Wang M, Wu Q, Qin B *et al.* Lightweight and manageable digital evidence preservation system on Bitcoin. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 33(3): 568–586 May 2018. DOI 10.1007/s11390-018-1841-4

Lightweight and Manageable Digital Evidence Preservation System on Bitcoin

Mingming Wang[1], Qianhong Wu[1], *Member, CCF, ACM, IEEE*, Bo Qin[2], Qin Wang[1]
Jianwei Liu[1], *Member, CCF, IEEE*, and Zhenyu Guan[1,*], *Member, IEEE*

[1] *School of Electronic and Information Engineering, Beihang University, Beijing 100191, China*
[2] *School of Information, Renmin University of China, Beijing 100872, China*

E-mail: {wangmingming, qianhong.wu}@buaa.edu.cn; bo.qin@ruc.edu.cn; wangqin0409@foxmail.com
{liujianwei, guanzhenyu}@buaa.edu.cn

Received November 20, 2017; revised March 26, 2018.

Abstract An effective and secure system used for evidence preservation is essential to possess the properties of anti-loss, anti-forgery, anti-tamper and perfect verifiability. Traditional architecture which relies on centralized cloud storage is depressingly beset by the security problems such as incomplete confidence and unreliable regulation. Moreover, an expensive, inefficient and incompatible design impedes the effort of evidence preservation. In contrast, the decentralized blockchain network is qualified as a perfect replacement for its secure anonymity, irrevocable commitment, and transparent traceability. Combining with subliminal channels in blockchain, we have weaved the transaction network with newly designed evidence audit network. In this paper, we have presented and implemented a lightweight digital evidence-preservation architecture which possesses the features of privacy-anonymity, audit-transparency, function-scalability and operation-lightweight. The anonymity is naturally formed from the cryptographic design, since the cipher evidence under encrypted cryptosystem and hash-based functions leakages nothing to the public. Covert channels are efficiently excavated to optimize the cost, connectivity and security of the framework, transforming the great computation power of Bitcoin network to the value of credit. The transparency used for audit, which relates to the proof of existence, comes from instant timestamps and irreversible hash functions in mature blockchain network. The scalability is represented by the evidence chain interacted with the original blockchain, and the extended chains on top of mainchain will cover the most of auditors in different institutions. And the lightweight, which is equal to low-cost, is derived from our fine-grained hierarchical services. At last, analyses of efficiency, security, and availability have shown the complete accomplishment of our system.

Keywords blockchain, covert channel, data auditing, digital evidence

1 Introduction

Digital evidence, as the derivatives from the vast E-commerce and network communication activities, has been playing an increasingly pivotal role in the areas of copyright protection, virtual property, commercial contracts and so on. For forensics and management, it on the one hand enjoys the convenience and repeatability, but on the other hand suffers its unique vulnerability. The risk of forgery, tampering, and file corruption all might make it totally valueless. Therefore, a stable, safe and reliable environment is essential for evidence preservation. In the meanwhile, the process of auditing, which builds the legal force towards digital evidence, is also crucial to be verifiable and regulatory under a reliable data network.

Currently, most of the evidence preservation systems are based on the third-party centralized storage structure, which may unavoidably result in the following problems:

Regular Paper

Special Section on Blockchain and Cryptocurrency Systems

This work is supported by the National Key Research and Development Program of China under Grant No. 2017YFB0802500, the National Natural Science Foundation of China under Grant Nos. 61772538, 61672083, 61370190, 61532021, 61472429, and 61402029, and the National Cryptography Development Fund of China under Grant No. MMJJ20170106.

[*] Corresponding Author

- *Single-Point Failure Problem.* Centralized institutions always take on enormous safety pressure. Once the central storage node is invaded, serious and irretrievable problems happened like information leakage and data forgery.

- *Opaqueness.* Highly centralized structure causes terrible transparency and absent confidence. It cannot be resolved due to centralized operations. People always doubt about whether the services are normative, authentic or trustful.

- *Inconsistent Policies.* Various systems specify different policies and regulations which are mutually incompatible and short of interworking. It results in negative impacts on the use of forensics as well as the system scalability.

In comparison, the decentralized blockchain network serves transparently reliable and verifiably secure environment. It safeguards data through large computing power. Trusted timestamp can be instantly attached to a newly generated block. Most importantly, trust problems could be avoided via distributing power from auditors. It proves the integrity, accuracy and timeliness needed in preservation.

The idea of employing blockchain network on preserving data is reasonable. Since Bitcoin becomes worldwide, people are trying to exploit the potential value of the blockchain. It consequently opens up the field of PoE (Proof of Existence) where various different services offer decentralized trusted timestamping service[1]. According to the construction mechanism, the services can be divided into the following two types.

- One type of services build directly on top of the mature blockchain network such as Bitcoin and Ethereum. The data is calculated and encoded into the blockchain in form of securely irreversible digests. Famous services are exemplified as OriginStamp①, Bitproof②, Proof of Existence[2], etc.

- The other type of services create their own decentralized network and establish their application ecology. Such services start a new chain demerged from the mainchain that they can embed more flexible functions and applications. There exist famous services like Factom[3], Florincoin (FLO)③.

Compared with the centralized design, the second approach has a greater flexibility in both rules and spaces. However, the short of enough computing power raises the risk of attacks and forgeries. Although innovative techniques such as pegged sidechain[4] and anchor transaction can establish contacts with blockchain from mature cryptocurrency in some way, more time and power are still needed to be devoted to networks for enough confidence. Due to the sensibility of the service, the situation of over-concentrated computing power tends to appear and thus violates the effect of creditability. More seriously, if the correlative service breaks down, users would lose all the proof data generated by the service.

Moreover, as for the preservation of digital evidence, the great transparency and the interoperability of mature blockchain network could figure out the problems of judicial cooperation. For example, within the current system, when international cooperation happens between multiple legal agencies, excessive energy would be invested to inquire the legal state of correlated evidence. Mass individual blockchains employed with different specificities could only exacerbate the cost. However, to some extent, if we extract the trust ingredient from the evidence-handling process and set up secure schemes to generate uniform proofs on the mature blockchain network, the whole process will be much more legible and equitable. The idea inspired us to construct the evidence network on top of origin proof service.

For such reasons, we decide to build our system directly on top of mature blockchain. We choose Bitcoin as the supported blockchain network which has the greatest computing power, extraordinary technical assistance and perfect feature of decentralization. In our research, we have found that the existing services employed on the mature blockchain are almost simple PoE services. Hence, the poor usability and the high limitations could hardly adapt the process of evidence preservation. The proof of audit procedure, data interoperability such as synchronization and recovery, expression of complex evidence types and relationships, and the ownership along with privacy all exist in our concern.

To achieve more usability and manageable properties, firstly, we excavate and analyze the covert channels in Bitcoin transaction structure to achieve the potential application value of Bitcoin system. Secondly, we design a scientific method to generate the PoE digest which preserves the existence of the evidence and its

①https://app.originstamp.org/home, Mar. 2018.

②https://www.crunchbase.com/organization/bitproof, Mar. 2018.

③http://florincoin.org/, Mar. 2018.

meta-data. Then we exploit suitable spaces to preserve the digest of evidence. Thirdly, an audit module is added to assist legal institutions for audit on chain. Cryptographic proofs are added to generate the auditing keys in Bitcoin ECDSA (Elliptic Curve Digital Signature Algorithm) scheme through traditional PKI key pair owned by the legal institution. Therefore the non-repudiative authentication can be protected by the blockchain network. Fourthly, we equip the transaction chain in Bitcoin with the evidence chain which possesses manageability and legibility. Moreover, we focus on the enhanced privacy concern of the service and bring forward the solution via subliminal channels in the ECDSA scheme. Finally, we build up a highly extensible framework with incorporated hierarchical service, which balances the cost and the efficiency of applications.

Through these ideas, we design and implement a lightweight digital evidence-preserving system which possesses the features of evidence-privacy, audit-transparency, function-scalability, and operation-lightweight. The framework presents an enjoyable solution to overcome the security and confidence problems which trouble the traditional schemes. It also has stronger expansibility and stability than existing PoE solutions.

2 Background

Digital evidence is a kind of probative information stored or transmitted in digital form in case of trial[5]. In existed standards such as UK ACPO guidelines[6] and ADAM principles[7], the integrity of the original data and the relevance among the different evidence are stressed. On the other hand, the audit trail and other records of the procedure in digital evidence should be preserved. Such principles inspire us with the aims to comply with the following rules when we design our system:

- making non-tampered proof for the origin evidence;
- revealing the relevance among evidence as possible;
- recording the audit process in a verifiable way while legal institutions handle the evidence.

The rigorist security is required in comparison with handling generic data from users. From the early time, related work is totally undertaken by the appointed legal institution. With the development of network technology, more third-party agencies come forth to reduce the pressure of power department. Crypto tools such as trusted timestamp[8] and fuzzy hash[9] make the process of evidence preservation more efficient. However, seldom solutions are found to perfectly solve the issues on the privacy of data and the vulnerability of centralized party.

In 2008, blockchain was conceptualized by Nakamoto[10]. And Bitcoin, the first digital currency to take blockchain as core technology, drew attention of the world by its decentralized framework and nice anonymity. Since Bitcoin accumulates incredibly massive computing power, the stability and the security of the system are easily affirmed by the public. From then on, numerous attentions and ideas are drawn to solve the record-keeping problems happened in life. Early in 2013, a rock band named 22HERTZ stored music copyrights on the Bitcoin OP_RETURN output script which sparked widespread discussion. And in 2015, Gipp *et al.* put forward a decentralized trust timestamp[1], and implemented the OriginStamp service[4], keeping the process away from the certificates and compromise issues.

Just like the operation of OriginStamp service, PoE is an irreversible record of context at specific time. Traditional verification relies on centralized authority, but it is gradually replaced by distributed blockchain nodes. The server of each node can collect the secure digests of files from users and calculate aggregated hashes which can ensure the integrity of all files. Then the server converts the hash to a corresponding Bitcoin address and broadcasts it as a transaction with little satoshis (the minimal unit of exchange in the Bitcoin system) to the targeted address. When the transaction is admitted by the network, all the connected files are stamped with a universally trusted timestamp.

As the promotion of blockchain technology, there appear quantities of record-keeping services built on specially generated blockchains (private or public). The outstanding schemes include Factom[3], Florincoin (FLO)[5], etc. Through abundant investigation and research on such services, we can conclude the connections between different construction mechanisms and specialities in Fig.1. Obviously, compared with the mature cryptocurrencies, these services could support larger volume capacity, more flexibility and specificity on design or rules. However, from our point of view, the

[4] https://app.originstamp.org/home, Mar. 2018.

[5] http://florincoin.org/, Mar. 2018.

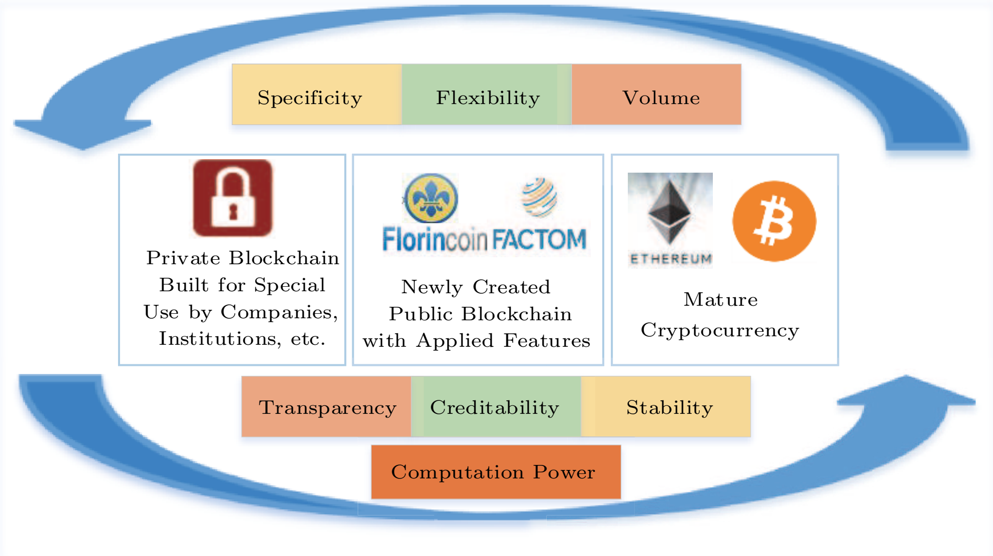

Specificity
Flexibility
Volume
Private Blockchain Built for Special Use by Companies, Institutions, etc.
Florincoin
FACTOM
Newly Created Public Blockchain with Applied Features
ETHEREUM
Mature Cryptocurrency
Transparency
Creditability
Stability
Computation Power

Fig.1. Decisions to make when employing the blockchain applications for evidence preservation.

services could suffer following disturbance when dealing with evidence forensic affairs.

- Due to the scrimpy computation power of the whole network, the newly generated chains could endure a long period time of vulnerability with the 51% attack.
- The sensibility of the application could lead to the over-concentrated distribution of nodes or computing power, which raises the risk of behaviours with negative creditability such as denial of service and privacy disclosure.
- Complex rules and high capacity demand build high entrance barriers for server nodes. Additionally, tremendous storage pressure decreases the stability of the service. Once there is a service crash, users would suffer a heavy loss.
- Individual services could do nothing good to the data interoperability, which in contrary is badly needed for judicial cooperation during the audit of evidence.

As a result, our service sets the mature blockchain network as the base of architecture. We design the system directly on top of Bitcoin due to its robustness, matureness and affordable computing-power cost. We generate proofs via constructing an auxiliary layer on the Bitcoin system. It can host evidence with stability, low risk, and high transparency. In order to make the system more practical, we refer to the restrictions of the Bitcoin system[11].

- *Speed.* With an average time of 10 minutes to generate each new block, users need to wait for a period of time to get proofs from evidence preserving service.
- *Cost.* With the continuously growing amount of transaction fee, the service will be costly for the general.
- *Block Space.* With the unresolved argument of Bitcoin expansion problem, it is more difficult for transactions to be accepted by the blockchain.

We choose covert subliminal communication as the solution to remit the restrictions of volume and cost, and at the same time greatly enhance the availability, scalability, and privacy compared with the original service. The concept of covert subliminal communication was raised by Simmonsli[12] in 1984, which focuses on confidentially transmission through open channels during communication. Simmonsli[12] put forward the solution by separating signature keys into two parts in shards to represent different sets of information, and the matching key has access to corresponding information. Further researches were proposed by Chapman and Davida[13] in 1997 and Bellare *et al.* in 2014[14]. Moreover, Kilroy and Richard[15] revealed that the U.S. Central Intelligence Agency stole the private keys of specified targets by constructing covert channels with backdoors in the TLS protocol. The covert channel makes our digital evidence transmit in secret channels with compete anonymity.

3 Design

In this section, we will dissect the challenges of the evidence preservation from top to bottom, including the core thought and the details of our design.

3.1 Entities and Demands

There exist four entities in reality during the procedure of evidence preservation. We firstly make discussions on their demands.

- *User*. A user will always want to get the permanent proof of existence and the audit for his/her evidence without leaking any confidential information he/she owns. Full transparency, manageable service and low policy barriers are also in his/her concern.
- *Legal Institution*. The legal institution has the right to offer audits to the related evidence which proves the legal effect. It is eager for the mutually secure schemes with high interoperability and efficiency.
- *Service Agency*. The service agency can be seen as a double-edged sword. On the one hand, it builds the bridges between users and legal institutions and helps users with the evidence management. In decentralized PoE architecture, it can also avoid troubles for users who do not own the token of certain cryptocurrency. On the other hand, it often becomes the weakest link when suffering from the cyber attack, opaque operations, and privacy issues. Hence, a proper framework is in need to insulate the privacy data from the agency, validate its action in trust at any time, and lower the risk of its storage pressure.
- *Malicious Attacker*. The malicious attacker will spare every effort to steal information, falsify data, and create disturbance under the system.

Our core thought is to weaken the trust-dependency, and make it usable from the general environment to the decentralized Bitcoin system. Using covert channels, our design maximizes the availability and the scalability on top of the Bitcoin blockchain, creating a manageable evidence chain with high efficiency and smooth interoperability. By encoding the cryptographic digests of evidence and the corresponding audit replies into the blockchain, our distributed network system will help users to conquer the apprehension on privacy and trust, and meanwhile remove agency's pressure in data recording and confidence. During the process, a trustful and convenient platform is provided for legal institutions to finish the audit work of massive evidence on chain. It saves much time to investigate the qualifications of agency servers, so that legal institutions can simply focus on the evidence. Malicious attackers cannot obtain anything precious from the service or the blockchain, and there is no chance for forging the existing proofs or the ongoing services. We summarize our design decisions in Table 1.

3.2 Technical Basis and Terminology

Before elaborating the concrete constructions, we need to briefly introduce the technical basis and terminology of the Bitcoin system.

As shown in Fig.2, the structure of blockchain provides Bitcoin's public ledger, an ordered and timestamped record of transactions. During generation, each new block collects the new transactions from dis-

Table 1. Summary of Design Decisions

Design Decision	Benefit	Challenge
Optimizing evidence audit procedure coordinates with the immutability, transparency and connectivity of blockchain	With notarization on mature blockchain, dishonest behaviors could be terminated from legal organizations. The integrity and non-repudiation of audit replies can be proved, which protects the legal effect of digital evidence.	Authentication scheme based on blockchain needs to be set up under different public key infrastructures. We should also keep the delivery of auditor information away from centralized risk.
Build complete and pluggable privacy protection scheme	Complete privacy protection terminates the chance of leakage, while the pluggable design enables users to activate partially shared information for better transparency and management characteristics.	Under rigorous circumstances, the negative impact on the disclosure of evidence digest should be considered. It is also tough to seek the proper bound between privacy and availability.
Bridge evidence chain to accommodate the relationship between various types of evidence. Additionally, establish convenient model of data interconnection.	The foundation of relationship network greatly raises the scalability and availability of service, layering isolate evidence preserved on the blockchain. Through this design, it is available for users or institutions to filter, recover or proceed trusteeship on the evidence. Also, the synchronization of trust entries between legal institutions becomes more efficient.	Restrictions from limited channels and costly transaction fees. Proper interconnection model for management to be explored between different types of users and service agents.
Find fine-grained service classification model to balance the cost and efficiency of the service	Users could make decisions between cost and efficiency with flexible strategies according to the importance and particularity of evidence.	Seeking for approaches to break the barrier of limited block space and linearly classify the requests from users.

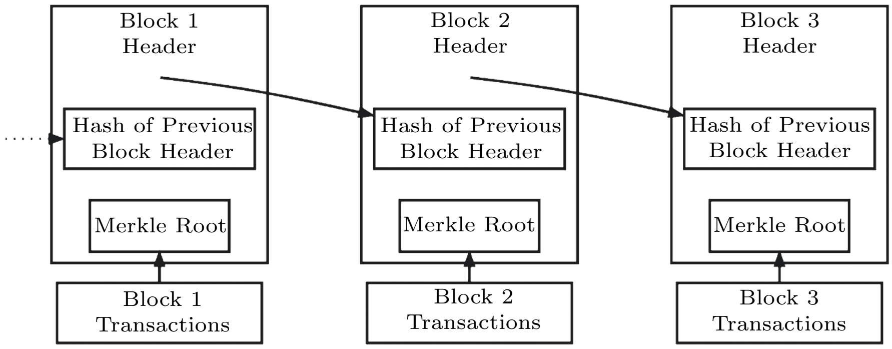


Fig.2. Simplified Bitcoin blockchain[2].

tributed network and generates the Merkle root as the integrity proof of such transaction data using the algorithm of Merkle Tree. The block header stored the Merkle root, mining nonce and the hash of previous block header, thus chaining the blocks together and keeping immutable record for the whole ledger. With the work of consensus mechanism, the generation time of each block stays as the unforgeable timestamp for all the data existed in the block.

We use Fig.3 to show the data structure of Bitcoin transactions and the connection between transactions. Except for coinbase transaction which is used for declaring mining reward, a Bitcoin transaction consists of a transaction identifier, meta-data (version, locktime, etc.), input(s), and output(s). We describe the definitions of each part as follows.

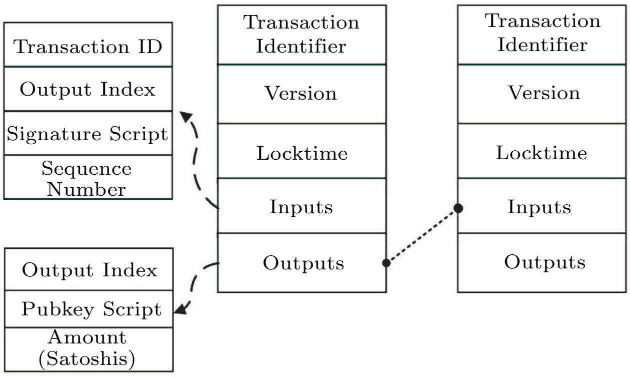


Fig.3. Data structure of Bitcoin transactions.

• A transaction identifier (abb. Txid) is used to uniquely identify a particular transaction (Tx), in the form of Sha256 hash of the transaction.

• Locktime field indicates the earliest time for the Tx may be added to a new block.

• Version number basically refers to consensus rules that Tx follows.

• The input in a transaction contains three fields: an outpoint, which consists of the Txid and the output index, and refers to a previous output and signature script for the spent transaction, a signature script (abb. ScriptSig), which satisfies the conditions left to the corresponding output, and a sequence number, which intends to allow the unconfirmed transactions to be updated. A transaction can contain multiple inputs.

• The output in a transaction contains three fields: an output index, which indicates the order of the output, a value field for transferring zero or more satoshis, and a public key script (abb. ScriptPubKey) for indicating what conditions must be fulfilled for those satoshis to be further spent. Standard types for ScriptPubkey include P2PKH, Pubkey, Nulldata, P2SH, and Multisig. A transaction can also have multiple outputs.

• Satoshi is the minimal denomination of Bitcoin value. One bitcoin (BTC) equals 10 000 000 satoshis.

Transactions are also chained together by spending an unspent transaction output (abb. UTXO) of previous Txs as an input in a new transaction. Bitcoin network chooses Elliptic Curve Digital Signature Algorithm (ECDSA), specifically the secp256k1 curve as its public key algorithm. The system uses secure hash algorithms and Base58 encoder to convert the public key to payment address of users, which is currently the most common way for users to exchange payment information.

3.3 Analysis on Covert Channels

The pivotal challenge of the transaction is to find cheap and efficient covert channels for transmission. For optimization, we excavate all the conceivable convert channels according to the data structure of the Bitcoin transaction and conduct analyses on capacity, transparency, security, cost, and multiplexing ability.

Thereinto, the transparency embodies in whether the accommodating information can be visible by the public network, the multiplexing ability indicates whether the channel can be parallelly used in one transaction, and the security is about the negative effect or special security value brought by the use of channel.

To measure the cost of channels, we embody the abstract conceptions into the transaction fees, and it becomes the central part we consider. Except for some negligibly influential factors, we define:

$$TxFee = NormalFee/Byte \times TxSize.$$

Here, $TxFee$ denotes the total expenses consumed by Tx, $TxSize$ denotes the bytes contained by Tx, and $NormalFee/Byte$ denotes the conversion rate between them, which can be recognized as a constant at some point in time.

According to the current Bitcoin market, we define the average transaction fee per byte as 450 satoshis. We also define the ratio of valid data and total data introduced to be the utilization ratio of the channel. The analyses of different subliminal channels are presented as follows. And Table 2 summarizes the results of our analyses.

Table 2. Analyses of Covert Channels in Bitcoin

Channel Name	Capacity	Transparency	Security	Multiplexing Ability	Cost to Hold a Sha256 Digest	Utilization Ratio (%)
LockTime	5	Y	No influence	N	–	100.0
SequenceNum	4	Y	No influence	Y	532 800	2.7
ScriptSig	31	N	Information hidden	Y	66 600	21.6
OP_RETURN	80	Y	No influence	N	19 350	87.9
Tx Amount	3	Y	No influence	Y	153 000	9.4
EcPubkey	32	N	No influence	Y	81 900	17.6
EcPrikey	32	N	Weak	Y	15 300	94.1

- *Locktime.* This free and flexible channel existing in the Bitcoin transaction represents the earliest time/block depth for confirmation. As Fig.4 shows, only the shadow part will lock Tx in reality. If we set the value within current time (Tnow)/block depth (Hnow), we have about 5-byte space to encode our data without delaying. Though it is not enough to host a full digest, we can use it to encode the related symbols. The channel keeps transparent and serves as the single field due to the rule.

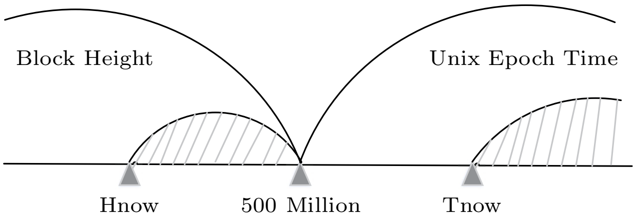


Fig.4. Construction of covert channel on Locktime field.

- *Sequence Number.* If Tx is not time-locked, then we have the whole 4-byte space constructed in this channel. The channel is usable because a Tx must process at least one input. However it is obviously uneconomical to introduce another input (148 bytes) on purpose to repeatedly use the channel.

- *ScriptSig.* We could encode 255-bit data to the ScriptSig field in each input by using the subliminal construction algorithm detailed in Subsection 3.7. The channel is completely concealed unless the corresponding Bitcoin private key leaks. The channel does not hold good transparency for trust entries, but can be designed to resolve privacy concern.

- *Tx Amount.* For each Txout less than 1 BTC there exists around 3-byte channel per transaction. We can transmit a 32-byte digest by adding Txouts up to 10. It is expensive to use this channel in a large scale. Also fragmented UTXOs would be generated, which is troublesome for sustained services.

- *OP_RETURN Script.* Standard OP_RETURN script has a capacity of 80 bytes where arbitrary data can be written. To store a 32-byte Sha256 digest, it would occupy 43-byte for extra. The channel holds high efficiency and perfect transparency. Only one such script can exist in each Tx.

- *EcPubkey.* We can map any Sha256 digest to an address which can be contained in Public-Key or P2PKH script. Since we cannot redeem UTXO without possessing the private key, a cost which equals at least 182-byte extra data is generated.

- *EcPrikey.* If we directly take the evidence digest as the private key, 34-byte generates in extra. However, security problems will come under such a simple procedure. For anyone who holds the evidence or digest, the private key remains as plaintext.

As we can see, the OP_RETURN script would be the best covert channel to transmit a Sha256 digest for a proof of existence service on the Bitcoin blockchain.

3.4 Extraction of Metadata

Due to the restrictions of capacity, it is uneconomical and catastrophic to carry all of the evidence data through the blockchain. To keep the integrity of the data, a single digest generated by secure hash algorithms is just enough. However, to achieve better efficiency and availability, we must focus on the particularity of digital evidence itself and carry out friendly design on algorithms and data structures. Our solution is shown as Fig.5.

For a piece of evidence consists of n files, we use Merkle tree to generate a Sha256 digest which covers the integrity of the origin evidence. For better availability in follow-on forensic work, we design three fields attached to the scheme: an information field which records the basic information of the evidence, such as name, generation time, usage and company, a structure field which explains the social relations of the evidences, and a secure field compatible for data generated by modern security tools, for example, a signature which represents the ownership, and a fuzzy hash which can be used for precise contrast. Information collected by above fields can be encoded to a metadata file. The final digests generated by our PoE service are computed as the Sha256 hash result of the connection by evidence digest and metadata digest.

Through the design, users can optionally bind the evidence with additive attributes and forensic friendly data. The metadata file will be sent to the server agency and legal institutions for further auditing work. Adaptive privacy can be implemented through the process. The metadata is seen as part of the evidence which will be preserved by the blockchain, and any inveracious information included would destroy the legal effect of the proof.

3.5 Secure Proof Schemes

The next step we take is to construct secure proof schemes to safeguard the life cycle of evidence. To prove the integrity and timeliness of the evidence, we could encode the final digest of evidence to the Nulldata field of Bitcoin transaction (with OP_RETURN script), which serves maximum transparency and efficiency. Once Tx is taken on by the newly generated block, users can receive the proof from any interface of the Bitcoin network by retrieving Txid. No matter who supplies the service, an individual wallet or a server agency, the process is steadily clear and unforgeable.

Other than PoE, the audit process of the evidence also deserves to get credible proof on the blockchain, which can prove the integrity, timeliness, and nonrepudiation of the corresponding audit result. To achieve the goal, a simple and obvious solution is to directly encode the signature generated by auditors' PKI to the blockchain. However, such a design is neither

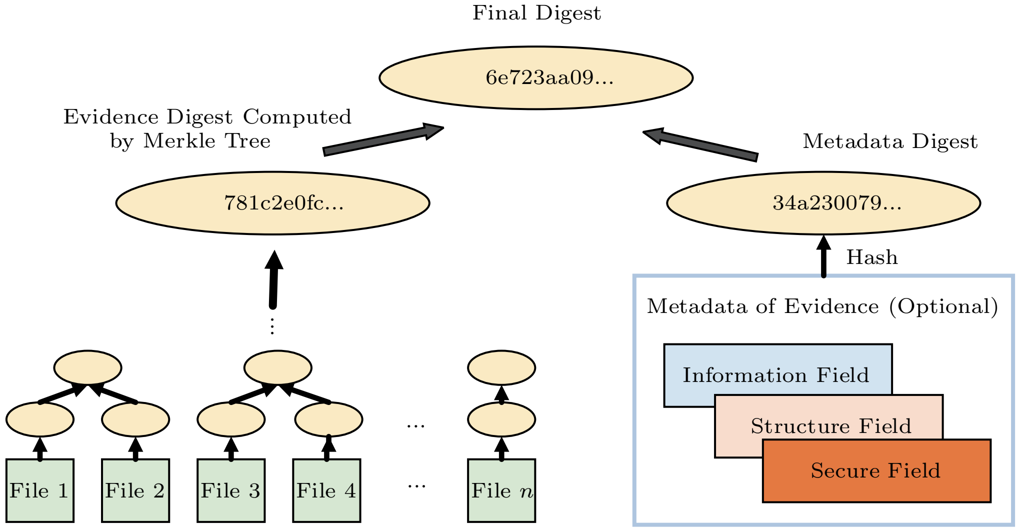


Fig.5. Extraction of metadata.

scalable nor manageable for interconnection. Since legislations are disparate in different areas, the auditor here represents not only a specific one but several certain legal institutions. Each evidence can obtain multiple proofs of audits from different agencies, and the final decision totally depends on the users' choice. For different PKI signature infrastructures, various and customized coding schemes are required, and for supervision and judicial cooperation, it is troublesome and wasteful to adapt to such many schemes.

Hence, we are looking forward to a unified and transparent approach to authenticate the audit reply on the blockchain. To avoid the negative effect of centralized architecture, we will not set up another centralized register institution. On the contrary, we will urge legal institutions to generate powerful proofs themselves. Actually, the Bitcoin key pair of the ECDSA scheme originally has the ability of authentication. Due to its anonymity, measures are needed to bind the identities of legal institutions with the specific public keys. We propose an idea to make a brief proof which can convert the probative force in the traditional PKI to the new generated audit keys in Bitcoin. The concrete algorithm is listed as follows.

ECDSA Audit Key Generation. To generate n pairs of audit keys on the Bitcoin network, an auditor $\mathcal{A}$ with associated key pair (P_{kA}, S_{kA}) from certain PKI does the followings.

- Generate n pairs of ECDSA key $[P_0, S_0, ..., P_{n-1}, S_{n-1}]$ via Bitcoin wallet randomly.
- Extract public keys $[P_0, P_1, ..., P_{n-1}]$ to construct a Merkle tree and finally compute the Merkle root M.
- Sign M with S_{kA} to generate signature Sig from PKI.
- Compute Sha256Hash(Sig) to get the digest H.
- Construct Bitcoin Tx with OP_RETURN script: $OP\_RETURN || H$, broadcast the transaction to the Bitcoin network, and record its Txid.
- Wait for Tx to be confirmed by a newly generated block, and get the proof $([P_0, P_1, ..., P_{n-1}], P_{kA}, Sig, Txid)$.
- Get n pairs of ECDSA audit keys $[P_0, S_0, ..., P_{n-1}, S_{n-1}]$.

Both server agencies and users can validate the identity of the auditor with the proof $([P_0, P_1, ..., P_{n-1}], P_{kA}, Sig, Txid)$ and it shares identical probative force with (P_{kA}, S_{kA}). While adversaries cannot forge any valid proof or audit key pairs of $\mathcal{A}$. When auditors finish the cognizancing process, audit replies will be generated to represent the legal force of users' evidence.

To give the audit reply an effective audit proof on blockchain, $\mathcal{A}$ should do the followings.

- Choose valid audit key pair (P_x, S_x), and sign the audit reply X to get the ECDSA signature Sig_X.
- Encode Sig_X to Der, generate a standard Bitcoin Tx with OP_RETURN script: $OP\_RETURN || Der(Sig_X)$ and broadcast it. Record its Txid.
- Wait for the Tx to be confirmed by a new block, and thus get the proof $(P_x, Txid)$.

Through the scheme, when supervision or legal cooperation happens, only a single time validation needs to be progressed through origin PKIs for each institution. The rest verifying procedure could be done easily on the Bitcoin blockchain. This greatly simplifies the workflow as well as building foundation for perfect data (trust entries) interconnection mode between legal institutions. Also it constitutes a naturally transparent supervision platform for the general public. Therefore any credit-corruption behaviours will be recorded and reminded forever by the powerful irreversible proofs. Furthermore, complicated cooperation modes such as hierarchical validation could be operated with the scalable scheme.

3.6 Construction of Evidence Chain

In reality, there are various types of evidence. Also the relationships between different types of evidence are complicated. If we take the proof we built as the representation of the evidence, encoding types, relationship and directivity to the evidence, then the service could be further manageable. Convenient and efficient interconnection mode can be created on top of our services.

To meet the target, firstly, we should resolve the need for directivity. By now, the generated PoE transactions of evidence are dispersed to the blockchain. Users can only retrieve them by Txids. And when the amount of evidence entries becomes large, it is troublesome for synchronization and management. As a result, we use the EcPubkey field as the directive solution. For the first time when some users or institutions join in the service, an ECDSA key pair (and its corresponding address) is(are) generated specially for direction. Every time an existence/audit proof is generated for the owner's evidence. A P2PKH script will be added to Tx to generate a dust UTXO (the minimal output value can be added to Tx) to the appointed address. Thus any entity can filter its trust entries efficiently on the

Bitcoin blockchain by simply using the wallet function (such as Bloom filtering[16]).

Due to the anonymity of Bitcoin address, privacy is respected during the procedure of directing. When there is data loss or service relocation, through corresponding address, the process of recovery and synchronization can be done real fast without special utilities or computing power. Additionally, the storage pressure of server agency is released by this measure.

Secondly, to distinguish different types of evidence and proofs, we use the transparent and free Locktime as the encoding channel, which has more than 2^{40} space to define.

Thirdly, we need to seek proper solutions to express the relationships between evidence. Through investigation, we divide the relationship into three fundamental types: the inheritance type (such as father contract and subcontracts), the version type (such as the copyright of product under timeline) and the audit type (such as evidence and its audit replies (proofs)). The design of our solution is shown as Fig.6, which constructs scalable evidence chain on top of the blockchain.

To present the inheritance relation, the best template could be the multiway tree. Enabling to achieve the construction, we define the Sequence Number field to be the identifier in the first input of evidence Tx. We add a 4-byte identifier field, a 2-byte seed number field and a 1-byte depth field to the OP_RETURN script. The Nulldata script can be expressed as

$$OP\_RETURN||FinalDigest||Seq\_Paprent|| \\ SeedNum||Depth\_Parent + 1.$$

During the generation of root evidence, it sets such three fields to zero. By searching all root evidence directing to the same user, it selects a different 4-byte random number as the identifier. When a child evidence enters, it firstly filters out the earliest parent Txs according to the identifier and depth. Then if the parent does exist, it filters out all the connected child evidence Txs on the blockchain and in pending pools to ensure the uniqueness of its seed number. Lastly, it computes $Indentifier = Sha256hash(Seq\_Paprent||SeedNum)\&(0xffff << 240)$ to detemine its sequence number. The design creates a theoretically $2^{32} \times 2^{8}$-field on inheritance tree for every user and greatly minishes the chance of collision.

To present version relationship, we chain the neighborhood Txs with the dust UTXO directed to the user address in the horizontal level. Thus, we can prove the versions by the nature timestamps of the Bitcoin blockchain. To express the audit relationship, audit Txs repeat the sequence identifier of the evidence Tx. Consequently, the design weaves a thoughtful and scalable evidence chain. Efficient and flexible interconnection mode is shown in Fig.7.

In each area, there exist hundreds of agency servers to commit proxy functions, which help users broadcast the evidence Txs without Bitcoin token. Users can also complete the task using individual Bitcoin wallet. Legal institutions share their proof information to agencies and users, and furthermore accomplish audit proofs on blockchain. Three roles contribute to wave the evidence chain together, while Bitcoin nodes main-

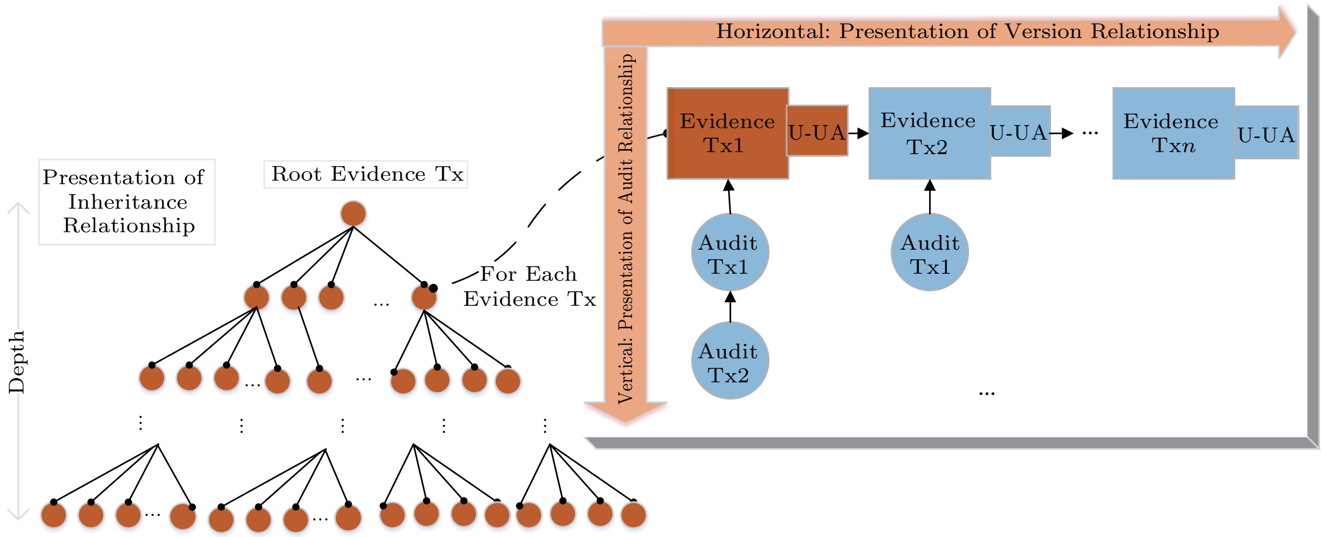

Horizontal: Presentation of Version Relationship
Vertical: Presentation of Audit Relationship
Evidence Tx1
U-UA
Evidence Tx2
U-UA
...
Evidence Txn
U-UA
Audit Tx1
Audit Tx2
Audit Tx1
...
Presentation of Inheritance Relationship
Root Evidence Tx
For Each Evidence Tx
Depth

Fig.6. Construction of evidence chain. U-UA: UTXO to user address.

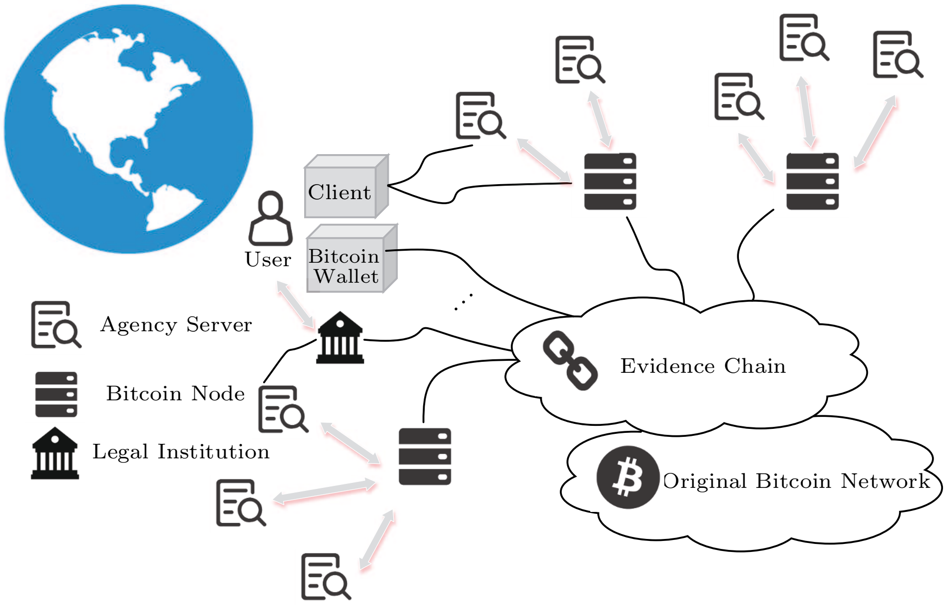


Fig.7. Interconnection mode of the service.

tain the whole network with great computation power. For legal cooperations around the world, participants can be encouraged to find a shortcut in supervision, synchronization, data recovery, and interaction. The network remains scalable and stable and thus would attract more attention and participation.

3.7 Enhanced Privacy Concern and Solutions

Constructing a system with simultaneously complete privacy and manageable properties is indeed a tough challenge. Under a rigorous environment, we have enhanced privacy concerns in specific situations. Suppose that an evidence is located by its meta-data and relationship on the evidence chain, once the adversary has stolen one or several "legendary" transcripts of the origin copy of evidence, he/she can easily make decisions to differentiate his/her masterwork according to the transparent digest on the blockchain.

The concern falls to the ground in reality, as for some confidential evidence, users would not like to publish the evidence digest to the public network. The owner only wants to share the digest with the corresponding auditor under specific time and circumstance. Then what can we do to fulfill his/her requests?

The easiest approach to solving the concern may be carried out by hashing the digest with salt. However, no one could guarantee the randomicity of the salt. For some users with ulterior motive, they can fake the digest of partitions in their evidence to be the random salt. Thereby, it is easy to conceal some information during the audit process. Another available approach is to import more complicated one-way translation schemes during the process of evidence preservation, while as long as the process is open and determinated, it cannot stop the adversary from trailing.

Encryption, as a good point, can protect the attackers from the digest. But declaring an uncertain key could also bring trouble for auditors. We construct an economical and secure solution to solve the concern. The core idea is to store the commitment of the key with subliminal channel constructed in the ScriptSig field. Concrete steps are as follows.

Privacy Enhanced Scheme. To generate a privacy-preserving digest from evidence E, user $\mathcal{A}$ should take the following steps.

1) Compute final digest of evidence H_E using methods in Subsection 3.4.

2) Get ECDSA key pair (d, Q) with balance only composed of dust UTXO from individual Bitcoin wallet.

3) Select a random 32-byte key K_r, to replace the random parameter k of ECDSA signature generation scheme and compute: $s = k^{-1}(e + dr) \bmod n$, where e is transferred from the message being signed, d indicates the ECDSA private key, and (r, s) is the generated signature.

4) If fails, return to step 3. Loop until $\mathcal{A}$ gets key K_{rx} and valid ScriptSig. Then the system includes

ScriptSig to the first input of Tx.

5) Using Advanced Encryption Standard (AES) to encrypt H_E with K_{rx}, the agency server encodes the result C_{HE} to the Op_turn script of Tx, and then broadcasts Tx and waits for confirmation.

6) When audit procedure happens, $\mathcal{A}$ shares the key pair (d, Q) with the legal institution, and the corresponding auditor validates the key pair using ScriptSig in Tx, recovers the key K_{rx} following the reverse procedure: $k = s^{-1}(e + dr) \bmod n$, decrypts C_{HE} by K_{rx}, and finally gets the accurate digest H_E of evidence E.

To enhance the security of the scheme, the encrypt algorithm in our scheme is ruled to eliminate the uncertainty brought by the algorithm itself. The shared key pair is required to be without the economic value for secure consideration. Our solution resolves the concern without any extra cost. The privacy-enhanced Tx remains the same form as the normal evidence Tx on the evidence chain, which further increases the privacy of the whole system.

4 Framework and Implementation

After a discussion on detailed design scheme, we plot the map of our framework and give analyses from the view of application. As shown in Fig.8, the system follows the Client-Server framework, with strong ability to ensure the information security. There exist three entities in our framework: client, server and auditor. They are based on modular design. The transactions on blockchain and audit replies of evidence accommodated in evidence chains are woven into a manageable network. The workflows of different entities are stated sequentially in detail. In addition, the natural formulation of special hierarchical services is discussed.

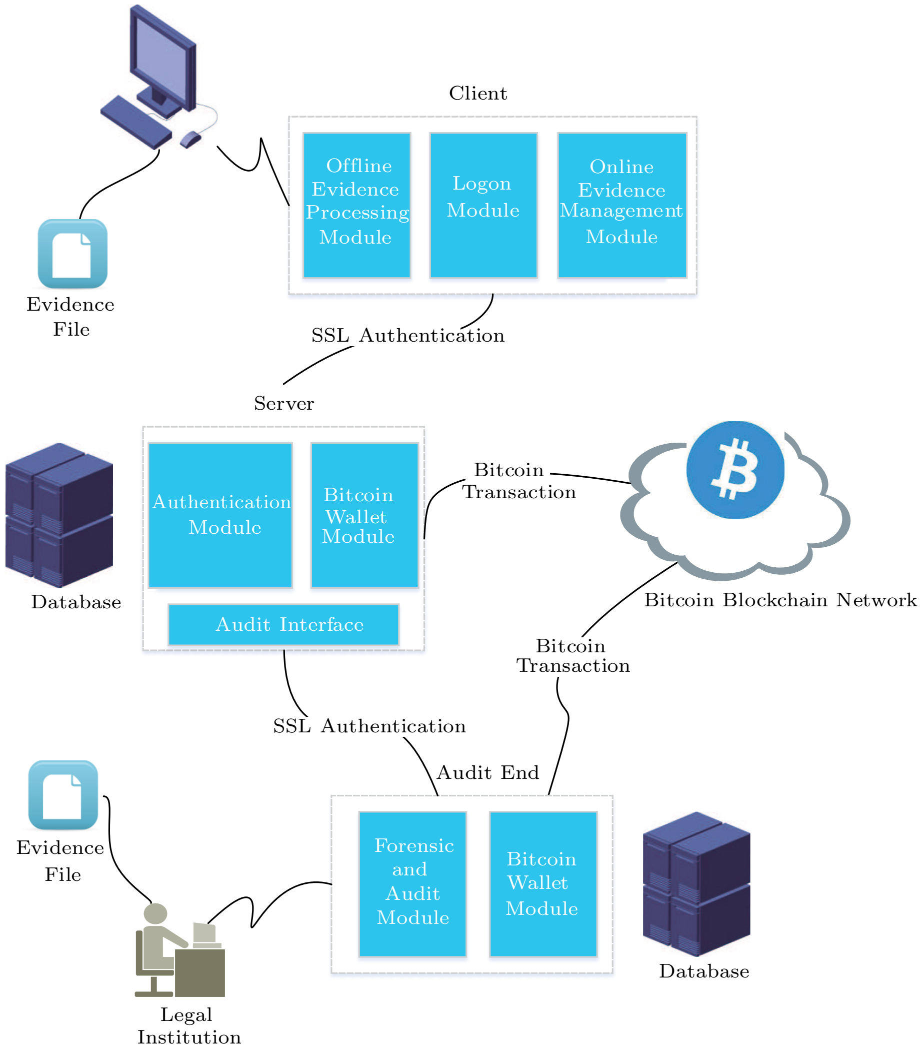


Fig.8. Overview of the system framework.

4.1 Client

The localization of clients is to assist users to manage evidence securely and efficiently with the help of server agency and evidence chain. The main workflow is shown in Fig.9. To keep the privacy of the entire system, the module for evidence processing is designed to operate offline through the whole process of meta-data extraction. Via the register and login module, users who do not have Bitcoin tokens can log in the client and register corresponding unique addresses from the agency server. The client will establish confidential communication to vicinal server agencies through the SSL protocol. Users are required to set a password which not only is for authentication, but also acts as the initial parameter in data encryption. For a comprehensive security, the system encrypts data in strict procedure. The key derivation function Argon2[17] we use here has perfect resistance to GPU cracking attacks, which can greatly improve the randomness of the AES key in system encryption. For online management module, the client calls public data interface from Bitcoin to synchronize the status of processed evidences and the evidence chain associated with users' addresses. After loading offline evidences, final digest, metadata and requested Tx scripts will be sent to the server agency through the SSL protocol. The individual user can also submit the corresponding address to activate trusteeship or recovery request. For the forensics model, users validate origin evidence file and meta-data file to inquire the evidence status on the blockchain in time, which provides a convenient and reliable entrance.

4.2 Server

The server acts as a succedaneum to help users broadcast and synchronize information on the public Bitcoin network. The provided service can be supervised through the current state of the Bitcoin network completely. It also acts as a bridge between users and auditors. Server agencies will collect auditor information and implement independent verification according to the audit proofs on Bitcoin blockchain. Vicious information will be filtered before the second verification of the client. Other than traditional servers, the server agent in our framework does not do anything sensitive and afford less storage pressure. The negative effects of centralized vulnerabilities have been totally restrained as every trust-needed operation builds clear and irreversible proofs on public blockchain. Even if a service collapse happens, there would be no negative effect on existing proofs and a quick recovery will be committed via the help of evidence chain. The workflow is shown in Fig.10.

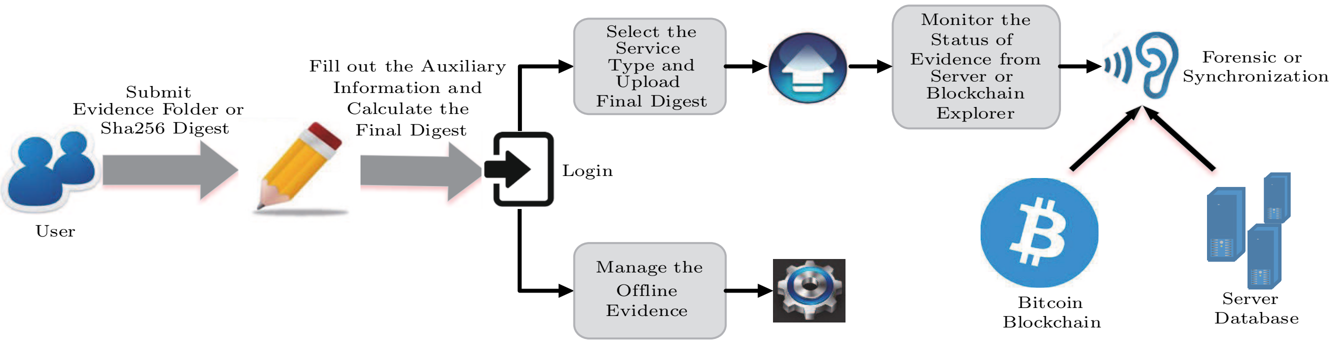


Fig.9. Client workflow.

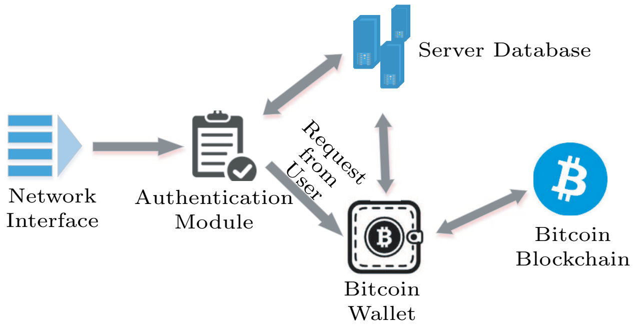


Fig.10. Server workflow.

4.3 Auditor

The auditor assists to efficiently and securely produce the audit proofs of the evidence preserved on blockchain from legal institutions. The results of audit work can be supervised and preserved permanently, and it terminates dishonest behaviors from legal organizations. Obeying the audit proof scheme in Subsection 3.5, the investigation of the evidence could be expanded through the meta-data and evidence chain. The procedure of validation gets huge simplification and perfect instantaneity. With great data interoperability and unified procedure for trust entries, the barriers of legal cooperation can be broken. The workflow of the auditor end is shown in Fig.11.

4.4 Hierarchical Service

In consideration of the balance between cost and efficiency, we bring a lightweight service with fine-grained hierarchical design into our system.

Inspired from the service of OriginStamp[6], for an extensive collection of evidence in limited time, the server can calculate the root digest of all the evidence submitted by lightweight services in advance and then upload the digest to the Bitcoin blockchain. As the transaction being received by the blockchain, a unified proof of existence will be verified for all the submitted evidence. The service could partly sacrifice the efficiency but greatly reduce the cost of the service.

The fundamental design in OriginStamp on such types of service is basically supported by donation and determination. As a result, it is totally free. However, users have to endure for large volume seed file and at least one day's confirmation delay, which greatly disrupts the quality of service. Additionally, if we follow the identical design, the property of evidence chain in our framework would be wasted for empty. Moreover, since there is little elasticity, the user's desire on the importance of evidence would be ignored by the service.

Hence, we reform the origin design with a fine-grained competitive mechanism, which introduces more flexibility and availability. The mechanism also offers the best adaptation to our designed evidence network. Firstly, we introduce an expectation field W in the meta-data which indicates the expectation for the importance of such evidence. Thus the evidence request expands to a tetrad $\{H,M,S,W\}$:

- H represents the final digest of the evidence;
- M indicates the metadata;
- S indicates the script parameters to construct the Tx;
- W represents the expectation of the evidence.

The value of W ranges from 0 to 1 000. When W equals 1 000, it shows that the evidence is urgent and valuable that users would commit total payment for the generated Tx fee. The agency server will complete and broadcast Tx immediately. When W equals 0, it shows users are more likely to choose free service. The server will broadly collect the requests for a steady time setting in time module and then calculate the aggregated hash to construct the PoE Tx. It is noted that the Tx generated from the free service does not flow to any directive address and also cannot get a unified audit proof associated to it. When W falls between 0 and 1 000, the user would only want to pay W thousandth of Tx fee. The server will take the following steps to finish the proof.

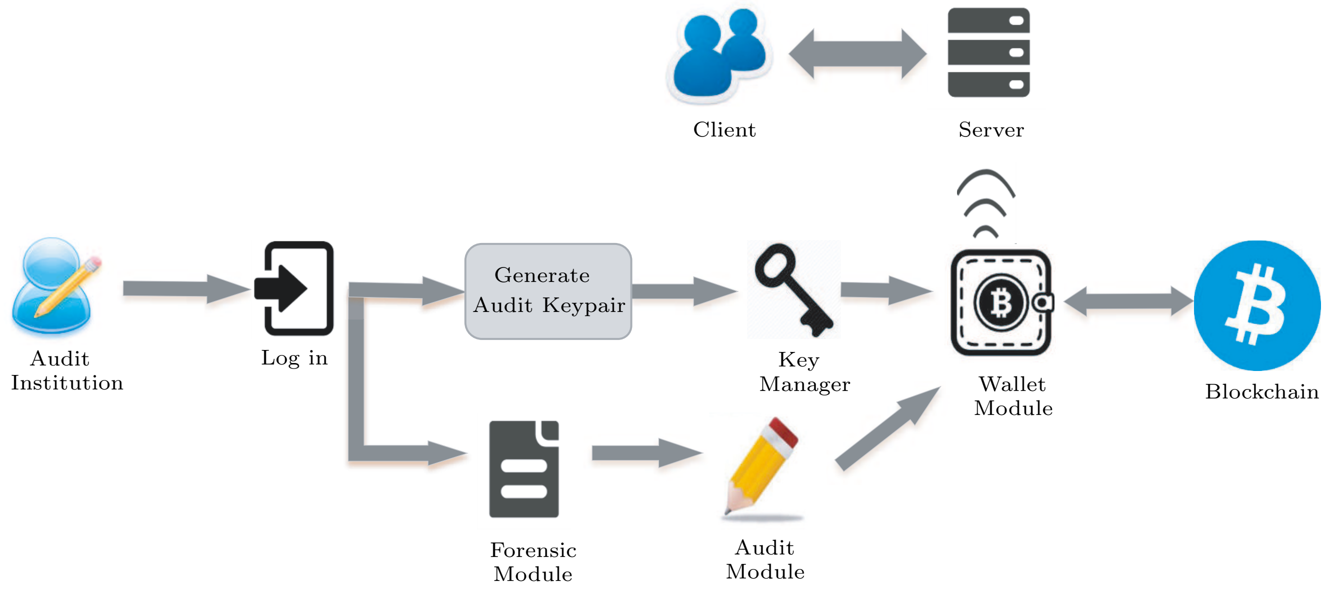


Fig.11. Auditor workflow.

[6] Originstamp. https://app.originstamp.org/home, Mar. 2018.

1) When initiated, the agency server virtualizes a container with volume $C_0 = 1\,000$.

2) Everytime when the server gets a valid request $\{H_i, M_i, S_i, W_i\}$, it calculates: $C_i = C_{i-1} - W_i$.

3) If $C_i > 0$, then the server continues to wait for the next request and the sequence number $i = i + 1$.

4) If $C_i < 0$, then the server gathers the above requests, and calculates the aggregated digest H_{agr} from $\{H_1, H_2 ... H_i\}$, the index x of maximum expectation W_x, and the sum of expectations W_{sum}.

5) The server constructs the proof with H_{agr}, completes the Tx with M_x and S_x, and calculates the autual cost for the i-th request $C_i^{'} = \frac{C \times W_i}{W_{\text{sum}}}$, where C is the additional cost to carry out an evidence Tx according to the current Tx fee.

6) Finally, the server broadcasts and synchronizes Tx, and offers each user involved a seed file recording H_i and W_i. Combining the seed file, the origin evidence and the evidence network on Bitcoin blockchain, users can validate the lightweight proof independently.

The role of the server agency would stay honestly since the expectation of each participant is preserved by the final digests. The design of the mechanism will greatly reduce the time for lightweight service with the group intelligence. For each user, a lower cost will be reached compared with his/her actual expectation; furthermore, the service would be operated with faster confirmation speed and smaller communication cost. It is also a fine-grained scheme due to flexible balance influenced by users' expectation. The specific workflow is shown in Fig.12.

5 Evaluation

In this section, we evaluate our framework from cost, efficiency and security. Then we introduce the concrete implementation of the paper. Eventually, we express our idea on the innovation of the project, and bring forward future prospection.

5.1 Cost

To estimate the whole scheme, firstly we estimate the cost of constructing a single evidence Tx. For a regular Bitcoin Tx, there exist at least one input and one output, which construct a fixed 192-byte data. For the proof itself, a Nulldata output donates 43-byte data. To weave the evidence network, another 7-byte data is introduced. In total, 242-byte data fee is needed. If the user activates lightweight service and sets the expectation to W_i, according to our scheme in Subsection 4.4, the actual cost can be reduced to $(\frac{242 \times W_i}{W_{\text{sum}}})$ bytes. To construct an evidence chain constructed with x Txs, for each evidence proof in timeline, 390-byte data is needed. Therefore the total data cost equals $(390x - 148)$ bytes.

For each audit proof, except for the fixed expending, we need a 72-byte Nulldata output to preserve the ECDSA signature. In all, 264-byte data is needed. Data cost can be transformed to the actual expense in dollar by the following formula:

$$\begin{aligned} TotalExpense = {} & 10^8 \times DataCost(byte) \times \\ & Txfee(satoshi/byte) \times \\ & ExchangeRate(dollar/BTC), \end{aligned}$$

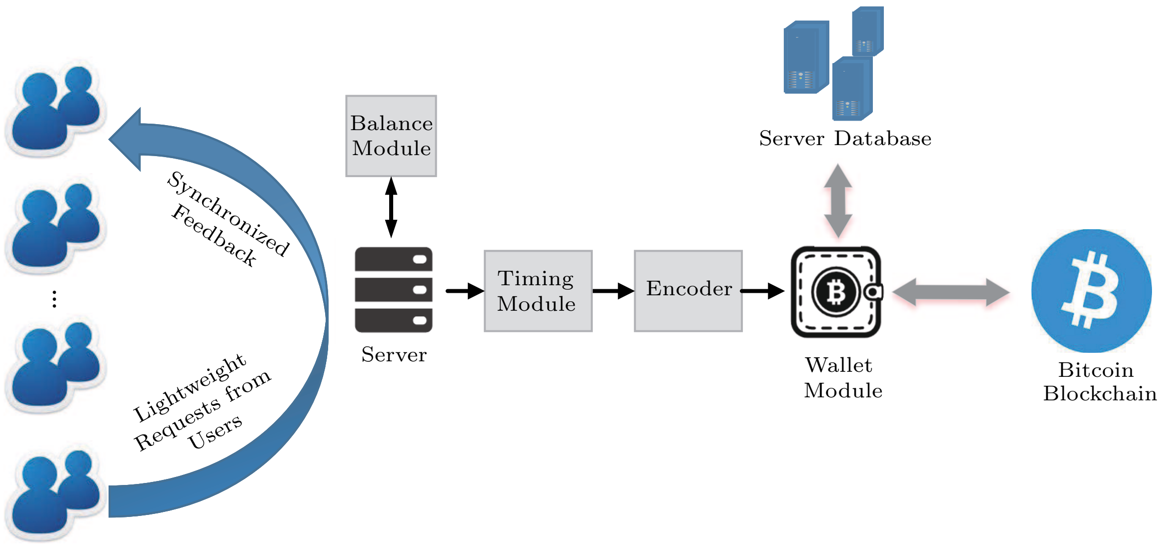


Fig.12. Lightweight service workflow.

where *DataCost* is the data consumption of Txs generated by service, $Txfee$ is the current transaction fees on Bitcoin blockchain, and *ExchangeRate* is the exchange rate of BTC to US dollar.

5.2 Efficiency

In terms of economic concern, our system excavates and analyzes covert channels in the Bitcoin transaction structure. Choosing the most economical channel for our proofs, in addition, we bring a lightweight service with an improved hierarchical design which commits fine-grained balance between cost and efficiency.

For evidence management, our scheme breaks restrictions from limited channels and costly transaction fee. Efficient evidence chains are constructed to offer shortcuts for recovery, synchronization and trusteeship of evidence. Furthermore brilliant interconnection model is built among different types of users and service agents.

Especially, our scheme greatly simplifies the audit and supervision process in forensic work. Effective and powerful audit proofs can be attached to the evidence on blockchain. Reliable and unified qualification proofs on legal institutions could also be easily set up for globally legal cooperations.

5.3 Security

We introduce the common security policy: protection, detection, and recovery as the core of the system security model, and we present assumptions in a mutually distrustful environment.

Ultimately, we have achieved all the destinations listed in our security model. 1) Except for 51% attack, no one can forge or tamper the proofs made in our system. Even if there is a collapse of the server, the proofs are still effective and easy to be tracked by the evidence chain we constructed. 2) There is no need for trust during the interaction among users. The cheating of the server and institutions will be easily discovered on transparent blockchain. 3) As data interworks with peer, planning a denial of service attack is extremely expensive. Any wallet could finish the submission according to the final digest. 4) The privacy of users is well-preserved due to the design of offline evidence handling and the enhanced encrypt schemes with the use of subliminal channel. 5) The censorship on identities of legal institutions will not be limited to centralized department. Users can confirm the identical proofs submitted by auditors and choose their trusting institutions to finish the audit work.

5.4 Supervision Features

Our system emerges in nice and efficient supervision features, and the mechanism of our design maximally transforms great computing power of Bitcoin network to notarial force in reality, which generates powerful and transparent proofs for the process of supervision. Hence legal force of involved forensic data can be maintained and protected. The setup of evidence network (proposed in Subsection 3.6) increases the efficiency of supervision. We elaborate the details in the following aspects.

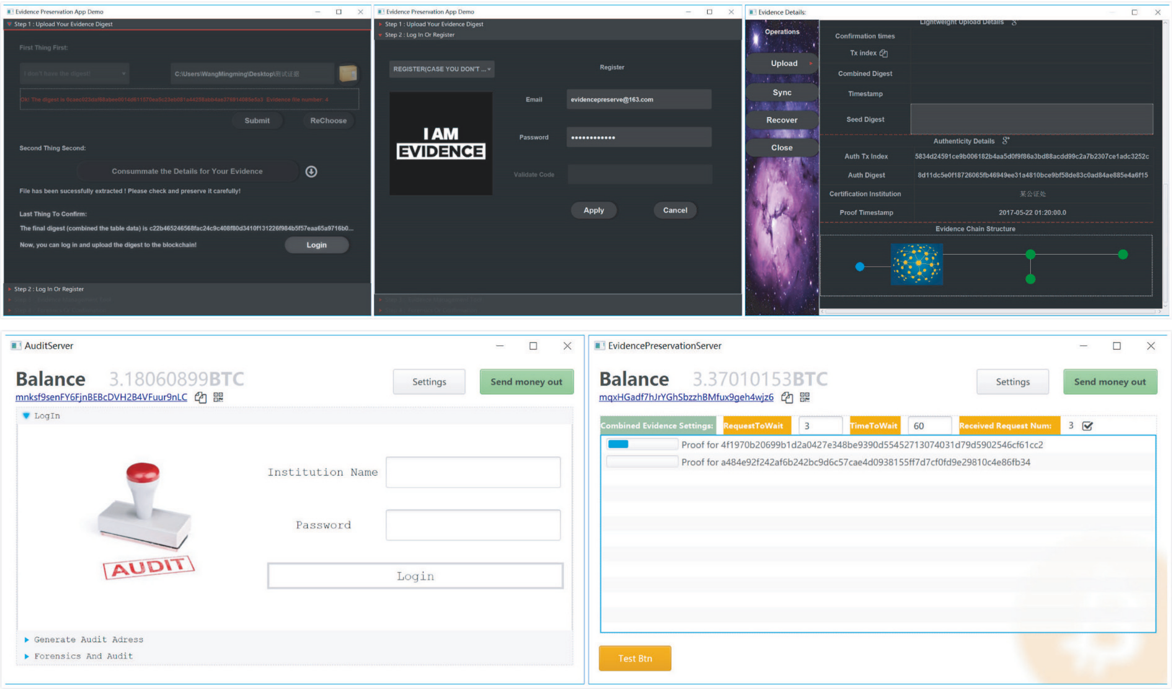
For the server, compared with traditional agencies, it does not conserve or handle any sensitive data from an original evidence file. Tasks are limited to help clients broadcast evidence Txs and synchronize the proof states from the Bitcoin blockchain. The validity and trustworthiness of all operations can be supervised through unattached Bitcoin query interfaces. It possesses the characteristics of transparency and lightweight, which is advantageous for supervisions from general public and associated institutions.

For legal institutions, it is convenient to carry out investigations on the existence, timeliness and social connections of evidence by indexing directive address from the evidence chain. General public can validate the existence and authentication of audit replies through the generation of audit proofs. When there exists supervision for legal cooperation between institutions, the apprehension from trust issues could be totally eliminated as unified and clear audit/evidence proofs are arranged on the evidence network. Also the process of recovery and synchronization could be managed rapidly.

For common users, the risk and doorsill of right maintenance have been greatly reduced. Fine-grained privacy can be proceeded though the scheme in Subsection 3.7 during the time of supervision.

5.5 Implementation

We have realized the implementation of our system, and it achieves the most features of our design. The software can be tested in both local regression network and Bitcoin public testing network — Testnet3. The software is developed by Java with the reference of the open source library Bitcoinj and follows the design pattern of MVC. All the functions of the system work as perfectly as scheduled during the test. The sketch map of our software is shown in Fig.13. More detailed in-



Evidence Preservation App Demo

Step 1 : Upload Your Evidence Digest

First Thing First:

I don't have the digest!

C:\Users\WangMingming\Desktop\测试证据

Ok! The digest is 0caec023daf68abee0014d611570ea5c23eb081a44258abb4ae376914085e5a3 Evidence file number: 4

Second Thing Second:

File has been sucessfully extracted ! Please check and preserve it carefully!

Last Thing To Confirm:

The final digest (combined the table data) is c22b465246568fac24c9c408f80d3410f131226f984b5f57eaa65a9716b0...

Now, you can log in and upload the digest to the blockchain!

Step 2 : Log In Or Register

Evidence Preservation App Demo

Step 1 : Upload Your Evidence Digest

Step 2 : Log In Or Register

REGISTER(CASE YOU DON'T ...

Register

Email: evidencepreserve@163.com

Password: ••••••••••••

Validate Code:

Evidence Details:

Lightweight Upload Details

Confirmation times:

Tx index:

Combined Digest:

Timestamp:

Seed Digest:

Authenticity Details

Auth Tx Index: 5834d24591ce9b006182b4aa5d0f9f86a3bd88acdd99c2a7b2307ce1adc3252c

Auth Digest: 8d11dc5e0f18726065fb46949ee31a4810bce9bf58de83c0ad84ae885e4a6f15

Certification Institution: 某公证处

Proof Timestamp: 2017-05-22 01:20:00.0

Evidence Chain Structure

AuditServer

Balance 3.18060899BTC

mnksf9senFY6FjnBEBcDVH2B4VFuur9nLC

LogIn

Institution Name:

Password:

Generate Audit Adress

Forensics And Audit

EvidencePreservationServer

Balance 3.37010153BTC

mqxHGadf7hJrYGhSbzzhBMfux9geh4wjz6

Combined Evidence Settings: RequestToWait 3 TimeToWait 60 Received Request Num: 3 ☑

Proof for 4f1970b20699b1d2a0427e348be9390d55452713074031d79d5902546cf61cc2

Proof for a484e92f242af6b242bc9d6c57cae4d0938155ff7d7cf0fd9e29810c4e86fb34

Fig.13. Software sketch map.

formation and further progress of the software can be obtained by accessing the Github link[7].

5.6 Future Goals

During our research, we have discovered great potential in evidence preservation in mature cryptocurrency system. By combining smart contracts and threshold signature on blockchain, a fine-grained forensic and management scheme can be designed. It would be convenient for internal management and forensics operation in large departments and institutions.

In addition, due to the activation of SegWit and the deployment of lightning network[18], we can use the time-lock transactions and micropayment channel to commit efficient proofs of existence scheme to preserve the evidence from the time of Tx broadcast to enough confirmations in block.

6 Related Work

There is a rich heritage of work in trusted timestamp and data auditing. Early in 1991, a complete and practical time-stamping scheme for documents in digital form was raised in [19]. However, centralized security risks and trust issues blocked the development of the field. After novel blockchain[10] designed by Nakamoto masterly figured out the consensus problems, abundant opportunities come forth. In [1], the concept of PoE was firstly brought forward with decentralized trusted timestamp. Pure timestamping services[8]~[10] were designed and released online on Bitcoin network. And further applications[20-22] appeared in the fields of video integrity, IoT (Internet of Things) and privacy-preserving abuse detection. PoE schemes timely solve credibility and integrity issues in such applications. Nevertheless, limitations in volume, speed and cost lower the quality of service. We partially avoid these issues by excavating covert channels to exploit suitable spaces preserving proof of evidence. Also we bring a competitive mechanism to form fine-grained hierarchical services, which balances the cost and the efficiency. Therefore a flexible commerce mode could be set up. Proofs generated by traditional PoE services are always scattered. We bridge evidence chains to accommodate the connections between evidences, which is much more organized in reality.

[7] https://github.com/Vivid-Wang/Clear-Evipreserve, Mar. 2018.

[8] https://app.originstamp.org/home, Mar. 2018.

[9] https://www.crunchbase.com /organization/bitproof, Mar. 2018.

[10] http://poex.io/, Mar. 2018.

There are also some solutions proposed to optimize the origin PoE services. In [23], standardized methods are brought out to optimize the process of hashes aggregation in proof generation. Our work in some sense extends this idea to fit forensic tooltips in extraction of metadata. In frameworks like [3], new blockchains are specially designed to preserve proofs and deploy more flexible rules, and anchor Txs are constructed to establish contacts with mature blockchain. Our work, however, is opposite to these efforts. To avoid the vulnerability of scrimpy computing power and over-concentrated distribution of nodes due to the sensibility of applications, we insist to deploy our service directly on mature blockchains. Efficient interconnection mode is exploited on basis of our evidence network, which offers perfect scalability and manageability.

Research in [24] summarizes the consequences and suitability of blockchain technology in law systems and forensic services. Our work specially contributes to consummating the procedure in audit, supervision and legal cooperations. The origin PoE service is adaptively extended to deal with the forensic work in evidence lifecycle.

7 Conclusions

In this paper, we proposed a lightweight digital evidence-preserving architecture built on top of the Bitcoin blockchain. The system possesses the feature of privacy-anonymity, audit-transparency, function-scalability, and operation-lightweight. It provides a perfect solution to the security and confidence problems compared with the traditional schemes. Due to the construction of evidence chain, the system also operates in sustainable stability and scalability, which is greatly propitious to the ecology of the application. The efficient interconnection mode offers shortcuts for recovery, synchronization and trusteeship of evidence, which makes the cooperation and supervision process simple and powerful. Users can easily obtain the proof of existence and the audit for their evidence, and in the meanwhile the system is able to conduct management on blockchain efficiently. Fine-grained privacy control and hierarchical schemes further improve the quality of service. As for the future work, the cost, efficiency and service scope will be further optimized by employing the newly proposed techniques including segregated witness, lightning network and so on. Our system will focus on providing comfortable service with strong security and perfect transparency.

References

[1] Gipp B, Meuschke N, Gernandt A. Decentralized trusted timestamping using the crypto-currency bitcoin. arXiv: 1502.04015, https://arxiv.org/abs/1502.04015, Mar. 2018.

[2] Lampson B W. A note on the confinement problem. *Communications of the ACM*, 1973, 16(10): 613-615.

[3] Snow P, Deery B, Lu J, Johnston D, Kirby P. Factom: Business processes secured by immutable audit trails on the blockchain. Whitepaper, Factom, 2014. https://raw.githubusercontent.com/FactomProject/FactomDocs/master/Factom_Whitepaper.pdf, Mar. 2018.

[4] Back A, Corallo M, Dashjr L, Friedenbach M, Maxwell G, Miller A, Poelstra A, Timon J, Wuille P. Enabling blockchain innovations with pegged sidechains. https://www.blockstream.com/sidechains.pdf, Mar. 2018.

[5] Casey E. Digital Evidence and Computer Crime. Elsevier Academic Press, 2004.

[6] Dinev T, McConnell A R, Smith H J. Research commentary: Informing privacy research through information systems, psychology, and behavioral economics: Thinking outside the "APCO" box. *Information Systems Research*, 2015, 26(4): 639-655.

[7] Adams R. The advanced data acquisition model (ADAM): A process model for digital forensic practice. *Journal of Digital Forensics, Security & Law*, 2013, 8(4): 25-48.

[8] Dan A, Iyengar A K, Kumar M. System and method for providing trusted services via trusted server agents: US Patent 6823456, 2004. http://www.freepatentsonline.com/6823456.html, Mar. 2018.

[9] Stein B. Fuzzy-fingerprints for text-based information retrieval. In *Proc. the 5th Int. Conf. Knowledge Management*, Oct. 2005, pp.572-579.

[10] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. 2008. https://bitcoin.org/bitcoin.pdf, Mar. 2018.

[11] Lemieux V L. Trusting records: Is blockchain technology the answer? *Records Management Journal*, 2016, 26(2): 110-139.

[12] Simmons G J. The prisoners' problem and the subliminal channel. In *Advances in Cryptology*, Chaum D (ed.), Springer, 1984, pp.51-67.

[13] Chapman M, Davida G. Hiding the hidden: A software system for concealing ciphertext as innocuous text. In *Proc. the 1st Int. Conf. Information and Communications Security*, November 1997, pp.335-345.

[14] Bellare M, Paterson K G, Rogaway P. Security of symmetric encryption against mass surveillance. In *Proc. the 34th Int. Cryptology Conf.*, August 2014.

[15] Kilroy Jr R J. No place to hide: Edward Snowden, the NSA, and the U.S. surveillance state. *Journal of Systems and Software*, 2016, 9(3): 99-102.

[16] Song H, Dharmapurikar S, Turner J, Lockwood J. Fast hash table lookup using extended bloom filter: An aid to network processing. *ACM SIGCOMM Computer Communication Review*, 2005, 35(4): 181-192.

[17] Biryukov A, Dinu D, Khovratovich D. Argon2: The memory-hard function for password hashing and other applications. 2015. https://password-hashing.net/argon2-specs.pdf, Mar. 2018.

[18] Poon J, Dryja T. The bitcoin lightning network: Scalable off-chain instant payments. Technical Report (DRAFT), Draft Version O.5.9.2, 2015. http://lightning.network/lightning-network-paper.pdf, Mar. 2018.

[19] Haber S, Stornetta W S. How to time-stamp a digital document. In *Proc. Conf. Theory and Appl. Cryptography*, May 1990, pp.437-455.

[20] Gipp B, Kosti J, Breitinger C. Securing video integrity using decentralized trusted timestamping on the bitcoin blockchain. In *Proc. the 10th Mediterranean Conf. Information Systems* (*MCIS*), September 2016.

[21] García-Recuero A, Burdges J, Grothoff C. Privacy-preserving abuse detection in future decentralised online social networks. In *Proc. the 11th Int. Workshop on Data Privacy Management*, September 2016, pp.78-93.

[22] Conoscenti M, Vetro A, de Martin J C. Blockchain for the Internet of Things: A systematic literature review. In *Proc. the 13th Int. Conf. Computer Systems and Appl.* (*AICCSA*), November 2016.

[23] Pedro Crespo A S, Garcia L I C. Stampery blockchain timestamping architecture (BTA). 2016. https://s3.amazonaws.com/stampery-cdn/docs/Stampery-BTA-v6-whitepaper.pdf, Mar. 2018.

[24] Hegadekatti K. Legal systems and blockchain interactions. 2017. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2893128, Mar. 2018.

Mingming Wang received his Bachelor of Engineering degree in electrical engineering from Beihang University, Beijing, in 2017. He is now pursuing his Ph.D. degree in information and communication engineering in Beihang University, Beijing. His research interests include blockchain, classic cryptography and game theory.

Qianhong Wu received his Ph.D. degree in cryptography from Xidian University, Xi'an, in 2004. Since then, he has been with Wollongong University (Australia) as an associate research fellow, with Wuhan University (Wuhan) as an associate professor, with Universitat Rovira i Virgili (Catalonia) as a research director, and now with Beihang University (Beijing) as a professor. He is a member of CCF, IACR, ACM, and IEEE. His current research interests include cryptography, data security and privacy, and information theory.

Bo Qin received her Ph.D. degree in cryptography from Xidian University, Xi'an, in 2008. Then, she was with Xi'an University of Technology (Xi'an) as a lecturer and with Universitat Rovira i Virgili (Catalonia) as a postdoctoral researcher. She is currently a lecturer in Renmin University of China, Beijing. Her research interests include pairing-based cryptography, data security and privacy, and VANET security. She has authored over 80 publications in well-recognized journals and conferences and served in the program committee of a number of international conferences in information security.

Qin Wang received his Bachelor of Engineering degree in electrical engineering from Northwestern Polytechnical University, Xi'an, in 2015. He is now pursuing his Master's degree in information and communication engineering in Beihang University, Beijing. His research interests include blockchain, classic cryptography and cloud security.

Jianwei Liu received his B.S. and M.S. degrees in electronic and information from Shandong University, Jinan, in 1985 and 1988, respectively. He received his Ph.D. degree in communication and electronic system from Xidian University, Xi'an, in 1998. He is now a professor of electronic and information engineering at Beihang University, Beijing. His current research interests include wireless communication network, cryptography, and information and network security.

Zhenyu Guan received his Ph.D. degree in electronic engineering from Imperial College London, UK, in 2013. Since then, he has joined Beihang University (Beijing) as a lecturer. He is a member of IEEE and IEICE. His current research interests include cryptography engineering, security of IoT, blockchain.