RESEARCH ARTICLE

A novel privacy-preserving biometric authentication scheme

Xuechun Mao[1], **Ying Chen**[1]*, **Cong Deng**[2], **Xiaqing Zhou**[3]

1 Department of Computer Science, Taizhou University, Taizhou, Zhejiang, China, **2** Hangzhou Jinritoutiao Technology Co., Ltd, Hangzhou, Zhejiang, China, **3** Medical Records Division, Taizhou Hospital, Taizhou, Zhejiang, China

* ychen222@foxmail.com

Citation: Mao X, Chen Y, Deng C, Zhou X (2023) A novel privacy-preserving biometric authentication scheme. PLoS ONE 18(5): e0286215. https://doi.org/10.1371/journal.pone.0286215

Editor: Pandi Vijayakumar, University College of Engineering Tindivanam, INDIA

Received: February 8, 2023

Accepted: May 10, 2023

Published: May 25, 2023

Data Availability Statement: All relevant data are within the paper and its Supporting information files.

Funding: This research is partially supported by the Natural Science Foundation of China (No. 61976149) and the Key Program of the Natural Science Foundation of Zhejiang province of China (No. LZ20F020002). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Most existing secure biometric authentication schemes are server-centric, and users must fully trust the server to store, process, and manage their biometric data. As a result, users' biometric data could be leaked by outside attackers or the service provider itself. This paper first constructs the EDZKP protocol based on the inner product, which proves whether the secret value is the Euclidean distance of the secret vectors. Then, combined with the Cuproof protocol, we propose a novel user-centric biometric authentication scheme called BAZKP. In this scheme, all the biometric data remain encrypted during authentication phase, so the server will never see them directly. Meanwhile, the server can determine whether the Euclidean distance of two secret vectors is within a pre-defined threshold by calculation. Security analysis shows BAZKP satisfies completeness, soundness, and zero-knowledge. Based on BAZKP, we propose a privacy-preserving biometric authentication system, and its evaluation demonstrates that it provides reliable and secure authentication.

Introduction

Biometric authentication has been popular in services such as access control to authenticate individuals based on their biometrics. In contrast to the passwords or tokens used in conventional authentication systems, biometric data, such as fingerprint, face, or gait, can physically authenticate an individual with high assurance. Biometric data is unique and safe for every individual. With the help of powerful computer and network technology, biometrics offer the advantages of uniqueness, permanence, and reliability.

However, biometric data is prone to diverse variations and distortions due to inherent and environmental noise. Thus, biometric authentication generally employs an error tolerance mechanism. Namely, the service provider uses a biometric vector w as a template stored in a central database, and then the user submits a new vector w' to the service provider for authentication. The user's authentication will succeed if w and w' are close enough under a certain distance metric.

This biometric authentication model above is server-centric, as the service provider is responsible for securing the templates and receiving the user's biometric information in plain text. This model has several unavoidable drawbacks. First, users must completely trust the service provider to protect their templates. However, malicious service providers may misuse the templates for profit, and security attacks may result in template leakage and violation of user privacy. Second, the templates must be recovered in plaintext for distance computation and comparison, which makes it possible for adversaries to spy on the registered or freshly submitted templates. To meet these challenges, it is necessary to propose a secure, privacy-preserving, and user-centered scheme [1, 2].

To address these issues, we propose a biometric authentication zero-knowledge proof (BAZKP) scheme that preserves the benefits of biometric authentication while enhancing its security. The BAZKP is a security, privacy-protection, and user-centered authentication scheme that combines biometric technology with a zero-knowledge proof (ZKP) protocol to authenticate users while maintaining anonymity. We also provide security proofs to show that our scheme guarantees privacy for users. Moreover, we have developed a privacy-preserving biometric authentication system to accomplish a complete life cycle of BAZKP.

Related work

Zero-Knowledge Proof (ZKP). Goldwasser et al. [3] first mentioned the ZKP in 1985, which enables a prover to convince a verifier that some statement is true without revealing anything more than the truth of the statement. Then, lots of research on ZKP protocols and many kinds of ZKP protocols were proposed, such as zk-SNARK and range proof. Maller et al. [4] proposed the first zk-SNARK with entirely succulent verification for general arithmetic circuits with SRS, known as Sonic. After that, Gabizon et al. proposed a more efficient zk-SNARK which is called PLONK [5]. To make improvements on the efficiency of zk-SNARK, Lately Bünz et al. proposed Super Sonic [6].

On the other hand, Brickell et al. [7] stated the first correlative algorithm of range proof. After that, in 1998, Chan et al. [8] proposed a kind of range proof whose security depends on modulus. It is called CTF. Using Lagrange's four-square theorem [9], Lipmaa [10] published a proof of any range for the first time. In 2005, based on Lagrange's four-square theorem, Groth [11] demonstrated that $4y+ 1$ could be represented as the sum of the squares of some three integers if y was a non-negative integer. Bootle et al. [12] improved the efficiency of ZKP by using the inner product method and recursion. In 2018, Bünz et al. proposed a range proof scheme which is called Bulletproofs [13]. Based on these works, Deng et al. [14] presented the Cuproof, a novel range proof scheme for any range. The time, communication costs and proof size of that scheme are maintained within a constant range regardless of how large the proof interval is.

Biometric authentication. The two primary types of biometrics applications [15] are biometric authentication and biometric identification. The existing research mainly focused on privacy-preserving biometric identification [16, 17], in which a server database is assumed to contain the templates of the registered users and their corresponding identities. Barni et al. [16] presented a privacy preserving protocol for fingerprint-based authentication. In 2018, Zhou et al. [18] proposed a user-centric biometric authentication system that allows users to use the proposed lightweight encryption approach to encrypt their templates. To improve the performance of each biometric system, Hammad et al. [19] proposed a secure multimodal biometric system using convolution neural network (CNN) and Q-Gaussian multi-support vector machine (QG-MSVM) based on different level fusion.

Privacy-preserving scheme. In recent years, privacy protection has become a crucial area of research and many researchers conducting extensive studies in this field. Zhang et al. [20] proposed a new privacy-preserving biometric identification scheme, in which they introduced perturb terms in each biometric data. Zhou et al. [18] proposed a user-centric biometric authentication scheme that enabled end-users to encrypt their own templates with the proposed light-weighted encryption scheme. Lee et al. [21] presented a new biometric authentication system based on blockchain which provided a decentralized and distributed mechanism for processing biometric authentication.

In addition, Azees et al. [22] proposed an efficient anonymous authentication scheme to avoid malicious vehicles entering into the VANET. After that, in 2022, Zhou et al. [23] introduced a security-enhanced solution for VANETs, whichj can resist a signature forgery attack. Liu et al. [24] proposed a novel privacy-preserving DSSE scheme for IIoTH system, which was the first DSSE scheme designed for personal health record (PHR) files database with forward security. Yang et al. [25] proposed a novel oblivious data sharing scheme employing the designed 1-out-of-n oblivious transfer protocol to achieve an efficient location-based service for users while effectively hiding location coordinates and protecting the privacy of users and servers. Wei et al. [26] presented a privacy-preserving implicit authentication framework using users' behavior features sensed by the mobile intelligent terminal based on the artificial intelligence methodology. An efficient affine cipher-based encryption technique is proposed by Azees et al. [27] to offer a high level of confidentiality with smaller key size compared to existing encryption techniques.

To address these security flaws, Subramani et al. [28] proposed a computationally efficient privacy-preserving anonymous authentication scheme for resource-limited WBAN. In 2022, Rajasekaran et al. [29] preserved the privacy and the anonymity of the end-users (patient/doctor) using an anonymous blockchain-based authentication scheme. To overcome IoHT imposes security challenges in maintaining patient data confidentiality and privacy, a novel blockchain-based privacy-preserving authentication scheme is proposed by Rajasekaran et al. [30] as an approach for achieving efficient authentication of the patient without the involvement of a trusted entity. Jegadeesan et al. [31] proposed a public key encryption based computationally efficient mutual authentication protocol for secure data transmission between incubator monitoring systems and doctors or administrators.

Contributions

The main contributions of this paper are summarized as follows:

- We employ the inner product to construct the Euclidean Distance Zero-knowledge Proof (EDZKP) protocol and prove that the protocol satisfies perfect completeness, perfect special honest verifier zero-knowledge, and computational witness extended emulation.
- By combining ZKP with biometric authentication, we construct a novel biometric authentication scheme called BAZKP. In this scheme, the EDZKP protocol guarantees that the secret value v is the Euclidean Distance between the secret vectors $\mathbf{w}$ and $\mathbf{w}'$, and the Cuproof protocol proves whether the secret value v is within the range $[0, e]$. We also provide security proof to show that our scheme guarantees privacy for users.
- Based on the proposed scheme, we present a privacy-preserving biometric authentication system enabling users to utilize their biometric data for authentication while preserving users privacy. We show its viability and operational efficacy by implementing the proposed system. The experimental data obtained from our tests show the system's potential for real-world deployment.

Preliminaries

Before we state our scheme, we first note some underlying tools. In this paper, $\mathcal{A}$ is a PPT adversary, which is a probabilistic interactive Turing machine that runs in polynomial time in the security parameter λ.

Assumptions

Groups of unknown order. In order to keep the soundness of our scheme, we use the RSA group $\mathbb{G}$ where the order of the group is unknown. The RSA group is generated by a trusted setup.

RSA group. In the multiplicative group $\mathbb{G}$ of integers modulo n where n is the product of the large primes p and q. The hardness of computing the order of the group $\mathbb{G}$ is as the same as the hardness of factoring n.

Assumption 1 (Discrete Log Relation Assumption) *For all PPT adversaries $\mathcal{A}$ and for all $j \geq 2$, there exists a negligible function $\mu(\lambda)$ such that*

$$\mathrm{P}\left[\begin{array}{lc} \mathbb{G} = \mathrm{Setup}(1^{\lambda}), & \\ g_1, \ldots, g_j \overset{\$}{\leftarrow} \mathbb{G}; & : \begin{array}{c} \exists a_i \neq 0, \\ \prod_{i=1}^{j} g_i^{a_i} = 1 \end{array} \\ a_1, \ldots, a_j \in \mathbb{Z}_n \leftarrow \mathcal{A}(g_1, \ldots, g_j) & \end{array}\right] \leqslant \mu(\lambda). \tag{1}$$

As Bünz et al. [13] stated, $\prod_{i=1}^{j} g_i^{a_i} = 1$ is a non-trivial discrete log relation between $g_1, \ldots, g_j$. The discrete log relation assumption ensures that an adversary can't find a non-trivial relation between randomly selected group elements. This assumption is equivalent to the discrete-log assumption when $j \geq 1$.

Assumption 2 (Order Assumption) *The Order Assumption holds for* Setup *if for any efficient adversary $\mathcal{A}$ there exists a negligible function $\mu(\lambda)$ such that*

$$\mathrm{P}\left[g_1 \neq 1 \wedge g_1^{a_1} = 1 : \begin{array}{l} \mathbb{G} \overset{\$}{\leftarrow} \mathrm{Setup}(\lambda), \\ \\ (g_1, a_1) \overset{\$}{\leftarrow} \mathcal{A}(\mathbb{G}), \\ \\ \text{where } a_1 \neq 0 \in \mathbb{Z}_n, \\ \\ \text{and } g_1 \in \mathbb{G} \end{array} \right] \leq \mu(\lambda). \tag{2}$$

Lemma 1 *The Order Assumption implies the Discrete Log Relation Assumption.*

Proof. We show that if adversary $\mathcal{A}_{Ord}$ breaks the Order Assumption, then we can construct $\mathcal{A}_{DL}$ which breaks the Discrete Log Relation Assumption with overwhelming probability. To get a vector $(g_1, g_2, \ldots, g_j) \in \mathbb{G}^j$ and a vector $(a_1, a_2, \ldots, a_j) \in \mathbb{Z}_n^j$ such that $g_1^{a_1} \cdot g_2^{a_2} \ldots g_n^{a_n} = 1$ where $g_i \neq 1$, $a_i \neq 0$ and $i \in \{1, 2, \ldots, j\}$, we run $\mathcal{A}_{Ord}$ for n times and it will outputs $g_j \in \mathbb{G}$ and $a_j \in \mathbb{Z}$ such that $g_j^{a_j} = 1$ for $j = 1, \ldots, n$. And it follows $\prod_{j=1}^{n} g_j^{a_j} = 1$.

Commitments

We adopt the following definitions from [14] for our notation.

Definition 1 (Commitments) *A non-interactive commitment scheme consists of a pair of probabilistic polynomial time algorithms* (Setup, Com). *The setup algorithm* pp ⟵ Setup(1^{λ}) *generates the public parameters pp with the security parameter λ. The commitment algorithm* Com_{pp} *defines a function* $\mathrm{M}_{pp} \times \mathrm{R}_{pp} \rightarrow \mathrm{C}_{pp}$ *for a message space* M_{pp}, *a randomness space* R_{pp} *and a commitment space* C_{pp} *determined by pp. For a message* $\mathrm{x} \in \mathrm{M}_{pp}$, *the algorithm draws* $r \overset{\$}{\leftarrow} \mathrm{R}_{pp}$ *uniformly at random, and computes the commitment* **com** = $\mathrm{Com}_{pp}(x, r)$.

Definition 2 (Homomorphic Commitments) *A homomorphic commitment scheme is a non-interactive commitment scheme such that* (M_{pp}, *), (R_{pp}, +), *and* (C_{pp}, +) *are all abelian* *groups, and for all* $\mathrm{x}_1, \mathrm{x}_2 \in \mathrm{M}_{pp}$, $\mathrm{r}_1, \mathrm{r}_2 \in \mathrm{R}_{pp}$, *we have*

$$\mathrm{Com}(x_1; r_1) * \mathrm{Com}(x_2; r_2) = \mathrm{Com}(x_1 + x_2; r_1 + r_2). \tag{3}$$

Here (M_{pp}, *) can be additive or multiplicative. For ease of notation, we drop *pp* in the subindex.

Definition 3 (Hiding Commitment) *A commitment scheme is said to be hiding if for every PPT adversary* $\mathcal{A}$ *there exists a negligible function* $\mu(\lambda)$ *such that*

$$\left| \mathrm{P}\left[b = b' \,\middle|\, \begin{array}{l} pp \leftarrow \mathrm{Setup}(1^{\lambda}); \\ (x_0, x_1) \in \mathrm{M}_{pp}^2 \leftarrow \mathcal{A}(pp), \\ b \xleftarrow{\$} (0,1), r \xleftarrow{\$} \mathrm{R}_{pp}, \\ \mathbf{com} = \mathrm{Com}(x_b; r), \\ b' \leftarrow \mathcal{A}(pp, \mathbf{com}) \end{array} \right] - \frac{1}{2} \right| \le \mu(\lambda), \tag{4}$$

where the probability is over b, r, Setup *and* $\mathcal{A}$. *If* $\mu(\lambda) = 0$ *then we say that the scheme is* perfectly hiding.

Definition 4 (Binding Commitment) *A commitment scheme is said to be binding if for every PPT adversary* $\mathcal{A}$ *there exists a negligible function* μ *such that*

$$\mathrm{P}\left[\begin{array}{l} \mathrm{Com}(x_0; r_0) = \mathrm{Com}(x_1; r_1), \\ x_0 \neq x_1 \end{array} \,\middle|\, \begin{array}{l} pp \leftarrow \mathrm{Setup}(1^{\lambda}), \\ x_0, x_1, r_0, r_1 \leftarrow \mathcal{A}(pp) \end{array} \right] \le \mu(\lambda), \tag{5}$$

where the probability is over Setup *and* $\mathcal{A}$. *If* $\mu(\lambda) = 0$, *then we say that the scheme is* perfectly binding.

In the following content, to ensure that the discrete log in the groups we used is intractable for PPT adversaries, the order of these groups is implicitly dependent on the security parameter.

Definition 5 (Pedersen Commitment) $\mathrm{M}_{pp}, \mathrm{R}_{pp} = \mathbb{Z}_n$ and $\mathrm{C}_{pp} = (\mathbb{G}, *)$ *being a multiplicative group.*

$$\mathrm{Setup} : g, h \xleftarrow{\$} \mathbb{G}, \tag{6}$$

$$\mathrm{Com}(x; r) = (g^x h^r). \tag{7}$$

Definition 6 (Pedersen Vector Commitment) $\mathrm{M}_{pp} = \mathbb{Z}_n^j, \mathrm{R}_{pp} = \mathbb{Z}_n$ and $\mathrm{C}_{pp} = (\mathbb{G}, *)$ *being a multiplicative group.*

$$\mathrm{Setup} : \mathbf{g} = (g_1, \ldots, g_j), h \xleftarrow{\$} \mathbb{G}, \tag{8}$$

$$\mathrm{Com}(\mathbf{x} = (x_1, \ldots, x_j); r) = h^r \mathbf{g}^{\mathbf{x}} = h^r \prod_i g_i^{x_i} \in \mathbb{G}. \tag{9}$$

The Pedersen vector commitment for the group $\mathbb{G}$ is perfectly hiding and computationally binding under the discrete logarithm assumption. In the definition, *r* is chosen at random.

Zero-knowledge arguments of knowledge

In our scheme, we construct the zero-knowledge arguments of knowledge. A zero-knowledge proof of knowledge means a prover can convince a verifier that some statements hold without revealing any information of the knowledge. An argument is a proof that holds when the prover is computationally bounded and certain computational hardness assumptions hold. The formal definitions are as follows.

Zero-knowledge arguments consist of three interactive algorithms (Setup, $\mathcal{P}$, $\mathcal{V}$), which run in probabilistic polynomial time. Setup is the common reference string generator, $\mathcal{P}$ is the prover and $\mathcal{V}$ is the verifier. The algorithm Setup produces a common reference string σ on inputting 1^{λ}. The transcript produced by $\mathcal{P}$ and $\mathcal{V}$ is denoted by $tr \leftarrow < \mathcal{P}(s), \mathcal{V}(t) >$, when they interact on the inputs s and t. We write $[\mathcal{P}(s), \mathcal{V}(t)] = b$ where $b = 0$ if verifier rejects, $b = 1$ if verifier accepts.

We let $\mathcal{R} \subset \{0,1\}^* \times \{0,1\}^* \times \{0,1\}^*$ be a polynomial-time-decidable ternary relation. Given a parameter σ, the w is a witness for a statement u only if $(\sigma, u, w) \in \mathcal{R}$. We define the CRS-dependent language

$$\mathcal{L}_{\sigma} = \{u | \exists w : (\sigma, u, w) \in \mathcal{R}\} \tag{10}$$

as the set of all the statements which have a witness w in the relation $\mathcal{R}$.

Definition 7 (Argument of Knowledge) *The triple* (Setup, $\mathcal{P}$, $\mathcal{V}$) *is called an argument of knowledge for relation R if it satisfies both the Perfect Completeness and Computational Witness-Extended Emulation as defined in* [13], *respectively.*

Definition 8 (Perfect Completeness) (Setup, $\mathcal{P}$, $\mathcal{V}$) *has perfect completeness if for all non-uniform polynomial time adversaries* $\mathcal{A}$

$$\mathrm{P}\left[\begin{array}{l|l} (\sigma, u, w) \notin \mathcal{R} & \sigma \leftarrow \mathrm{Setup}(1^{\lambda}) \\ \\ or\ [\mathcal{P}(\sigma, u, w), \mathcal{V}(\sigma, u)] = 1 & (u, w) \leftarrow \mathcal{A}(\sigma) \end{array}\right] = 1. \tag{11}$$

Definition 9 (Computational Witness-Extended Emulation) *For every deterministic polynomial time* $\mathcal{P}^*$ *there exists an expected polynomial time emulator* ε *such that for every pair of interactive adversaries* $\mathcal{A}_1$ and $\mathcal{A}_2$, *there exists a negligible function* $\mu(\lambda)$

$$\left| \begin{array}{l} \mathrm{P}\left[\begin{array}{l|l} & \sigma \leftarrow \mathrm{Setup}(1^{\lambda}), \\ & (u, s) \leftarrow \mathcal{A}_2(\sigma), \\ \mathcal{A}_1(tr = 1) & tr \leftarrow \langle \mathcal{P}^*(\sigma, u, s), \\ & \mathcal{V}(\sigma, u)\rangle \end{array}\right] - \\ \\ \mathrm{P}\left[\begin{array}{l|l} \mathcal{A}_1(tr) = 1 & \sigma \leftarrow \mathrm{Setup}(1^{\lambda}), \\ \wedge(tr \text{ is accepting} & (u, s) \leftarrow \mathcal{A}_2(\sigma), \\ \Rightarrow (\sigma, u, w) \in \mathcal{R}) & (tr, w) \leftarrow \varepsilon^{\mathcal{O}}(\sigma, u) \end{array}\right] \end{array} \right| \leq \mu(\lambda), \tag{12}$$

where the oracle is given by $\mathcal{O} = \langle \mathcal{P}^*(\sigma, u, s), \mathcal{V}(\sigma, u)\rangle$, *and permits rewinding at each round* *after the prover commits and resuming with fresh randomness for the verifier from this point onwards, then we say* (Setup, $\mathcal{P}$, $\mathcal{V}$) *has witness-extended emulation.*

Definition 10 (Public Coin) *An argument of knowledge* (Setup, $\mathcal{P}$, $\mathcal{V}$) *is called public coin if all messages sent from the verifier to the prover are chosen uniformly at random and independent of the prover's messages, i.e., the challenges correspond to the randomness* ρ.

Definition 11 (Perfect Special Honest-Verifier Zero-Knowledge) *A public coin argument of knowledge* (Setup, $\mathcal{P}$, $\mathcal{V}$) *is a perfect special honest verifier zero knowledge (SHVZK) argument of knowledge for* $\mathcal{R}$ *if there exists a probabilistic polynomial time simulator* $\mathcal{S}$ *such that for every pair of interactive adversaries* $\mathcal{A}_1$ and $\mathcal{A}_2$:

$$\mathrm{P}\left[(\sigma,u,w)\in\mathcal{R}\text{ and }\mathcal{A}_1(tr)=1\;\middle|\;\begin{array}{l}\sigma\leftarrow\mathrm{Setup}(1^\lambda)\\ (u,w,\rho)\leftarrow\mathcal{A}_2(\sigma),\\ tr\leftarrow\langle\mathcal{P}(\sigma,u,w),\mathcal{V}(\sigma,u;\rho)\rangle\end{array}\right] \tag{13}$$

$$=\mathrm{P}\left[(\sigma,u,w)\in\mathcal{R}\text{ and }\mathcal{A}_1(tr)=1\;\middle|\;\begin{array}{l}\sigma\leftarrow\mathrm{Setup}(1^\lambda),\\ (u,w,\rho)\leftarrow\mathcal{A}_2(\sigma),\\ tr\leftarrow\mathcal{S}(u,\rho)\end{array}\right]$$

where ρ *is the public coin randomness used by the verifier. The "transcript" can be simulated by S without knowing w.*

In this definition, the adversary chooses a distribution over statements and witnesses. However, the adversary still cannot distinguish between the simulated and the honestly generated transcripts for valid statements and witnesses.

Now we define range proofs. Range proofs are the proofs that the prover knows an opening to a commitment in which the committed value is in a certain range. Range proofs can be used to state that an integer commitment is for a positive number or when two homomorphic commitments are added together, it will not overflow when they are taken modulo the given prime, and these two homomorphic commitments are the commitments to the elements in a prime field.

Definition 12 (Zero-Knowledge Range Proof) *Given a commitment scheme* (Setup, Com) *over a message space* M_{pp} *which is a set with a total ordering, a zero-knowledge range proof is a SHVZK argument of knowledge for the relation*:

$$\mathcal{R}_{\text{Range}}:(pp,(\mathbf{com},l,r),(x,\rho))\in\mathcal{R}_{\text{Range}}\leftrightarrow\mathbf{com}=\mathrm{Com}(x;\rho)\;\wedge\;(l\le x<r). \tag{14}$$

Theorem 1 (Lagrange's three-square theorem) *If x is a positive integer, then* $4x + 1$ *can be written as the sum of three squares.*

The proof for Theorem 1 is given in [9, 11] offered an efficient and simple algorithm for finding three such squares. Theorem 1 also means writing $4x + 1$ as the sum of three squares implies that x is non-negative.

Notation

Let $[N]$ denote the set $\{1, \ldots, N-1\}$. Let p and q denote two prime numbers. Let $\mathbb{G}$ denote the multiplicative group of integers modulo n, where n is the product of p and q, i.e. $\mathbb{G}$ is a RSA group. Let $\mathbb{Z}_n$ denote the ring of integers modulo n. Let $\mathbb{Z}$ denote the set of all integers. Let $\mathbb{G}^j$ and $\mathbb{Z}_n^j$ be vector spaces of dimension j over $\mathbb{G}$ and $\mathbb{Z}_n$, respectively. Let $\mathbb{Z}_n^*$ denote $\mathbb{Z}_n \setminus \{0\}$. Group elements which represent commitments are capitalized. For example, $C = g^a h^\alpha$ is a Pedersen commitment to a for $g, h \in \mathbb{G}$. $x \xleftarrow{} {}^{\mathbb{Z}_n^*}$ means the uniform sampling of an element from $\mathbb{Z}_n^*$. In this paper, $\mathbf{a} \in \mathbb{F}^j$ is a vector with elements $a_1, \ldots, a_j \in \mathbb{F}$. For an element $c \in \mathbb{Z}_n$ and a vector $\mathbf{a} \in \mathbb{Z}_n^j$, we denote by $\mathbf{b} = c \cdot \mathbf{a} \in \mathbb{Z}_n^j$ the vector with $b_i = c \cdot a_i$. For the two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{F}^j$, let $\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^{j} a_i \cdot b_i$ denote the inner product and $\mathbf{a} \circ \mathbf{b} = (a_1 \cdot b_1, \ldots, a_j \cdot b_j) \in \mathbb{F}^j$ the Hadamard product, respectively. We define vector polynomials $P(x) = \sum_{i=0}^{d} \mathbf{p_i} \cdot x^i \in \mathbb{Z}^j[x]$ where each coefficient $\mathbf{p_i}$ is a vector in $\mathbb{Z}^j$. The inner product between two vector polynomials $l(x)$ and $r(x)$ is defined as

$$\langle l(x), r(x) \rangle = \sum_{i=0}^{d} \sum_{j=0}^{i} \langle \mathbf{l_i}, \mathbf{r_j} \rangle \cdot x^{i+j} \in \mathbb{Z}[x] \tag{15}$$

Let $\mathbf{a} \| \mathbf{b}$ denote the concatenation of two vectors: if $\mathbf{a} \in \mathbb{Z}_n^j$ and $\mathbf{b} \in \mathbb{Z}_n^m$ then $\mathbf{a} \| \mathbf{b} \in \mathbb{Z}_n^{j+m}$. For $0 \leq \ell \leq s$, we use Python notation to denote slices of vectors:

$$\mathbf{a}_{[:\ell]} = \mathbf{a}_{[0:\ell]} = (a_1, \ldots, a_\ell) \in \mathbb{F}^\ell, \tag{16}$$

$$\mathbf{a}_{[\ell:]} = \mathbf{a}_{[\ell:s]} = (a_{\ell+1}, \ldots, a_s) \in \mathbb{F}^{s-\ell}. \tag{17}$$

Let $t(x) = \langle \mathbf{l}(x), \mathbf{r}(x) \rangle$, then the inner product is defined such that $t(x) = \langle l(x), r(x) \rangle$ holds for all $x \in \mathbb{Z}_n$. For vectors $\mathbf{g} = (g_1, \ldots, g_j) \in \mathbb{G}^j$ and $\mathbf{a} \in \mathbb{Z}_n^j$ we write $C = \mathbf{g}^{\mathbf{a}} = \prod_{i=1}^{j} g_i^{a_i} \in \mathbb{G}$. For $1 \leq u$ we set $\vec{\mathbf{u}} = (1, 2, 3, \ldots, u) \in \mathbb{Z}^u$.

Biometric Authentication Zero-Knowledge Proof (BAZKP) scheme

In this section, we introduce our BAZKP scheme. First, we state the EDZKP protocol and the Cuproof protocol, which are essential components of the BAZKP scheme.

EDZKP protocol

In the EDZKP protocol, The vector $\mathbf{w,w',d}$ is a secret vector encrypted by the prover based on its biometric data, where $\mathbf{w}, \mathbf{w}' \in \mathbb{Z}^m$ and $\mathbf{d} = \mathbf{w}' - \mathbf{w}$. The secret value v is the Euclidean Distance between $\mathbf{w}$ and $\mathbf{w}'$, i.e. $v = \langle \mathbf{d}, \mathbf{d} \rangle$. We will use the EDZKP protocol to prove the following relationship:

$$\{(g, h, V \in \mathbb{G};\ \ v, r \in \mathbb{Z}_n;\ \ \mathbf{w}, \mathbf{w}', \mathbf{d} \in \mathbb{Z}^m):\ \ V = h^r g^v,\ \ \langle \mathbf{d}, \mathbf{d} \rangle = v,\ \mathbf{d} = \mathbf{w} - \mathbf{w}'\}. \tag{18}$$

Theorem 2 (EDZKP protocol) *The EDZKP protocol presented in this section satisfies perfect completeness, perfect special honest verifier zero-knowledge, and computational witness extended emulation.*

Proof. The proof for Theorem 2 is given in S2 Appendix.

EDZKP Protocol ($V \in \mathbb{G};\ \ v \in \mathbb{Z}_n;\ \ \mathbf{w}, \mathbf{w}', \mathbf{d} \in \mathbb{Z}^m$)

1. $\mathcal{P}$ gets $\mathbf{d}, \mathbf{w}, \mathbf{w}'$, where $\langle \mathbf{d}, \mathbf{d} \rangle = v, \mathbf{d} = \mathbf{w} - \mathbf{w}'$;
2. $\mathcal{P}$ selects $\alpha, \beta, \theta, \phi \overset{\$}{\leftarrow} \mathbb{Z}_n$;
3. $\mathcal{P}$ computes $D = h^{\alpha}\mathbf{g}^{\mathbf{d}}\mathbf{h}^{\mathbf{d}} \in \mathbb{G}, W = h^{\beta}\mathbf{g}^{\mathbf{w}}\mathbf{h}^{\mathbf{w}} \in \mathbb{G}, W' = h^{\theta}\mathbf{g}^{\mathbf{w}'}\mathbf{h}^{\mathbf{w}'} \in \mathbb{G}, K = h^{\phi} \in \mathbb{G}$;
4. $\mathcal{P}$ selects $\mathbf{s}_L, \mathbf{s}_R \overset{\$}{\leftarrow} \mathbb{Z}_n^m, \rho \overset{\$}{\leftarrow} \mathbb{Z}_n$;
5. $\mathcal{P}$ computes $S = h^{\rho}\mathbf{g}^{\mathbf{s}_L}\mathbf{h}^{\mathbf{s}_R} \in \mathbb{G}$;
6. $\mathcal{P} \rightarrow \mathcal{V} : D, W, W', S, K$;
7. $\mathcal{V}$ selects $y, z, c \overset{\$}{\leftarrow} \mathbb{Z}_n^*$
8. $\mathcal{V} \rightarrow \mathcal{P} : y, z, c$;
9. $\mathcal{P}$ selects $\tau_1, \tau_2 \overset{\$}{\leftarrow} \mathbb{Z}_n$;
10. $\mathcal{P}$ computes $T_i = g^{t_i}h^{\tau_i} \in \mathbb{G}$, where t_i can be computed without knowing $\mathbf{l}$ and $\mathbf{r}$, i.e. t_i is the coefficient of $\langle l(x), r(x) \rangle$,
respectively. $i = \{1, 2\},\ \ e = c(\theta + \alpha - \beta) + \phi \in \mathbb{Z}$;
11. $\mathcal{P} \rightarrow \mathcal{V} : T_1,\ \ T_2,\ \ e$;
12. $\mathcal{V}$ selects $x \overset{\$}{\leftarrow} \mathbb{Z}_n^*$;
13. $\mathcal{V} \rightarrow \mathcal{P} :\ \ x$;
14. $\mathcal{P}$ computes $\mathbf{l} = l(x) = \mathbf{d}z - \mathbf{y} + \mathbf{s}_L x\ \ \in \mathbb{Z}^m,\ \ \mathbf{r} = r(x) = \mathbf{d}z + \mathbf{y} + \mathbf{s}_R x\ \ \in \mathbb{Z}^m,\ \ \hat{t} = \langle \mathbf{l}, \mathbf{r} \rangle = t_0 + t_1 x + t_2 x^2 \in \mathbb{Z}$;
15. $\mathcal{P}$ computes $\tau_x = \tau_2 \cdot x^2 + \tau_1 \cdot x + z^2 r\ \ \in \mathbb{Z},\ \ \mu = \alpha z + \rho x\ \ \in \mathbb{Z}$;
16. $\mathcal{P} \rightarrow \mathcal{V} : \tau_x,\ \ \mu,\ \ \hat{t},\ \ \mathbf{l},\ \ \mathbf{r}$;
17. $\mathcal{V}$ computes $P = D^z \cdot S^x \cdot \mathbf{g}^{-\mathbf{y}} \cdot \mathbf{h}^{\mathbf{y}}\ \ \in \mathbb{G}$;
18. $\mathcal{V}$ checks these equations : $P \overset{?}{=} h^{\mu} \cdot \mathbf{g}^{\mathbf{l}} \cdot \mathbf{h}^{\mathbf{r}} \in \mathbb{G},\ \ g\hat{t}h^{\tau_x} \overset{?}{=} V^{z^2}g^{-\delta(y)} \cdot T_1^x \cdot T_2^{x^2} \in \mathbb{G},\ \ \hat{t} \overset{?}{=} \langle \mathbf{l}, \mathbf{r} \rangle \in \mathbb{Z},\ \ W^c h^e \overset{?}{=} W'^c \cdot D^c \cdot K$.

Cuproof protocol

One of the effective range proof protocols is the Cuproof [14] protocol. This protocol can maintain time, communication overhead, and proof size within a fixed range regardless of how large the proof interval is. We use the Cuproof protocol in BAZKP scheme to convince the verifier that the secret number v is in $[a, b]$. Based on Lagrange three-square theorem, we can find $a, b \in \mathbb{Z}_n$ and $\mathbf{u} = (u_1, u_2, u_3, u_4, u_5, u_6) \in \mathbb{Z}_n^6$ such that the following conditions hold

$$\begin{cases} u_1^2 + u_2^2 + u_3^2 = 4v - 4a + 1 = v_1 \in \mathbb{Z}, \\ u_4^2 + u_5^2 + u_6^2 = 4b - 4v + 1 = v_2 \in \mathbb{Z}. \end{cases} \tag{19}$$

In this protocol, the $\delta(y) = \langle \mathbf{y}, \mathbf{y} \rangle \in \mathbb{Z}$ and $\mathbf{y} \in \mathbb{Z}^6$. We will prove the following relations:

$$\{(g, h \in \mathbb{G}, \mathbf{V} \in \mathbb{G}^2) : V_j = h^{r_j} g^{v_j} \ \forall\, j \in \{1, 2\}, V = g^v h^r \wedge v \in [a, b]\} \tag{20}$$

Theorem 3 (Cuproof protocol) *The Cuproof protocol presented in this section satisfies perfect completeness, perfect special honest verifier zero-knowledge, and computational witness extended emulation.*

Proof. The proof for Theorem 3 is given in [14]

```
Cuproof ($V \in \mathbb{G};\ v,\ a,\ b \in \mathbb{Z}$):
  1. $\mathcal{P}$ computes $v_1 = 4v - 4a + 1, v_2 = 4b - 4v + 1 \ \in \mathbb{Z}$;
  2. $\mathcal{P}$ selects $\alpha \overset{\$}{\leftarrow} \mathbb{Z}_n$;
  3. $\mathcal{P}$ computes $A = h^{\alpha} \prod_{j=1}^{2} \mathbf{g}_{[3(j-1):3j]}^{j \cdot \mathbf{d}_{[3(j-1):3j]}} \cdot \prod_{j=1}^{2} \mathbf{h}_{[3(j-1):3j]}^{j \cdot \mathbf{d}_{[3(j-1):3j]}} \in \mathbb{G}$;,
  4. $\mathcal{P}$ selects $\mathbf{s}_L, \mathbf{s}_R \overset{\$}{\leftarrow} \mathbb{Z}_n^6, \rho \overset{\$}{\leftarrow} \mathbb{Z}_n$;
  5. $\mathcal{P}$ computes $S = h^{\rho} \mathbf{g}^{\mathbf{s}_L} \mathbf{h}^{\mathbf{s}_R} \in \mathbb{G}$;
  6. $\mathcal{P} \rightarrow \mathcal{V} : A, S$;
  7. $\mathcal{V}$ selects $y, z \overset{\$}{\leftarrow} \mathbb{Z}_n^*$;
  8. $\mathcal{V} \rightarrow \mathcal{P} : y, z$;
  9. $\mathcal{P}$ selects $\tau_3, \tau_4 \overset{\$}{\leftarrow} \mathbb{Z}_n$;
  10. $\mathcal{P}$ computes computes $T_i = g^{t_i} h^{\tau_i} \in \mathbb{G}$ where $t_i$ can be computed without knowing $\mathbf{l}'$ and $\mathbf{r}'$, i.e. $t_i$ is the coefficient of $\langle l(x), r(x) \rangle$
      respectively. $i = \{3, 4\}, r_1 = 4r \in \mathbb{Z}, r_2 = -4r \ \in \mathbb{Z}$;
  11. $\mathcal{P} \rightarrow \mathcal{V} : T_3,\ T_4$;
  12. $\mathcal{V}$ selects $x \overset{\$}{\leftarrow} \mathbb{Z}_n^*$;
  13. $\mathcal{V} \rightarrow \mathcal{P} :\ x$;
  14. $\mathcal{P}$ computes $\mathbf{l}' = z \cdot \sum_{j=1}^{2} j \cdot (\mathbf{0}^{3(j-1)} \parallel \mathbf{d}_{[3(j-1):3j]} \parallel \mathbf{0}^{3(2-j)}) - \mathbf{y} + \mathbf{s}_L x \in \mathbb{Z}^6$,
      $\mathbf{r}' = z \cdot \sum_{j=1}^{2} j \cdot (\mathbf{0}^{3(j-1)} \parallel \mathbf{d}_{[3(j-1):3j]} \parallel \mathbf{0}^{3(2-j)}) + \mathbf{y} + \mathbf{s}_R x \in \mathbb{Z}^6,\ \hat{t}' = \langle \mathbf{l}', \mathbf{r}' \rangle \in \mathbb{Z}$;
  15. $\mathcal{P}$ computes $\tau_x' = \tau_4 x^2 + \tau_3 x + z^2 \sum_{j=1}^{2} j^2 \cdot r_j \in \mathbb{Z},\ \mu' = \alpha z + \rho x \ \in \mathbb{Z}$;
  16. $\mathcal{P} \rightarrow \mathcal{V} : \tau_x',\ \mu',\ \hat{t}',\ \mathbf{l}',\ \mathbf{r}'$;
```

17. $\mathcal{V}$ computes : $P = A^z \cdot S^x \cdot \mathbf{g}^{-\mathbf{y}} \cdot \mathbf{h}^{\mathbf{y}} \in \mathbb{G},\ V_1 = V^4 \cdot g^{-4a} \cdot g = g^{4v-4a+1} h^{4r} = g^{v_1} h^{r_1} \in \mathbb{G},\ V_2 = g^{4b} \cdot V^{-4} \cdot g = g^{4b-4v+1} h^{-4r} = g^{v_2} h^{r_2} \in \mathbb{G},\ \mathbf{V} = (V_1, V_2) \in \mathbb{G}^2$;

18. $\mathcal{V}$ checks these equations : $P \overset{?}{=} h^{\mu'} \cdot \mathbf{g}^{\mathbf{l}'} \cdot \mathbf{h}^{\mathbf{r}'} \in \mathbb{G},\ g\hat{t}' h^{\tau_x'} \overset{?}{=} \mathbf{V}^{z^2 \cdot (\vec{\mathbf{2}} \circ \vec{\mathbf{2}})} \cdot g^{-\delta(y)} \cdot T_3^x \cdot T_4^{x^2} \in \mathbb{G},\ \hat{t}' \overset{?}{=} \langle \mathbf{l}', \mathbf{r}' \rangle \in \mathbb{Z}$;

Our scheme

The main idea of BAZKP scheme is set out below. Given three vectors $\mathbf{w},\ \mathbf{w}',\ \mathbf{d} \in \mathbb{Z}^m$, the secret vectors $\mathbf{w}$ and $\mathbf{w}'$ are biometric data of $\mathcal{P}$, then the vectors $\mathbf{w}$, $\mathbf{w}'$ and $\mathbf{d}$ satisfy the following relation:

$$\mathbf{w} = \mathbf{w}' + \mathbf{d}. \tag{21}$$

The objective of BAZKP scheme is to persuade the verifier $\mathcal{V}$ that the secret vectors $\mathbf{w}$,$\mathbf{w}'$ belong to the same person, that is, the Euclidean Distance between $\mathbf{w}$, $\mathbf{w}'$ is smaller than e. Let $v = \langle \mathbf{d}, \mathbf{d} \rangle \in \mathbb{Z}$, v is the Euclidean Distance between $\mathbf{w}$ and $\mathbf{w}'$, the goal of this scheme is to prove that the secret value $v \in [0, e]$.

In detail, we will prove the following relations:

$$\{(v \in \mathbb{Z}; \mathbf{w}, \mathbf{w}', \mathbf{d} \in \mathbb{Z}^m; g, h, V \in \mathbb{G}) : V = g^v h^r, \mathbf{w} = \mathbf{w}' + \mathbf{d}, v = \langle \mathbf{d}, \mathbf{d} \rangle, v \in [0, e]\}. \tag{22}$$

First, we use Pedersen Commitment and Pedersen Vector Commitment to commit v, $\mathbf{w}$,$\mathbf{w}'$, $\mathbf{d}$. Then, let $\mathcal{P}$, $\mathcal{V}$ run the EDZKP protocol and the Cuproof protocol to prove the equation $\mathbf{w} = \mathbf{w}' + \mathbf{d}$, $v = \langle \mathbf{d}, \mathbf{d} \rangle$, $v \in [0, e]$. The $\mathcal{V}$ outputs "accept" only when the two protocols are ran successfully.

BAZKP scheme $(V \in \mathbb{G};\ v,\ 0,\ e \in \mathbb{Z};\ \mathbf{w}, \mathbf{w}', \mathbf{d} \in \mathbb{Z}^m)$

1. $\mathcal{P}$ computes $V = g^v h^r \in \mathbb{G}$ and sends V to $\mathcal{V}$;
2. $\mathcal{P}$, $\mathcal{V}$ run EDZKP protocol $(V \in \mathbb{G};\ v \in \mathbb{Z};\ \mathbf{w}, \mathbf{w}', \mathbf{d} \in \mathbb{Z}^m)$.
3. $\mathcal{P}$, $\mathcal{V}$ run Cuproof protocol $(V \in \mathbb{G};\ v,\ 0, e \in \mathbb{Z})$.
4. $\mathcal{V}$ outputs "accept" only when that two protocols are ran successfully.

Theorem 4 (BAZKP Scheme) *The BAZKP scheme presented in this section satisfies perfect completeness, perfect special honest verifier zero-knowledge, and computational witness extended emulation.*

Proof. Because the EDZKP protocol and the Cuproof protocol satisfy perfect completeness, perfect special honest verifier zero-knowledge, and computational witness extended emulation. Therefore, our BAZKP scheme has perfect completeness, perfect special honest verifier zero-knowledge, and computational witness extended emulation.

This scheme can also be transformed into a NIZK scheme by using the Fiat-Shamir heuristic. For example, we set $y = Hash(D||W)$,$z = Hash(D||W')$,$c = Hash(D||S)$,$x = Hash(D||K)$ in EDZKP protocol. And we set $y = Hash(A||S)$, $z = Hash(\tau_x||\mu)$, $x = Hash(T1||T2)$ in Cuproof protocol.

Privacy-preserving biometric authentication system

In this section, we construct a privacy-preserving biometric authentication system based on BAZKP scheme and evaluate its comprehensive performance using fingerprints. Before anything else, we provide some crucial background information on biometric authentication. Then, we demonstrate how to build biometric authentication systems that protect user privacy using our suggested scheme.

Background

The first critical step in biometric authentication is efficiently transforming biometric traits into feature vectors that are easy to compute. We adopt fingerprint vectors *w* with length 299, generated from the fingerprint minutiae proposed in [32]. The conversion process consists of two sequential stages: generation of the minutia cylinder-code with the MCC SDK and the kernel principal component analysis (KPCA) transformation. Our experiments are carried out using FVC 2002 and FVC 2004. Each of these datasets has a Set A and Set B. We used Set B to generate the fixed-length vector and Set A for the verification. In the following discussion, we assume that a sensor can capture the user's biometric trait and transform it into a fixed-length vector on the local side.

System construction

A privacy-preserving biometric authentication system is composed of four phases: setup, registration, query and authentication.

1. **Setup**: The user output the public parameters (g, h, $\mathbf{g}$, $\mathbf{h}$).
2. **Registration**: The user receives the binary fingerprint vector $\mathbf{w}$ from the sensor and stores it in his database. Then, the user generates the commitment of $\mathbf{w}$ (we use W to represent the commitment of $\mathbf{w}$) and sends the W to the cloud server. The cloud server stores the tuple (g, h, $\mathbf{g}$, $\mathbf{h}$, W).
3. **Query**: The user uses the sensor to get a new binary fingerprint vector $\mathbf{w}'$. Then, The user generates $\mathbf{d}$ by using $\mathbf{d} = \mathbf{w}{-}\mathbf{w}'$, where the user's local database contains $\mathbf{w}$. Then the user sends $\mathbf{d}$ and $(A, S, T_1, T_2, \tau_x, \ \mu, \ \hat{t}, \ \mathbf{l}, \ \mathbf{r}, T_3, T_4, \tau_x', \ \mu', \ \hat{t}', \ \mathbf{l}', \ \mathbf{r}')$ to the cloud server.
4. **Authentication**: Once the cloud sever receives W, it finds the tuple (W, g, h, $\mathbf{g}$, $\mathbf{h}$). The server runs the BAZKP scheme. Then, the server will output an authentication result.

Fig 1 shows the setup and registration phase, Fig 2 shows the query and authentication phase.

Security analysis

In this section, we briefly analyse the security strength of our proposed system with respect to various security attacks. First, we prove that the BAZKP scheme satisfies perfect completeness, perfect special honest verifier zero-knowledge, and computational witness extended emulation. As a result, our proposed system based on BAZKP scheme also satisfies these properties. Referring to the security analysis in [27–31], we prove that our proposed system can effectively

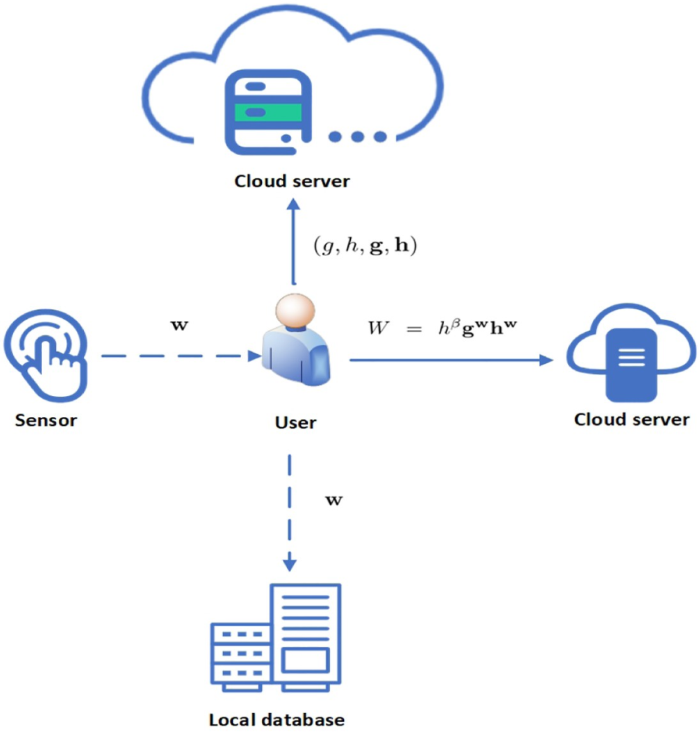
Cloud server
$(g, h, \mathbf{g}, \mathbf{h})$
Sensor
$\mathbf{w}$
User
$W = h^{\beta}\mathbf{g}^{\mathbf{w}}\mathbf{h}^{\mathbf{w}}$
Cloud server
$\mathbf{w}$
Local database

Fig 1. The setup and registration phase.

https://doi.org/10.1371/journal.pone.0286215.g001

withstand various types of attacks, including replay attacks, man in the middle attacks, impersonation attacks, and message modification attacks. Additionally, our system also supports non-traceability, further enhancing its security capabilities.

Replay attack and man in the middle attack. In our proposed system, the cloud server receives the tuple (W, g, h, $\mathbf{g}$, $\mathbf{h}$) before authenticating the user's identity. Once the data is received, the cloud server maps it to the timestamp and user identity and stores them in the

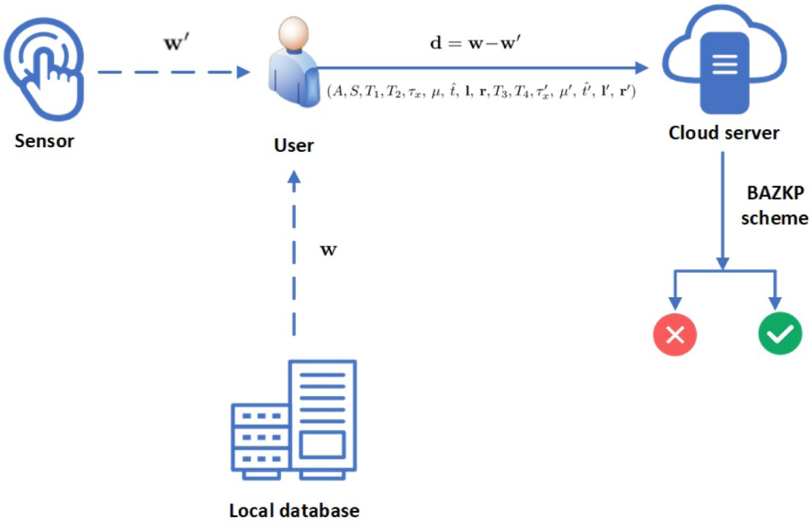
Sensor
$\mathbf{w}'$
User
$\mathbf{d} = \mathbf{w} - \mathbf{w}'$
$(A, S, T_1, T_2, \tau_x, \mu, \hat{t}, \mathbf{l}, \mathbf{r}, T_3, T_4, \tau'_x, \mu', \hat{t}', \mathbf{l}', \mathbf{r}')$
Cloud server
BAZKP scheme
$\mathbf{w}$
Local database

Fig 2. The query and authentication phase.

https://doi.org/10.1371/journal.pone.0286215.g002 database. This approach allows the cloud server to defend against replay attacks by verifying if the mapped data in the database matches the data sent by the user and if the time difference between them falls within a legal interval.

Additionally, the proposed system enables mutual authentication to be performed securely between the user and the cloud server, ensuring the system can withstand man in the middle attacks.

Impersonation attack and message modification attack. Suppose a malicious attacker sends the tuple $(A, S, T_1, T_2, \tau_x, \ \mu, \ \hat{t}, \ \mathbf{l}, \ \mathbf{r}, T_3, T_4, \tau'_x, \ \mu', \ \hat{t'}, \ \mathbf{l}', \ \mathbf{r}', W, W')$ to the cloud server where $W' \neq h^{\theta}\mathbf{g}^{\mathbf{w}'}\mathbf{h}^{\mathbf{w}'}$. Once the attacker pass the authentication with overwhelming probability then it means the BAZKP scheme can not meet soundness. Since we have proven that the BAZKP scheme meets soundness, a message modification attack is impossible in the proposed system. Therefore our proposed system can withstand the message modification attack and impersonation attack.

Non-traceability. In this proposed system, we utilize Pedersen Vector Commitment to preserve the secret values, specifically $W = h^{\theta}\mathbf{g}^{\mathbf{w}}\mathbf{h}^{\mathbf{w}}$, based on the hardness of the DL assumption. Consider a scenario where an attacker has captured the value **w** from W. However, it is impossible for the attacker to compute **w** from W due to the difficulty of solving the DL assumption. Additionally, the user selects random numbers $\alpha, \beta, \theta, \phi \in \mathbb{Z}_n$ for each execution in our proposed system, and the BAZKP scheme also meets zero-knowledge. Therefore, our scheme supports non-traceability.

Implementation and performance analysis

In this section, we present the results of our evaluation and prototype implementation. Then, compare it with other existing related biometric authentication schemes.

Implementation

As a user-centric biometric authentication system, our system's performance is evaluated on a personal laptop and a mobile phone. In the simulation, we utilize a mobile phone with Android 13 operating system, 3123 MHz CPU, and 8 GB RAM. We also use a personal laptop with macOS 11.2.2, Apple M1, and 8 GB RAM. The Java library UJMP and the C++ library Armadillo are used for server emulation of mobile phones and computers, respectively. At the same time, the client is written using React.

We noticed that performance depends on many factors. The system may not operate at its best due to the above choice. The size of the two primes p, q is set to 1024 bits. A Pedersen hash function over an RSA group whose modulo $n = p^*q$ is benchmarked. We generate witness refer to literature [9, 10]. We also set $e = 7000$, in other words, the user can pass the identity authentication only when the Euclidean distance between **w** and **w**′ is less than or equal to 4.

The system uses sensors to capture the user's fingerprints and generate the corresponding vectors. The sensor captures the fingerprint feature points and generates the fingerprint vectors based on the horizontal and vertical coordinates of these fingerprint points, as shown in Fig 3.

The final data is the average of the data we obtained by doing 10000 experiments. It is possible to conclude that our scheme's communication cost is equal to $11\mathbb{G} + 7\mathbb{Z}_n + (2m + 19)\mathbb{Z}$ through computation and analysis, indicating that it is kept constant.

Fig 4 shows the line charts of the proving time and the verification time of the EDZKP protocol, respectively. Fig 5 shows the line charts of the proving time and the verification time of the BAZKP scheme (not including witness generation), respectively. Table 1 shows our system's proving time and verification time under the different Euclidean distance. The

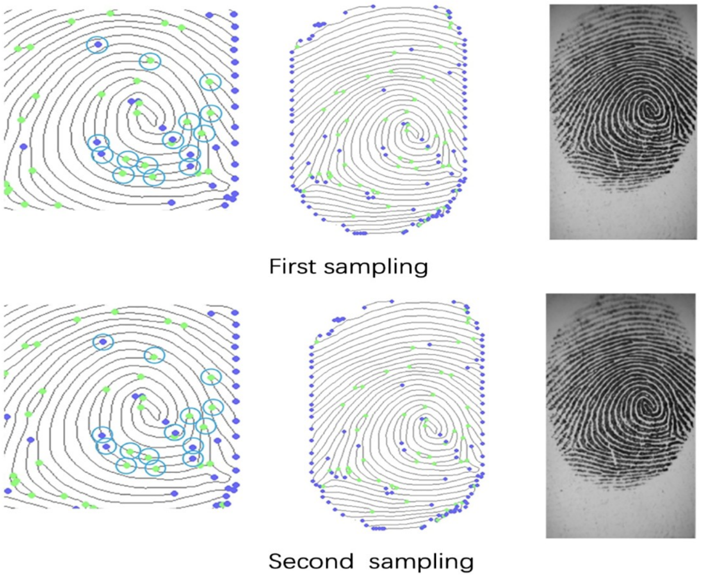

Fig 3. Details of fingerprint sampling.

https://doi.org/10.1371/journal.pone.0286215.g003

simulation results show that the proving and verification time of the BAZKP scheme remains almost constant with the increase of the Euclidean distance. From this, we can conclude that no matter how extensive the range is, the proving time and the verification time remain almost unchanged.

Performance analysis

Finally, we compare the BAZKP scheme with other similar schemes. Table 2 shows the comparison results between our scheme, literature [20], literature [18], and literature [21]. Lee et al. [21] proposed a biometric authentication scheme that divides the biometric template into

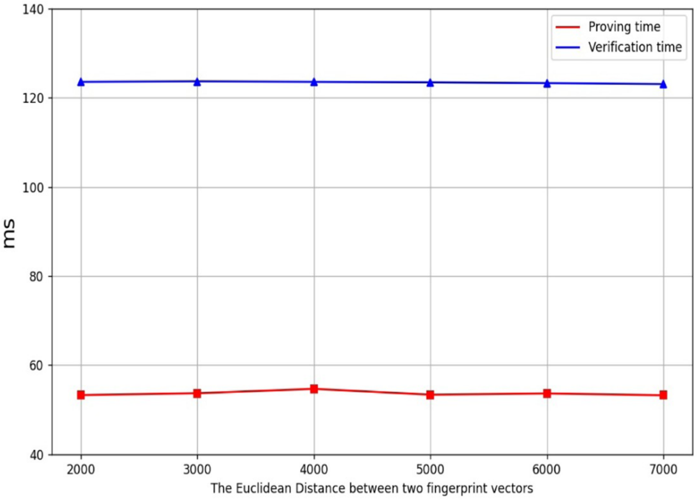

Fig 4. The time cost of EDZKP protocol.

https://doi.org/10.1371/journal.pone.0286215.g004

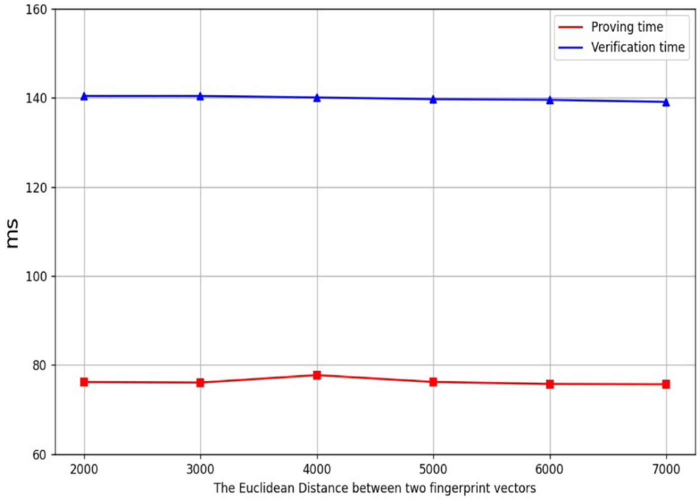

Fig 5. The time cost of BAZKP scheme.

https://doi.org/10.1371/journal.pone.0286215.g005

Table 1. BAZKP Performance.

Length of biometric vector	Euclidean distance	Timing (ms)	
		Prove	Verify
299	2000	76.14	140.44
299	3000	76.03	140.43
299	4000	77.70	140.11
299	5000	76.17	139.71
299	6000	75.70	139.57
299	7000	75.64	139.10

https://doi.org/10.1371/journal.pone.0286215.t001

Table 2. Comparison results of similar schemes.

Schemes	Universal	Untrusted Setup	Assumption	Time Complexity
Literature [20]	•	○	–	$O(n^2)$
Literature [18]	•	○	–	$O(n^2)$
Literature [21]	○	○	–	$O(n)$
Our scheme	•	○	RSA+DL	$O(n)$

Here *n* represents the length of the fingerprint vector. A black circle for a universal denotes that the scheme can be promoted, and a white circle for an untrusted setup denotes that the scheme is updatable. DL stands for discrete log.

https://doi.org/10.1371/journal.pone.0286215.t002

fragments and stores them on the blockchain. However, compared with our scheme, the real-time performance of this scheme is low. Similarly, Zhou et al. [18] presented a user-centric biometric authentication scheme, but it encountered an exponential increase in time complexity with the increase in *n*. In contrast, our solution offers a significant advantage regarding time complexity. Compared with the biometric authentication scheme in [20], our scheme achieves the goal of privacy protection while improving efficiency. In addition, the security of this scheme is based on the RSA assumption and the discrete log assumption, which is relatively secure compared to other schemes.

Conclusions and future work

In this paper, we combine biometrics with ZKP protocol to introduce a biometric authentication scheme for privacy-preserving named BAZKP. The security analysis shows that the scheme satisfies the perfect completeness, perfect special honest verifier zero-knowledge, and computational witness extended emulation. The evaluation results show that, no matter how big the interval of the range proof is, the proof time and the verification time are nearly constant. Finally, based on the BAZKP scheme, we build a user-centric biometric authentication system.

We will focus on improving the efficiency and safety of the BAZKP scheme in the future. One approach will be to leverage a more efficient proof-of-zero algorithm to refine the accuracy of biometric templates while preserving privacy. Another approach will be to employ multi-factor authentication, which combines biometric data with other authentication methods, such as passwords or tokens, to boost security. Additionally, combining the BAZKP scheme with blockchain technology can provide additional security and transparency to the authentication process.

Supporting information

S1 Data.
(ZIP)

S1 Appendix.
(PDF)

S2 Appendix.
(PDF)

Author Contributions

Conceptualization: Ying Chen, Xiaqing Zhou.

Data curation: Cong Deng.

Formal analysis: Cong Deng.

Funding acquisition: Ying Chen.

Software: Cong Deng, Xiaqing Zhou.

Validation: Xiaqing Zhou.

Writing – original draft: Xuechun Mao.

Writing – review & editing: Ying Chen.

References

1. Jain AK, Nandakumar K, Ross A. 50 Years of Biometric Research: Accomplishments, Challenges, and Opportunities. Pattern Recognition Letters. 2016; 79:80–105. https://doi.org/10.1016/j.patrec.2015.12.013
2. Sarkar A, Singh BK. A Review on Performance, Security and Various Biometric Template Protection Schemes for Biometric Authentication Systems. Multimedia Tools and Applications. 2020; 79 (37):27721–27776. https://doi.org/10.1007/s11042-020-09197-7
3. Goldwasser S, Micali S, Racko C. The Knowledge Complexity of Interactive Proof Systems. SIAM Journal on Computing. 1989; 18(1):186–208. https://doi.org/10.1137/0218012
4. Maller M, Bowe S, Kohlweiss M, Meiklejohn S. Sonic: Zero-Knowledge SNARKs from Linear-Size Universal and Updateable Structured Reference Strings; 2019.

5. Gabizon A, Williamson ZJ, Ciobotaru O. PLONK: Permutations over Lagrange-bases for Oecumenical Noninteractive arguments of Knowledge; 2019.
6. Bunz B, Fisch B, Szepieniec A. Transparent SNARKs from DARK Compilers. Annual International Conference on the Theory and Applications of Cryptographic Techniques; 2020: 677–706.
7. Brickell EF, Chaum D, Damgard IB. Gradual and Verifiable Release of a Secret (Extended Abstract). Advances in Cryptology—CRYPTO'87; 1988: 156–166.
8. Chan A, Frankel Y, Tsiounis Y. Easy Come Easy Go Divisible Cash. Advances in Cryptology—EUROCRYPT'98; 1998: 561–575.
9. Rabin MO, Shallit JO. Randomized Algorithms in Number Theory. Communications on Pure and Applied Mathematics. 1986; 39(1):239–256. https://doi.org/10.1002/cpa.3160390713
10. Lipmaa H. On Diophantine Complexity and Statistical Zero-Knowledge Arguments. Advances in Cryptology—ASIACRYPT 2003; 2003: 398–415.
11. Groth J. Non-interactive Zero-Knowledge Arguments for Voting. Applied Cryptography and Network Security; 2005: 467–482.
12. Bootle J, Cerulli A, Chaidos P. Efficient Zero-Knowledge Arguments for Arithmetic Circuits in the Discrete Log Setting. Advances in Cryptology—EUROCRYPT 2016; 2016: 327–357.
13. Bunz B, Bootle J, Boneh D. Bulletproofs: Short Proofs for Confidential Transactions and More. 2018 IEEE Symposium on Security and Privacy; 2018: 315–334.
14. Deng C, Tang X, You L. Cuproof: A Novel Range Proof with Constant Size; 2021
15. Jain A, Bolle R, Pankanti S. Introduction to Biometrics; 1996.
16. Barni M, Bianchi T, Catalano D. Privacy-Preserving Fingercode Authentication. Proceedings of the 12th ACM Workshop on Multimedia and Security; 2010: 231–240.
17. Blanton M, Gasti P. Secure and Efficient Protocols for Iris and Fingerprint Identification. Computer Security; 2011: 190–209.
18. Zhou K, Ren J. PassBio: Privacy-Preserving User-Centric Biometric Authentication. IEEE Transactions on Information Forensics and Security. 2018; 13(12):3050–3063. https://doi.org/10.1109/TIFS.2018.2838540
19. Hammad M, Liu Y, Wang K. Multimodal Biometric Authentication Systems Using Convolution Neural Network Based on Different Level Fusion of ECG and Fingerprint. IEEE Access. 2019; 7:26527–26542. https://doi.org/10.1109/ACCESS.2018.2886573
20. Zhang C, Zhu L, Xu C. PTBI: An Efficient Privacy-Preserving Biometric Identification based on Perturbed Term in the Cloud. Information Sciences. 2017; 409:56–67. https://doi.org/10.1016/j.ins.2017.05.006
21. Lee YK, Jeong J. Securing Biometric Authentication System using Blockchain. ICT Express. 2021; 7 (3):322–326. https://doi.org/10.1016/j.icte.2021.08.003
22. Azees M, Vijayakumar P, Deboarh LJ. EAAP: Efficient Anonymous Authentication With Conditional Privacy-Preserving Scheme for Vehicular Ad Hoc Networks. IEEE Transactions on Intelligent Transportation Systems. 2017; 18(9):2467–2476. https://doi.org/10.1109/TITS.2016.2634623
23. Zhou X, Luo M, Vijayakumar P. Efficient Certificateless Conditional Privacy-Preserving Authentication for VANETs. IEEE Transactions on Vehicular Technology. 2022; 71(7):7863–7875. https://doi.org/10.1109/TVT.2022.3169948
24. Liu Y, Yu J, Fan J. Achieving Privacy-Preserving DSSE for Intelligent IoT Healthcare System. IEEE Transactions on Industrial Informatics. 2022; 18(3):2010–2020. https://doi.org/10.1109/TII.2021.3100873
25. Yang H, Vijayakumar P, Shen J. A Location-based Privacy-Preserving Oblivious Sharing Scheme for Indoor Navigation. Future Generation Computer Systems. 2022; 137(1):42–52. https://doi.org/10.1016/j.future.2022.06.016
26. Wei F, Vijayakumar P, Kumar N. Privacy-Preserving Implicit Authentication Protocol Using Cosine Similarity for Internet of Things. IEEE Internet of Things Journal. 2021; 8(7):5599–5606. https://doi.org/10.1109/JIOT.2020.3031486
27. Azees M, Vijayakumar P, Karuppiah M. An Efficient Anonymous Authentication and Confidentiality Preservation Schemes for Secure Communications in Wireless Body Area Networks. Wireless Networks. 2021; 27:2119–2130. https://doi.org/10.1007/s11276-021-02560-y
28. Subramani J, Maria A, Rajasekaran AS. Lightweight Privacy and Confidentiality Preserving Anonymous Authentication Scheme for WBANs. IEEE Transactions on Industrial Informatics. 2022; 18(5):3484–3491. https://doi.org/10.1109/TII.2021.3097759
29. Rajasekaran AS, Azees M. An Anonymous Blockchain-Based Authentication Scheme for Secure Healthcare Applications. Security and Communication Networks. 2022. https://doi.org/10.1155/2022/2793116

30. Rajasekaran AS, Maria A, Rajagopal M. Blockchain Enabled Anonymous Privacy-Preserving Authentication Scheme for Internet of Health Things. Sensors. 2022; 23(1):240. https://doi.org/10.3390/s23010240 PMID: 36616838

31. Jegadeesan S, Dhamodaran M, Azees M. Computationally Efficient Mutual Authentication Protocol for Remote Infant Incubator Monitoring System. Healthcare Technology Letters. 2019; 6(4):92–97. https://doi.org/10.1049/htl.2018.5006 PMID: 31531222

32. Jin Z, Lim MH, Teoh ABJ. Generating Fixed-Length Representation From Minutiae Using Kernel Methods for Fingerprint Authentication. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2016; 46(10):1415–1428. https://doi.org/10.1109/TSMC.2015.2499725