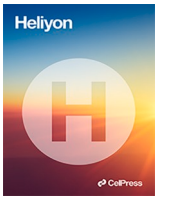
Heliyon 9 (2023) e17084

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Blockchain-based fake news traceability and verification mechanism

Xiaowan Wang [a,b], Huiyin Xie [c], Shan Ji [d,*], Liang Liu [d], Ding Huang [e]

[a] Beijing Normal University, Beijing, 100000, China
[b] Xi'an University of Posts & Telecommunications, Xi'an, 710000, China
[c] Yunnan University, Kunming, 650091, China
[d] Nanjing University of Aeronautics and Astronautics, Nanjing, 210000, China
[e] School of Computer, University of Information Science and Technology, Nanjing, 210000, China

ARTICLE INFO

Keywords:
Fake news data
Blockchain
Polynomial commitment

ABSTRACT

The rapid development of the Internet and Internet of Things has rapidly introduced human society into the information age, and the way of fake news production has been updated, which has greatly affected the normal life of human beings. In order to identify worthless fake news and trace massive fake news data from unknown sources, and share valuable news data to fully disseminate effective real news, news owners usually store news data in cloud. Users of IoT terminals can access news data on demand without storing it locally. However, the authenticity of the fictive newspaper numbers source, which is easy to destroy, and the social media platform. Besides, when massive news data is saved on cloud server, the news owners have to at the risk of lose physical control over news data and it will face the risk of fake news being disseminated and real news being falsified. Thus, this paper proposes a novel mechanism for secure storage of news data using blockchain technology. Firstly, traceability and verification of fake news data is improved by the cooperative storage model on and off the chain. Secondly due to the inability of past polynomial commitment to update the commitment, we will be a hindrance to use polynomial commitment to build a secure authentication protocol. Therefore, in this paper, we design the update algorithm for polynomial commitment in order to be able to guarantee the consistency of on-chain and blockchain database news data.

1. Introduction

Today, information dissemination is characterized by immediate, interactive, global, and massive data. With the popularization of the Internet, the development of communication technology and Internet of Things (IoT) [1–3], the prosperity of social media platforms, the rapid dissemination of news information, diversification of public access to information channels, many IoT terminal devices need to obtain news data from the cloud. More and more individual users are producing and publishing news data as self-media and releasing valuable news data on social platforms, which are characterized by a large volume of data, multiple data types, high content value and fast processing speed, bringing many conveniences to the development of news dissemination. The world has entered the era of omni media. Domestic resource integration and multimedia channel advantages are at the forefront of development.

* Corresponding author.
E-mail address: shanji2022082022@163.com (S. Ji).

https://doi.org/10.1016/j.heliyon.2023.e17084
Received 31 March 2023; Received in revised form 3 June 2023; Accepted 7 June 2023
Available online 15 June 2023

Therefore, in this context, a new model of news production, crowdsourced journalism, has arisen. Users have become the owners and publishers of news, choosing their own topics, producing content, and editing it for publication. Thus, the authenticity and reliability of news data is particularly important. Unreliable data sources, exposed communication links and insecure storage environments provide conditions for the generation and spread of fake news [4].

Currently, there is a lack of consensus on a universally accepted definition of fake news. Previous studies have often linked terms and concepts like fake news to the dissemination of deceptive information.

It is seen as an "intentional news article" (deceptive news) [5], news articles or information published and disseminated through the media, regardless of the means, the motives behind it" and fake news, disinformation [6], misinformation [7], satirical news [8], [9]. Fake news is widely regarded as a significant menace to democracy, journalism, and the fundamental principles of freedom of expression. It erodes public trust in government [10], [11]. While fake news is not a new phenomenon [12], [13], but questions such as why it has become a topic of global interest and why it is increasingly attracting public attention are common which is very relevant at this point. The main reason is that fake news can be created and published online faster and cheaper than traditional news media such as newspapers and television [14]. The rise and popularity of social media has also played an important role in this surge of interest [15], [16]. By August 2018, approximately 68% of Americans relied on social media as a primary source of news. However, the echo chamber effect prevalent on social media platforms often leads to the amplification and reinforcement of biased information [17].

The storage, processing and mining of news data has always been the main content of news-based data management. 5G technology improve the data transmission speed while simultaneously solving the problem of computing power and storage capacity of localized news data [18], [19].

News data owners outsource important news data to cloud servers for storage [20] Since news owners lose actual control over the news data, cloud database servers can modify the news data at will, which result in news data being always at risk, greatly increasing the probability of fake news. Therefore, the consequent problem of tracing and verifying fake news has become particularly prominent [21], [22]. Due to the storage and computing limitations of IoT devices, news data owners choose to store their data in cloud servers. However, when numerous IoT end devices query the news data in the cloud server, it is particularly important to ensure the integrity of the data [23].

Blockchain technology has many features, such as decentralization, traceability, and non-tamperability, which can better provide confidentiality and integrity for news data, and realize data rights confirmation and other services [24], [25]. However, due to the massive, diverse and streaming characteristics of news data. In the face of massive and growing news data, the traditional blockchain architecture will be difficult to meet the storage and performance requirements, so directly storing data on the blockchain is not a good choice. In order to ensure the timeliness, security and traceability while storing massive news data, it is urgent to design a new blockchain storage architecture. For this reason, this article has done the following:

- This paper designs a news data security storage model based on blockchain, which stores news metadata on the blockchain, and stores the complete data in the blockchain database server. It can store a large amount of news data while ensuring the security of the data to detect the fake news, and can quickly update the news data.
- An updatable polynomial commitment is designed, which can query and update one or more positions in the commitment, which greatly reduces the computational cost and improves the efficiency.
- A secure storage scheme for constructing news data using updatable polynomial commitments. It enables the news information verifier in the model to query, verify and update massive news data in batches. The integration with the blockchain guarantees the integrity and consistency of on-chain news metadata and blockchain database specific news data.
- The update records of news data are recorded in the blockchain, which realizes the traceability of news data, to filter fake news.

The remainder of this paper is structured as follows: Section 2 provides an overview of the related research in the fields of news data traceability, verification, and polynomial commitment.

Section 3 highlights the security challenges associated with news data, shedding light on the existing problems. Section 4 presents an elaborate explanation of our proposed mechanism for blockchain-based traceability and verification of fake news. This section encompasses key aspects such as the on-chain process of news data, a secure storage model based on blockchain technology, verification of news authenticity using polynomial commitment, and comprehensive security proofs. Finally, in Section 5, we draw conclusions based on our findings and contributions presented in this paper.

2. Related work

2.1. News data traceability and verification

The risks of sharing news data in cloud computing have received increasing attention. After some owners of confidential data outsource news data to cloud storage services and share it with others, data owners largely lose control over the data [26], [27]. New data traceability technology is a traceability technology proposed for data, aiming to reproduce the historical state and evolution of data throughout its life cycle. In 2016, To protect the identity privacy of members in the system and ensure the traceability of identities, Yang et al. [28] designed an efficient public audit solution. But when people use the browser's web front-end interface, it is difficult to build trust with cloud service providers and supervise their fulfillment of service agreements. In 2017, Dawle et al. [29] designed a database intrusion detection mechanism to enhance the security of the database, which uses SQL injection on the website to record all the activities of intruders. Ramachandran [30] implements a cloud data procurement framework through blockchain, which

is secure, tamper-proof, low-cost source data, and improves privacy and usability. Simultaneously, it has been acknowledged that a blockchain-based framework for automated verification of source records is currently lacking. To address this gap, they proposed solution incorporates smart contracts and an Open-Source Model (OPM). This innovative approach leverages cloud-based verification scripts to enable seamless and dependable source capture, verification, and management processes. In 2018, Banerjee [31] proposed a blockchain-based solution to increase product sourcing in two ways, one is independent tracking through blockchain, where sources and provenance like organizations act as mapping platforms to facilitate source tracking for customers by using global maps. The other solution is based on custom source solutions, where software solution companies provide software solutions to customers based on their needs with explicit validation of products and industries. In 2019, To contribute anonymous and traceable data across groups, Huang et al. [32] et al. combined blockchain technology to enable members of a system composed of different groups to easily share data without the participation of a third party.

2.2. Polynomial commitment

Polynomial commitment scheme has been an important field of research in the cryptographic area. Chase [33] introduced unpredictable commitment to construct ZKS in 2005, which eventually formed commitment schemes for commitment message vectors. In 2018, Bünz et al. [34] proposed a technique for constructing polynomial commitment scheme based on multi-round interaction protocols. Vlasov [35] proposed a new efficient and transparent construction of polynomial commitment schemes in 2019. The polynomial commitment scheme enables the prover to generate a commitment to a polynomial of a predetermined order d using a string. This commitment can be subsequently used by the verifier to verify the calculated value of the committed polynomial at a specific point. In 2020, Boneh et al. [36] proposed to perform the computation based on two additive groups, for which they need the Q-DOG assumption. To compute efficiently, their protocol uses a pairing mapping, which effectively maps elements from two additive groups G_1, G_2 to another multiplicative group G_t. The authors proposed the Q-DLOG assumption, which is a generalization of the discrete logarithm assumption. In the same year, Boneh et al. [37] extended the additionally proposed Halo method to any polynomial commitment scheme with a more general nature, that is, the proposed scheme demonstrates the capability to aggregate commitments in a linear combination, resulting in a compact commitment that can later be efficiently opened to reveal the linear combination.

3. Problem statement

With the development of communication technology, crowdsourced journalism [38], [39] has become an innovative experiment in changing existing production models. The combination of crowdsourcing and journalism is mainly reflected in the development of user-generated content, such as video, audio and reporting leads provided by the audience. The social and economic value of news data has been brought into full play. However, in the process of news data collection, the quality of news data cannot be guaranteed due to unknown data sources (official or unofficial) and invisible communication links [40] Besides, when self-publishing news owners store news data in the servers of service providers [41], [42], they inevitably face communication and storage security issues. Malicious unregulated third parties or unauthorized malicious groups may cause damage to the security of news stored in the cloud. The integrity and consistency of news data will be greatly challenged.

Information security requires guaranteeing the confidentiality, availability, and integrity of news data. 1) Confidentiality: Unauthorized people cannot access news data created by news owners 2) Availability: News data provided by news owners is valuable, timely and accurate. 3) Integrity: Ensuring that news data is not tampered with without authorization or can be detected at time after being tampered to solve the lack of reciprocity of management system privacy protection schemes, so that malicious attackers can use some tricks to detect the real location of users to solve the problem of ensuring data confidentiality. In 2019, Saad et al. [43] hypothesized that a new blockchain system could overcome existing challenges and limit the spread of fake news in the network. To this end, we examine the information flow within social networks and develop an efficient detection system that minimizes resource requirements while ensuring optimal performance. Balouchestani et al. [44] propose a new method for using blockchain called SANUB, which not only provides anonymous news publishing, news evaluation, journalist verification, fake news detection, and proof of news ownership and other functions, but also detect fake news. Song et al. [45] propose an architecture that can realistically archive content on social media using blockchain technology.

However, when news data is stored in an untrusted third-party cloud server, although the past schemes have stored news data on the blockchain to solve the problem of secure data storage, there is no blockchain for storing news. The limitation of data capacity is studied. Therefore, we have designed a on-chain and off-chain collaboration blockchain-based solution for the secure storage of news data. By leveraging the tamper-evident and traceability features of blockchain technology, we successfully establish the integrity and non-repudiation of news data. This ensures that the information remains unaltered and can be traced back to its origin, thereby enhancing the overall credibility and accountability of the news data. Besides, to reduce the blockchain storage limitation of huge amount of news data stored in the blockchain [46], [47], we designed a collaborative storage structure to store abstract of news data on the chain and concrete data under the chain, thus increasing throughput of news data. Full news data is stored by storing it in a database on a blockchain database cloud server and saving metadata news data representing proofs of data in on-chain. We store the commitment value of the polynomial commitment generated by the data update as metadata into the block on the chain, so that the off-chain data is bound to the commitment on the chain. Contact to achieve the purpose of binding off-chain and on-chain, thus ensuring consistency between on-chain metadata of news data and blockchain database concrete news data.

4. Blockchain-based fake news traceability and verification mechanism

We propose a blockchain-based news data security storage model. The digests generated by dynamically updated news data are stored on the blockchain as news metadata, and the specific news data is saved in the blockchain database cloud storage server. To enhance the storage capacity of the blockchain, we adopt a collaborative approach that combines on-chain storage and blockchain database. By leveraging the decentralized, traceable, and tamper-proof nature of blockchain technology, we can ensure the utmost security of news data storage. In this model, the integrity and consistency of extensive news data stored in the off-chain cloud server and the metadata stored on-chain are meticulously maintained. This approach guarantees the overall reliability and integrity of the news data throughout the storage process. We design an on-chain and blockchain database verification mechanism for news data blockchain based on polynomial commitments. The corresponding blockchain database news data is regularly verified to verify the integrity and consistency of on-chain and blockchain database massive news data in cloud server.

4.1. News data on-chain process

Introduce a few terms related to blockchain first. Associated with each epoch of the blockchain is a unique puzzle *ID* that is known to all nodes, denoted as ID_p. During the working process of Bitcoin, all nodes involved in mining try to solve a Scratch-off puzzle, as it is the key to the consensus mechanism. The Scratch-off puzzle can be simply defined by two algorithms (Guess, Wining).

(1) Guess: Guessing algorithm, also known as ticket generation algorithm. Input ID_p, this algorithm can randomly generate a candidate solution to the Scratch-off puzzle, which is often referred to as ticket.
(2) Wining: Verification algorithm. Input ID_p, Z, ticket, if the ticket is indeed the solution to the Scratch-off puzzle, the algorithm outputs 1, otherwise, the algorithm outputs 0. Input Z determines the difficulty of solving the Scratch-off puzzle, and in some solutions, Z can be described as Hardness.

The scheme proposed is based on the Scratch-off puzzle mechanism described above. Corresponding tickets can be generated through the retrievability proof, which can be used to decide who wins the proof between nodes. The workflow of the news storage scheme based on the improved retrievability proof contains four stages: parameter setting, transaction request, node consensus and news on-chain, and its workflow is shown in Fig. 1.

(1) Parameter setting stage: This stage refers to the parameter setting algorithm of the retrievability proof. First, an encoding algorithm is applied to divide some news ∂ into a data block set ∂', which contains $N = \rho n$ data blocks, denoted as $\overrightarrow{\partial_i}$. The node holding the public key pk can choose a subset S_∂ of the data block set to store, assuming that each node stores ι data blocks. Using the private key sk and pseudo-random number family, this scheme can generate a tag, denoted as t. The final data blocks set ∂' can form a set $\{i, \overrightarrow{\partial_i}, t\}$, in which i satisfies $0 \le i \le n-1$.
(2) Transaction request stage: The purpose of this stage is that the node generates a ticket through a proof of retrievability, instead of the Guess algorithm of the above Scratch-off puzzle, which can be used as a proof that it has mined the block. Ticket is a set of information related to the proof and verification of news ∂, and generates a signature σ. In the Permacoin scheme, an efficient

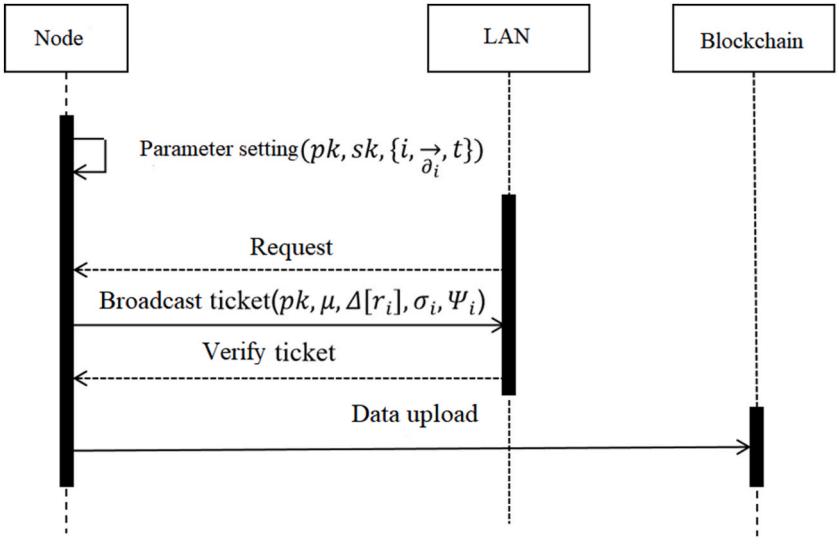


Fig. 1. News secure storage solution workflow.

multi-use hash signature scheme called Floating-Preimage Signature (FPS) is employed. The public key of the FPS scheme corresponds to the root of a Merkle tree. This key comprises a collection of randomly generated strings that are accessible through a pseudo-random function, facilitating the selection of a subset of leaves for signature display. Within this scheme, the FPS mechanism is utilized to sign news articles. Specifically, the first k iterations involve a Scratch-off process, while the subsequent (k+1) iterations serve as rewards in the form of bitcoins upon successful mining.

Let ID_p be a public puzzle identifier that does not require repeated computation, the node generates random challenges set $((i,v_i),r_i)$, where $i \in C$, $C \in [0, n-1]$. This scheme generates proofs without selecting random strings, using the property of updatable polynomial commitment. This scheme can generate proofs ψ_{r_i}. The final obtained ticket can be expressed as follow equation (1):

$$\forall i = 1, 2, \ldots k, ticket = (pk, \mu, \Delta[r_i], \sigma_i, \psi_{r_i}) \tag{1}$$

where ψ_{r_i} is the proof generated by the node's public key, σ_i represents the signature, $\mu = \sum_{i \in C} v_i F[r_i], 0 \le i \le n-1$. Finally, the node broadcasts its ticket in the network for other nodes to verify.

(3) Node consensus: All nodes compete to generate an appeal ticket through the Guess algorithm until a winning node appears. The winning node's ticket needs to be verified by all nodes to reach a consensus, and there are only two results: verification success and failure.
(4) News on-chain: In the final verification stage, after giving a ticket, all nodes need to verify whether Scratch-off is executed correctly during the mining process. After the winning node's ticket is verified by other nodes, the ticket is issued by the winning node and added to the blockchain, and he will also receive a fixed amount of coin rewards.

4.2. News data storage and validation model

As an important social resource, massive news data produced by media companies or governments has been valued by the whole society and even the country for its data security storage. The authenticity of news data will not only have a great impact on individuals, but also have important implications for the development of society. Hence, there is a pressing societal need to safeguard the vast amounts of news data stored within untrusted cloud servers. To address this concern, this paper proposes a blockchain-based news data security storage model that involves several key entities. These entities comprise news owners, responsible for creating and managing news data, news verifiers who verify the authenticity of the news, offline chain storage servers for secure storage, and online data blocks that hold the news data in a decentralized manner. Together, these entities work collaboratively to establish a robust and secure framework for storing news data while ensuring its integrity and authenticity. Fig. 2 describes the diagram of system model.

News Owners: In this proposed system, news data owners, who are members of the blockchain system, generate a public key (PK) and a private key (SK) during the key generation and setup stage. These keys are used to create signatures for the initial news data.

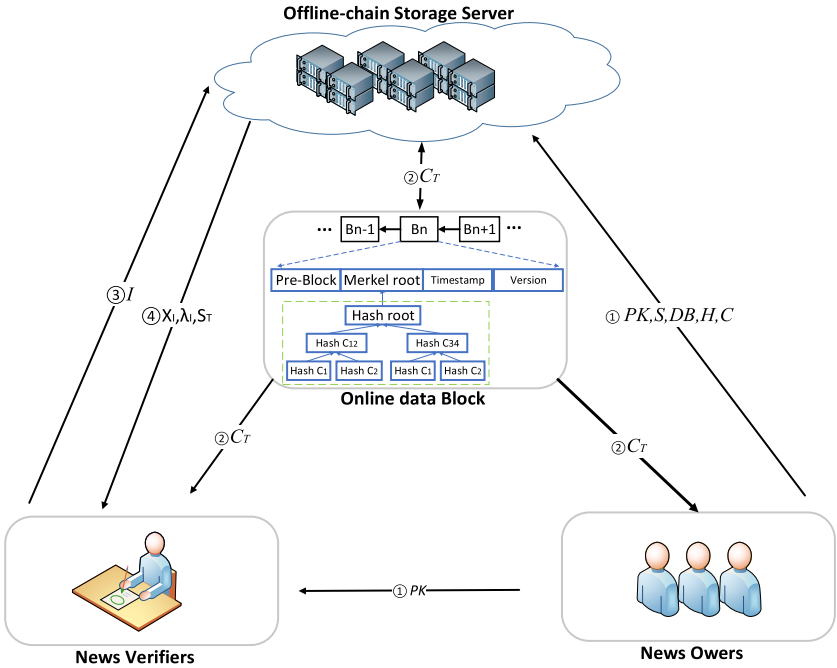


Fig. 2. System model diagram.

Once verified by other nodes, the signatures are stored on the blockchain, while the complete news data is stored off-chain. Multiple values of news data can be simultaneously recorded or updated, with corresponding digital signatures generated. These signatures and commitments are then combined and stored on the blockchain, undergoing further verification by other nodes. Following this, news owners can update the news data, which is stored off-chain, while the corresponding news records in the cloud server are modified accordingly.

Offline-chain Storage Server (OffCSS): The owners of the news data upload complete news data to the blockchain database cloud storage server. When initial storage of the blockchain database news data or data update is performed, the original news data or the proof created by the update algorithm needs to be stored in the blockchain database cloud database. Furthermore, the storage cloud server will assume an additional responsibility of computing proofs. This entails generating a proof in response to a query from the respective news data verifier. The proof is then transmitted back to the verifier, providing them with the necessary evidence or validation they require for the specific news data in question.

Online Data Block (OnDB): News data owner will memorize the abstract created from news data in the on-chain data block, that is, if news data in the blockchain database cloud server database is modified, inserted or deleted, the cloud server will first verify the digest of news and store it in the blockchain after the verification is passed, which is saved in the same way like the traditional blockchain data storage.

News Verifiers: The news data verifiers inquire news data in blockchain database cloud storage server, gets the respective data, after that, perform verifying algorithm to proof the validation of news data. For the news data in cloud database, news data verifiers will send 1 to n query requests $(I_1, I_2, ..., I_n)$ at each time and news data cloud servers perform the query algorithm, then send the news information set of the respective location and create the relevant verifications to news verifiers. After getting the news proofs and information, the data verifiers perform the verifying algorithm and return the news if the proofs are valid.

4.3. Verification of news authenticity based on polynomial commitment

1) Random selection by news owners: Let $\varnothing(x) \in Z_P[x]$, and polynomial $\varnothing(x)$. Note that $\varnothing(k_i) = v_i$. After that, the authority generates $\overline{\varnothing}(x) \in Z_P[x]$ and a commitment for the $\varnothing(x)$ as $C = g^{\varnothing(s)} \bullet P^{\overline{\varnothing}(s)} \in G$. Authority is the owners of news.

2) The news verifier uses the PK and the current keys of blockchain system to verify the news data. First, the news verifier provides a query index i. Then, the offline server extracts the corresponding k_i and v_i, $1 \le i \le n$ from the offline database and sends them to news owner. Subsequently, the authority computes auxiliary polynomials $\varphi_{k_i}(x) = \frac{\varnothing(x)-\varnothing(k_i)}{x-k_i}$ and $\overline{\varphi_{k_i}}(x) = \frac{\overline{\varnothing}(x)-\overline{\varnothing}(k_i)}{x-k_i}$. Finally, the authority computes auxiliary parameters $\gamma_{k_i} = g^{\varphi_{k_i}(s)} p^{\overline{\varphi_{k_i}}}(s)$, $C_{k_i} = g^{\varnothing(k_i)} p^{\overline{\varnothing}(k_i)}$ and $\eta_{k_i} = g^{k_i}$. The proof of the query data record $\tau = (k_i, \varnothing(k_i), \overline{\varnothing}(k_i), \gamma_{k_i}, C_{k_i}, \eta_{k_i})$.

4.4. Updatable polynomial commitment and fake news verification mechanism

Since the original polynomial commitment scheme does not have the function of witness updatability and it is not scalable and not suitable for certain application scenarios where elements need to be added and deleted. This paper first proposes an updatable polynomial commitment that can achieve efficient update of the witness.

4.4.1. Definition

The proposed updatable polynomial commitment algorithm is based on a bilinear mapping implementation, and an updatable polynomial commitment scheme can be defined by seven algorithms $(Setup, Commit, Open, VerifyPoly, Creatwitness, VerifyEval, Update)$.

(1) *Setup*: parameter setting algorithm. The algorithm generates the algebraic structure $\mathscr{G}$ and polynomially committed $\langle PK, SK \rangle$. This step of the algorithm is run by a trusted or decentralized authority, and it should be noted that the remaining steps of the scheme do not require the use of the private key.

(2) *Commit*: commitment algorithm. Input PK and polynomial $F(x)$, output the commitment $\mathscr{C}$ of the $F(x)$, and a decommit information $\mathscr{D}$, which is not needed in some schemes.

(3) *Open*: open commitment algorithm. Outputs the polynomial $F(x)$ and the decommit information $\mathscr{D}$ used in constructing the commitment.

(4) *VerifyPoly*: polynomial verification algorithm. Verify whether $\mathscr{C}$ is a commitment of polynomial $F(x)$. The verification passes outputs 1, indicating that the verification passes, otherwise it outputs 0.

(5) *Creatwitness*: witness generation algorithm. Output $\langle i, F(i), \omega_i \rangle$, ω_i is the witness of the polynomial $F(x)$ at the index i of the value $F(i)$, witness is like the proof.

(6) *VerifyEval*: witness verification algorithm. Verify whether ω_i is a witness to the value of the $F(x)$ at the index i of value $F(i)$. Likewise, if the verification process is successful, the algorithm will produce an output of 1; otherwise, it will produce an output of 0.

(7) *Update*: update algorithm. The update algorithm contains two operations: add and delete. The algorithm checks whether the index i^+/i^- of the witness to be added/deleted is in the index set I, and if the verification is successful, the witness set W can be updated by adding/deleting the corresponding witness.

A secure polynomial commitment scheme has to satisfy three security properties: correctness, binding and hiding, and specifically

defined as follows.

(1) Correctness: The correctness of an updatable polynomial commitment means that for a commitment $\mathscr{C}$ output by the Commit algorithm and all polynomials $F(x) \in Z_P[\mathscr{X}]$, it needs to satisfy:
   ① The output of the Open algorithm of the polynomial commitment can be verified by the *VerifyPoly* algorithm
   ② The output $\langle i, F(i), \omega_i \rangle$ of the *Creatwitness* algorithm of the polynomial commitment can be verified by the *VerifyEval* algorithm.

(2) Binding: The binding of updatable polynomial commitments is further divided into evaluation binding and polynomial binding, for any arbitrary B, its specific definitions are as follows two probability formulas (2) and (3):

$$Pr\begin{bmatrix} PK \leftarrow Gen(1^k), (\mathscr{C}, \langle F(x), F(x)^{'}\rangle) \leftarrow B(PK): \\ VerifyPoly(PK, \mathscr{C}, F(x)) = 1 \wedge \\ VerifyPoly(PK, \mathscr{C}, F(x)^{'}) = 1 \wedge \\ F(x) \neq F(x)^{'} \end{bmatrix} = \varepsilon(\kappa) \tag{2}$$

$$Pr\begin{bmatrix} PK \leftarrow Gen(1^k), (\mathscr{C}, \langle i, F(i), \omega_i \rangle, \langle i, F(i)', \omega_i' \rangle) \\ \leftarrow B(PK): \\ VerifyPoly(PK, \mathscr{C}, i, F(i), \omega_i) = 1 \wedge \\ VerifyPoly(PK, \mathscr{C}, i, F(i)', \omega_i') = 1 \wedge \\ F(x) \neq F(x)^{'} \end{bmatrix} = \varepsilon(\kappa) \tag{3}$$

(3) Hiding: The hiding of an updatable polynomial commitment means that the constructed scheme needs to satisfy:
   ① Computational hiding: for any unqueried index j. There is no adversary B can compute $F(j)$ in a non-negligible probability.
   ② Unconditional hiding: for any unqueried index j, no computationally infinite adversary B has any information about $F(j)$.

4.4.2. Specific scheme

An updatable polynomial commitment scheme consists of seven algorithms: *Setup*, *Commit*, *Open*, *VerifyPoly*, *Creatwitness*, *VerifyEval*, *Update*. The detail of the scheme is described as follows:

(1) *Setup*: Trusted authority computes two groups G and G_T of prime order p (with k bit safety), such that there is a symmetric bilinear pair $e: G \times G \rightarrow G_T$, and the t-SDH assumption holds. The updatable polynomial commitment to represent the generation of bilinear groups with: $\mathscr{G} = \langle e, G, G_T \rangle$. A generator selected $g \in_R G$, and μ is the *SK*. The algorithm generates a set of $\langle g, g^{\mu}, \ldots, g^{\mu^n} \rangle$, outputs the public key $PK = \langle \mathscr{G}, g, g^{\mu}, g^{\mu^2} \ldots, g^{\mu^n} \rangle$.

(2) *Commit*: Input public key *PK*, calculate the commitment of the polynomial $F(x) \in Z_P[\mathscr{X}]$ (the number of polynomials is less than or equal to t) $\mathscr{C} = g^{F(\mu)} \in G$, because $F(x) = \sum_{j=0}^{deg\,(F)} F_j x^j$, the polynomial commitment of the final output of the algorithm can be expressed as follow equation (4) :

$$\mathscr{C} = \prod_{j=0}^{\deg\,(F)} \left(g^{\mu^j}\right)^{F_j} \tag{4}$$

(3) *Open*: Output the promised polynomial $F(x)$.

(4) *VerifyPoly*: Verify whether the formula $\mathscr{C} = g^{F(\mu)}$ holds. If the equation $F(x) = \sum_{j=0}^{deg\,(F)} F_j x^j$ for $F(x) = \sum_{j=0}^{deg\,(F)} F_j x^j$ is valid, otherwise, the algorithm outputs 0. Note that the algorithm works if and only if $deg\,(F) \leq t$.

(5) *Creatwitness*: takes *PK*, index i as input, and check whether $i \in I$. Calculate equation (5) as follow:

$$f_i(x) = \frac{F(x) - F(i)}{x - i} \in Z_P[\mathscr{X}] \tag{5}$$

Output $\langle i, F(i), \omega_i \rangle$, $\omega_i = g^{f_i(\mu)}$, similar to the method for calculating commitments. Then, generate the set $W = \{ \omega_1, \omega_2, \ldots, \omega_t \}$ through ω_i.

(6) *VerifyEval*: Verify that $F(i)$ as the evaluation of the polynomial promised by $\mathscr{C}$ at the index i, and check whether $i \in I$ is satisfied, otherwise the algorithm terminates. If the verification of equation (6) holds.

$$e(\mathscr{C}, g) = e\left(\omega_i, g^{\mu} / g^{i}\right) e(g, g)^{F(i)} \tag{6}$$

(7) *Update*: The witness update algorithm includes witness addition and deletion algorithms.

Witness addition algorithm: Input public key *PK*, index set I, witness set W and index to be added i^+, check whether $i^+ \in I$, otherwise the algorithm outputs 1. Calculate equation (7) as follow:

$$f_{i^+}(x) = \frac{F(x) - F(i^+)}{x - i^+} \in Z_P[\mathscr{X}] \tag{7}$$

Where $\omega_{i^+} = g^{f_{i^+}(\mu)}$. Then update the new witness ω_{i^+} to W, and update the index set $I \cup \{i^+\}$.

Witness deletion algorithm: Input the public key PK, index set I, witness set W and the index to be deleted i^-, if $i^- \in I$ output 1, otherwise output 0. Update the index set $I/\{i^-\}$, delete the corresponding witness ω_{i^-} in the witness set W.

4.4.3. Security proofs

Within this paper, we commence by establishing the correctness of the updatable polynomial commitment. Subsequently, we proceed to demonstrate the binding and hiding properties of the updatable polynomial commitment, substantiating each aspect sequentially.

(1) Correctness

The correctness of the updatable polynomial commitment requires the derivation of whether the following equation (8) holds.

$$e\left(\omega_i, g^{\mu} / g^{i}\right) e(g,g)^{F(i)} = e\left(g^{f_i(\mu)}, g^{(\mu - i)}\right) e(g,g)^{F(i)} = e(g,g)^{f_i(\mu)(u-i)} e(g,g)^{F(i)}$$

$$= e(g,g)^{f_i(\mu)(u-i)+F(i)} = e(g,g)^{F(u)} \tag{8}$$

(2) Binding

The binding of updatable polynomial commitment is divided into polynomial binding and evaluation binding.

Polynomial binding: Suppose there exists an adversary that can break the polynomial binding property of polynomial commitment by two polynomials $F_1(x)$ and $F_2(x)$. This paper constructs an algorithm $\mathscr{E}$ to compute a private key $SK = \mu$ using an adversary, and the updatable polynomial commitment scheme constructed in this paper is homomorphic. For the adversary-generated $F_1(x)$ and $F_2(x)$, the corresponding commitment is computed:

$$\mathscr{C} = g^{F_1(\mu)} = g^{F_2(\mu)} \tag{9}$$

For a polynomial $F_3(\mu) = F_1(\mu) - F_2(\mu) \in Z_P[\mathscr{X}]$, the corresponding commitment is as follow equation (10):

$$\mathscr{C}_{F_3(\mu)} = g^{F_3(\mu)} = \frac{g^{F_1(\mu)}}{g^{F_2(\mu)}} = 1 \tag{10}$$

Therefore $F_3(\mu) = 0$, through the factorization of $F_3(\mu)$, it can be seen that μ is the root of the polynomial $F_3(\mu)$, $\mathscr{E}$ can find $SK = \mu$, solve the n-SDH problem instance.

Evaluation binding: Suppose there is an adversary who can break the committed evaluation binding and compute two witness groups accepted by VerifyPoly $\langle i, F(i), \omega_i \rangle$, $\langle i, F^*(i), \omega_i^* \rangle$. This paper will describe how an adversary construct an algorithm $\mathscr{E}$ that can break the n-SDH assumption. The algorithm $\mathscr{E}$ generates an n-SDH problem instance $\langle \mathscr{G}, g, g^{\mu}, g^{\mu^2}, \ldots, g^{\mu^n} \rangle$ as the public key to the adversary, and the adversary outputs a promise $\mathscr{C}$ and two witness groups $\langle i, F(i), \omega_i \rangle$, $\langle i, F^*(i), \omega_i^* \rangle$, get equation (11):

$$e(\mathscr{C}, g) = e\left(\omega_i, g^{\mu - i}\right) e(g,g)^{F(i)} = e\left(\omega_i^*, g^{\mu - i}\right) e(g,g)^{F(i)} \tag{11}$$

For $f_i = \log_g \omega_i$, $f_i^* = \log_g \omega_i^*$, it can be calculated the followed two equations (12) and (13):

$$f_i(\mu - i) + F(i) = f_i^*(\mu - i) + F^*(i) \tag{12}$$

$$\frac{f_i - f_i^*}{F_i^*(i) - F(i)} = \frac{1}{\mu - i} \tag{13}$$

Then, the algorithm $\mathscr{E}$ calculates the followed equation (14):

$$\left(\frac{\omega_i}{\omega_i^*}\right)^{\frac{1}{F_i^*(i) - F(i)}} = g^{\frac{f_i - f_i^*}{F_i^*(i) - F(i)}} = g^{\frac{1}{\mu - i}} \tag{14}$$

The algorithm $\mathscr{E}$ then returns a $\langle -i, g^{\frac{1}{\mu - i}} \rangle$ as the solution to the problem instance q-SDH. It becomes apparent that the probability of successfully solving the given instance is identical to that of the adversary, while the computational time required is a constant factor greater than the time required by the adversary.

(3) Hiding

Hiding: Assuming that there is an adversary that can break the hiding of the updatable polynomial commitment, given t valid witness tuples $\langle i, F(i), \omega_i \rangle$, the polynomial can correctly calculate $F(x)$. This paper introduces how to use an adversary to generate an

algorithm $\mathscr{E}$ to break the DL assumption. Let $\langle g, g^{\mu}\rangle$ as example of discrete logarithm problem that the algorithm $\mathscr{E}$ needs to solve, the algorithm $\mathscr{E}$ randomly selects a number $\mu \in Z_p^*$ to generate the public key PK to the adversary, $PK = \langle \mathscr{G}, g, g^{\mu}, g^{\mu^2}, \ldots, g^{\mu^n}\rangle$. Algorithm $\mathscr{E}$ sets $\langle \tau, F(\tau)\rangle$ as the polynomial $F(x)$ at the index τ. Then assume $F(0) = \mu$, the answer to the DL instance, then evaluate $g^{F(x)}$ using n+1 exponents: $\langle 0, g^{\mu}\rangle$ and other selected tuples $\langle \tau, g^{F(\tau)}\rangle$. Finally, the algorithm $\mathscr{E}$ calculates the testimony of the evaluation of t choices $\langle \tau, F(\tau)\rangle$ as followed equation (15) shows:

$$\omega_{\tau} = \left(g^{F(\mu)} / g^{F(\tau)}\right)^{\frac{1}{\mu-\tau}} \tag{15}$$

The algorithm $\mathscr{E}$ then the public key PK and t witness tuple $\langle \tau, F(\tau), \omega_{\tau}\rangle$ were send to the adversary. Then $F(x)$ was generated by adversary, the algorithm $\mathscr{E}$ generates a solution of DL instance, constant term $F(0)$. Obviously, the likelihood of successfully solving the discrete logarithm (DL) instance is equivalent to that of the adversary, while the computational time required is only marginally greater than the adversary's.

5. Conclusion

This paper discusses the research on the storage, traceability, and verification of fake news data in the context of unreliable news from unknown sources and crowdsourced journalism data stored in the untrusted cloud server have the risk of be tampering. We proposes a novel mechanism for secure storage of news data using blockchain technology, Built upon blockchain technology, our proposed scheme employs a collaborative storage approach that combines on-chain and off-chain blockchain database modes. This storage mechanism effectively addresses the challenge of limited blockchain storage capacity. Specifically, the abstracts of news data are stored on the blockchain, while the detailed news data are securely stored on an off-chain blockchain database hosted on a cloud server. By adopting this approach, we significantly enhance storage capabilities, ensuring that both abstract and concrete news data can be accommodated without limitations within the blockchain ecosystem. Besides, in order to realize the news data consistency on the chain and off the chain to build a secure authentication protocol, we design an updatable polynomial commitment based on the polynomial commitment, and compare the commitment generated by the news data update with the commitment corresponding to the commitment. The signature is bound, so that the summary data on the chain is consistent with the specific news data off the chain. Furthermore, we conduct a comprehensive analysis of the proposed scheme for updatable polynomial commitment, focusing on its binding, correctness, hiding, and overall security. The results of our analysis affirm that our scheme not only provides a high level of security but also exhibits exceptional efficiency in its operations.

Funding statement

This work was supported by the National Key R&D Program of China (No.2021YFB2700500), the National Natural Science Foundation of China (No. 62072249), the National Key R&D Program of Guangdong Province (No. 2020B0101090002), Shaanxi Province Social Science Foundation Project: Research and Respond Mechanism Design on the Internet Public Opinion of Shaanxi Province University based on the Meta Media Big Data (Project No: 2019N007), and the Natural Science Foundation of Jiangsu Province (No. BK20200418, BE2020106).

Author contribution statement

Xiaowan Wang: Wrote the paper; Analyzed and interpreted the data. Huiyin Xie: Performed the experiments. Shan Ji: Analyzed and interpreted the data. Liang Liu: Conceived and designed the experiments. Ding Huang: Contributed reagents, materials, analysis tools or data.

Data availability statement

The data that has been used is confidential.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[[1] W. Yu, F. Liang, X.F. He, W.G. Hatcher, G. William, et al., A survey on the edge computing for the Internet of Things, IEEE Access 6 (2017) 6900–6919.
[2] C.L. Li, J. Zhang, X.M. Yang, Y.L. Luo, Lightweight blockchain consensus mechanism and storage optimization for resource-constrained IoT devices, Inf. Process. Manag. 58 (4) (2021), 102602.
[3] Z.H. Qian, Y.J. Wang, IoT technology and application, Acta Electonica Sinica 40 (5) (2012) 1023.
[4] Y.J. Ren, D. Huang, W.H. Wang, X.F. Yu, BSMD: a blockchain-based secure storage mechanism for big spatio-temporal data, Future Generat. Comput. Syst. 138 (1) (2023) 328–338.
[5] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, J. Econ. Perspect. 31 (2) (2017) 211–236.
[6] N. Kshetri, J. Voas, The economics of "fake news", IT Professional 19 (6) (2017) 8–12.

[7] A. Kucharski, Post-truth: study epidemiology of fake news, Nature 540 (7634) (2016) 525.
[8] V.L. Rubin, Y.M. Chen, N.J. Conroy, Deception detection for news: three types of fakes, Proc. Assoc. Inform. Sci. Technol. 52 (1) (2015) 1–4.
[9] P. Fraga-Lamas, T.M. Fernández-Caramés, Fake news, disinformation, and deepfakes: leveraging distributed ledger technologies and blockchain to combat digital deception and counterfeit reality, IT Professional 22 (2) (2020) 53–59.
[10] D. Pogue, How to stamp out fake news, Sci. Am. 316 (2) (2017) 24, 24.
[11] R. Zafarani, X. Zhou, K. Shu, H. Liu, Fake news research: theories, detection strategies, and open problems, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 3207–3208.
[12] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (6380) (2018) 1146–1151.
[13] E.C. Tandoc Jr., Z.W. Lim, R. Ling, Defining "fake news" A typology of scholarly definitions, digital journalism 6 (2) (2018) 137–153.
[14] K. Shu, A. Sliva, S.H. Wang, J.L. Tang, H. Liu, Fake news detection on social media: a data mining perspective, ACM SIGKDD Explorations Newsletter 19 (1) (2017) 22–36.
[15] A. Olteanu, C. Castillo, F. Diaz, E. Kiciman, Social data: biases, methodological pitfalls, and ethical boundaries, Frontiers in Big Data 2 (2019) 13.
[16] R. Zafarani, M.A. Abbasi, H. Liu, Social Media Mining: an Introduction, Cambridge University Press, 2014.
[17] K.H. Jamieson, J.N. Cappella, Echo Chamber: Rush Limbaugh and the Conservative Media Establishment, Oxford University Press, 2008.
[18] C.P. Ge, W. Susilo, J. Baek, Z. Liu, J.Y. Xia, et al., Revocable attribute-based encryption with data integrity in clouds, IEEE Trans. Dependable Secure Comput. 21 (3) (2021) 1–12.
[19] J. Allen, B. Howland, M. Mobius, D. Rothschild, D.J. Watts, Evaluating the fake news problem at the scale of the information ecosystem, Sci. Adv. 6 (14) (2020) eaay3539.
[20] Y.J. Ren, F.J. Zhu, S.P. Kumar, T. Wang, J. Wang, et al., Data query mechanism based on hash computing power of blockchain in internet of Things, Sensors 20 (1) (2020) 1–22.
[21] X.C. Zhang, A.A. Ghorbani, An overview of online fake news: characterization, detection, and discussion, Inf. Process. Manag. 57 (2) (2020), 102025.
[22] D.M.J. Lazer, M.A. Baum, Y. Benkler, A.J. Berinsky, K.M. Greenhill, et al., The science of fake news, Science 359 (6380) (2018) 1094–1096.
[23] C.P. Ge, W. Susilo, Z. Liu, J.Y. Xia, L.M. Fang, et al., Secure keyword search and data sharing mechanism for cloud computing, IEEE Trans. Dependable Secure Comput. 6 (18) (2021) 2787–2800.
[24] Y.J. Ren, F. Zhu, J. Wang, P. Sharma, U. Ghosh, Novel vote scheme for decision-making feedback based on blockchain in internet of vehicles, IEEE Trans. Intell. Transport. Syst. 23 (2) (2022) 1639–1648.
[25] W. Shahid, B. Jamshidi, S. Hakak, H. Isah, W.Z. Khan, et al., Detecting and mitigating the dissemination of fake news: challenges and future research opportunities, IEEE Transactions on Computational Social Systems (2022) 1–14.
[26] B. Han, X. Han, H. Zhang, J. Li, X. Cao, Fighting fake news: two stream network for deepfake detection via learnable SRM, IEEE Transactions on Biometrics, Behavior, and Identity Science 3 (3) (2021) 320–331.
[27] C.P. Ge, W. Susilo, J. Baek, Z. Liu, J.Y. Xia, et al., A verifiable and fair attribute-based proxy Re-encryption scheme for data sharing in clouds, IEEE Trans. Dependable Secure Comput. 19 (5) (2022) 2907–2919.
[28] G.Y. Yang, J. Yu, W.T. Shen, Q.Q. Su, Z.J. Fu, et al., Enabling public auditing for shared data in cloud storage supporting identity privacy and traceability, J. Syst. Software 113 (2016) 130–139.
[29] Y. Dawle, M. Naik, S. Vande, N. Zarkar, Database security using intrusion detection system, Int. J. Sci. Eng. Res. 8 (2) (2017) 30–34.
[30] A. Ramachandran, D. Kantarcioglu, Using blockchain and smart contracts for secure data provenance management, CoRR 1709 (2017), 10000.
[31] A. Banerjee, Blockchain technology: supply chain insights from ERP, Adv. Comput. 111 (2018) 69–98.
[32] H. Huang, X. Chen, J. Wang, Blockchain-based multiple groups data sharing with anonymity and traceability, Sci. China Inf. Sci. 63 (3) (2020) 1–13.
[33] M. Chase, A. Healy, A. Lysyanskaya, T. Malkin, L. Reyzin, Mercurial Commitments with Applications to Zero-Knowledge Sets, in: Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2005, pp. 422–439.
[34] B. Bünz, J. Bootle, D. Boneh, A. Poelstra, P. Wuille, et al., Bulletproofs: short proofs for confidential transactions and more, SP, IEEE Symposium on Security and Privacy (2018) 315–334.
[35] A. Vlasov, K. Panarin, Transparent Polynomial Commitment Scheme with Polylogarithmic Communication Complexity, Cryptology ePrint Archive, 2019.
[36] D. Boneh, J. Drake, B. Fisch, A. Gabizon, Efficient Polynomial Commitment Schemes for Multiple Points and Polynomials, Cryptology ePrint Archive, 2020.
[37] D. Boneh, J. Drake, B. Fisch, A. Gabizon, Halo infinite: proof-carrying data from additive polynomial commitments, Annual International Cryptology Conference 12825 (2021) 649–680.
[38] H. Jeff, The Rise of Crowdsourcing, Wired, 2006.
[39] A. Tanja, Motivation factors in crowdsourced journalism: social impact, social change, and peer learning, Int. J. Commun. (2015) 3523–3543.
[40] X. Dong, U. Victor, L. Qian, Two-path deep semisupervised learning for timely fake news detection, IEEE Trans. Comput. Soc. Syst. 7 (6) (2020) 1386–1398.
[41] Y.J. Ren, Y. Leng, J. Qi, P.K. Sharma, J. Wang, et al., Multiple cloud storage mechanism based on blockchain in smart homes, Future Generat. Comput. Syst. 115 (2021) 304–313.
[42] C.P. Ge, Z. Liu, J.Y. Xia, L.M. Fang, Revocable identity-based broadcast proxy Re-encryption for data sharing in clouds, IEEE Trans. Dependable Secure Comput. 3 (18) (2021) 1214–1226.
[43] M. Saad, A. Ahmad, A. Mohaisen, Fighting fake news propagation with blockchains, Conf. Commun. Netw. Security (2019) 1–4.
[44] A. Balouchestani, M. Mahdavi, Y. Hallaj, D. Javdani, SANUB: a new method for sharing and analyzing news using blockchain, in: 16th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology, 2019, pp. 139–143.
[45] G. Song, S. Kim, H. Hwang, K. Lee, Blockchain-based notarization for social media, IEEE Int. Conf. Consum. Electron. (2019) 1–2.
[46] R.F. Sari, A. Ilmananda, D. Romano, Social trust-based blockchain-enabled social media news verification system, J. Univers. Comput. Sci. 27 (9) (2021) 979–998.
[47] Y.J. Ren, Y. Leng, Y.P. Cheng, J. Wang, Secure data storage based on blockchain and coding in edge computing, Math. Biosci. Eng. 16 (4) (2019) 1874–1892.