Kalos: Hierarchical-auditable and Human-binding Authentication Scheme for Clinical Trial

Chang Chen, Zelong Wu, Guoyu Yang, Qi Chen, Wei Wang, *Member, IEEE* and Jin Li, *Senior Member, IEEE*

***Abstract*—Clinical trials are crucial in the development of new medical treatment methods. To ensure the correctness of clinical trial results, medical institutes need to collect and process large volumes of participant data, which has prompted research on privacy preservation and data reliability. However, existing solutions struggle to resolve the trade-off between them due to the trust gap between the physical and digital worlds, limiting their practicality. To tackle the issues above, we present Kalos, a novel authentication scheme for clinical trials. Kalos leverages diversified cryptographic tools, such as card-based anonymous credential and zero-knowledge proof to achieve authentication with visual verification and selective disclosure of attributes. It has properties such as unforgeability, blindness, privacy preservation, and human-binding that support hierarchical auditability and data de-duplication to enhance the reliability of clinical trials. We then provide the security and performance analysis of Kalos to show its potential to be deployed in the medical consumer electronics scenario. The computational cost of the smartcard is irrespective of the number of certified attributes, and the total computational cost of Kalos is within tens of milliseconds with the commonly used number of attributes.**

***Index Terms*—Clinical trial, smartcard, de-duplication, anonymous credentials, privacy-preserving**

I. Introduction

Clinical trials play a crucial role in medical research, providing scientific evidence for evaluating the effectiveness and safety of new drugs or treatment methods. The advancement of pharmacology has led to an increasing number of new drugs to be evaluated and the exploration of new indications, consequently resulting in a significant rise in the number of clinical trials [1]. Each year, millions of subjects participate in different trials [2].

Recruiting subjects is one of the primary challenges in clinical trials, which usually occupies one-third of the study duration [3]. To improve recruitment efficiency, research institutions often provide compensation to encourage active participation [4]. However, some individuals, motivated by these profits, may join “multiple overlapping trials” within a short period. Such behavior may not only adversely affect the health of the overlapping trial participants [4], [5], but also introduce errors or even premature termination of studies due to heightened adverse events. The presence of overlapping trial participants significantly impacts the reliability of clinical trial data, and thus warrants serious attention.

To ensure the data reliability of clinical trials, data de-duplication is usually performed in two different phases: before or after the clinical trial. Before the trial, unique identifiers can be generated for participants via entity resolution. This process evaluates the similarity between records with different identifiers based on information like birthday, name, surname, and phone number, thus achieving data de-duplication [6], [7], [8]. Overlapping trial participants can also be identified through registry information upon registration [4], [9]. After the trial, participants' personal information can be used to compare with the history data to check if they have recently participated in other trials [5]. However, these techniques rely on participants' private data, posing significant security risks. The approach of simply using identifiers to anonymize participants may fail for de-duplication across different medical institutions because the principles for generating identifiers are not standardized, and the same participant may be assigned different identifiers in different clinical trials. Although most clinical trial participants consent to share their data with data scientists, the sharing of private data may still act as a barrier to the recruitment of potential participants [10].

A. Motivation and Contributions

Numerous studies focus on data privacy preservation, protecting data privacy at the source is a more reliable approach as the upper bound of privacy disclosure is lower [11], [12]. Therefore, we aim to develop a novel authentication scheme for clinical trial that ensures data reliability while protecting privacy along the following dimensions: 1) In the physical world, authenticating trial participants through biometric feature without involving third parties. But avoid introducing heavy dependency of specific devices [13], [14] for better availability; 2) In the digital world, minimizing the information disclosure to protect participants' private attributes. Additionally, providing hierarchical auditability can improve the scheme's practicality [15], [16]. With this property, one or more medical institutions can cooperate to find overlapping

Chang Chen is with the Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, Beijing Jiaotong University, Beijing 100044, China, and also with the Guangdong Provincial Key Laboratory of Blockchain Security, Guangzhou University, Guangzhou 510006, China (e-mail: cc.blockchain@bjtu.edu.cn).

Wei Wang is with the Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, Beijing Jiaotong University, Beijing 100044, China (e-mail: wangwei1@bjtu.edu.cn).

Zelong Wu, Guoyu Yang, Qi Chen, and Jin Li are with the Institute of Artificial Intelligence, Guangzhou University, Guangzhou 510700, China, and also with the Guangdong Provincial Key Laboratory of Blockchain Security, Guangzhou University, Guangzhou 510006, China (e-mail: 2112206151@e.gzhu.edu.cn, yangguoyu1020@163.com; chenqi.math@gmail.com; lijin@gzhu.edu.cn).

This work was supported in part by the National Natural Science Foundation of China, under Grant U23A20307, U21A20463, 62132018, 62272118, 62172117, and in part by Beijing Natural Science Foundation, under Grant L221014.

Corresponding authors: Wei Wang; Qi Chen

trial participants through privacy-insensitive authentication records. Misbehaved participants will be held accountable by revealing their real identities with the help of authorities.

So we propose Kalos to cover the aforementioned goals. Our main contributions are:

1) We present a new public key encryption scheme that supports verifiable encryption and equality test. Anyone can test whether two ciphertexts, even encrypted with different public keys, are encryptions of the same message.
2) We propose a novel hierarchically auditable authentication scheme based on a novel human-binding card-based credential system with minimum hardware dependency. It allows subjects be visually authenticated to participate in clinical trials with privacy preserved but will be identified and held accountable once misbehave.

The rest of the article is organized as follows. In Section I-B, we discuss the related works. In Section II, we introduce the main building blocks including preliminaries and the schemes we proposed. The problem definition and the detailed construction are described in Section III and Section IV, respectively. In Sections V and VI, we give the security analysis and performance evaluation. Conclusion and future work are reported in Section VII.

B. Previous Work

Clinical trials involve extensive data throughout the initiation, recruitment, trial, and analysis phases, raising significant concerns about privacy issues. Angeletti *et al.* [17] focused on protecting participants' privacy during the recruitment phase. They designed a framework for digital clinical trials that includes authenticated wearable devices. Throughout the recruitment phase, participants' data will be stored locally. Digital signatures and public blockchain technology ensure the authenticity and completeness of the collected data. However, this solution relies on specific wearable devices to collect and store data and can only prevent data duplication by submitting participants' privacy data to medical institutions after recruitment. Yuan *et al.* [18] explored how searchable encryption can protect the privacy of clinical research networks without compromising usability. Their results demonstrated that this technology achieves zero compromise in accuracy for privacy-protecting cohort discovery tasks, but it is only applicable during participant screening. Hripcsak *et al.* [19] introduced a shift and truncate method to obscure data and reduce the possibility of identifying participants through public data. This method maintains the relative temporal relationship between clinical trial events while hiding the actual dates. Maritsch *et al.* [20] addressed the re-identification risks of clinical trial participants by analyzing existing publications and defining K-anonymity-based standards. Movahedi *et al.* [21] proposed a comprehensive privacy protection protocol to derive causal relationships from randomized controlled trials. They integrated three privacy protection technologies, such as PSI, MPC, and DP into the processes of user recruitment, data aggregation, and statistical analysis to ensure mutual privacy protection. However, these technologies focus on protecting data privacy during post-trial data usage. Tucker *et al.* [22] examined how to share clinical trial data with third-party researchers while maintaining privacy, and suggested principles for data anonymization (e.g., re-encoding patient identifiers, obscuring dates) and controlled data access (e.g., legally binding sharing agreements, secure "vault" systems). Such principle is applicable only when the recruitment is completed and the participants are willing to submit their full private data to the medical institution. To de-duplicate participants in clinical trials, Emam *et al.* [5] developed a probabilistic check protocol using homomorphic encryption and a central database to screen participants interviewed by phone and on-site. However, ciphertext data must be decrypted to respond to the queries. In summary, the challenge of protecting participant privacy while supporting data de-duplication during clinical trial remains strict. New technologies are needed to ensure data reliability and privacy security of the participants' data during the clinical trial.

II. Building Blocks

A. Preliminaries and Notation

1) Bilinear Groups: Bilinear groups are a set of three groups $\mathbb{G}_1$, $\mathbb{G}_2$, and $\mathbb{G}_T$ of order p along with a map, called pairing, $\boldsymbol{e} : \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ that is:

- bilinear: for any $g \in \mathbb{G}_1$, $\tilde{g} \in \mathbb{G}_2$, and $a, b \in \mathbb{Z}_p$, $\boldsymbol{e}(g^a, \tilde{g}^b) = \boldsymbol{e}(g, \tilde{g})^{ab}$;
- non-degenerate: for any $g \in \mathbb{G}_1^*$ and $\tilde{g} \in \mathbb{G}_2^*$, $\boldsymbol{e}(g^a, \tilde{g}^b) \neq 1_{\mathbb{G}_T}$;
- efficient: for any $g \in \mathbb{G}_1$ and $\tilde{g} \in \mathbb{G}_2$, $\boldsymbol{e}(g^a, g^b)$ can be efficiently computed.

2) Computational Assumptions:

- Discrete Logarithm (DL) assumption: Given $(g, g^a) \in \mathbb{G}^2$, the DL assumption in the group $\mathbb{G}$ states that there is no probabilistic polynomial time ($\mathcal{PPT}$) algorithm that can recover a with nonnegligible advantage.
- Decisional Diffie–Hellman (DDH) assumption: Given $(g, g^a,\ g^b, g^c) \in \mathbb{G}^4$, the DDH assumption in the group $\mathbb{G}$ states that there is no $\mathcal{PPT}$ algorithm that can decide whether $c = a \cdot b$ or c is random with nonnegligible advantage.
- q-Strong Diffie– Hellman (q-SDH) assumption: Given a (q+2) tuple $(g, g^x, g^{x^2}, \ldots, g^{x^q})$, the q-SDH assumption in the group $\mathbb{G}$ states that there is no $\mathcal{PPT}$ algorithm that can output a pair $(c, g^{\frac{1}{x+c}})$ where $c \in \mathbb{Z}_p$ with nonnegligible advantage.

3) Zero-Knowledge Protocols: Zero-knowledge protocol enables a prover to convince a verifier that a statement is true without revealing anything except the validity of the statement. An interactive zero-knowledge proof system is called a Sigma protocol (Σ-protocol) if it contains 3 phases between the prover P and the verifier V as below:

- (Commit) P sends a first message a to V;
- (Challenge) V sends a random challenge e to P;
- (Response) P replies with a second message z.

For any NP relation $(x, \omega) \in R$, where ω is the witness of the statement x, a valid Sigma protocol is required to satisfy standard completeness and the variants of soundness and zero-knowledge as below:

- Completeness. If P and V follow the protocol, then V always accepts.
- Special soundness. For any $x \in X$ and any pair of accepting transcripts (a, c, r), (a, c', r') with $c \neq c'$, there exists a $\mathcal{PPT}$ extractor outputs a witness ω for x.
- Special honest-verifier zero-knowledge (SHVZK). There exists a $\mathcal{PPT}$ simulator $\mathcal{S}$ such that for any $x \in X$ and challenge c, $\mathcal{S}$ produces conversations (a, c, r) with the same probability distribution as conversations between honest P and V.

We also use Signature of Knowledge (SoK) as non-interactive zero-knowledge proof (following the definition in [23] and [24]) while designing the authentication protocol.

4) Card-based Anonymous Credential with BBS+ Signature: A card-based Anonymous Credential (cbAC) system [25] is an anonymous credential scheme with visual holder authentication. It contains three interactive procedures (Setup,Join,Present) which will be executed between an issuer, and arbitrary tamper-resistant smartcards, holders, and verifiers. In their settings, the verifier can visually verify that the picture on the smartcard (hereinafter referred to as card) matches the individual. The holder must cooperate with the card to produce valid proof of knowledge of the selectively disclosed attributes and the corresponding BBS+ signature. Both of them will receive the shared state from a trusted card issuer $\mathcal{F}_{\mathsf{cardAuth}}$, consisting of a Pseudo-Random Function (PRF) key K and a non-hiding commitment $Q = h_1^m$ to attribute m contributed by the card.

We briefly review the standard BBS+ signature: Let $h_0, \ldots, h_\ell \in \mathbb{G}_1^{\ell+1}$ be the generators. The issuer randomly chooses $\gamma \leftarrow\!\!\$\ \mathbb{Z}_p^* (\overset{\mathsf{def}}{=} \mathbb{Z}_p \backslash \{1\})$ and set $(\gamma, \omega = \tilde{g}^\gamma)$ as the secret-public key pair. Given messages $\boldsymbol{m} = (m_1, \ldots, m_\ell) \in \mathbb{Z}_p$, the issuer randomly chooses $e, s \leftarrow\!\!\$\ \mathbb{Z}_p^2$ and computes $A = (gh_0^s \prod_{i=1}^{\ell} h_i^{m_i})^{\frac{1}{e+\gamma}}$. The BBS+ signature can be verified by checking whether $\boldsymbol{e}(A, \omega\tilde{g}^e) = \boldsymbol{e}(gh_0^s \prod_{i=1}^{\ell} h_i^{m_i}, \tilde{g})$ holds. BBS+ signature satisfies the EUF-CMA security if the q-SDH problem is hard in the bilinear group [26].

5) Twisted ElGamal Encryption: Twisted ElGamal [27] is modified from the standard ElGamal encryption algorithm, it switched the roles of key encapsulation and session key, and lifted the message m on a new generator. Its ciphertext has the same structure as Pedersen commitment, so it can easily connect with existing zero-knowledge proof systems (such as sigma protocol [28]). Twisted ElGamal is IND-CPA secure under the DDH assumption. We recall the algorithm proposed in [27] as follows:

- $\mathsf{TE.Setup}(1^\lambda)$: Run $(\mathbb{G}_1, h_0, p) \leftarrow \mathsf{GroupGen}(1^\lambda)$, pick $h_1 \leftarrow\!\!\$\ \mathbb{G}_1^* (\overset{\mathsf{def}}{=} \mathbb{G}_1 \backslash \{1\})$, set $pp = (\mathbb{G}_1, h_0, h_1, p)$ as global public parameters. The randomness and message spaces are $\mathbb{Z}_p$.
- $\mathsf{TE.KeyGen}(pp)$: On input pp, choose $sk \leftarrow\!\!\$\ \mathbb{Z}_p$, set $pk = h_0^{sk}$.
- $\mathsf{TE.Encrypt}(pk, m)$: Pick $r \leftarrow\!\!\$\ \mathbb{Z}_p$,compute $X = pk^r$, $Y = h_0^r h_1^m$, output $C = (X, Y)$.
- $\mathsf{TE.Decrypt}(sk, C)$: Parse $C = (X, Y)$, compute $h_1^m = Y / X^{sk^{-1}}$, recover m from h_1^m.

6) Public-Key Encryption with Equality Test: Also denoted as PKEET, is a primitive proposed by Yang *et al.* [29] that can categorize ciphertexts, even those encrypted with different public keys, with the same underlying messages into one cluster. Here we recall the definition of PKEET:

- $\mathsf{PKEET.Setup}$: Pick $x \leftarrow\!\!\$\ \mathbb{Z}_p$ and compute $y = g^x$. Set $pk = y$ and $sk = x$.
- $\mathsf{PKEET.Encrypt}(m, y)$: Let $m \in \mathbb{G}_1^*$, pick $r \leftarrow\!\!\$\ \mathbb{Z}_p^*$, compute $U = g^r$, $V = m^r$, $W = \mathsf{H}(U, V, y^r) \oplus\ m||r$. The ciphertext is $C = (U, V, W)$.
- $\mathsf{PKEET.Decrypt}(C, x)$: To decrypt a ciphertext $C = (U, V, W)$, compute $m||r \leftarrow \mathsf{H}(U, V, U^x) \oplus W$. If $(m \in \mathbb{G}_1^* \wedge r \in \mathbb{Z}_p^* \wedge U = g^r \wedge V = m^r)$, return m; otherwise, return $\perp$.
- $\mathsf{PKEET.Test}(C_1, C_2)$: Given two ciphertexts $C_1 = (U_1, V_1, W_1)$ and $C_2 = (U_2, V_2, W_2)$, if $\boldsymbol{e}(U_1, V_2) = \boldsymbol{e}(U_2, V_1)$, return 1; otherwise, return 0.

B. Twisted ElGamal with Equality Test

At a high-level overview, we need a construction that supports the hiding of sensitive information while supporting the equivalence comparison. The first requirement ensures that the subjects can only selectively disclose the required attributes, while the second gives the medical institutes the ability to deduplicate the subjects according to their authentication information. Specifically, we use twisted ElGamal to better adapt to the BBS+ signature based cbAC systems without introducing excessive knowledge proofs for consistency between them.

Now, we extend the capabilities of the Twisted ElGamal (TE) encryption algorithm by combining it with PKEET to get Twisted ElGamal with Equality Test (TEET). This allows the consistency of the plaintext to be checked without decrypting the corresponding ciphertext. We use a non-hiding commitment as the input and output of the encryption and decryption methods of TEET. Formally, this new primitive consists of five algorithms as below:

- $\mathsf{TEET.Setup}(1^\lambda)$: Same as $\mathsf{TE.Setup}$.
- $\mathsf{TEET.KeyGen}(1^\lambda)$: Same as $\mathsf{TE.KeyGen}$.
- $\mathsf{TEET.Encrypt}(pk, m; r, v)$: Let $Q = h_1^m$. Compute $X = pk^r$, $Y = h_0^r Q$, $U = Q^v$, $K = h_1^v$, output $ct = (X, Y, U, K)$.
- $\mathsf{TEET.Decrypt}(ct)$: Parse $ct = (X, Y, U, K)$, compute $h_1^m = Y / X^{sk^{-1}}$, recover m from h_1^m.
- $\mathsf{TEET.Test}(ct_1, ct_2)$: Given two TEET ciphertexts $ct_1 = (X_1, Y_1, U_1, K_1)$ and $ct_2 = (X_2, Y_2, U_2, K_2)$, if $\boldsymbol{e}(U_1, K_2) = \boldsymbol{e}(U_2, K_1)$, return 1; otherwise, return 0.

However, the randomness we used is identical between (X, Y) and (U, K), while the former can be treated as a TE ciphertext. We add additional proof to ensure the same Q was used in Y and U without disclosing the value of Q, so that the equality of different TEET ciphertexts can only be tested with the same approach of PKEET. This can help us to integrate other researches related to PKEET, such as authorization policies [30] or scenarios of outsourced computation [31], to enhance the functionality and expand application scenarios.

We define the relation mentioned above as L_{valid}, where

$r_1 = -rv$:

$$\begin{aligned} L_{valid} = \{(pk, h_0, h_1, X, Y, U, K) | \exists\ r, m, v, r_1\ s.t.& \\ X = pk^r \wedge& \\ Y = h_0^r h_1^m \wedge& \\ U = K^m \wedge& \\ U = Y^v h_0^{r_1}\}& \end{aligned}$$

Sigma protocol for L_{valid}. To prove L_{valid} in zore knowledge, we design a Sigma protocol $\Sigma_{valid} = (\mathsf{Setup}, \mathsf{P}, \mathsf{V})$ for L_{valid} to prove that the twisted ElGamal encryption ciphertext $C = (X, Y)$ and a pair for equality test (U, K) consist of the same value h_1^m. The Setup algorithm of Σ_{valid} is the same as that of the twisted ElGamal. On statement (h_0, h_1, X, Y, U, K), P and V interact as below:

1. P picks $a, b, c, d \leftarrow\!\!\$\ \mathbb{Z}_p^4$, sends $A = pk^a$, $B = h_0^a h_1^b$, $C = K^b$, $D = M^c h_1^d$ to V.
2. V picks $e \leftarrow\!\!\$\ \mathbb{Z}_p$ and sends it to P as the challenge.
3. P computes $z_1 = a + er$, $z_2 = b + em$, $z_3 = c + ev$, $z_4 = d + er_1$ using witness $w = (r, m_1, v, r_1)$, then sends (z_1, z_2, z_3, z_4) to V. V accepts if and only if the following three equations hold simultaneously:

$$pk^{z_1} = AX^e \tag{1}$$

$$h_0^{z_1} h_1^{z_2} = BY^e \tag{2}$$

$$K^{z_2} = CU^e \tag{3}$$

$$Y^{z_3} h_0^{z_4} = DU^e \tag{4}$$

Lemma 1: Σ_{valid} is a public-coin SHVZK proof of knowledge for L_{valid}.

Proof: We prove that all three properties required for Σ_{valid} are met.

Perfect completeness is obvious from a simple calculation.

To show *special soundness*, we fix the initial message (A, B, C, D), suppose there are two accepting transcripts (e, z_1, z_2, z_3, z_4) and $(e', z_1', z_2', z_3', z_4')$ with $e \neq e'$, the witness can be extracted as below. From (1), we have $z_1 = a + er$ and $z_1' = a + e'r$, which implies $r = (z_1 - z_1')/(e - e')$. And so as from (2), (3) and (4), we have $z_2 = b + em$, $z_2' = b + e'm_1$, $z_3 = c + ev$, $z_3' = c + e'v$, $z_4 = d + er_1$, $z_4' = d + e'r_1$, which imply $m = (z_2 - z_2')/(e - e')$, $v = (z_3 - z_3')/(e - e')$ and $r_1 = (z_4 - z_4')/(e - e')$.

To show *special HVZK*, for a fixed challenge e, the simulator $\mathcal{S}$ works as below: picks $z_1, z_2, z_3, z_4 \leftarrow\!\!\$\ \mathbb{Z}_p^4$, then computes $A = pk^{z_1}/X^e$, $B = h_0^{z_1} h_1^{z_2}/M^e$, $C = K^{z_2}/U^e$, $D = M^{z_3} h_0^{z_4}/U^e$. It is obvious that $(A, B, C, e, z_1, z_2, z_3, z_4)$ is an accepting transcript, and it is distributed as in the real protocol.

This proves Lemma 1. ■

C. Joint Signature of Knowledge for TEET

In this section, we present a signature of knowledge scheme for TEET that requires two parties to contribute (Abbreviated as JSoK-TEET, Joint Signature of Knowledge for TEET). With this new primitive, we can adapt the properties of TEET into a cbAC system introduced in [25].

The main idea is as follows. Since the secret value uid (denoted as m in TEET) is only stored in the card, the holder (only has $Q = h_1^{\mathsf{uid}}$) must cooperate with the card to produce a valid zero-knowledge proof. Based on this observation, we modify Σ_{valid} by splitting and distributing the functionality of P to the card and its holder, as well as using the Fiat–Shamir heuristic [32] to get a non-interactive zero-knowledge proof. To reduce the computational burden on the card and protect the privacy of uid and Q, we directly use $Q = h_1^m$ as the input of the $\mathsf{TEET.Encrypt}$ algorithm. Now the Holder can independently generate TEET ciphertext and collaborate with the Card to generate a valid non-interactive zero-knowledge proof. This modification does not affect the security of the TEET algorithm, as the original algorithm also calculates the ciphertext based on $Q = h_1^m$ after inputting m. And Q is a secret value shared by the Holder and Card. The new protocol is shown in Figure 1.

P chooses random numbers r and v, and calculates the TEET encryption of $Q = h_1^m$ (X, Y, U, K) for initiation. To prove that the ciphertext was constructed correctly, the P calculates D and sends pk, K, D, r to Card. Then the Card calculates A, B, C and the challenge e accordingly. Besides, to avoid being exploited by the adversaries, Card additionally selects n_C' and generates r' with its PRF key K to calculate z_1', a masked value of z_1. Finally, P generates the full proof and sends it together with pk, n_C' and the ciphertext (X, Y, U, K) to V.

III. Problem Definition

A. System Model

As shown in Figure 2, Kalos involves four entities: subjects, issuers, medical institutes, and auditors.

1) **Subjects:** Subjects are entities (Composed of credential holder $\mathcal{H}$ and card $\mathcal{C}$) who obtain identities along with attributes from the issuer and authenticate themselves to the medical institutes.
2) **Issuers:** Denoted as $\mathcal{I}$. By interacting with holders, the issuer confirms their legal identification (i.e., ID cards or passports) and signs the attributes so that the holder can jointly prove the possession of a credential with the help of the card.
3) **Medical institutes:** Denoted as $\mathcal{M}$. Medical institutes look forward to recruiting clinical trial volunteers to meet certain criteria. Therefore, when verifying the identity of a subject, the organization will first visually compare the subject to the image of the person on the card. Then $\mathcal{M}$ will interact with $\mathcal{H}$ and $\mathcal{C}$ to determine whether the subject fulfills the required attributes. The authentication information will be used for de-duplication or stored for subsequent identity tracing. It can be instantiated as set of medical electronics.
4) **Auditors:** Denoted as $\mathcal{AU}$. Auditors are parties who are involved in cases when law enforcement or other authorized entities require the ability to extract the real identity of a malicious subject.
5) **Blockchain:** Here, we model the blockchain as a publicly accessible, globally consistent, and tamper-resistant bulletin board [33], [34]. All the public parameters, such as the bilinear group, public key of $\mathcal{I}$ or $\mathcal{AU}$, etc. We omit the statements related to blockchain operations for simplicity.

Card (m)		Holder $(r, v, r_1, ct = (X, Y, U, K))$		$(n_{\mathcal{V}})$ Verifier
$a, b \leftarrow\$ \mathbb{Z}_p^2$		$c, d \leftarrow\$ \mathbb{Z}_p^2$		
$A \leftarrow pk^a$	$\xleftarrow{pk, K, D, r, n_{\mathcal{V}}}$	$D \leftarrow Y^c h_0^d$		
$B \leftarrow h_0^a h_1^b$				
$C \leftarrow K^b$				
$n'_{\mathcal{C}} \leftarrow \{0,1\}^{\lambda}$		$r' = \mathsf{PRF}_K(n'_{\mathcal{C}})$		
$r' = \mathsf{PRF}_K(n'_{\mathcal{C}})$		$e = e' \oplus r'$		calculate
$e = \mathsf{H}(A \Vert B \Vert C \Vert D \Vert n'_{\mathcal{C}} \Vert n_{\mathcal{V}})$	$\xrightarrow{e', z'_r, z'_m, n'_{\mathcal{C}}}$	$z_r = z'_r \oplus r'$		$A' = pk^{z_r} / X^e$
$e' = e \oplus r'$		$z_m = z'_m \oplus r'$		$B' = h_0^{z_r} h_1^{z_m} / Y^e$
$z'_r = (a + e \cdot r) \oplus r'$		$z_v = c + e \cdot v$	$\xrightarrow{\pi, n'_{\mathcal{C}}, pk, ct}$	$C' = K^{z_m} / U^e$
$z'_m = (b + e \cdot m) \oplus r'$		$z_{r_1} = d + e \cdot r_1$		$D' = Y^{z_v} h_0^{z_{r_1}} / U^e$
		$\pi = (e, z_r, z_m, z_v, z_{r_1})$		$e' = \mathsf{H}(A' \Vert B' \Vert C' \Vert D' \Vert n'_{\mathcal{C}} \Vert n_{\mathcal{V}})$
				check if
				$e = e'$

Fig. 1. Cooperative version of Σ_{valid}.

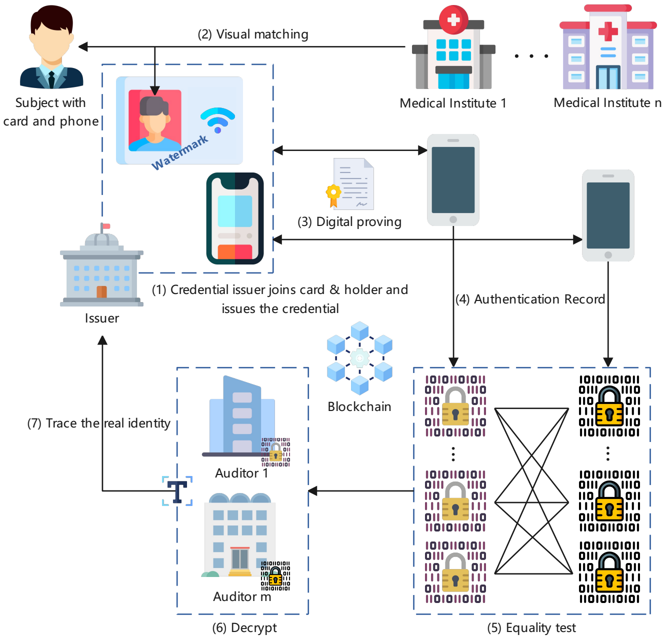

Fig. 2. Overview of the proposed authentication scheme.

B. Clinical Trials

The clinical trial is "the most definitive tool for evaluation of the applicability of clinical research". Following the guidance [35] for developing the clinical trials protocol, there are several critical steps such as "Background Study", "Objectives", "Design of the study", etc. Defining the study population in the protocol is an integral part of posing the primary question. In reporting the study, the investigator must clarify what population was studied and how they were selected. To ensure the meaningfulness of clinical trials, the study population and their selection criteria (typically including Inclusion Criteria and Exclusion Criteria) must be clearly outlined in the reporting of trial results. This helps delineate the boundaries of the trial's applicability for other researchers and enables them to reproduce and evaluate the clinical trial. Statistics show that over the past 20 years, the average number of recruitment criteria for clinical trials has remained around 30, with minimal variation [3].

In Kalos, we only consider "Inclusion Criteria" for simplicity, because detailed examinations are always needed to confirm "Exclusion Criteria". The attributes of a subject will be split into static attributes (i.e., gender, and blood type) and dynamic attributes (i.e., vaccinated with 2-3 doses of COVID-19 inactivated vaccine, or infected with COVID-19). A subject can get static attributes in government departments and dynamic attributes in healthcare organizations. To answer specific clinical trial recruitment, a subject needs to prove that some attributes it possesses satisfy the requirements of $\mathcal{M}$ while keeping other attributes secret.

C. Design Goals

We present the design goals of our authentication system:

1) **Unforgeability:** A $\mathcal{PPT}$ adversary $\mathcal{A}$ without the legal authentication on public/private attributes cannot forge a credential to convince medical institutes that it is valid, i.e., it cannot forge a credential to pass the verification (see the $VerifyCred$ phase mentioned below).
2) **Blindness:** The credential issuer learns nothing about the subjects' private attributes (e.g., their static attributes) except that these attributes satisfy the requiring statement C. Blindness implies the minimal disclosure of subject credentials for satisfying the statement (see the $IssueCred$ phase mentioned below).
3) **Privacy Preservation:** The medical institute learns nothing about the subject's attributes except those that were requested (see the $ShowCred$ phase mentioned below).
4) **Human Bidning:** This property means the authentication capability is binding to the holder and cannot be delegated to other persons with different physical appearances than the holder.
5) **Hierarchical Auditability:** This property means that the medical institutes and the auditors can audit the

authentication record of the subject to different extents. Specifically, medical institutes can link different authentication records belonging to the same subjects, while the auditor can trace the real identity with the issuer's help.

D. High Level Description

Kalos comprises a set of $\mathcal{PPT}$ algorithms, including $\{Setup$, $KeyGen$, $IssueCred$, $ShowCred$, $VerifyCred$, $Compare$, $Trace\}$.

1) **Setup:** Input a security parameter λ, it outputs a set of system parameters.
2) **KeyGen:** Input the system parameters $para$, it invokes xx sub algorithms UserKeyGen, IPKeyGen, AuditorKeyGen to generate the corresponding key pairs for subjects, issuers, and medical institutes.
3) **IssueCred:** Input the attribute set $\mathbf{m} = (m_1, \ldots, m_r)$ and relevant proofs, the issuer checks if they are valid, then signs a credential cred.
4) **ShowCred:** A subject blinds the credential cred, generates a claim C, a TEET ciphertext ct, and proofs that they are valid and satisfy the required attributes.
5) **VerifyCred:** The medical institute verifies whether the subject's proof is valid. If so, store the verification record and conduct the trial accordingly.
6) **Test:** To find out if overlapping participants exist in some clinical trials, the medical institutes can conduct equality tests between the records. Even the records from different medical institutes or encrypted by different auditor's public keys can be tested normally.
7) **Trace:** It is an interactive protocol between the medical institutes and the auditors. The auditors can trace the real identity of the participant by decrypting the record.

IV. Construction

A. Detailed Construction

In this section, we first present the main construction of Kalos. Then we describe the signature of knowledge in detail. The main parameters in Kalos are listed in Table I.

1) **Setup:** Given a security parameter λ as an input, it generates the bilinear group $\mathcal{G} = (\mathbb{G}_1, g, \mathbb{G}_2, \tilde{g}, \mathbb{G}_T, q, \boldsymbol{e})$. Let $\mathsf{H} : \{0,1\}^* \to \mathbb{Z}_p$ be a hash function. $\mathcal{I}$ chooses generators $(h_0, \ldots, h_\ell) \in \mathbb{G}_1^{\ell+1}$, where the number of attributes a credential will preserve is $\ell - 1$.
2) **KeyGen:**
   a) SubjectKeyGen: The auditor $\mathcal{AU}$ accesses to a trusted card issuer $\mathcal{F}_{\mathsf{cardAuth}}$. Then the card $\mathcal{C}$ (with holder's visual information printed on it) will receive a unique identifier uid and a PRF key K, and the holder $\mathcal{H}$ will receive $(\mathcal{C}, Q = h_1^{\mathsf{uid}}, K)$. Then $\mathcal{AU}$ stores Q and the subject's information together in list $\mathcal{L}_{\mathcal{S}}$.
   b) IssuerKeyGen: The issuer $\mathcal{I}$ randomly selects $x \leftarrow_{\$} \mathbb{Z}_p$, and calculates $\omega = \tilde{g}^x$. The secret key is x and the public key is ω.
   c) AuditorKeyGen: Every $\mathcal{AU}$ randomly selects $sk \leftarrow_{\$} \mathbb{Z}_p$, and calculates $pk = h_0^{sk}$. The secret key is sk and the public key is pk.

TABLE I
Definition of parameters

Field	Description
λ	Security parameter
$\mathbb{G}_1, \mathbb{G}_2$	Additive cyclic groups of order p (λ-bit prime)
$\mathbb{G}_T$	Multiplicative cyclic groups of order p
$\boldsymbol{e}$	Bilinear Pairing
$g, h_0, \ldots, h_\ell$	Generators of $\mathbb{G}_1$
$\tilde{g}$	Generators of $\mathbb{G}_2$
K	PRF key hold by the card and its holder
uid	Secret attribute preserved by the card
Q	Secret commitment hold by the holder ($Q = h_1^{\mathsf{uid}}$)
x, ω	Key pair of the issuer
sk, pk	Key pair of the auditor
ct	TEET ciphertext
$\mathbf{a}_H, \mathbf{a}_I$	Attribute list used in credential issuance
$\boldsymbol{a}_H, \boldsymbol{a}_M$	Attribute list used in credential presentation
$n_{\mathcal{C}}, n_{\mathcal{I}}, n_{\mathcal{H}}, n_{\mathcal{V}}$	Nonce to prevent message replay
$\mathbf{m}$	Set of the $\ell - 1$ attributes a credential preserves
$\sigma = (A, e, s)$	Content of BBS+ signature
$\mathsf{cred} = (\mathcal{C}, Q, K, \sigma, \mathbf{m})$	Content of a credential
$\mathcal{L}_{\mathcal{S}}$	List of the "tag–information" of the subjects
$\mathcal{L}_{\mathcal{C}}$	List of the credentials belonging to a subject
$\mathcal{L}_{\mathcal{R}}$	List of the verification records
$\pi_1, \pi_2, \pi_3, \pi_4, \pi_5$	Noninteractive zero-knowledge proofs

3) **IssueCred:** A subject will receive the credential with the attributes $\mathbf{a}_H = \{(i, m_i) : i \in H\}$ (contribute by Holder) and $\mathbf{a}_I = \{(i, m_i) : i \in I\}$ (contribute by Issuer) for $i \in [1, \ell], m_i \in \mathbb{Z}_p^*$ through the following steps. The outline of this process is shown in Figure 3.
   a) $\mathcal{I}$ randomly selects a number $n_{\mathcal{I}} \leftarrow \{0,1\}^\lambda$, then sends it to $\mathcal{C}$.
   b) $\mathcal{C}$ randomly selects $n_{\mathcal{C}} \leftarrow \{0,1\}^\lambda$, calculates $r = \mathsf{PRF}_K(n_{\mathcal{C}})$ with the shared pseudo-random function key K, and $B = h_1^{\mathsf{uid}} h_0^r$. Then $\mathcal{C}$ sends $n_{\mathcal{C}}$ and B to $\mathcal{I}$, along with a signature of knowledge $\pi_1 = \mathsf{SoK}\{(\mathsf{uid}, r) : B = h_1^{\mathsf{uid}} h_0^r\}(n_{\mathcal{I}})$.
   c) $\mathcal{I}$ returns $\perp$ if π_1 is not valid. Otherwise, sends $n_{\mathcal{I}}$ and $n_{\mathcal{C}}$ to $\mathcal{H}$.
   d) $\mathcal{H}$ randomly selects $s' \leftarrow \mathbb{Z}_p$, calculates $C = h_0^{s'} \prod_{i \in H} h_{i+1}^{m_i}$, and $r = \mathsf{PRF}_F(n_{\mathcal{C}})$ with the shared pseudo-random function key K. Then $\mathcal{H}$ sends C and H to $\mathcal{I}$, along with a signature of knowledge $\pi_2 = \mathsf{SoK}\{(s', \{m_i\}_{i \in H}) : C = h_0^{s'} \prod_{i \in H} h_{i+1}^{m_i}\}(n_{\mathcal{I}})$.
   e) $\mathcal{I}$ returns $\perp$ if π_2 is not valid. Otherwise, randomly selects $e \leftarrow \mathbb{Z}_p \backslash \{x\}, \tilde{s} \leftarrow \mathbb{Z}_p$, calculates $A = (g \cdot h_0^s \cdot B \cdot C \cdot \prod_{i \in I} h_{i+1}^{m_i})^{1/(e+x)}$. Then sends $\mathbf{a}_I, A, e, \tilde{s}$ to $\mathcal{H}$.
   f) $\mathcal{H}$ sets $\mathbf{m} = \mathbf{a}_H \oplus \mathbf{a}_I = (m_2, \ldots, m_\ell)$. Aborts if $\boldsymbol{e}(A, w\tilde{g}^e) \neq \boldsymbol{e}(g \cdot h_0^{\tilde{s}+s'+r} \cdot Q \cdot \prod_{i=2}^{\ell} h_{i+1}^{m_i}, \tilde{g})$. Then stores $\mathsf{cred} = (\mathcal{C}, Q, K, \sigma = (A, e, \tilde{s} + s' + r), \mathbf{m})$ to a list of credentials $\mathcal{L}_{\mathcal{C}}$.
4) **ShowCred & VerifyCred:** In our settings, this algorithm will achieve two goals: 1) prove the possession of the required attributes $\boldsymbol{a}_M$; and 2) generate valid audit information. The outline of this process is shown in Figure 4.
   a) $\mathcal{M}$ visually verifying that the picture on the card matches the individual provides the missing link in the verification chain.
   b) $\mathcal{H}$ randomly selects $n_{\mathcal{H}} \leftarrow \{0,1\}^\lambda$ and sends it to $\mathcal{M}$. Then $\mathcal{M}$ sends $n_{\mathcal{H}}, n_{\mathcal{M}}$ to $\mathcal{C}$, where $n_{\mathcal{M}} \leftarrow \{0,1\}^\lambda$.

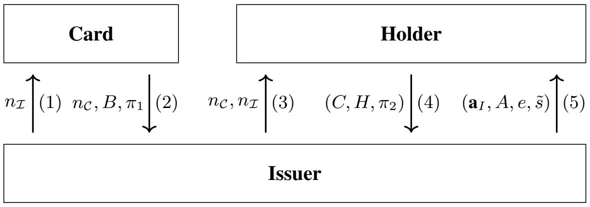
Card
Holder
$n_{\mathcal{I}}$ (1)
$n_{\mathcal{C}}, B, \pi_1$ (2)
$n_{\mathcal{C}}, n_{\mathcal{I}}$ (3)
(C, H, π_2) (4)
$(\mathbf{a}_I, A, e, \tilde{s})$ (5)
Issuer

Fig. 3. Outline of credential issuance between Card, Holder, and Issuer. Skeleton borrowed from [25].

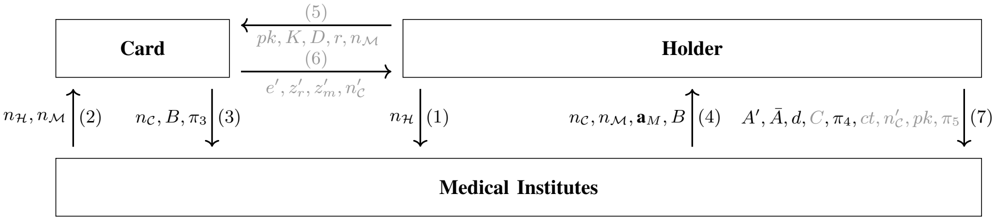
(5)
$pk, K, D, r, n_{\mathcal{M}}$
(6)
$e', z_r', z_m', n_{\mathcal{C}}'$
Card
Holder
$n_{\mathcal{H}}, n_{\mathcal{M}}$ (2)
$n_{\mathcal{C}}, B, \pi_3$ (3)
$n_{\mathcal{H}}$ (1)
$n_{\mathcal{C}}, n_{\mathcal{M}}, \mathbf{a}_M, B$ (4)
$A', \bar{A}, d, C, \pi_4, ct, n_{\mathcal{C}}', pk, \pi_5$ (7)
Medical Institutes

Fig. 4. Outline of credential presentation between Card, Holder, and Medical Institutes. Skeleton borrowed from [25]. The gray text denotes the messages we used to generate TEET ciphertext and the corresponding noninteractive zero-knowledge proof.

c) $\mathcal{C}$ randomly selects $n_{\mathcal{C}} \leftarrow \{0,1\}^{\lambda}$, calculates $n = n_{\mathcal{C}} \| n_{\mathcal{H}}$, $r = \mathsf{PRF}_K(n)$, $B = h_1^{\mathsf{uid}} h_0^r$. Then $\mathcal{C}$ sends $n_{\mathcal{C}}$ and B to $\mathcal{M}$, along with a signature of knowledge $\pi_3 = \mathsf{SoK}\{(\mathsf{uid}, r) : B = h_1^{\mathsf{uid}} h_0^r\}(n_{\mathcal{M}})$.

d) $\mathcal{M}$ returns $\perp$ if π_3 is not valid. Otherwise, sends $n_{\mathcal{C}}, n_{\mathcal{M}}, \boldsymbol{a}_M$, B to $\mathcal{H}$.

e) $\mathcal{H}$ calculates $n = n_{\mathcal{C}} \| n_{\mathcal{H}}$, parses $\boldsymbol{a}_M = \{(i, m_i) : i \in M\}$. Then determines the element $\mathsf{cred} = (\mathcal{C}, Q, K, \sigma, \mathbf{m})$ in $\mathcal{L}_{\mathcal{C}}$ such that: $r = \mathsf{PRF}_K(n)$, $B = Q \cdot h_0^r$ and $\boldsymbol{m}[i] = m_i$ for all $i \in M$. Returns $\perp$ if no record is found. Otherwise, defines $m_{i+1}' = m_i$ for $i \in [\ell]$ and thus $\boldsymbol{a}_M = \{(i, m_{i+1}') : i \in M\}$. Parses σ as (A, e, s). Sets $r = \mathsf{PRF}_K(n)$, $\boldsymbol{a}_H$ as the hiding attributes ($H = \{2, \ldots, \ell\} \backslash M$). Randomly selects $r_1 \leftarrow \mathbb{Z}_p^*$, $r_2, \overline{r} \leftarrow \mathbb{Z}_p^2$, and sets $r_3 = r_1^{-1}$, $s' = s - r_2 r_3 - r$, $\overline{s} = s' - \overline{r}$. Calculates $A' = A^{r_1}$, $b = g_1 h_0^s Q \prod_{i=2}^{\ell} h_i^{m_i'}$, $\bar{A} = A'^{-e} b^{r_1}$, $d = b^{r_1} h_0^{-r_2}$, $C = h_0^{\overline{r}} \prod_{i \in H} h_i^{m_i'}$ and $\pi_4 \leftarrow \mathsf{SoK}\{(e, \overline{s}, r_2, r_3, \{m_i'\}_{i \in H}) : A'^{-e} h_0^{r_2} = \bar{A}/d \wedge d^{-r_3} h_0^{s'} C = g_1^{-1} B^{-1} \prod_{i \in M} h_i^{-m_i'}\}(n_{\mathcal{M}})$.

f) $\mathcal{H}$ randomly selects $v \leftarrow_{\$} \mathbb{Z}_p^2$, sets $Y = B$, calculates $X = pk^r, U = Q^v, K = h_1^v$, and sets $ct = (X, Y, U, K)$ as the audit information. Then $\mathcal{H}$ interacts with $\mathcal{C}$ to generate a signature of knowledge $\pi_5 \leftarrow \mathsf{SoK}\{(r, \mathsf{uid}, v, r_1 = -rv) : X = pk^r \wedge Y = h_0^r h_1^{\mathsf{uid}} \wedge U = K^{\mathsf{uid}} \wedge U = Y^v h_0^{r_1}\}(n_{\mathcal{M}})$. Then sends $(A', \bar{A}, d, C, \pi_4, ct, n_{\mathcal{C}}', pk, \pi_5)$ to $\mathcal{M}$.

g) $\mathcal{M}$ returns $\perp$ if π_4, π_5 are not valid or $\boldsymbol{e}(\bar{A}, \omega) \neq \boldsymbol{e}(A', \tilde{g})$. Otherwise, stores $(A', \bar{A}, d, \pi_4, ct, n_{\mathcal{C}}', pk, \pi_5)$ to a list of verification records $\mathcal{L}_{\mathcal{R}}$.

5) **Test:** Given a verification record R, iterate every element $\mathsf{R}_i \in \mathcal{L}_{\mathcal{R}}$. Check if the equation $\boldsymbol{e}(U, K_i) = \boldsymbol{e}(U_i, K)$ holds, where $(U, K) \in \mathsf{R}$ and $(U_i, K_i) \in \mathsf{R}_i$. If so, it indicates that these two records belong to the same subject.

6) **Trace:** After checking the authentication record provided by $\mathcal{M}$s and being convinced, $\mathcal{AU}$ can obtain $Q = h_1^{\mathsf{uid}}$ by partially decrypting the audit information. Then the real identity of the subject can be found from $\mathcal{L}_{\mathcal{S}}$ with the help of $\mathcal{I}$.

B. Details of SoK

1) $\pi_1 = \mathsf{SoK}\{(\mathsf{uid}, r) : h_1^{\mathsf{uid}} h_0^r = B\}(n_{\mathcal{I}})$.

a) $\mathcal{C}$ randomly selects $r_{\mathsf{uid}}, r_r \leftarrow_{\$} \mathbb{Z}_p^2$, calculates $\overline{B} = h_1^{r_{\mathsf{uid}}} h_0^{r_r}$, $e = \mathsf{H}(B \| \overline{B} \| n_{\mathcal{I}})$, $z_{\mathsf{uid}} = r_{\mathsf{uid}} + e \cdot \mathsf{uid}$, $z_r = r_r + e \cdot r$, and sends $(e, z_{\mathsf{uid}}, z_r)$.

b) The proof will be accepted if $e = \mathsf{H}(B \| h_1^{z_{\mathsf{uid}}} h_0^{z_r} / B^e \| n_{\mathcal{I}})$ holds.

2) $\pi_2 = \mathsf{SoK}\{(s', \{m_i\}_{i \in H}) : C = h_0^{s'} \prod_{i \in H} h_{i+1}^{m_i}\}(n_{\mathcal{I}})$.

a) $\mathcal{H}$ randomly selects $r_{s'}, \{r_{m_i}\}_{i \in H} \leftarrow_{\$} \mathbb{Z}_p^{|H|+1}$, calculates $\overline{C} = h_0^{r_{s'}} \prod_{i \in H} h_{i+1}^{r_{m_i}}$, $e = \mathsf{H}(C \| \overline{C} \| n_{\mathcal{I}})$, $z_{s'} = r_{s'} + e \cdot s'$, $\{z_{m_i} = r_{m_i} + e \cdot m_i\}_{i \in H}$, and sends $(e, z_{s'}, \{z_{m_i}\})$.

b) The the proof will be accepted if $e = \mathsf{H}(C \| h_0^{z_{s'}} \prod_{i \in H} h_{i+1}^{z_{m_i}} / C^e \| n_{\mathcal{I}})$ holds.

3) The proof process of π_3 is consistent with π_1.

4) $\pi_4 \leftarrow \mathsf{SoK}\{(e, \overline{s}, r_2, r_3, \{m_i'\}_{i \in H}) : A'^{-e} h_0^{r_2} = \bar{A}/d \wedge d^{-r_3} h_0^{s'} C = g_1^{-1} B^{-1} \prod_{i \in M} h_i^{-m_i'}\}(n_{\mathcal{M}})$.

a) $\mathcal{H}$ randomly selects $r_e, r_{\bar{s}}, r_{r_2}, r_{r_3} \leftarrow_{\$} \mathbb{Z}_p^4$, calculates $\Theta \leftarrow A'^{-r_e} h_0^{r_{r_2}}$, $\Xi \leftarrow d^{-r_{r_3}} h_0^{r_{\bar{s}}}$, $e = \mathsf{H}(\Theta \| \Xi \| n_{\mathcal{I}})$, $z_e = e + c \cdot e$, $z_{\bar{s}} = r_{\bar{s}} + c \cdot \bar{s}$, $z_{r_2} = r_{r_2} + c \cdot r_2$, $z_{r_3} = r_{r_3} + c \cdot r_3$, and sends $(e, z_e, z_{\bar{s}}, z_{r_2}, z_{r_3})$.

b) Accept the proof if $e = \mathsf{H}(A'^{-z_e} h_0^{z_{r_2}} / (\frac{\bar{A}}{d})^c \| d^{-z_{r_3}} h_0^{z_{\bar{s}}} C^e / (g_1^{-1} B^{-1} \prod_{i \in V} h_i^{-m_i'})^c \| n_{\mathcal{I}})$ holds.

5) $\pi_5 \leftarrow \mathsf{SoK}\{(r, \mathsf{uid}, v, r_1 = -rv) : X = pk^r \wedge Y = h_0^r h_1^{\mathsf{uid}} \wedge U = K^{\mathsf{uid}} \wedge U = Y^v h_0^{r_1}\}(n_{\mathcal{M}})$.

a) $\mathcal{H}$ randomly selects $c, d \leftarrow_{\$} \mathbb{Z}_p^2$, calculates $D = Y^c h_0^d$. Then sends $pk, K, D, r, n_{\mathcal{M}}$ to $\mathcal{C}$.

b) $\mathcal{C}$ randomly selects $a, b \leftarrow_{\$} \mathbb{Z}_p^2$ and $n_{\mathcal{C}}' \leftarrow \{0,1\}^{\lambda}$, calculates $A \leftarrow pk^a$, $\overline{B} \leftarrow h_0^a h_1^b$, $C \leftarrow K^b$, $r' = \mathsf{PRF}_K(n_{\mathcal{C}}')$, $e = \mathsf{H}(A \| \overline{B} \| C \| D \| n_{\mathcal{C}}' \| n_{\mathcal{M}})$, $e' = e \oplus r'$, $z_r' = (a + e \cdot r) \oplus r'$, $z_{\mathsf{uid}}' = (b + e \cdot \mathsf{uid}) \oplus r'$. Then returns $(e', z_r', z_{\mathsf{uid}}', n_{\mathcal{C}}')$ to $\mathcal{H}$.

c) $\mathcal{H}$ calculates $r' = \mathsf{PRF}_K(n'_{\mathcal{C}})$, $e = e' \oplus r'$, $z_r = z'_r \oplus r'$, $z_{\mathsf{uid}} = z'_{\mathsf{uid}} \oplus r'$, $z_v = c + e \cdot v$, $z_{r_1} = d + e \cdot r_1$. Then sends $(e, z_r, z_{\mathsf{uid}}, z_v, z_{r_1})$.

d) The proof will be accepted if $e = \mathsf{H}(pk^{z_r}/X^e \| h_0^{z_r} h_1^{z_{\mathsf{uid}}}/Y^e \| K^{z_{\mathsf{uid}}}/U^e \| Y^{z_v} h_0^{z_{r_1}}/U^e \| n'_{\mathcal{C}} \| n_{\mathcal{M}})$ holds where $Y = B = h_1^{\mathsf{uid}} h_0^r$.

V. Security Analysis

Kalos achieves our design goals including unforgeability, blindness, hierarchical auditability, and human-binding. Similar to the security proof in [36], we reduce these properties to the security of BBS+ signature and zero-knowledge proof.

Lemma 2: Kalos is unforgeable if BBS+ signature is secure under the q-SDH assumption (see Section II-A2) and the card is tamper-resistant.

Sketch: When $\mathcal{PPT}$ adversary $\mathcal{A}$ tries to forge a valid credential, there are two cases: 1) it forges a BBS+ signature of the issuer. Doing so contradicts the q-SDH assumption; 2) it modifies the uid in the card and uses the corresponding BBS+ signature. However, this is contradictory to the tamper-resistant property of a card. □

Lemma 3: Kalos is blind if π_2 is a signature of knowledge protocol satisfying zero-knowledge.

Sketch: In $IssueCred$ phase, the subject sends the signature of knowledge $\pi_2 = \mathsf{SoK}\{(s', \{m_i\}_{i \in H}) : C = h_0^{s'} \prod_{i \in H} h_{i+1}^{m_i}\}(n_{\mathcal{I}})$ of his private attributes $\{m_i\}_{i \in H}$. Due to SoK's properties, the zero-knowledge property of π_2 ensures that commitment does not reveal information about these attributes. □

Lemma 4: Kalos is privacy-preserving if π_4 is a signature of knowledge protocol satisfying zero-knowledge.

Sketch: In $ShowCred$ phase, the subject sends the signature of knowledge $\pi_4 \leftarrow \mathsf{SoK}\{(e, \overline{s}, r_2, r_3, \{m_i\}_{i \in H}) : A'^{-e} h_0^{r_2} = \bar{A}/d \wedge d^{-r_3} h_0^{s'} C = g_1^{-1} B^{-1} \prod_{i \in M} h_i^{-m_i}\}(n_{\mathcal{M}})$ to the medical institute. This ensures that all unused attributes are protected. Besides, the construction of commitment $C = h_0^{\overline{r}} \prod_{i \in H} h_i^{m'_i}$ satisfies perfect hiding. So the holder's unused attributes cannot be inferred by $\mathcal{PPT}$ adversary. □

Lemma 5: Kalos is human-binding if π_3 and π_4 are non-interactive SoK protocols satisfying soundness and the card is tamper-resistant.

Sketch: In $ShowCred$ phase, the physical appearance will be visually checked first. Then the subject sends the proof $\pi_3 = \mathsf{SoK}\{(\mathsf{uid}, r) : B = h_1^{\mathsf{uid}} h_0^r\}(n_{\mathcal{M}})$ and $\pi_4 \leftarrow \mathsf{SoK}\{(e, \overline{s}, r_2, r_3, \{m_i\}_{i \in H}) : A'^{-e} h_0^{r_2} = \bar{A}/d \wedge d^{-r_3} h_0^{s'} C = g_1^{-1} B^{-1} \prod_{i \in M} h_i^{-m_i}\}(n_{\mathcal{M}})$ to the medical institute. As the SoK proofs are sound, the holder proves that he is indeed the owner of the card that can provide the missing link in the verification chain. □

Lemma 6: Kalos is hierarchically auditable if π_5 is non-interactive signature of knowledge protocols satisfying soundness.

Sketch: In $Test$ phase, as proof $\pi_5 \leftarrow \mathsf{SoK}\{(r, m_1, v, r_1 = -rv) : X = pk^r \wedge Y = h_0^r h_1^{m_1} \wedge U = K^{m_1} \wedge U = Y^v h_0^{r_1}\}(n_{\mathcal{M}})$. Then sends $(A', \bar{A}, d, C, \pi_4, ct, \pi_5)$ to $\mathcal{M}$ is sound, then we know that ct is indeed the TEET ciphertext of uid. Medical institutes can test the equality of different records while auditors can decrypt them correctly. Then, the subject's real identity can be traced with the issuer's help. □

TABLE II
Execution time of each operation

Notions	Description	Values
T_{cmp}	Compare two Pairings	0.6927ms
T_{mul}	Multiplication operation in $\mathbb{G}$	0.2547ms
T_{add}	Add operation in $\mathbb{G}$	0.014 ms
T_{hash}	Hash operation	0.5262 ms
$\lvert\mathbb{G}_1\rvert$	Bit length of an element in $\mathbb{G}_1$	384 bits
$\lvert\mathbb{G}_2\rvert$	Bit length of an element in $\mathbb{G}_2$	768 bits
$\lvert\mathbb{G}_T\rvert$	Bit length of an element in $\mathbb{G}_T$	576 bits
$\lvert\mathbb{Z}_p\rvert$	Bit length of an element in $\mathbb{Z}_p$	256 bits

TABLE III
Computation Cost of Kalos

Algorithms		Computation cost
KeyGen	$\mathcal{C}\&\mathcal{H}$	0
	$\mathcal{I}$	1 T_{mul}
	$\mathcal{AU}$	1 T_{mul}
Issue	$\mathcal{C}$	4 T_{mul} + 2 T_{add} + 1 T_{hash}
	$\mathcal{H}$	2 T_{cmp} + (3 + 2 $\lvert\mathbf{a}_H\rvert$ + ℓ) $\cdot$ T_{mul} + (2 + 2 $\lvert\mathbf{a}_H\rvert$ + ℓ) $\cdot$ T_{add} + 1 T_{hash}
	$\mathcal{I}$	(6 + ℓ) $\cdot$ T_{mul} + (7 + ℓ) $\cdot$ T_{add} + 2 T_{hash}
Show	$\mathcal{C}$	(4 T_{mul} + 2 T_{add} + 1 T_{hash}) + (4 T_{mul} + 1 T_{add} + 1 T_{hash})
	$\mathcal{H}$	(10 + 5 + $\lvert\boldsymbol{a}_H\rvert$ + ℓ) $\cdot$ T_{mul} + (4 + 1 + $\lvert\boldsymbol{a}_H\rvert$ + ℓ) $\cdot$ T_{add} + 1 T_{hash}
	$\mathcal{M}$	1 T_{cmp} + (10 + 10 + $\lvert\boldsymbol{a}_M\rvert$) $\cdot$ T_{mul} + (14 + 6 + $\lvert\boldsymbol{a}_M\rvert$) $\cdot$ T_{add} + (1 + 1) $\cdot$ T_{hash}
Test	$\mathcal{M}$	$\lvert\mathcal{L}_{\mathcal{R}}\rvert$ T_{cmp}
Trace	$\mathcal{AU}$	1 T_{mul} + 1 T_{add}

VI. Performance Analysis

In this section, we conduct performance tests on the proposed construction on a personal computer (Dell with an i5-9600K CPU, 16GB RAM, and Windows 10 x86_64 operating system). We use a bilinear pairing $\boldsymbol{e} : \mathbb{G}_1 \times \mathbb{G}_2 \rightarrow \mathbb{G}_T$ with the 256-bit order p to achieve 128 bits security level. The testing program is written in Golang and developed based on the bbs[1] and mathlib[2] libraries.

The execution time of each operation is given in Table II. We analyze the efficiency of Kalos by counting different operations, i.e., T_{cmp}, T_{mul}, T_{add}, and T_{hash}. We also give the computation cost analysis of Kalos in Table III. In $Issue$ phase h and i refer to the number of attributes held by H and I. In $Show$ phase h and m refer to the unused number of attributes held by H and the number of attributes required by I. We ignore the computation cost related to $\mathcal{F}_{\mathsf{cardAuth}}$, so the total computation cost in $SubjectKeyGen$ phase is 0. We use gray text to mark the additional cost introduced by our scheme in Table III and IV compared with [25].

We also test the time consumption of different processes during the credential issuance and presentation process. The overall computational overhead of the protocol is in milliseconds. In the credential issuance phase (shown in Figure 6), we set $h = i$, and the computational overhead increases with the number of attributes. In the credential presentation phase (shown in Figure 5), we vary the ratio of the number

[1] https://github.com/trustbloc/bbs-signature-go/tree/main

[2] https://github.com/IBM/mathlib

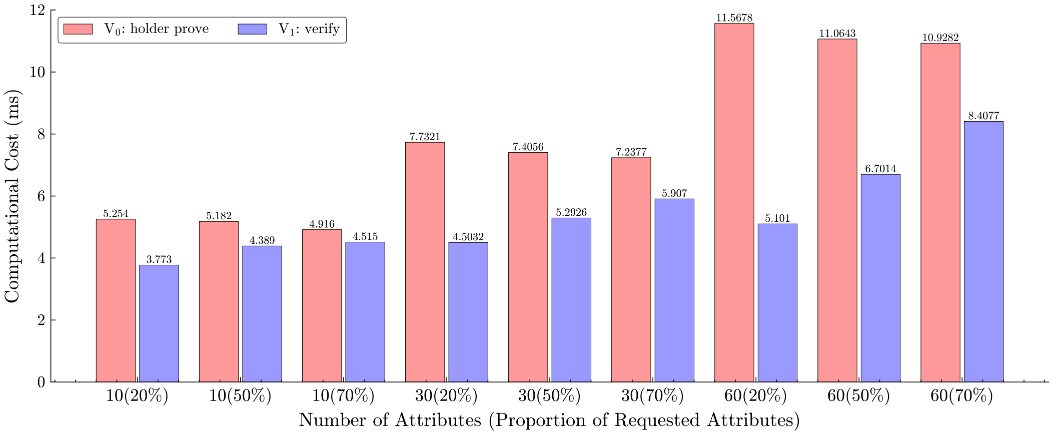

Fig. 5. Computational cost for different operations during the credential presentation. Notion V$_0$ and V$_1$ denote steps e)+f), d)+g) in $ShowCred\&VerifyCred$ phase respectively. We omitted step c) and the overhead of $\mathcal{C}$ in step f) here because the cost is static (0.558ms and 0.547ms on average).

of requested attributes m (20%, 50%, and 70%). When the total number of attributes (ℓ) is fixed, the computational overhead for proof generation by the holder decreases as the number of requested attributes increases, while the verification overhead increases. The entries with a value of 60 (50%) in Figure 5 can be considered a benchmark. When 30 attributes are requested, the holder and the verifier need 11.06ms and 6.7ms respectively, and the computation cost of the holder will increase with larger ℓ. We omit the computational consumption related to the card in both figures because they were fixed (between $0.5 \sim 0.6$ ms).

For the communication cost of Kalos, we show it in Table IV. The communication overhead is mainly concentrated on the phases related to the transfer of attributes, while the costs for other phases are fixed.

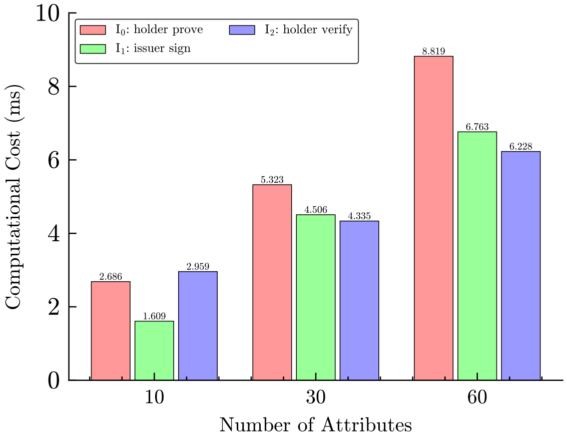

Fig. 6. Computational cost for different operations during the credential issuance. Notion I$_0$, I$_1$, and I$_2$ denote steps d), c)+e), f) in $IssueCred$ phase respectively. We omitted step b) here because the cost is static (0.558ms on average).

TABLE IV
COMMUNICATION COST OF KALOS

Algorithms		Communication cost	Specific values
Issue	$\mathcal{I} \rightarrow \mathcal{C}$	1 $\lvert\mathbb{Z}_p\rvert$	256 bits
	$\mathcal{C} \rightarrow \mathcal{I}$	4 $\lvert\mathbb{Z}_p\rvert$ + 1 $\lvert\mathbb{G}_1\rvert$	1408 bits
	$\mathcal{I} \rightarrow \mathcal{H}$	2 $\lvert\mathbb{Z}_p\rvert$	512 bits
	$\mathcal{H} \rightarrow \mathcal{I}$	(2 + $\lvert\mathbf{a}_H\rvert$) $\cdot$ $\lvert\mathbb{Z}_p\rvert$ + $\lvert\mathbb{G}_1\rvert$ + $\lvert\mathbf{a}_H\rvert \cdot \lvert int\rvert$	896 + 320$\cdot\lvert\mathbf{a}_H\rvert$ bits ($\lvert int\rvert$ = 64 bits)
	$\mathcal{I} \rightarrow \mathcal{H}$	(2 + $\lvert\mathbf{a}_I\rvert$) $\cdot$ $\lvert\mathbb{Z}_p\rvert$ + $\lvert\mathbb{G}_1\rvert$	896 + 256$\cdot\lvert\mathbf{a}_I\rvert$ bits
Show	$\mathcal{H} \rightarrow \mathcal{M}$	1 $\lvert\mathbb{Z}_p\rvert$	256 bits
	$\mathcal{M} \rightarrow \mathcal{C}$	2 $\lvert\mathbb{Z}_p\rvert$	512 bits
	$\mathcal{C} \rightarrow \mathcal{M}$	4 $\lvert\mathbb{Z}_p\rvert$ + $\lvert\mathbb{G}_1\rvert$	1408 bits
	$\mathcal{M} \rightarrow \mathcal{H}$	(2 + $\boldsymbol{a}_M$) $\cdot$ $\lvert\mathbb{Z}_p\rvert$ + $\lvert\mathbb{G}_1\rvert$	896 + 256$\cdot\lvert\boldsymbol{a}_M\rvert$ bits
	$\mathcal{H} \rightarrow \mathcal{C}$	2 $\lvert\mathbb{Z}_p\rvert$ + 3 $\lvert\mathbb{G}_1\rvert$	1664 bits
	$\mathcal{C} \rightarrow \mathcal{H}$	4 $\lvert\mathbb{Z}_p\rvert$	1024 bits
	$\mathcal{H} \rightarrow \mathcal{M}$	(5 + 6) $\cdot$ $\lvert\mathbb{Z}_p\rvert$ + (3 + 5) $\cdot$ $\lvert\mathbb{G}_1\rvert$	5888 bits

VII. Conclusion

In this work, we proposed Kalos, a privacy-preserving authentication scheme with hierarchical auditability and human binding for clinical trials. Our first key contribution was to present a public key encryption scheme (TEET) to realize verifiable encryption and equality test. We then combine TEET with a card-based anonymous credential system to protect the participants' privacy and supervise their potential malicious behavior in clinical trials. The security and performance analysis ensures reasonable security assurance and cost. In future work, we will focus on issues such as adapting to the W3C Decentralized Identifiers standard (DIDs) [37] and considering credential revocation to improve the usability of Kalos.

References

[1] E. D. Namiot, D. Smirnovová, A. V. Sokolov, V. N. Chubarev, V. V. Tarasov, and H. B. Schiöth, "The international clinical trials registry platform (ictrp): data integrity and the trends in clinical trials, diseases, and drugs," *Frontiers in Pharmacology*, vol. 14, p. 1228148, 2023.

[2] C. Leuker, L. Samartzidis, R. Hertwig, and T. J. Pleskac, "When money talks: Judging risk and coercion in high-paying clinical trials," *PloS one*, vol. 15, no. 1, p. e0227898, 2020.

[3] M. Hutson, "How ai is being used to accelerate clinical trials." *Nature*, vol. 627, no. 8003, pp. S2–S5, 2024.

[4] G. M. Zanini, "A new job: research volunteer?" *Swiss medical weekly*, vol. 135, no. 2122, pp. 315–315, 2005.

[5] K. El Emam, H. Farah, S. Samet, A. Essex, E. Jonker, M. Kantarcioglu, and C. C. Earle, "A privacy preserving protocol for tracking participants in phase i clinical trials," *Journal of Biomedical Informatics*, vol. 57, pp. 145–162, 2015.

[6] A. Garcia, J. Lee, V. Balasubramanian, R. Gardner, S. E. Gummidipundi, G. Hung, T. Ferris, L. Cheung, S. Desai, C. B. Granger, *et al.*, "The development of a mobile app-focused deduplication strategy for the apple heart study that informs recommendations for future digital trials," *Stat*, vol. 11, no. 1, p. e470, 2022.

[7] H. Köpcke, A. Thor, and E. Rahm, "Evaluation of entity resolution approaches on real-world match problems," *Proceedings of the VLDB Endowment*, vol. 3, no. 1-2, pp. 484–493, 2010.

[8] O. Benjelloun, H. Garcia-Molina, D. Menestrina, Q. Su, S. E. Whang, and J. Widom, "Swoosh: a generic approach to entity resolution," *The VLDB Journal*, vol. 18, pp. 255–276, 2009.

[9] D. Boyar and N. M. Goldfarb, "Preventing overlapping enrollment in clinical studies," *Journal of Clinical Research Best Practices*, vol. 6, no. 4, pp. 1–4, 2010.

[10] M. M. Mello, V. Lieou, and S. N. Goodman, "Clinical trial participants' views of the risks and benefits of data sharing," *New England journal of medicine*, vol. 378, no. 23, pp. 2202–2211, 2018.

[11] M. F. Ayub, X. Li, K. Mahmood, S. Shamshad, M. A. Saleem, and M. Omar, "Secure consumer-centric demand response management in resilient smart grid as industry 5.0 application with blockchain-based authentication," *IEEE Transactions on Consumer Electronics*, 2023.

[12] P. Verma, V. Tripathi, and B. Pant, "Secure hashgraph for healthcare: Strengthening privacy and data security in patient records," *IEEE Transactions on Consumer Electronics*, 2024.

[13] D.-S. Kim, S.-Y. Lee, B.-S. Kim, S.-C. Lee, and D.-H. Chung, "On the design of an embedded biometric smart card reader," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 573–577, 2008.

[14] S. Keykhaie and S. Pierre, "Mobile match on card active authentication using touchscreen biometric," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 4, pp. 376–385, 2020.

[15] D. Rui-zhong, W. Yuan, and W. Zi-yuan, "Real-time audit scheme based on multilevel roles in a medical cloud environment," in *2022 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 2022, pp. 1–7.

[16] L. Li, J. Liu, L. Cheng, S. Qiu, W. Wang, X. Zhang, and Z. Zhang, "Creditcoin: A privacy-preserving blockchain-based incentive announcement network for communications of smart vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 7, pp. 2204–2220, 2018.

[17] F. Angeletti, I. Chatzigiannakis, and A. Vitaletti, "Privacy preserving data management in recruiting participants for digital clinical trials," in *Proceedings of the first international workshop on human-centered sensing, networking, and systems*, 2017, pp. 7–12.

[18] J. Yuan, B. Malin, F. Modave, Y. Guo, W. R. Hogan, E. Shenkman, and J. Bian, "Towards a privacy preserving cohort discovery framework for clinical research networks," *Journal of biomedical informatics*, vol. 66, pp. 42–51, 2017.

[19] G. Hripcsak, P. Mirhaji, A. F. Low, and B. A. Malin, "Preserving temporal relations in clinical data while maintaining privacy," *Journal of the American Medical Informatics Association*, vol. 23, no. 6, pp. 1040–1045, 2016.

[20] F. Maritsch, I. Cil, C. McKinnon, J. Potash, N. Baumgartner, V. Philippon, and B. G. Pavlova, "Data privacy protection in scientific publications: process implementation at a pharmaceutical company," *BMC medical ethics*, vol. 23, no. 1, p. 65, 2022.

[21] M. Movahedi, B. M. Case, J. Honaker, A. Knox, L. Li, Y. P. Li, S. Saravanan, S. Sengupta, and E. Taubeneck, "Privacy-preserving randomized controlled trials: A protocol for industry scale deployment," in *Proceedings of the 2021 on Cloud Computing Security Workshop*, 2021, pp. 59–69.

[22] K. Tucker, J. Branson, M. Dilleen, S. Hollis, P. Loughlin, M. J. Nixon, and Z. Williams, "Protecting patient privacy when sharing patient-level data from clinical trials," *BMC medical research methodology*, vol. 16, pp. 5–14, 2016.

[23] M. Chase and A. Lysyanskaya, "On signatures of knowledge," in *Advances in Cryptology-CRYPTO 2006: 26th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 2006. Proceedings 26*. Springer, 2006, pp. 78–96.

[24] X. Zhou, D. He, J. Ning, M. Luo, and X. Huang, "Aadec: Anonymous and auditable distributed access control for edge computing services," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 290–303, 2022.

[25] J. Hesse, N. Singh, and A. Sorniotti, "How to bind anonymous credentials to humans," in *USENIX Security Symposium*, 2023.

[26] D. Boneh and X. Boyen, "Short signatures without random oracles and the sdh assumption in bilinear groups," *Journal of cryptology*, vol. 21, no. 2, pp. 149–177, 2008.

[27] Y. Chen, X. Ma, C. Tang, and M. H. Au, "Pgc: decentralized confidential payment system with auditability," in *Computer Security–ESORICS 2020: 25th European Symposium on Research in Computer Security, ESORICS 2020, Guildford, UK, September 14–18, 2020, Proceedings, Part I 25*. Springer, 2020, pp. 591–610.

[28] S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof-systems," in *Providing sound foundations for cryptography: On the work of shafi goldwasser and silvio micali*, 2019, pp. 203–225.

[29] G. Yang, C. H. Tan, Q. Huang, and D. S. Wong, "Probabilistic public key encryption with equality test," in *Topics in Cryptology-CT-RSA 2010: The Cryptographers' Track at the RSA Conference 2010, San Francisco, CA, USA, March 1-5, 2010. Proceedings*. Springer, 2010, pp. 119–131.

[30] S. Ma, Q. Huang, M. Zhang, and B. Yang, "Efficient public key encryption with equality test supporting flexible authorization," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 3, pp. 458–470, 2014.

[31] W. Susilo, F. Guo, Z. Zhao, Y. Jiang, and C. Ge, "Secure replication-based outsourced computation using smart contracts," *IEEE Transactions on Services Computing*, 2023.

[32] A. Fiat and A. Shamir, "How to prove yourself: Practical solutions to identification and signature problems," in *Advances in Cryptology—CRYPTO'86: Proceedings 6*. Springer, 1987, pp. 186–194.

[33] C. H. Lau, F. Yan, and S. Chan, "On the resilience of authentication schemes in iot networks with different structural topologies," *IEEE Transactions on Consumer Electronics*, 2023.

[34] R. Kumar, D. Javeed, A. Aljuhani, A. Jolfaei, P. Kumar, and A. N. Islam, "Blockchain-based authentication and explainable ai for securing consumer iot applications," *IEEE Transactions on Consumer Electronics*, 2023.

[35] L. M. Friedman, C. D. Furberg, D. L. DeMets, D. M. Reboussin, and C. B. Granger, *Fundamentals of clinical trials*. Springer, 2015.

[36] Z. Bao, D. He, M. K. Khan, M. Luo, and Q. Xie, "Pbidm: Privacy-preserving blockchain-based identity management system for industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1524–1534, 2022.

[37] W3C, "Decentralized identifiers (dids) v1.0," accessed: 2024-05-16. [Online]. Available: https://w3c.github.io/did-core/