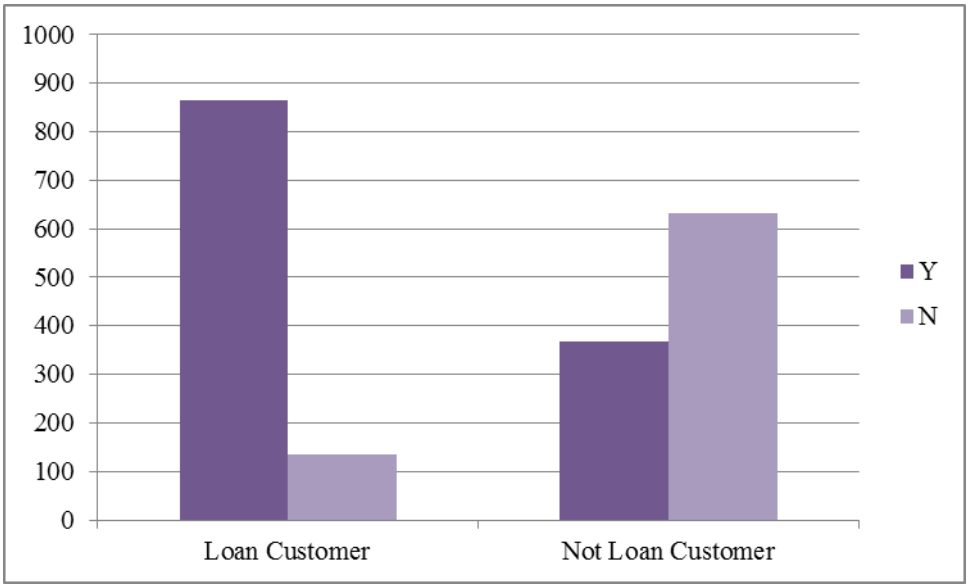

Fig. 5 Loan Customer toward Disclosure Preference

It can be concluded that customers with loans are more likely to agree to disclosure information.

2. The general conclusion is that customers with smaller ages and higher education levels are more willing to disclosure information.

First, the age of customers is divided into: under 20 years old, 20-30 years old, 30-40 years old, 40-50 years old, and over 50 years old. At present, the age structure of FRCU is 1:2:2:3:2. According to the number of people of all ages, stratified random sampling

corresponding customers to conduct telephone interviews, communicate and understand their willingness to disclosure the personal information to achieve better banking services. Secondly, the customer education background is divided into: doctor, master, undergraduate, high school, junior high school. The current FRCU academic structure is 1:2:9:6:2. According to the total number of students in each academic period, stratified random sampling corresponding customers to conduct telephone interviews, communicate and understand their willingness to disclosure the personal information to achieve better banking services. The ratios of agreement among different ages and education are shown in Fig. 6.

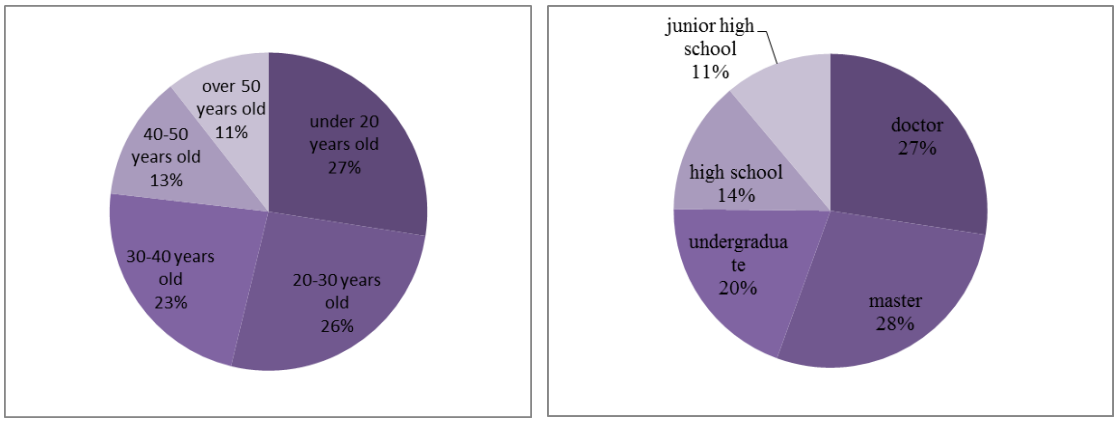

Fig. 6 Age and Education toward Disclosure Preference

It can be concluded that customers with younger age and higher education levels are more likely to agree to disclosure information. Then from these field experiments it implies the conclusion that the collaborative filtering model can work well by that the similar customer have the similar data privacy schemes.

6.2 Laboratory Experiments

In this section, we have conducted a realistic experiment to verify the proposed framework. This experiment was conducted on a PC with a 4-core CPU, 8GB Memory connected with LAN with bandwidth 100Mbps.

1. Data privacy use cases

We choose some frequently-used financial operations to evaluate the data privacy preservation of the entire framework. Table 3 gives the analysis of these scenarios.

Table 3 Privacy Preservation Analysis

Scenarios	Operations	Privacy preservation analysis
Add the new customer	1. Create a new customer in bank A; 2. Set the attribute disclosure; 3. Save and consensus; 4. Bank B synchronizes the smart ledger.	Bank B can get the new customer.
	1. Create a new customer in bank A; 2. Set the attribute non-closure; 3. Save and consensus; 4. Bank B synchronizes the smart ledger.	Bank B cannot get the new customer.
Update a customer	1. Choose a disclosure customer in bank A; 2. Change the age attribute; 3.Save and consensus; 4. Bank B synchronizes the smart ledger.	Bank B can see the new age data.
	1. Choose a disclosure customer in bank A and money attribute only disclosure to bank C; 2. Change the age and money attribute; 3.Save and consensus; 4. Bank B synchronizes the smart ledger; 5. Bank C synchronized the smart ledger.	Bank B can see the new age data but cannot see the new money data. Bank C can see all changed data.
	1. Choose a non-disclosure customer in bank A; 2. Change the age attribute; 3.Save and consensus; 4. Bank B synchronizes the smart ledger.	Bank B cannot see the new age data.
	1. Choose a non-disclosure customer in bank A; 2. Change the schemes to the disclosure; 3. Change the age attribute; 4. Save and consensus; 5. Bank B synchronizes the smart ledger.	Bank B can see the new age data.

As shown in Table 3, our proposed framework can ensure the data preservation.

2. Collaborative filtering model's recommending scheme capabilities.

The experiment was placed in a bank test environment and the desensitized customer data disclosure schemes are selected. The number of customers in the dataset is 20000, and e_{ij} covers 100 attributes, wherein each determined policy is recorded as one line, a total of 450 thousand data items. The training set and the test set are divided according to 4:1. The evaluation indicators use the classic precision, recall and improved precision.

$$\mathrm{Pr}\,ecision = \frac{|\cap(S_{prediction}, S_{reference})|}{|S_{prediction}|}$$

$$\mathrm{Re}\,call = \frac{|\cap(S_{prediction}, S_{reference})|}{|S_{reference}|}$$

$$\mathrm{Im}\,proved\,\mathrm{Pr}\,ecision = \frac{|\cap_{inSet(p_1 \sim p_t)}(S_{prediction}, S_{reference})|}{|S_{prediction}|}$$

The improved precision means that schemes are confirmed type is correct. This indicator shows the model can generate the same type schemes although not same schemes. The results are shown in Table 4.

Table 4 Results of Models

Precision	Improved Precision	Recall
83%	100%	21%

The experimental results show that the precision is high, and the schemes that need to be confirmed are correctly calculated. Although the recall rate is low, since the data has a default policy guarantee, the security of the customer data can be guaranteed by the nudging operation.

3. Blockchain security and throughput abilities.

In this part, we choose the blockchain private information encryption test and the blockchain interface permission control test verify the security of the chosen blockchain. Table 5 shows a comparison of the technical specifications of Fabric and Ethereum.

Table 5 Comparison of the technical specifications of Fabric and Ethereum

Indicator	Fabric	Ethereum
Concesus	Kafka	PoW
Smart Contract	Chainnode	EVM

Database	Go-leveldb	Go-leveldb
Access control	CA	N/A
Throughput	Average：10000	<100
	Peak：30000	
Transaction delay	<300ms	14s

Fig. 7 shows the throughput of the blockchain under different node counts on a PC with a quad-core CPU, 8GB Memory, where we choose transactions per second (TPS) as the performance metric. It can be seen that the blockchain has a higher TPS and can meet the needs of data on-chain and off-chain. As a normal bank would have average 1000 accounting transactions per second, so the proposed blockchain system can fulfill the real-time requirement in this banking scenarios.

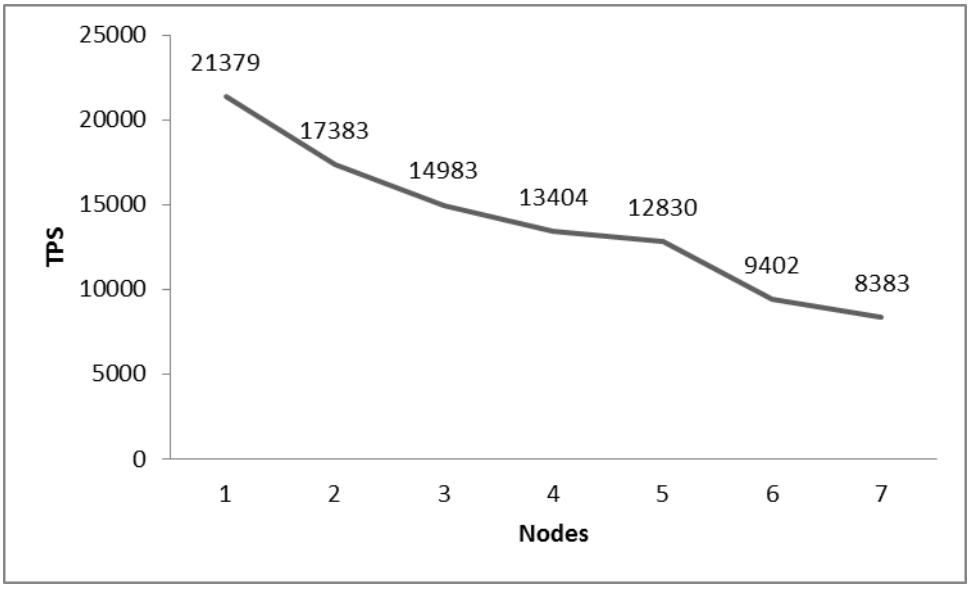

Fig. 7 Throughput of the Blockchain

We next evaluate the fault tolerance of the proposed blockchain system via conducting several experiments. In particular, we adopt four nodes in our experiments. First, we shut down two nodes and keep two nodes running. We then initiate a legal transfer request in our blockchain system. Fig. 8 shows that normal transactions can still

be processed even when two nodes are shut down. process after a downtime that does not affect normal transactions as shown in Fig.8.

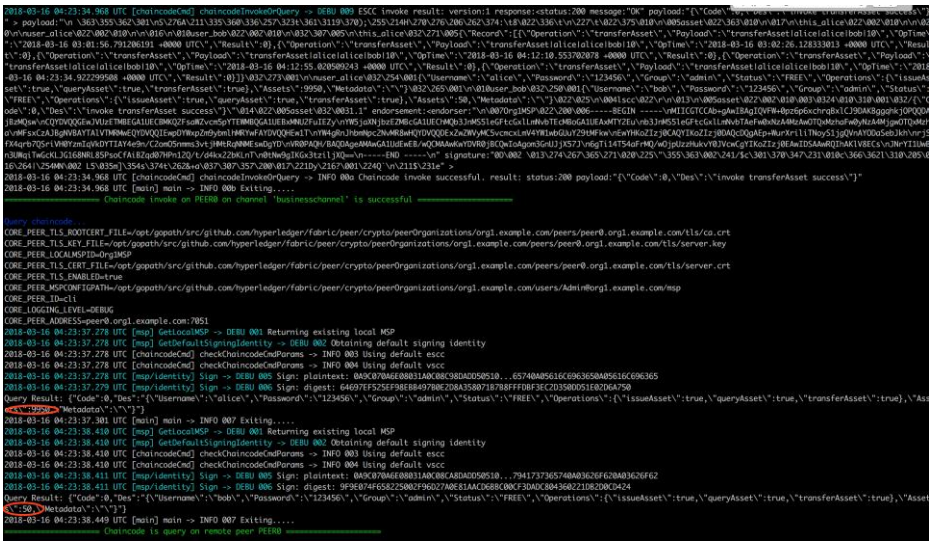

Fig. 8 Fault tolerance under 2 nodes

We next shut down another node and only one node is running. Fig. 9 then shows that the consensus cannot reach in this case. This result implies that at least two nodes are the necessity for the system.

Fig. 9 Fault tolerance under 1 node

The blockchain interface permission control test procedure is:

(1) Create an account and assign the query operation permission to the account.

(2) Use this account to perform the query operation and the query is successful.

(3) Change the query permission of the account to null.

(4) Use this account to perform the query again. The query fails and the prompt does not have permission.

The blockchain private information encryption test procedure is as follows:

(1) Create an account.

(2) Private data is encrypted using openssl.

(3) View the encrypted cipher text.

(4) Use openssl to decrypt the cipher text, consistent with the original text before encryption.

The results are shown in Fig. 10 and Fig. 11, respectively.

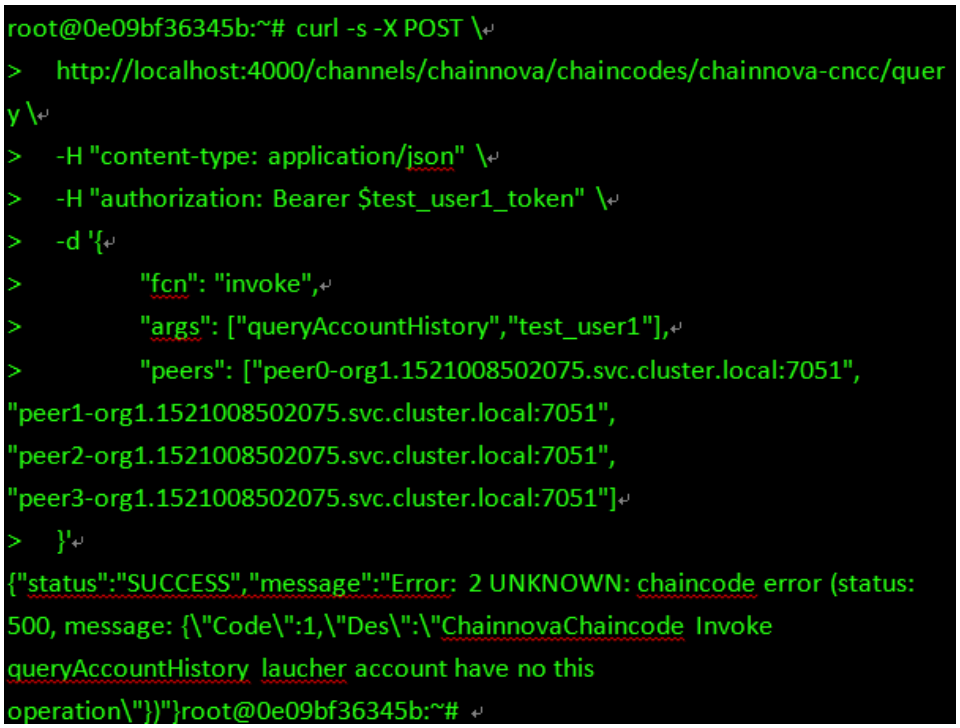

Fig. 10 The Return Message of Permission Reject

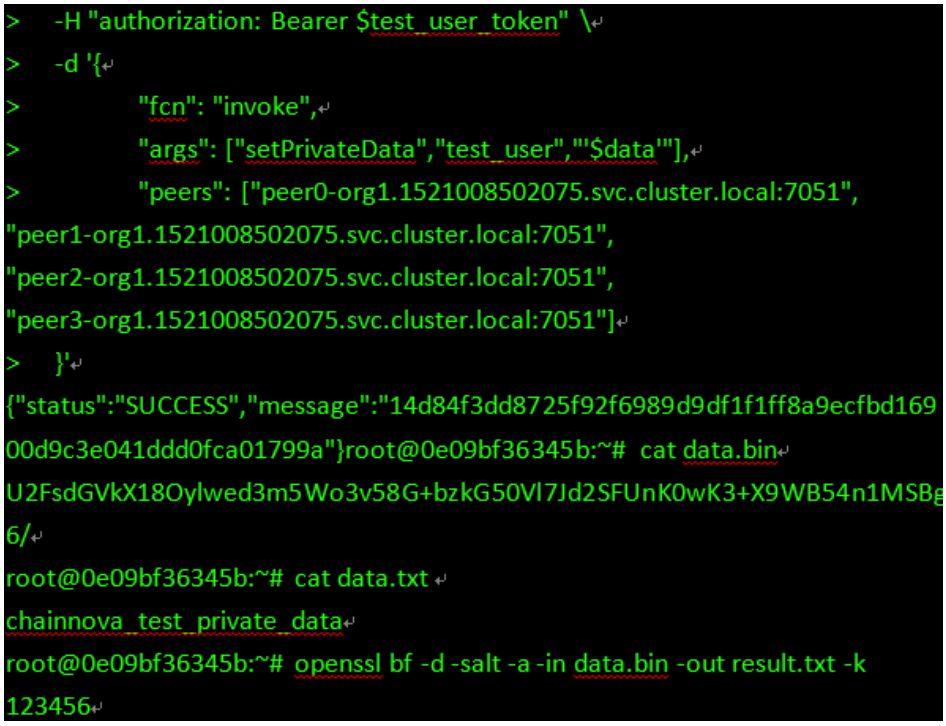

Fig. 11 The Return Message of Decryption

Experiments verify that the chosen blockchain in this paper meets the security requirements.

7. Conclusions

Customer data privacy management is one of the most important components in open banking. How to deal with customer profiles with respect to customers right on privacy is still a challenge. In this paper, we design a data privacy management framework according to the characteristics of banking data. We propose the customer-strategy model of collaborative filtering algorithm and the confirmation of default data disclosure schemes based on the Nudge Theory. Finally, we implement a blockchain-based financial data privacy management prototype. The experiments show that the proposed framework meets the reality in banking data privacy management.

Our future work will incorporate the testing of existing secure and scalable blockchain and the design of a layered architecture for financial applications, such as

loan management [39], with hybrid blockchain and feature engineering technologies. The nudging schemes will be enhanced to consider the aspects of different product services [40]. In addition, we would strengthen the management of third-party data users in the financial blockchain, ensuring that the deleted data required by the customer can be completely erased or frozen in the databases of the third parties.

Acknowledgements

This work is partially funded by the Fujian Fumin Foundation and is partially supported by the Science and Technology Planning Project of Guangdong Province (No. 2017A050501035), Science and Technology Program of Guangzhou (No. 201807010058), and the Deanship of Scientific Research, King Saud University through research group number RG-1435-051.

References

[1] Avital M, Hedman J, Albinsson L, et al. Smart Money: Blockchain-Based Customizable Payments System. Dagstuhl Reports, 2017, 7(3): 104-106.

[2] Wang H, Guo H. Achieving fairness in wireless environment//Emerging Technologies: Frontiers of Mobile and Wireless Communication, 2004. Proceedings of the IEEE 6th Circuits and Systems Symposium on. IEEE, 2004, 1: 117-120.

[3] Wang H, Guo H, Lin M, et al. A new dependable exchange protocol. Computer communications, 2006, 29(15): 2770-2780.

[4] Voigt P, von dem Bussche A. The EU General Data Protection Regulation (GDPR).

[5] Kröner M. API deep dive: Who will thrive in an open banking world? Why meeting regulatory requirements is not enough for banks to remain relevant. Journal of Digital Banking, 2018, 2(3): 198-203.

[6] Omohundro S. Cryptocurrencies, smart contracts, and artificial intelligence. AI matters, 2014, 1(2): 19-21.

[7] Gaetani E, Aniello L, Baldoni R, et al. Blockchain-based database to ensure data integrity in cloud computing environments. 2017.

[8] Liu P T S. Medical record system using blockchain, big data and tokenization. International Conference on Information and Communications Security. Springer, Cham, 2016: 254-261.

[9] Conoscenti M, Vetro A, De Martin J C. Blockchain for the Internet of Things: A systematic literature review. Computer Systems and Applications (AICCSA), 2016 IEEE/ACS 13th International Conference of. IEEE, 2016: 1-6.

[10] Hou H. The Application of Blockchain Technology in E-Government in China. Computer Communication and Networks (ICCCN), 2017 26th International Conference on. IEEE, 2017: 1-4.

[11] Frey R, Wörner D, Ilic A. Collaborative Filtering on the Blockchain: A Secure Recommender System for e-Commerce. 2016.

[12] Manset D. Big Data and Privacy Fundamentals: Toward a “Digital Skin”//The Digitization of Healthcare. Palgrave Macmillan, London, 2017: 241-255.

[13] Gatteschi V, Lamberti F, Demartini C, et al. Blockchain and smart contracts for insurance: Is the technology mature enough? Future Internet, 2018, 10(2): 20.

[14] Kosba A, Miller A, Shi E, et al. Hawk: The blockchain model of cryptography and privacy-preserving smart contracts. 2016 IEEE Symposium on Security and Privacy (SP 2016), IEEE, 2016: 839-858.

[15] Atzori M. Blockchain technology and decentralized governance: Is the state still necessary? 2015.

[16] Yue X, Wang H, Jin D, et al. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. Journal of medical systems, 2016, 40(10): 218.

[17] Azaria A, Ekblaw A, Vieira T, et al. Medrec: Using blockchain for medical data access and permission management. International Conference on Open and Big Data (OBD), IEEE, 2016: 25-30.

[18] Sikorski J J, Haughton J, Kraft M. Blockchain technology in the chemical industry: Machine-to-machine electricity market. Applied Energy, 2017, 195: 234-246.

[19] Sreehari P, Nandakishore M, Krishna G, et al. Smart will converting the legal testament into a smart contract. Networks & Advances in Computational Technologies (NetACT), 2017 International Conference on. IEEE, 2017: 203-207.

[20] Sharples M, Domingue J. The blockchain and kudos: A distributed system for educational record, reputation and reward. European Conference on Technology Enhanced Learning. Springer, Cham, 2016: 490-496.

[21] Peters G W, Panayi E. Understanding modern banking ledgers through blockchain technologies: Future of transaction processing and smart contracts on the internet of money. Banking Beyond Banks and Money. Springer, Cham, 2016: 239-278.

[22] Agrawal R, Srikant R. Privacy-preserving data mining. ACM Sigmod Record. ACM, 2000, 29(2): 439-450.

[23] Dai W, Qiu M, Qiu L, et al. Who moved my data? privacy protection in smartphones. IEEE Communications Magazine, 2017, 55(1): 20-25.

[24] Zyskind G, Nathan O. Decentralizing privacy: Using blockchain to protect personal data. Security and Privacy Workshops (SPW), 2015 IEEE. IEEE, 2015: 180-184.

[25] Antonopoulos A M. Mastering Bitcoin: unlocking digital cryptocurrencies. O'Reilly Media, Inc., 2014.

[26] Fan J, YI LT S. Research on the technologies of Byzan-tine system. Journal of Software, 2013, 24(6): 1346-1360.

[27] Maffè C A C. Future of the CIO: Towards an Enterpreneurial Role. CIOs and the Digital Transformation. Springer, Cham, 2018: 61-68.

[28] Dinh T T A, Liu R, Zhang M, et al. Untangling blockchain: A data processing view of blockchain systems. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(7): 1366-1385.

[29] Vukolić M. Rethinking permissioned blockchains. Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts. ACM, 2017: 3-7.

[30] Elli Androulaki, Artem Barger, Vita Bortnikov, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Christopher Ferris, Gennady Laventman, Yacov Manevich, et al. 2018. Hyperledger fabric: a distributed operating system for permissioned blockchains. In Proceedings of the Thirteenth EuroSys Conference. ACM, 30.

[31] G Feng, L Zhu, S Meng, K Sharif, Z Wan. A blockchain-based privacy-preserving payment mechanism for vehicle-to-grid networks. IEEE Networks, 2018(99):1-9.

[32] Guo Y, Liang C. Blockchain application and outlook in the banking industry. Financial Innovation, 2016, 2(1): 24.

[33] Wilk J. Mind, nature and the emerging science of change: An introduction to metamorphology. Metadebates on Science. Springer, Dordrecht, 1999: 71-87.

[34] Corrales M, Jurcys P, Kousiouris G. Smart Contracts and Smart Disclosure: Coding a GDPR Compliance Framework. 2018.

[35] Popescul A, Pennock D M, Lawrence S. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 2001: 437-444.

[36] Ma S, Ye D. Research on computing minimum entropy based attribute reduction via stochastic optimization algorithms. Moshi Shibie yu Rengong Zhineng/Pattern Recognition and Artificial Intelligence, 2012, 25(1):96-104.

[37] Mainelli M, Smith M. Sharing ledgers for sharing economies: an exploration of mutual distributed ledgers (aka blockchain technology). 2015.

[38] Watanabe H, Fujimura S, Nakadaira A, et al. Blockchain contract: A complete consensus using blockchain. 2015 IEEE 4th Global Conference on Consumer Electronics (GCCE 2015), IEEE, 2015: 577-57.

[39] Wang H, Guo C, Cheng S. LoC - A New Financial Loan Management System based on Smart Contracts. Future Generation Computer Systems. In press.

[40] Wang H, Ma S, Dai H-N, A Rhombic Dodecahedron Topology for Human-Centric Banking Big Data, IEEE Transactions on Computational Social Systems.IEEE. In press.