# A Privacy-Preserving Protocol to Prevent Fraudulent SPID Registrations

Francesco Buccafurri[1,*], Carmen Licciardi[1], Matteo Cecilia[2], Marco De Iacovo[2] and Rocco Mammoliti[2]

[1]University Mediterranea of Reggio Calabria, Reggio Calabria, Italy
[2]Poste Italiane, Roma, Italy

**Abstract**

This paper introduces a privacy-preserving identity registration protocol for SPID (the Italian Digital Identity System) aimed at further strengthening the system against fraudulent or duplicate identity registrations while maintaining full user privacy. Although identity theft attempts involving digital identity systems are relatively rare, some recent cases have shown that even well-regulated federated frameworks may benefit from additional mechanisms that support coordinated, privacy-preserving verification across Identity Providers (IdPs). Such scenarios do not indicate systemic issues, but rather highlight the importance of continuously evolving security measures as digital identity adoption grows. The proposed protocol relies on an Oblivious Pseudorandom Function (OPRF) construction that enables a central authority to detect repeated or failed Know Your Customer (KYC) procedures without learning users' personal data or enabling cross-domain correlation. Each IdP generates domain-specific pseudonyms derived from secret salts, while the central authority maintains a privacy-preserving blind registry. The protocol is lightweight, compatible with existing SPID processes, and compliant with privacy and regulatory requirements, providing an additional layer of protection for both citizens and providers.

**Keywords**

Digital Identity, Impersonation, eIDAS, SPID

## 1. Introduction

In recent years, some isolated cases of identity theft have been reported, showing that even consolidated identity frameworks, such as the Italian system SPID, may occasionally encounter attempts by malicious actors to reuse personal information across different providers. Such situations do not indicate systemic issues, but rather highlight that, in federated settings where multiple IdPs independently perform the Know Your Customer (KYC) procedure, the absence of a coordination mechanism may leave room for abuse.

The user's identity verification procedure of SPID–the Know Your Customer (KYC) procedure– is regulated and well-defined. Although this process is robust, a determined attacker who manages to acquire another person's data may still attempt to register with a different IdP, as no cross-provider mechanism is currently designed to detect such duplications. However, any information sharing mechanism must strictly comply with privacy requirements, which specify that each IdP may access personal data only for the users it directly serves.

In this paper, we face the above problem by introducing a cryptographic protocol that allows a central authority to detect identity reuse and fully preserves user privacy. The core construction, presented in a general federated authentication setting in [1], is based on an Oblivious Pseudorandom Function (OPRF) that enables IdPs to obtain domain-specific pseudonyms while keeping user information hidden from all other entities.

In our protocol, each IdP acts as an OPRF client, while a central authority (*Central Trusted Service*–CTS) plays the role of OPRF server. The CTS maintains a blind registry that tracks whether a pseudonym has

*Joint National Conference on Cybersecurity (ITASEC & SERICS 2026), February 09-13, 2026, Cagliari, IT*
*Corresponding author.
bucca@unirc.it (F. Buccafurri); carmen.licciardi@unirc.it (C. Licciardi); matteo.cecilia@posteitaliane.it (M. Cecilia); marco.deiacovo@posteitaliane.it (M. D. Iacovo); rocco.mammoliti@posteitaliane.it (R. Mammoliti)
0000-0002-0877-7063 (F. Buccafurri); 0009-0002-9981-1569 (C. Licciardi)

previously been involved in a successful or failed KYC attempt. This enables the detection of repeated or fraudulent registration attempts while preserving privacy and ensuring that:

- no party ever learns both the user's identity information and the cryptographic secrets used by the CTS and the IdPs,
- IdPs operate within separate cryptographic domains defined by private salts,
- failed KYC attempts are tracked without revealing user identities or linking activity across IdPs.

The structure of the paper is the following. In Section 2 the Italian SPID system is described. Section 3 introduces the cryptographic primitives used in our construction, including OPRFs and the RSA implementation. Section 4 describes the system model, the involved entities, and the notation adopted throughout the paper. In Section 5 we describe the notation adopted throughout the paper. Section 6 presents the full protocol specification, covering first-time identity enrollment, subsequent registrations, and the cooperative blind global check mechanism. Section 7 discusses the reaction logic used to interpret KYC outcomes in a privacy-preserving manner. Section 8 defines the threat model and provides an informal security analysis. Related work is discussed in Section 9. Finally, Section 10 concludes the paper and outlines potential directions for future work.

## 2. The SPID IdP Functions

In this section, we describe how the SPID system works with specific focus on the responsibilities and activities of the Identity Provider.

### 2.1. Identity Provider and Registration Procedure (KYC)

SPID is the Public Digital Identity System that enables citizens to access online services offered by public administrations and participating private entities through single sign-on credentials. The system is regulated and supervised by the Italian Digital Agency (AgID). Identity Providers (IdPs) are various entities, accredited by AgID, responsible for issuing digital identities and managing their use for authentication and service access. The activation of a digital identity requires a Know Your Customer (KYC) process and is structured in several phases.

The first phase is **Registration**, which may occur either prior to or concurrently with the subsequent identification process, and involves entering personal and contact information such as first name, last name, tax code, type and number of identity document, e-mail address, mobile phone number, and so forth. At this stage, the Provider certifies the verification of the contact details (e-mail and mobile number) and checks the relevant attributes against available sources (e.g., SCIPAFI).

The second phase is to **Upload document** such as the identity card, passport, or driver's license, along with the health insurance card (which, in some cases, can be replaced by tax code or an equivalent document).

The subsequent phase, which is of paramount importance, is the **Identification (KYC)**, which may be carried out through the following modalities: **In Person Identification** or **Electronic Identification**.

The first mode may be conducted in person (In-Person KYC) through proprietary networks or third-party entities. Alternatively, it may be performed remotely (Remote KYC) via a synchronous video call with an operator; by using an electronic identity document (CIE or passport) combined with an operator's asynchronous verification, or by submitting a paper identity card or driver's license together with a confirmation bank transfer, followed by an operator's asynchronous verification.

The second mode may be carried out through the use of a digital signature (KYC with Digital Signature), specifically a Qualified Electronic Signature issued by a QTSP (Qualified Trust Service Provider). Alternatively, the process may be completed by using an Electronic Identity Document, either through the CIE with PIN or through the TS/CNS with PIN. Finally, as another option, it may be performed by leveraging pre-existing identities, either through mechanisms pre-authorized by AgID or subject to authorization by individual Providers.

The final phase consists in the **creation and delivery of the security credentials**, which may take different forms depending on the required security level and on the specific features of the selected Identity Provider's service:

- **First security level**: based on a single authentication factor, such as a password;
- **Second security level**: multi-factor credentials combining at least two factors from the three categories of knowledge (for example, a password), possession (for example, a One-Time Password generated by a secure device or received via SMS, an activated app and a PIN code), and inherence (for example, biometric authentication);
- **Third security level**: two-factor authentication systems based on digital certificates, whose private keys are securely stored on dedicated hardware devices.

It is important to note that second-level security credentials are by far the most widespread and commonly used by users.

### 2.1.1. General Model and Selected Critical Issues

SPID was designed as a system with public governance (AgID) and with both public and private operators responsible for managing the services related to the issuance and use of Digital Identities. No central node was implemented, and citizens were granted the option to obtain multiple identities issued by different Providers. While these features have contributed to the development of the System, enabling it to achieve a leading position at the European level, they have also brought to light certain critical issues.

The structural possibility within the SPID system to allow the creation of multiple digital identities for the same tax code across different Providers operates without any mechanism to verify whether an identity has already been issued for an individual. This lack of information underlies the so-called *double SPID fraud*, in which a malicious actor may activate a second SPID identity using data and documents obtained fraudulently (e.g., by exploiting specific weaknesses in process implementations), without the first provider, or the citizen—being aware of it. Since the regulatory framework permits multiple SPID identities for the same tax code, issued by different providers, no technical constraint exists to prevent the creation of more than one SPID. Moreover, the regulation does not require that the e-mail address and mobile phone number be registered in the name of the identity holder: it is sufficient that their possession be verified and that these attributes are not already associated with another digital identity held with the same Provider. As a result, a potential attacker who has fraudulently passed the KYC checks could use contact information under their own control, preventing the victim from receiving notifications.

## 2.2. Types of responses in the event of issues

The issues that may arise, from a security standpoint, can affect two main areas: the **activation of the digital identity** and the **use of the credentials**.

Under the category of **Fraudulent Activation of the Digital Identity** fall the following cases:

- Attempts to circumvent the controls carried out by in-person identification networks, for example by exploiting possible inconsistencies or heterogeneities among the various third-party networks responsible for identification.
- Attempts to bypass the controls in remote identification procedures, by using digital identification instruments (such as the Health Insurance Card or the National Service Card) or by exploiting new techniques to target legacy systems, for instance through the use of operator-assisted webcam sessions.

Under the category of **Unauthorized Use of Credentials**, the main risks include:

- Phishing or social engineering attacks aimed at capturing the user's credentials.

- Exploitation of vulnerabilities in credential management systems, which may allow unauthorized access.

Such issues may be detected through various channels, including AgID or judicial authorities, the user themselves, or internal monitoring systems.

The measures available to the Provider include the revocation of the digital identity or its precautionary suspension in cases where further verification is required.

Additional security measures applicable to the use of credentials include: blocking access after a limited number of unsuccessful attempts, restricting the number of apps that may be associated with each profile, and revoking the identity after 24 months of inactivity.

### 2.3. Areas for improvement

The absence of a centralized verification system means that, at present, fraud detection occurs exclusively in a reactive manner—namely, following the Provider's ex-post monitoring activities, reports filed by the individual targeted by the fraud attempt or by the unlawful use of their identity, notifications from the authorities, or anomalies detected within connected services. Moreover, no mechanisms are currently in place to prevent the creation of fraudulent identities in cases in which a perpetrator successfully passes the KYC process. It is important to note that any such mechanism must comply with GDPR requirements.

This scenario highlights significant areas for improvement that could serve as a catalyst for a substantial strengthening of a system that is now essential for accessing online services (over 1.2 billion logins, accounting for more than 90% of all online access to public administration services), both in the public and private sectors.

In the following section, we propose a cryptographic solution to address the problem described above. Our construction provides strong guarantees that double SPID fraud cannot occur, even in scenarios where a malicious actor is able to successfully pass the KYC procedure.

We start by introducing the necessary background on the cryptographic primitives used in our protocol.

## 3. Background

### 3.1. Oblivious Pseudorandom Functions

Given any value $x$, and a function $F$, an *Oblivious Pseudorandom Function (OPRF)* enables a client to obtain $F_K(x)$ from a server holding the secret key $K$, such that:

- The client learns only $F_K(x)$, not $K$;
- The server learns nothing about $x$.

The OPRF guarantees that from the perspective of the client, $F_K(x)$ is computationally indistinguishable from the output of a random function, provided the server is honest.

### 3.2. RSA-Based Instantiation

Let the server hold an RSA key pair $K = (N, e, d)$, where $N$ is the module, $e$ is the public key, and $d$ is the private key. The OPRF can be instantiated as:

$$F_K(x) = H'(x^d \bmod N)$$

where $H'$ is a hash mapping RSA outputs into a fixed domain. To compute $F_K(x)$ obliviously, the client performs:

$$\begin{aligned} X &= H(x) \cdot r^e \bmod N, \\ Y &= X^d \bmod N, \\ F_K(x) &= Y \cdot r^{-1} \bmod N. \end{aligned}$$

This ensures that $F_K(x) = (H(x))^d \bmod N$ without revealing $x$ to the server.

## 4. System Model and Entities

The protocol involves three entities:

- **User**: owns personal data including a fiscal code ($CF$) and undergoes KYC at an IdP. For simplicity, we assume that $CF$ is the only personal information used by the IdP to identify a user.
- **Identity Provider (IdP$_i$)**: verifies KYC, blinds $CF$, and applies a domain-specific RSA transformation using its private key $t_i$.
- **Central Trusted Service CTS**: holds the master RSA key $(N, e, d)$, evaluates blinded inputs, issues blind tokens, and maintains a pseudonymous registry of KYC states.

Each IdP$_i$ possesses a private RSA key $t_i$ and a corresponding public key $P_i = (N_i, e_i)$ defining its domain transformation. The CTS stores only $P_i$ values to support domain coordination without exposure to private salts.

## 5. Notation

This section introduces the notation used throughout the paper. The system involves three main actors: the user, each Identity Provider (IdP), and the Central Trusted Service (CTS). All computations performed during identity enrollment use the RSA modulus $N$ of the CTS, ensuring mathematical compatibility of the cooperative RSA-OPRF construction. The IdPs use their own RSA moduli $N_i$ only during the exponent-exchange phase of the cooperative global check, never during enrollment.

The user is identified, for cryptographic purposes, by a hashed representation $x = H(CF)$, where $H$ is a cryptographic hash function. This value is revealed only to the IdP that performs the registration procedure. Whenever it must be sent to the CTS, it is first blinded using a random blinding factor $r \in \mathbb{Z}_N^*$, ensuring that no party other than the user learns $H(CF)$.

Each IdP $i$ possesses an RSA public modulus $N_i$ with exponent $e_i$ and the corresponding private exponent $t_i$. These keys are used to derive a domain-specific User Pseudonym $UP_i$ through a cooperative RSA-OPRF evaluation jointly performed with the CTS. The exponents $(e_i, t_i)$ are also used during the cooperative check phase, where the IdP proves the value $r^{t_i} \bmod N_i$ needed to unblind identifiers belonging to its domain.

The CTS holds its own RSA key pair $K = (N, e, d)$, used to evaluate blinded inputs through $Eval_K(X) = X^d \bmod N$. Since all OPRF computations are performed modulo $N$, every $UP_i$ produced during enrollment also lies in the domain $\mathbb{Z}_N$.

After a successful KYC verification at IdP $i$, the CTS issues a signature over $(UP_i, pk_{u_i})$, producing a token $Token_i$. This token binds the pseudonym to the user's public key while preserving privacy. During subsequent registrations at any IdP, the user proves knowledge of the corresponding private key via a challenge–response mechanism.

The blind registry maintained by the CTS stores, for each pseudonym, a minimal status value $status_{KYC} \in \{ok, alarm\}$, used to drive the reaction logic described later in the paper. No additional information is ever stored or revealed.

The complete meaning of the above notations, summarized in Figure 1, will be clearer after the description of the protocol in the next section.

## 6. Protocol Definition

In this section, we describe the privacy-preserving protocol we propose to counter fraudulent SPID registrations. The protocol unfolds across three distinct phases, corresponding to the possible states in which a user may interact with an Identity Provider. First, a user with no prior SPID credentials

| Symbol | Meaning |
|---|---|
| $CF$ | User's fiscal code. |
| $x = H(CF)$ | Hash of the fiscal code used for cryptographic processing. |
| $r$ | Blinding factor in $\mathbb{Z}_N^*$. |
| $t_i,\ (N_i, e_i)$ | RSA private exponent $t_i$ and public key of IdP $i$. |
| $K = (N, e, d)$ | RSA key pair of the CTS. |
| $Blind(x) = x \cdot r^e \bmod N$ | RSA blinding function (CTS modulus). |
| $Eval_K(X) = X^d \bmod N$ | OPRF evaluation performed by the CTS. |
| $UP_i = (H(CF))^{d \cdot t_i} \bmod N$ | User Pseudonym for IdP $i$ (always modulo $N$). |
| $Token_i = \mathsf{Sign}_{CTS}(UP_i, pk_{u_i})$ | signature of CTS binding user public key $pk_{u_i}$ to $UP_i$. |
| $status_{KYC} \in \{ok,\ alarm\}$ | Status stored in the blind registry. |

**Figure 1:** Notation used throughout the protocol.

undergoes a first-time enrollment, during which a domain-specific $UP$ and the user's initial token are generated. Subsequent registrations reuse any previously issued token to prove prior successful verification, while producing new domain-separated identifiers. Finally, when no valid token can be presented, the IdP must determine—without revealing or learning sensitive information—whether the same user has already been registered elsewhere. This is achieved through a cooperative blind global check involving all IdPs and the CTS. The following subsections describe these three phases in detail.

## 6.1. First-Time Identity Enrollment

Before a user can obtain a SPID identity for the first time, they must undergo an initial enrollment procedure at an Identity Provider (IdP). This phase is special because the user does not yet possess any cryptographic token issued by the CTS, and therefore cannot prove prior verification. Consequently, the IdP must rely solely on the outcome of its own KYC process and on the cooperative RSA-OPRF interaction with the CTS to generate the user's first pseudonym.

The objective of this phase is twofold: (i) to bind the user's real-world identity (validated through KYC) to a domain-specific $UP_1$, and (ii) to let the CTS issue the user's first signed token. Importantly, the token is issued only after the user proves possession of the private key corresponding to the public key $pk_{u_1}$ included in the message to be signed. This ensures that the token cannot be transferred or misused by an attacker.

The steps of the first-time enrollment are as follows.

1. The user provides $CF$ to $\text{IdP}_1$ and successfully completes the KYC procedure.
2. $\text{IdP}_1$ computes $x = H(CF)$, selects a random blinding factor $r$, and computes the blinded input $Blind(x) = x \cdot r^e \bmod N$.
3. The IdP applies its domain transformation by computing $X_1 = Blind(x)^{t_1} \bmod N$ and sends $X_1$ to the CTS.
4. The CTS evaluates the OPRF by computing $Y_1 = X_1^d \bmod N$ and returns it to $\text{IdP}_1$.
5. $\text{IdP}_1$ unblinds:
$$UP_1 = Y_1 \cdot (r^{t_1})^{-1} \bmod N.$$
6. The user sends $(UP_1, pk_{u_1})$ to the CTS, and proves possession of the private key associated with $pk_{u_1}$ through a challenge-response protocol. Only if this proof succeeds, the CTS stores $(UP_1, status = ok)$ in its registry and issues to the user the signed token $Token_1$.

## 6.2. Subsequent Registration

After the user has successfully completed at least one SPID enrollment, the CTS has already issued to them a signed token $Token_j$ for some $j$. Any future enrollment at a different Identity Provider (IdP)

follows a similar cryptographic workflow to the first-time registration, with two important differences: (i) the user is expected to prove that they have already been verified at least once, and (ii) the IdP must rely on this proof to distinguish between a new legitimate enrollment and a potential impersonation attempt.

To register again at an IdP, the user begins by undergoing the same RSA-OPRF procedure to compute a new domain-specific $UP_i$, exactly as in the initial enrollment. Before the IdP proceeds with KYC, the user must present a previously issued token $Token_j$ *together with a proof of possession of the corresponding private key*. This is achieved through a standard challenge-response mechanism: the IdP sends a fresh challenge and the user returns its signature using the private key associated with the public key contained in $Token_j$. Since all valid tokens bind the same user public key $pk_{u_j}$ to distinct pseudonyms, *any previously issued token is acceptable*, regardless of its age or of which IdP issued it.

The subsequent registration then proceeds as follows.

1. IdP$_i$ computes $UP_i$ for the user as in the first-time enrollment.
2. The user presents a previously obtained token $Token_j$ and proves possession of the corresponding private key through a challenge-response exchange. If the token is invalid or the proof fails, the procedure halts and $(UP_i, status = alarm)$ is recorded in the CTS registry.
3. If the proof succeeds, IdP$_i$ performs its KYC verification. If KYC fails, the CTS records $(UP_i, status = alarm)$.
4. If KYC succeeds, the CTS records $(UP_i, status = ok)$ and issues to the user a new signed token $Token_i$.

This mechanism allows IdPs to verify that the user attempting a subsequent registration has already been successfully enrolled at least once, while preserving unlinkability: the IdP cannot determine when, where, or how many times the user was previously enrolled, and the CTS never learns the user's fiscal code or personal information.

## 6.3. Cooperative Blind Global Check Protocol

When a user attempts a registration at IdP$_i$ without presenting a valid token, the IdP must determine whether the same fiscal code has already been successfully registered at any other IdP, without ever learning any identifier outside its own domain. This is achieved through a cooperative blind global check involving IdP$_i$, all other IdPs, and the CTS. All computations are performed without exposing $H(CF)$, any $UP$ belonging to another IdP, or any sensitive exponent.

The protocol consists of three conceptual steps, even though multiple entities contribute to the computation.

### Blinded Submission

IdP$_i$ begins by computing the blinded value

$$X = H(CF) \cdot r^e \bmod N,$$

where $r \in \mathbb{Z}_N^*$ is random and known only to IdP$_i$. The value $X$ hides $H(CF)$ from the CTS and from all other IdPs. IdP$_i$ sends $X$ to the CTS.

### Distributed Domain Transformations

To determine whether the same user has been registered under any IdP$_j$, the CTS forwards $X$ to each IdP$_j$. Each IdP$_j$ applies its domain exponent and returns:

$$T_j = X^{t_j} \bmod N.$$

Since all computations are modulo $N$, these values are compatible with the OPRF evaluation performed by the CTS. The CTS computes

$$Y_j = T_j^d \bmod N$$

for every IdP$_j$ and returns $Y_j$ to IdP$_i$. At this point, the CTS has evaluated the blinded input under every IdP domain without learning any information about the user's identity. The algebraic form of the blinded values may still reflect domain-specific exponents, allowing the CTS to distinguish transformations associated with different IdPs; however, this does not enable linking registration attempts belonging to the same user nor does it compromise user privacy.

### Exponent Exchange

To locally unblind $Y_j$, IdP$_i$ must obtain the value $r^{t_j}$, but IdP$_j$ should not learn anything about $r$ or $H(CF)$. The following RSA-based exchange allows IdP$_i$ to obtain $r^{t_j}$ blindly, using IdP$_j$'s own modulus $N_j$ only for this auxiliary step.

For each IdP$_j$:

1. IdP$_i$ samples a random $k \in \mathbb{Z}_{N_j}^*$ and computes
$$m = k^{e_j} \cdot r \bmod N_j.$$
2. IdP$_i$ sends $m$ to IdP$_j$.
3. IdP$_j$ computes
$$s = m^{t_j} \bmod N_j = k \cdot r^{t_j} \bmod N_j$$
and returns $s$ to IdP$_i$.
4. IdP$_i$ recovers
$$r^{t_j} = s \cdot k^{-1} \bmod N_j.$$

This step reveals no information about $H(CF)$ or about the $UPs$ across domains, while giving IdP$_i$ exactly the quantity needed for unblinding.

### Local Unblinding and Registry Match

Given $Y_j$ and $r^{t_j}$, IdP$_i$ reconstructs the $UP$ for the domain of IdP$_j$:

$$UP_j = Y_j \cdot (r^{t_j})^{-1} \bmod N.$$

IdP$_i$ sends all reconstructed $\{UP_j\}$ to the CTS. The CTS checks whether any of them appears in its registry with status *ok*.

If such a match exists, this indicates that the attempted registration corresponds to an identity already verified at another IdP, and IdP$_i$ must reject the current procedure. In that case, $(UP_i,\ status = alarm)$ is inserted into the CTS registry.

If no match exists, IdP$_i$ proceeds with enrollment as for a first-time registration.

## 7. Fraud Reaction Logic

In this section, we propose a reaction policy to adopt in case of anomalous events. The design rationale is stated in the next subsection.

## 7.1. Design Rationale

The proposed reaction mechanism is based on a *victim-centric* security model, consistent with privacy-by-design principles and with the guarantees of the cooperative RSA-OPRF protocol described in this work. A fundamental observation is that, in a multi-IdP system, the party attempting to register at an IdP is never assumed to be the attacker. Instead, the attacker is the entity misusing a stolen fiscal code (CF) elsewhere in the system. Therefore, any failure or inconsistency observed during a registration attempt must be treated either as (i) a potential impersonation attempt against the identity, or (ii) a benign technical or procedural issue. For this reason, no user is ever penalized, labelled, blacklisted, or prevented from future registrations.

Furthermore, because $UPs$ are domain-separated and all operations are blind, neither the IdPs nor the CTS can determine which user triggered an ALARM event. Although domain-specific algebraic patterns may still be distinguishable by the CTS, these do not reveal user identities nor enable correlation of events across IdPs. Thus, reaction policies must avoid leaking information and must not alter observable behaviour across IdPs. The system must provide protection without revealing correlations.

The reaction mechanism therefore distinguishes between:

- **ALARM events with high certainty of active attack**, where a fraudulent attempt is detected and effectively blocked by the protocol itself;
- **ALARM events with ambiguous interpretation**, where the system cannot determine whether the failure is due to an attack or to a legitimate user experiencing technical issues or having lost the token.

This distinction guides how the system reacts in a way that protects the identity without compromising usability or violating privacy guarantees.

## 7.2. Reaction Policy

Let $(UP_i, status)$ denote the pseudonymous record created during a registration attempt at Identity Provider IdP$_i$. The CTS stores $status \in \{ok, alarm\}$.

### 7.2.1. Category A: High-Certainty Attack Signals

Certain ALARM events indicate strong evidence of a fraudulent attempt, in particular:

1. the cooperative blind check reveals an existing $UP_j$ for the same CF with status $ok$;
2. the token presented by the user is structurally invalid, forged, or cryptographically inconsistent;
3. the protocol messages show signs of active manipulation.

In these cases, the system can safely assume that an attacker is attempting to reuse or forge an identity, because the legitimate user would possess a consistent token and would not produce manipulated protocol messages. Therefore:

- the current registration procedure is immediately aborted;
- the CTS records $(UP_i, alarm)$;
- no additional information is revealed to the IdP or to the user;
- no future KYC penalties or restrictions are imposed on the legitimate user;
- the event is logged internally for system-wide governance, monitoring, anomaly detection, and IdP auditing.

### 7.2.2. Category B: Ambiguous Risk Signals

Other ALARM events arise from conditions that cannot be distinguished from benign user behaviour:

1. failure of KYC during the first registration attempt at an IdP;

2. failure of KYC during a later registration attempt even in the presence of a valid token;
3. absence of a valid token (e.g., the user lost the credential or changed devices);
4. protocol inconsistencies compatible with technical failure rather than malicious manipulation.

These events indicate potential identity theft but do not provide evidence strong enough to distinguish attacker from victim. In such cases, the system must ensure robust protection while maintaining fairness and privacy. The reaction is therefore:

- the current registration attempt may be aborted, depending on the IdP's local KYC policy;
- the CTS records $(UP_i, alarm)$;
- any future registration attempt at any IdP that corresponds to the same CF (detected through the cooperative blind check) must undergo hardened KYC;
- hardened KYC continues to apply until a valid registration produces a new record $(UP_k, ok)$, thus resolving the ambiguity in favour of the legitimate user.

This policy ensures that legitimate users are always able to successfully register, possibly with stronger verification if needed, while attackers systematically fail to overcome the reinforced checks. The system never penalizes users, never reveals cross-IdP information, and remains fully privacy-preserving.

### 7.3. Governance Considerations

ALARM events are also used by the CTS for privacy-preserving system governance. High-certainty ALARM events (Category A) may indicate coordinated attacks, weaknesses in specific IdPs, or anomalies requiring intervention. Ambiguous ALARM events (Category B) support statistical monitoring but do not trigger punitive actions. In all cases, no IdP learns whether a user is associated with ALARM events at other IdPs, and no user is ever prevented from future registration attempts.

## 8. Threat Model and Security Analysis Sketch

In this section, we define the threat model considered in our work and provide an informal overview of the security guarantees offered by the protocol.

### 8.1. Threat Model

In our threat model, we assume that both the CTS and the Identity Providers (IdPs) behave as honest-but-curious entities: they follow the prescribed protocol but may attempt to infer additional information about users' identities or about potential correlations between registration attempts. In particular, each of these entities may try to reconstruct whether two protocol interactions correspond to the same user, or to extract information about the fiscal code from the blinded values they receive. Although the construction prevents such inferences, this assumption is essential because CTS and IdPs are internal system components and therefore represent realistic passive adversaries.

The protocol also prevents linking the same user across different IdPs. However, domain-level transformations may remain distinguishable at a purely algebraic level: for example, the CTS may observe that different $UP_j$ values follow structurally different patterns induced by the domain-specific exponents $t_j$. This leakage does *not* compromise user privacy, as it does not enable the CTS to identify the underlying user or to correlate two UP values as belonging to the same person; what it reveals is only the IdP domain that generated each pseudonym. We return to this point in the security analysis.

We do not consider scenarios involving collusion among IdPs or between an IdP and the CTS. The adversary is able to participate in the registration procedures, manipulate protocol messages within the limits imposed by the underlying cryptographic primitives, and potentially pass weak KYC checks. The adversary's objectives include:

- obtaining a second SPID identity for the same individual,

- circumventing the CTS's blind detection of duplicate registrations,
- reusing or forging verification tokens,
- inferring information about the victim's registration status at other IdPs.

### 8.2. Security Analysis Sketch

Although not a formal proof, the following arguments illustrate why the construction achieves fraud detection while preserving user privacy.

**Privacy of User Data.** The CTS and the IdPs never learn $CF$ or $H(CF)$ in the clear. All data sent to the CTS is blinded as $X = H(CF) \cdot r^e \bmod N$, and the blinding factor $r$ is never revealed. Domain separation through the private exponents $t_i$ prevents any IdP from correlating identifiers belonging to another domain, and the CTS sees only opaque pseudonyms.

**Integrity and Non-Transferability of Tokens.** Each token binds a domain-specific identifier $UP_i$ to the user's public key via a signature. Before issuing a token, the CTS verifies proof of possession of the private key corresponding to $pk_{u_i}$, ensuring that tokens cannot be transferred or reused by an attacker. Standard RSA assumptions make token forgery infeasible, and no IdP can create the identifier of another IdP.

**Blind Detection of Duplicate Registrations.** If an attacker attempts to reuse another user's fiscal code, the cooperative blind global check reconstructs all domain-specific $UP_j$ identifiers and detects any entry with status *ok*. An attacker who does not own a valid token cannot unblind OPRF outputs and therefore cannot falsify $UP_j$. Failed or suspicious attempts are recorded as *alarm* events without penalizing the legitimate user.

**Non-Linkability Across Domains and Domain-Level Leakage.** The CTS never sees unblinded identifiers unless explicitly provided during a legitimate registration process. Blinded OPRF evaluations and domain-separated exponents ensure that neither IdPs nor the CTS can link multiple registration attempts belonging to the same user. However, because the pseudonyms $UP_j$ are deterministically derived from domain-specific exponents $t_j$, the CTS can distinguish–and therefore *cluster*–the pseudonyms by their IdP of origin. This leakage is an inherent consequence of using domain-specific RSA exponents and does not reveal any information about the underlying user: the CTS may infer *which* IdP produced a pseudonym, but not *whether two pseudonyms correspond to the same individual.* Moreover, this leakage does not assist any malicious objective of the CTS under the honest-but-curious model: learning the IdP domain of origin does not help reconstruct $H(CF)$, does not alter the blindness of the OPRF interaction, and does not enable cross-IdP linkability of user identities. As discussed in the conclusion, this form of leakage is benign in practice and compatible with the privacy goals of the protocol, and can be removed entirely through extended variants of the construction.

In summary, the protocol enforces blind global consistency: it can detect fraudulent or duplicate identity registrations across IdPs without revealing, correlating, or exposing the user's personal information, while leaking only limited and benign metadata about the IdP domain associated with each pseudonym.

## 9. Related Work

Research on privacy-preserving identity systems spans several decades, beginning with the foundational notions of blind signatures [2] and pseudonymous transactions [3]. These ideas were later strengthened in anonymous credential systems such as the Camenisch-Lysyanskaya construction [4], which enables unlinkable and non-transferable credentials. While these schemes support privacy-preserving authentication, they do not provide mechanisms for detecting duplicate or fraudulent identity registrations.

Privacy in federated identity management has been widely studied. Systems such as PRIMA [5] aim to protect users from cross-provider tracking by issuing unlinkable identifiers across domains.

Broader analyses of the privacy risks in federated identity systems, including correlation and metadata leakage, are discussed in [6]. These works focus on protecting authentication flows, assuming that the underlying user identity has already been correctly established.

Recent work on OPRF-based anonymous token systems provides another relevant foundation. The Privacy Pass architecture [7] and subsequent OPRF standardization efforts [8] show how blind signatures and OPRFs can be combined to produce unlinkable tokens. More advanced schemes such as Anon-Tokens [9] and OPRF-based adaptive anonymous credentials [10] further illustrate the practicality of OPRF-assisted credential issuance. However, these approaches target anonymous authorization and rate-limiting, not identity proofing or coordinating KYC outcomes across multiple providers.

From a regulatory and operational perspective, digital identity frameworks such as NIST SP 800-63-3 [11] and national systems like SPID [12] highlight the challenges of identity proofing and the inability of current architectures to detect duplicated registrations without violating privacy.

In contrast to prior work, our protocol directly addresses the enrollment phase, introducing a cooperative RSA-OPRF mechanism that allows a central authority to detect duplicate or fraudulent registrations without ever learning the user's personal data or enabling cross-provider correlation. To our knowledge, this is the first approach that provides global KYC consistency while maintaining full domain separation and privacy guarantees compatible with federated identity systems such as SPID.

## 10. Conclusion

We presented a cooperative RSA-OPRF protocol for privacy-preserving identity registration in federated systems such as SPID. The construction enables a central authority to detect duplicate or fraudulent registrations without ever learning the user's fiscal code in cleartext or acquiring information that would allow correlation across IdPs. Through blinded OPRF evaluation, domain-specific exponents, and signed tokens, the protocol ensures global consistency of KYC outcomes while preserving strong privacy guarantees. The reaction logic introduces a victim-centric approach that avoids blacklisting and maintains usability, while still supporting strengthened KYC verification when risk signals emerge.

The protocol presented in this paper is intentionally kept lightweight and easy to deploy, as one of the main goals of this work is to support the rapid adoption of a privacy-preserving duplicate-registration detection mechanism in real identity federations. As discussed in the security analysis, the current construction leaks a limited amount of domain-level metadata to the CTS. Specifically, the CTS can distinguish the originating Identity Provider of each pseudonym. This leakage is benign, as it does not reveal user identifiers, does not enable cross-IdP correlation, and does not weaken any privacy guarantee for end users. Still, it represents an area where further hardening is possible. To this end, we note that the authors have already explored enhanced protocol variants that eliminate this leakage entirely, which achieve full domain unlinkability while preserving efficiency and compatibility with the operational constraints of real-world federations. Although these extensions go beyond the scope of the present paper, they appear sufficiently efficient and realistic to be considered for future integration in deployed systems.

Although our results show that privacy and global KYC consistency can be simultaneously achieved, several open directions remain. A first avenue concerns integrating revocation mechanisms. For example, legitimate users may need to invalidate compromised tokens, yet such revocation must preserve unlinkability and avoid introducing new correlation vectors. Besides revocation, the alignment of the protocol with the validity time of SPID registrations should be managed. Another direction is developing a formal security analysis to fully assess the security of the solution. Moreover, we plan to test the scalability of the solution with many IdPs.

Finally, aligning our protocol with emerging European digital identity systems (eIDAS 2.0, wallet-based identities) represents a promising long-term direction. For example, tokens can be provided as verifiable credentials to include in the national wallet. Overall, the protocol contributes a practical, privacy-preserving solution to identity fraud detection, strengthening the foundational layer of the SPID system while maintaining strict protection of user information.

## Acknowledgments

This work is supported by Agenzia per la Cybersicurezza Nazionale under the programme for promotion of XL cycle PhD research in cybersecurity – C36E24000080005. The views expressed are those of the authors and do not represent the funding institutions.

## Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

## References

[1] F. Buccafurri, C. Licciardi, A privacy-preserving information-sharing protocol for federated authentication, arXiv preprint arXiv:2512.01832 (2025).
[2] D. Chaum, Blind signatures for untraceable payments, in: Advances in Cryptology: Proceedings of Crypto 1982, Springer, 1983, pp. 199–203.
[3] D. Chaum, Security without identification: Transaction systems to make big brother obsolete, Communications of the ACM 28 (1985) 1030–1044.
[4] J. Camenisch, A. Lysyanskaya, An efficient system for non-transferable anonymous credentials with optional anonymity revocation, in: Advances in Cryptology – CRYPTO 2001, Springer, 2001, pp. 93–118.
[5] H. Asghar, H. Lee, S. Habib, G. Russello, Prima: Privacy-preserving federated identity management, in: Network and Distributed System Security Symposium (NDSS), 2016.
[6] G. Danezis, S. Lewis, R. Anderson, Privacy issues in identity management systems, in: DIM '05: Proceedings of the 2005 Workshop on Digital Identity Management, ACM, 2005, pp. 35–43.
[7] A. Davidson, J. Iyengar, N. Sullivan, F. Valsorda, C. Wood, Privacy pass: Bypassing internet challenges anonymously, IETF Draft draft-ietf-privacypass-architecture, 2020.
[8] H. Krawczyk, C. Wood, Oblivious pseudorandom functions (oprfs) and verifiable oprfs (voprfs), IETF Draft draft-irtf-cfrg-voprf, 2023.
[9] P. Silde, M. Strand, Anon-tokens: Privacy-preserving authorization with rsa blind signatures, in: European Symposium on Research in Computer Security (ESORICS), Springer, 2022, pp. 3–25.
[10] R. Baseri, A. Jain, A. Sahai, Adaptive anonymous credentials from oprfs, in: IEEE Symposium on Security and Privacy, 2024.
[11] P. A. Grassi, M. E. Garcia, J. L. Fenton, Digital Identity Guidelines, Technical Report NIST SP 800-63-3, National Institute of Standards and Technology, 2017.
[12] Agenzia per l'Italia Digitale, SPID Technical Rules and Identity Proofing Guidelines, Technical Report, AgID, 2022.