Secure User Identification Without Privacy Erosion

Stefan Brands*

INDIVIDUALS ARE INCREASINGLY CONFRONTED with requests to identify themselves when accessing services provided by government organizations, companies, and other service providers. At the same time, traditional transaction mechanisms are increasingly being replaced by electronic mechanisms that underneath their hood automatically capture and record globally unique identifiers. Taken together, these interrelated trends are currently eroding the privacy and security of individuals in a manner unimaginable just a few decades ago. Privacy activists are facing an increasingly hopeless battle against new privacy-invasive identification initiatives: the cost of computerized identification systems is rapidly going down, their accuracy and efficiency is improving all the time, much of the required data communication infrastructure is now in place, forgery of non-electronic user credentials is getting easier all the time, and data sharing imperatives have gone up dramatically. This paper argues that the privacy vs. identification debate should be moved into less polarized territory. Contrary to popular misbelief, identification and privacy are not opposite interests that need to be balanced: the same technological advances that threaten to annihilate privacy can be exploited to save privacy in an electronic age. The aim of this paper is to clarify that premise on the basis of a careful analysis of the concept of user identification itself. Following an examination of user identifiers and its purposes, I classify identification technologies in a manner that enables their privacy and security implications to be clearly articulated and contrasted. I also include an overview of a modern privacy-preserving approach to user identification.

DE PLUS EN PLUS ON DEMANDE AUX INDIVIDUS de s'identifier afin d'accéder aux services fournis par les organismes gouvernementaux, les entreprises et d'autres fournisseurs. En même temps, peu à peu les mécanismes de transaction traditionnels sont remplacés par des mécanismes électroniques qui automatiquement, mais subrepticement, saisissent et enregistrent globalement les identificateurs uniques. Ces nouvelles tendances, ensemble, érodent la vie privée et la sécurité des personnes d'une façon inimaginable il y a quelques décennies à peine. Les personnes activement engagées dans la protection des renseignements personnels se heurtent désespérément à ces nouvelles mesures d'identification qui portent atteinte à la vie privée : le coût des systèmes d'identification automatisés baisse rapidement; leur fiabilité et leur efficacité s'améliorent sans cesse; l'infrastructure généralement requise pour la communication des données existe déjà; la contrefaçon des justificatifs d'identité non numérisés devient de plus en plus facile; et le partage de renseignements s'impose de plus en plus comme une exigence. L'article soutient que le débat entre la vie privée et l'identification devrait se faire en territoire moins polarisé. Contrairement à la croyance populaire erronée, l'identification et la vie privée ne constituent pas des intérêts opposés à peser afin de trouver un juste équilibre : les progrès technologiques mêmes qui menacent d'annihiler la vie privée peuvent être exploités afin de préserver la confidentialité à l'ère de la numérisation. L'article cherche à élucider cette thèse à partir d'une analyse minutieuse de la notion même d'identification de l'utilisateur. Après une étude des identificateurs utilisateurs et de leurs objets, l'auteur classifie les technologies d'identification, en cherchant à bien dégager leurs répercussions sur le plan de la vie privée et de la sécurité et à permettre la comparaison de ces techniques. L'auteur décrit également sommairement un nouveau concept d'identificateur utilisateur qui préserve la confidentialité des renseignements.

---

* Adjunct Professor in Cryptology, School of Computer Science, McGill University, Montreal, email: <brands@credentica.com>, <brands@cs.mcgill.ca>.

207 1. INTRODUCTION

208 2. USER IDENTIFICATION AND AUTHENTICATION

208 *2.1. Identification Purposes*
209 *2.2. Examples of User Identifiers*
209 *2.3. Locality and Heterogeneity of Identifiers*
210 *2.4. Authentication*

210 3. NON-CERTIFIED USER IDENTIFIERS

211 *3.1. Privacy of Users*
211 *3.2. Security of Users*
212 *3.3. Security of Relying Parties*

212 4. CERTIFIED USER IDENTIFIERS

213 *4.1. Security of Relying Parties*
214 *4.2. Security and Privacy of Users*

215 5. A MODERN APPROACH TO USER IDENTIFICATION

223 6. FINAL WORD

Secure User Identification Without Privacy Erosion

Stefan Brands

1. INTRODUCTION

INDIVIDUALS ARE INCREASINGLY CONFRONTED with requests to securely identify themselves when accessing services. Much of the growing demand for secure user identification is driven by efficiency imperatives in client relationship management, by a post-9/11 demand for increased security, and by the explosive rise of "phishing" attacks and identity theft. As a result, more and more corporations and governments are rolling out secure identification tokens to the users of their services. This shift towards more secure user identification in turn provides strong incentives for individuals and organizations alike to rely on the same identification tokens for accessing multiple services, within organizational domains and increasingly across them. Namely, enrolling individuals and token issuance are costly and time-consuming processes, and users prefer not to be burdened with having to manage many identification tokens.

Taken together, these interrelated trends are currently eroding the privacy of individuals in a manner unimaginable just a few decades ago. As the traditional segmentation of identity domains disappears, individuals lose all control over the extent to which corporations, governments, and rogue hackers can electronically monitor their communications and transactions. Privacy activists are facing an increasingly hopeless battle: the cost of computerized identification systems is rapidly going down, their accuracy and efficiency continues to improve, electronic data communication networks have become commonplace, forgery of paper and plastic user credentials is getting easier all the time, and data sharing imperatives have gone up dramatically.

This paper argues that the privacy vs. identification debate is based on fundamental misconceptions. *Secure identification and privacy are not opposites that must be balanced.* The same technological advances that threaten to annihilate privacy can be exploited to rescue privacy, while

simultaneously addressing legitimate security and data sharing objectives of corporations and government organizations.

The remainder of this paper is organized as follows. In Part 2, I will take a closer look at the notions of user identification and authentication. In Parts 3 and 4, I will analyse and contrast the privacy and security implications of non-certified and certified user identifiers. This analysis sets the stage for Part 5, which provides a non-technical overview of cryptographic advances that enable one to do secure user identification without eroding privacy.

*

2. USER IDENTIFICATION AND AUTHENTICATION

AN IDENTIFIER IS A PIECE OF INFORMATION that names or indicates a person, a process, an application, a location (such as a place on earth or a CPU memory address), a tangible object (such as a book, a text file, or a device), or any other type of entity or grouping of entities. *User identifiers* are identifiers that represent users (i.e., individuals or groups of individuals) in their interactions with relying parties. Users may present their identifiers verbally, on paper, on plastic cards, or in any other appropriate manner. Electronic user identifiers are electronically presented over data communication channels by user-operated computing devices such as PCs, laptops, mobile phones, and smartcards.

2.1. Identification Purposes

Within a designated *context*, user identifiers enable relying parties to distinguish between the individuals they interact with; this is known as *identification*. Examples of a context include a sphere of activity, a geographical region, a communication platform, an application, and a logical or physical domain.

Within their designated context, user identifiers serve one or more of the following purposes for relying parties:

- Identifier as contact address: To enable relying parties to contact users, now or later, to deliver or retrieve services, goods, or information.
- Identifier for security: To enable relying parties to blacklist users who have engaged in unwanted behavior (so as to be able to deny them access) and perhaps to enable relying parties to trace such users for accountability reasons.[1]
- Identifier as registration proof: To enable relying parties to infer that users have been pre-approved in some sense in an enrollment process.
- Identifier as account index: To enable relying parties to build or consult indexed user accounts (also known as records, profiles, dossiers, and so on) that contain user-related information.

In each of the latter two cases, relying parties can use the additional user

1. To trace a person means to unambiguously determine that person's "real identity" through a discovery process that starts with information that is linked to that person.

information they learn to offer better personalized services, to improve business aspects (e.g., better inventory management or direct marketing), or to make better access control decisions.

2.2. Examples of User Identifiers

To appreciate the contextual nature of user identifiers, consider the designated contexts and purposes of the following user identification methods:

- Birth names, corporate names, nicknames, and author pseudonyms;
- E-mail addresses, telephone numbers, postal box numbers, and URLs;
- Fingerprints, iris or retina scans, and DNA samples;
- User account identifiers with ISPs, banks, utility companies, and so on;
- Credit cards, debit cards, calling cards, and loyalty tokens;
- Employee badges, sports club membership cards, and hotel key cards;
- Social security numbers, health insurance numbers, passports, and driver licences;
- Online usernames (*e.g.*, for instant messaging and chat rooms), cookies, and SSL certificates; and
- MAC addresses, IP addresses, smartcard serial numbers, Bluetooth identifiers, GSM IMEI numbers, RFID tag identifiers, and other addresses of networked user devices.

2.3. Locality and Heterogeneity of Identifiers

As the examples illustrate, users traditionally are represented in their interactions with relying parties by a plurality of "local" user identifiers with incompatible formats:

- The symbols that make up an identifier must be meaningful to the relying party. Humans are not good at memorizing and recognizing binary strings, while computers are not designed to handle non-numerical data. Thus, traditionally the encoding of user identifiers depends on their context.
- A larger namespace is required for distinguishing between increasing numbers of users, but lengthier identifiers are less desirable for human processing. Thus, traditionally the length of user identifiers is related to their designated context.
- In the absence of agreed-upon standards for encoding and generating identifiers, relying parties are likely to use proprietary formats for user identifiers. They may even do so deliberately in order to counter impersonation attacks based on leveraging user identifiers from other domains with a similar look and feel.

Thus, the traditional heterogeneity of communication platforms and the lack of connectivity have historically created an abundance of *single-domain* user identifiers that are designed to be relied on by only one or a few relying parties that all trust each other not to violate each other's security and privacy interests.

2.4. Authentication

Authentication, in its most general form, is a process for gaining confidence that something is, in fact, what it appears to be. In everyday life, people continually authenticate people, objects, and other entities around them, by either consciously or subconsciously assessing clues that provide evidence of authenticity. In communication and transaction settings, authentication typically refers to the process of confirming a claimed *identity*. An identity is claimed when a user presents to a relying party a user identifier that uniquely represents the user in the relying party's context. Identity authentication involves verifying that the presenter of the identifier is authorized to do so.

Single-factor identity authentication ascertains that the presenter possesses something associated with the presented user identifier that is not generally accessible. This can be something the user knows (such as a password or a cryptographic key), something the user has (such as a chip card), or something the user is (*i.e.*, a user biometric). Each of these three single-factor authentication methods has limitations: what the user knows may be guessed, forgotten, or shared with others; what the user has may be costly, faulty, lost, stolen, or replicated; and what the user is cannot be revoked, does not permit privacy, and requires human involvement. To strengthen the process of identity authentication, several single-factor methods may be combined, resulting in multi-factor authentication.

When authenticating claimed identities, relying parties implicitly place trust in the authenticity of the clues themselves. In cases where relying parties are effectively agents or proxies of the user, users can be trusted with creating their own identifiers and using their own authentication method. For example, a self-generated username and password suffice to protect access to one's own computer.

In many communication and transaction systems, however, the security interests of relying parties and users are not aligned. In these cases, relying parties require clues that originate from trusted parties, also referred to as *issuers*, in order to gain confidence in the authenticity of presented user identifiers.

Users and relying parties each have their own security and privacy concerns at stake with respect to how user identifiers are formed and protected. In the next three parts, I analyse and contrast the privacy and security implications of three fundamentally different approaches to user identification.

*

3. NON-CERTIFIED USER IDENTIFIERS

NON-CERTIFIED USER IDENTIFIERS are generated by their own users (or by computers that represent them). Self-generated online usernames for use in chat rooms are an example.

3.1. Privacy of Users

User privacy is about the ability of users to minimize what user-related information relying parties can learn beyond what users choose to explicitly disclose. The difficulty of statistically correlating presented user identifiers is proportional to the quality of the randomness used to generate them. To maximize privacy within a designated context, identifiers must be generated independently and uniformly at random from the set of identifiers used within that context. Randomly self-generated user identifiers cannot be *linked* to other identifiers of the same user and cannot be *traced* to the user's real identity.

Of course, any user-related information that relying parties can capture when user identifiers are presented increases their linking and tracing capabilities; this includes not only any personal information that users choose to disclose when presenting their identifiers, but also any circumstantial information (such as the user's location and communication device characteristics) that relying parties may be able to capture.

By minimizing the disclosure of additional information (especially of invariant personal information) and by using different self-generated identifiers in different contexts, users can minimize what relying parties can infer about them from the information they choose to disclose.

3.2. Security of Users

What prevents another party from presenting someone else's user identifier? Generating user identifiers at random from a large name space reduces the risk of accidental identifier collisions, but does not address interception of presented identifiers nor misuse by relying parties. In many contexts, impersonation constitutes an unacceptable security risk to users; at the very least, users should have control over such delegated use.

Secure transport of user identifiers (*e.g.*, physical protection of a plastic user token or encryption of an electronic identifier at the data transport layer) suffices to protect against outsiders who attempt to steal or copy identifiers in transit, but does not stop relying parties from replaying presented user identifiers at other relying parties for their own benefit.[2] Even if there is only a single relying party, from the perspective of the user it may be unacceptable if that relying party is able to cause a false entry in an audit log by replaying a presented user identifier.

The most secure way to counter replay attacks is for users to bind their self-generated identifiers to something that cannot be stolen or copied and that must be verified at presentation time. One effective technique to accomplish this is to digitally sign each self-generated identifier with the private key of a randomly self-generated one-time key pair. Correspondingly, relying parties should request that any presentation of these user identifiers be accompanied by a response (generated with the private key associated with the identifier) to

2. This impersonation attack may or may not constitute a security threat to relying parties themselves.

a fresh challenge message. A side benefit of these identifiers is that users can digitally sign documents as well as messages representative of their actions with respect to presented identifiers. For example, each time when withdrawing money from a bank account the account holder could sign the details of the withdrawal; this enables the user to detect and refute any fraudulent account withdrawals by hackers or insiders.

3.3. Security of Relying Parties

While non-certified self-generated identifiers offer privacy and security to their users, they offer no security to relying parties. Whether or not relying parties have security concerns of their own depends on the designated context and purpose(s) of presented user identifiers. Most transactions involve some kind of security risk for relying parties relating to user authentication, especially applications that rely on access control or secure data sharing between relying parties; in these cases, presented identifiers are relied on as secure account indices.

Even when user identifiers serve merely as contact addresses, liability issues may arise for relying parties that deliver information or goods to the wrong users. More generally, relying parties cannot securely rely on non-certified user identifiers regardless of their context and purpose, for the following reasons:

- Anyone can self-generate and present user identifiers;
- Users can self-generate any number of uncorrelated identifiers; and,
- Users can arbitrarily share self-generated identifiers with others.

As a consequence, relying parties cannot securely perform any of the purposes of user identifiers—they are completely dependent on the honesty of users.

*

4. CERTIFIED USER IDENTIFIERS

CERTIFIED USER IDENTIFIERS ARE, as the name implies, endowed by or on behalf of relying parties with one or more *security guarantees* and optional *attributes* (*i.e.*, information of any kind). Typical security guarantees are:

- The format of the identifier has been approved in some sense;
- The identifier is unique within the specified context;
- The user identifier cannot be transferred to other parties (non-transferability);
- The user of the identifier has been approved in some sense;
- The identifier cannot be cloned because it is locked up in a tamper-resistant device;
- No more than x certified identifiers have been issued to the user; and,
- The "real identity" of the user has been associated with the identifier.

During the certification process, a certified user identifier may optionally be bound to one or more attributes, such as:

- An expiry date or a more general specification of the lifetime of the identifier;

- The maximum number of authorized uses;
- The designated use context;
- One or more designated purposes;
- Details about the strength of the certification process; and,
- Personal information relating to the user.

Attributes can be certified along with the identifier itself or be stored in an online account indexed by an account pointer that is certified along with the identifier. The latter approach allows for dynamic attributes, which change over the lifetime of the token. The former approach is suitable only for static attributes, but has the advantage of allowing relying parties to verify attributes off-line at presentation time.

4.1. Security of Relying Parties

When a user identifier is issued to a user by the relying party itself, its security guarantees and any optional attributes need not be tied to the identifier itself; instead, the relying party can simply keep track of which security guarantees and attributes are associated with the identifier and do a local look-up when presented later on with the identifier.

When user identifiers are relied on by parties other than their issuer, their certification is accomplished by specifying them on *identification tokens* that bear *authenticity marks* that are hard to forge. Authenticity marks are made as follows:

- For non-electronic user identifiers that are specified on paper documents, plastic cards, and other physical tokens, authenticity marks take the form of seals, handwritten signatures, intaglio printing, special paper, watermarks, security threads, color-shifting ink, holograms, and so on. Well-known examples of non-electronic certified identifiers are passports, health insurance cards, driver's licences, employee access cards, credit cards, and debit cards.
- For electronic user identifiers, authenticity marks can be established in one of two ways. If relying parties trust the tamper-resistance of the user's computing device, the user identifier and any attributes can be stored in the device together with a secret key that is used to authenticate the presentation of the identifier and its associated attributes. If the user's device is not tamper-resistant, the identifier and any associated attributes must be *cryptographically certified* by a message authentication code or a digital signature. In practice, tamper-resistance and cryptographic certification may be combined to ensure a degree of fall-back security in case one of the two mechanisms gets compromised.

To ensure non-transferability, at token-issuing time one or more biometrical characteristics must also be certified. Photographs and handwritten signatures are the predominant methods, but more secure methods such as fingerprints, hand geometry scans, and iris scans are coming into vogue. Relying parties can compare these certified biometric characteristics with fresh samples taken at

presentation time. This technique applies to both non-electronic and electronic identifiers; in the latter case, the biometric clues can be stored within the device chip itself. To prevent a user from replaying a biometric scan of another user, users and relying parties must be in physical proximity.[3]

Relying parties can also ensure a degree of *non-repudiability* by requiring users to sign documents as well as statements that attest to the actions they undertake with respect to their certified identifiers:

- For non-electronic identifiers, relying parties can compare fresh handwritten signatures with signatures that appear on documents issued by trusted third parties.
- For electronic identifiers, users can be asked to digitally sign data. This requires user identifiers to be bound at certification time to user-generated asymmetric key pairs. At presentation time, users must present their certified identifier together with their public key, and use their private key to digitally sign messages. X.509 certificates are a well-known example of this approach.

In addition, when the presentation of user identifiers requires physical proximity, relying parties may also rely on witnesses, photographs, and sound recordings.

4.2. Security and Privacy of Users

While certified identifiers offer all kinds of protection to relying parties vis-à-vis users, they offer no privacy and security for users:

- All actions of users can be traced and linked on the basis of their identifiers, particularly when users must strongly identify themselves at token issuance time (as required for non-transferability and other security guarantees, as well as for the specification of vetted user-related attributes);
- Relying parties can make discriminatory service decisions for any user on the basis of any account information that they can associate with that user;
- Relying parties can arbitrarily blacklist targeted users to lock them out of services;
- Issuers can impersonate targeted users by forging certified identifiers under their names, possibly without risk of detection;
- Any attribute information that is certified along with a user identifier can be used to discriminate against its user.

Users may not be overly concerned about these powers when their certified identifiers are relied on only by one service or a few services within the same organizational domain. When the same user identifiers are relied on across many services, especially across organizational domains, there is much more reason for concern.

---

3. One way around this is to rely on a tamper-resistant biometric scanner at the user's end that is able to detect replay. Alternatively, for biometric voice scanning the user can be asked from remote to utter a fresh challenge word or sentence. Both methods have significant security and flexibility limitations.

*

5. A MODERN APPROACH TO USER IDENTIFICATION

THE THIRD AND LAST CATEGORY of technologies for user identification is the result of advances in modern cryptography. This work has its roots in the seminal work of David Chaum in 1980s on the paradigm of "security without identification."[4] Chaum's so-called blind signatures are of no interest, however, to relying parties that require the strong security guarantees of certified identifiers. For example, blind signatures cannot be revoked, are unsuitable to convey optional attributes, and cannot be tied to tamper-resistant user devices without sacrificing their privacy guarantees.

Nevertheless, Chaum's work has been a valuable source of inspiration for cryptographic advances in the 1990s that have succeeded in closing the gap between user privacy and security for relying parties. This work has resulted in user identification methods that combine all the benefits of the best self-generated user identifiers (as described in Part 3) with all the benefits of the strongest possible certified user identifiers (as described in Part 4), while eliminating all of their drawbacks.

The basic idea is to securely embed a "master identifier" that is unique to that particular user into all of a user's ID Tokens and to do so in such a manner that (1) the user can present his ID Tokens to relying parties without disclosing any information about the embedded master identifier, while (2) relying parties can leverage the presence of the embedded master identifier to achieve any of the security guarantees of conventional certified identifiers. Relying parties can securely leverage the presence of embedded master identifiers in various ways without learning them. For example:

- All of a user's ID Tokens can be blacklisted on the basis of any one of them without knowing their embedded master identifier, while fully preserving the unlinkability and untraceability of the ID Tokens.
- By appending user-confidential information to the embedded master identifier, the user can be discouraged from lending or copying his ID Tokens. Namely, it is not possible to present an ID Token without knowing the embedded master identifier.
- Users can selectively disclose any property of the attributes that have been bound to their ID Tokens, while unconditionally hiding any other information.
- An ID Token can be issued in such a manner that its embedded identifiers can be computed by third parties if and only if the ID Token is presented more than once.

---

4. David Chaum presents a non-technical overview in the following works: David Chaum, "Security Without Identification: Card Computer to Make Big Brother Obsolete" (1985) 28:10 Communications of the ACM 1030, <http://www.chaum.com/articles/Security_Wthout_Identification.htm> and David Chaum, "Achieving Electronic Privacy" (1992) 267:2 Scientific American 96, <http://www.chaum.com/articles/Achieving_Electronic_Privacy.htm>.

- When presenting an ID Token, its user can (on a voluntary per-case basis) attach a verifiable encryption of the embedded master identifier to enable a designated "escrow agency" to retroactively trace the user if needed.
- Relying parties can securely share *unlinkable* account information that pertains to the same user by directing assertions about account data in protected form through the user. Protected account assertions cannot be transferred among colluding users en route: all protected assertions embed the master identifier of the ID Token that is hooked up to the source account, and this link is verified at the destination account.[5]

In the non-technical overview that follows below, the new user identifiers are referred to as *ID Tokens*. Figures 1 to 6 illustrate how ID Tokens can be used to build a secure yet privacy-preserving *identity management* infrastructure. The figures show a user, called Bob, using a smartcard to access services.

5. For an in-depth technical overview of the new paradigm and hundreds of relevant literature references, see Stefan A. Brands, *Rethinking Public Key Infrastructure and Digital Certificates: Building in Privacy* (Cambridge: MIT Press: 2000), <http://www.credentica.com/technology/book.html>.

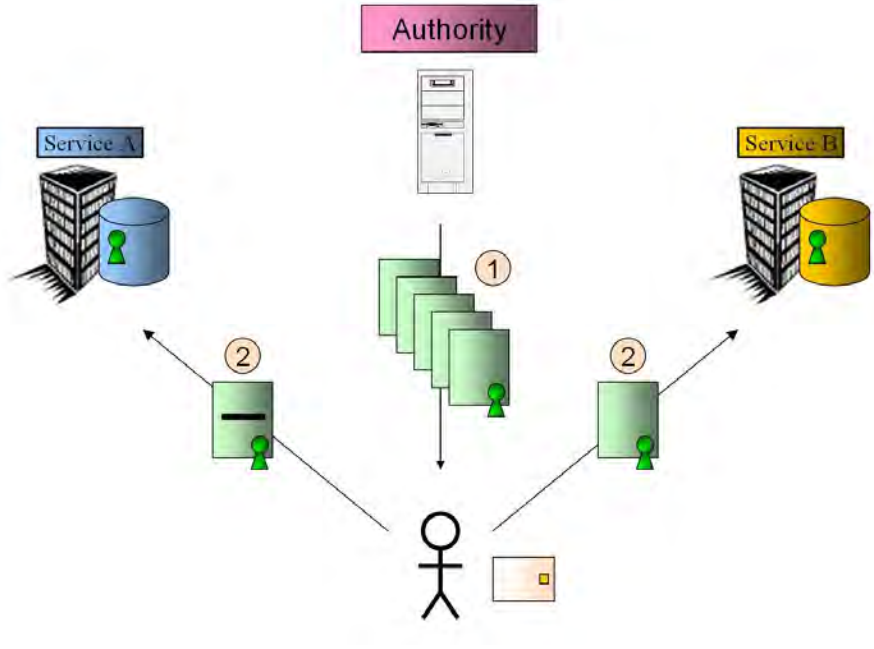

Figure 1: **Step 1:** In an enrollment phase, Bob's smartcard retrieves ID Tokens from an Authority that is trusted by the services that Bob plans to access. The Authority endows Bob's ID Tokens with relevant security guarantees and cryptographically embeds a unique random number into all of them. The Authority may also endow Bob's ID Tokens with attribute information to enable service providers to make better informed decisions about Bob. The Authority never gets to see the certified identifiers that Bob obtains; from a privacy perspective, each of Bob's identifiers is the equivalent of a unique randomly self-generated number. **Step 2:** The first time that Bob accesses a service, his smartcard transmits a fresh ID Token to the service. In doing so, Bob's card can selectively hide any irrelevant attribute information that the Authority may have tied to the D Token. The invisibly embedded number remains unconditionally hidden. Bob's card uses a different ID Token at each service.

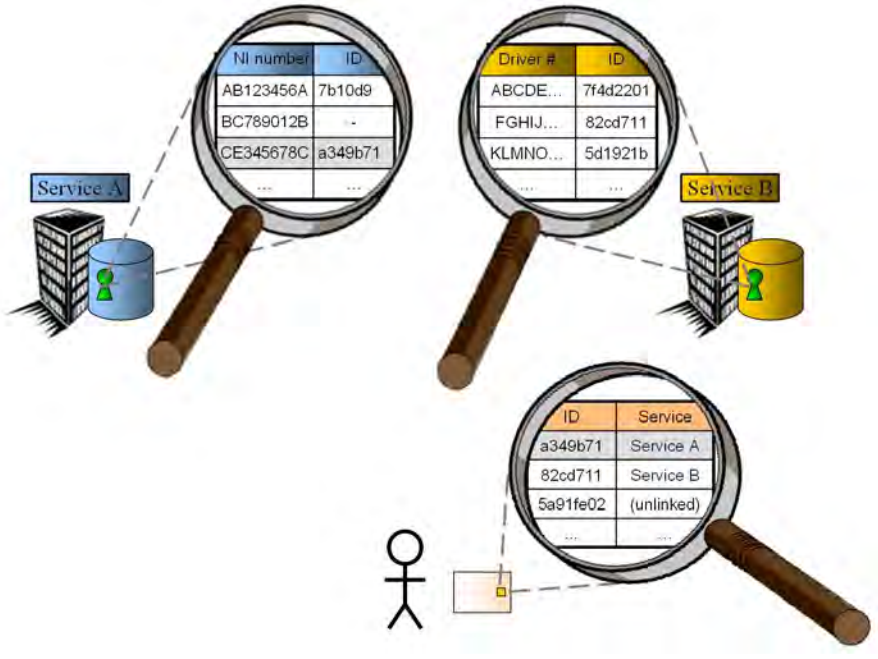

Figure 2: Each service associates the certified identifier it receives from Bob with the legacy account information it holds on Bob (indexed, in this example, by his national insurance number at Service A and by his driver number at Service B). Likewise, Bob's smartcard keeps track of which user identifier it has hooked up to which service. Because Bob's identifiers are the equivalent of randomly self-generated numbers, they are unlinkable and untraceable. As a result, Service A, Service B, and the Authority (even when pooling all their data) do not gain any linking, tracing, and profiling powers over Bob. In effect, the services have simply associated a fresh random number to their pre-existing accounts on Bob. Service A and B continue to know Bob exactly as they used to know him—under his national insurance number and his driver number, respectively. However, through the embedded master identifier that is invisibly present in each of Bob's ID Tokens, Bob's accounts with Service A and B are now securely connected.

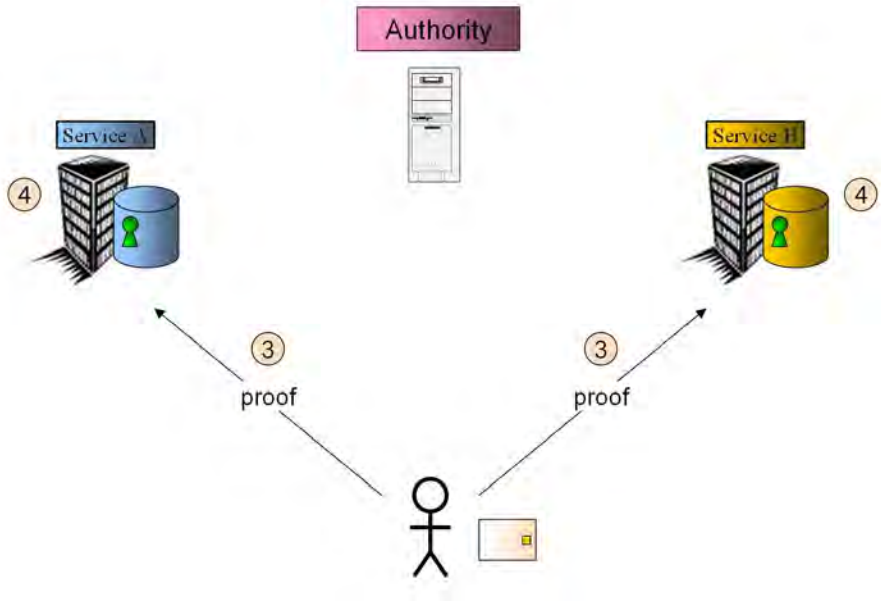

Figure 3: **Steps 3 and 4:** In subsequent visits to a service, Bob's smartcard authenticates to that service using the identifier associated with his user account at that service. To authenticate, Bob's smartcard generates a cryptographic proof-of-possession of a private key that corresponds to the hooked-up identifier. This proof cannot be forged by anyone—generating a proof requires knowledge of the identifier's private key, which never leaves Bob's smartcard. The service verifies Bob's authenticity by verifying the submitted cryptographic proof. Assuming Bob needs to authenticate to his own smartcard only once in order to operate the card, Bob enjoys the convenience of a single sign-on experience at the various services.

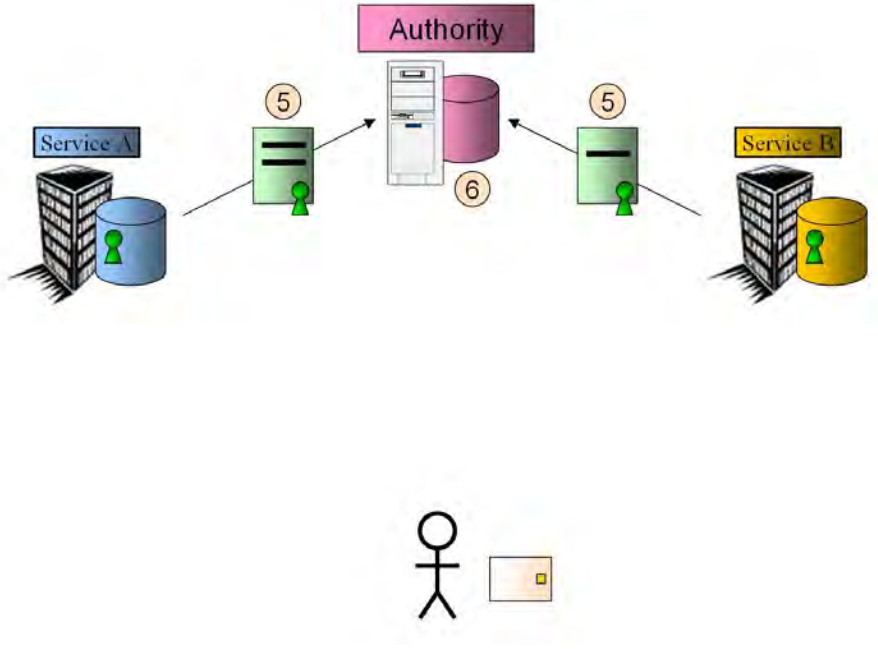

Figure 4: This figure shows how, on top of the basic system in Figures 1, 2, and 3, the services can collect non-repudiable audit trails that are not privacy invasive. **Step 5:** Services A and B can forward non-repudiable digital audit trails to the Authority (or any other auditing body); they can capture these whenever Bob interacts with them using the appropriate ID Tokens. The services can optionally censor the audit data prior to forwarding it, so as to protect Bob's privacy interests or their own privacy and autonomy interests vis-à-vis auditors. **Step 6:** The Authority can keep the audit trails and verify the validity of transactions. In case of a dispute, censored data can be uncensored with the help of the appropriate services. The Authority cannot trace and link the actions of Bob across the services on the basis of the non-repudiable audit trails, unless Bob has chosen to specifically enable this. (Bob can decide to do so on a per-transaction basis.)

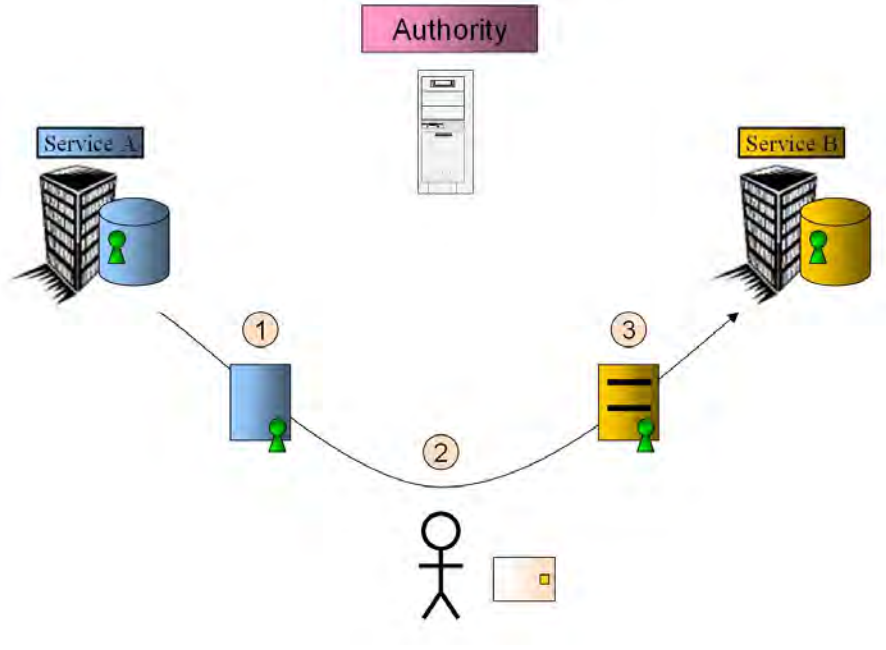

Figure 5: This figure shows how, on top of the basic system outlined in Figures 1, 2, and 3, the services can securely share account data about Bob, even though they do not know him under the same identifier. **Step 1:** Service A makes an "assertion" about Bob based on the account information it has on Bob, and sends the assertion in digitally protected form to Bob (or a representative that Bob may designate on a case-by-case basis). **Step 2:** Bob's smartcard sanitizes the protected assertion by "randomizing" any information that would otherwise lead to an increase in linking and profiling powers by Service A and B over him. **Step 3:** Bob's smartcard selectively discloses to Service B the minimal assertion information needed by Service B. For example, if Service A has asserted Bob's city of residence and street name, Bob could disclose to Service B only the fact that his city is in a certain region of the country, without revealing the city name. Service B can verify on its own the origin of the data (Service A), its integrity (Bob did not modify it), and the fact that it relates to Bob—even though Service A and Service B do not know Bob under the same identifier. The fact that the transferred assertion information truly pertains to Bob and has not been transferred by Bob to another user can be verified by leveraging the presence of the embedded master identifier that is invisibly present in all of Bob's certified identifiers.

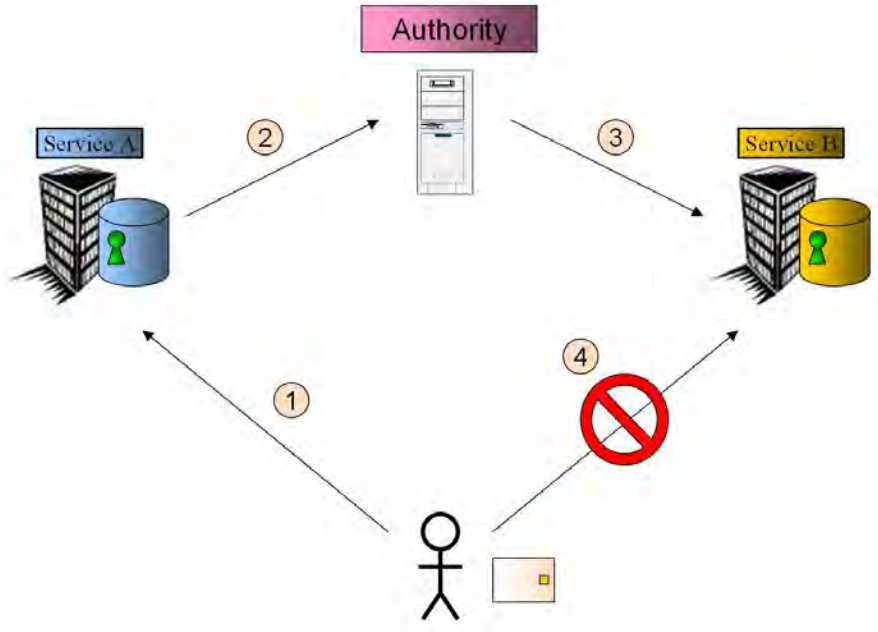

Figure 6: This figure shows how, on top of the basic system outlined in Figures 1, 2, and 3, Service B could revoke access to Bob in case Bob commits a fraud at Service A, even though they cannot correlate the identities they manage in their own domains. **Step 1:** Suppose Bob commits a fraud at Service A, and it is legitimate to be able to deny Bob access to Service B. **Step 2:** Service A informs the Authority about the fraud, and provides revocation information associated with the certified identifier that Bob uses at Service A. **Step 3:** The Authority broadcasts the revocation information to Service B and any other service for which cross-domain blacklisting has explicitly been enabled. **Step 4:** When Bob tries to access Service B he can be denied access, even though he is known under a different unlinkable user identifier at that service. This is accomplished by leveraging the master identifier that is invisibly present in Bob's user identifiers at Service A and Service B.

The cross-domain revocation feature described in Figure 6 does not impinge on Bob's privacy. Firstly, cross-domain revocation by Service A and Service B is possible only if Bob's certified identifiers at these services have been issued in a manner that requires the *cooperation* of Bob's smartcard. For services that have no legitimate reason to be able to revoke Bob's access across their domains, Bob's card would hook up ID Tokens that do *not* allow for cross-domain revocation. Secondly, in order for Service B in Figure 6 to be able to deny access to Bob in case Bob has abused his access rights at Service A, Service B must ask *each* user that requests access to its service to submit a cryptographic proof that the invisibly embedded number in their ID Token with Service B is *different* from the revoked embedded number. If the embedded number of an access requestor is on the blacklist, no valid cryptographic proof can be created. Other important points to note are: (a) entries on the revocation list are meaningless random numbers to everyone, (b) the list of revoked numbers *must* be sent to each user who is requesting access, and so each user sees that they are asked to prove they are *not* on the list, and (c) proving that one is *not* on the revocation list does *not* invade one's privacy.

*

6. FINAL WORD

CONTRARY TO POPULAR MISBELIEF, identification and privacy are not opposite interests that need to be balanced. Advances in modern cryptography allow for the construction of compact user identifiers that combine all the benefits of non-certified self-generated identifiers with those of certified user identifiers while eliminating all of their respective drawbacks.

It may be too much to ask that legislators, systems designers, and privacy activists intimately familiarize themselves with these modern technologies for user identification. However, it is important that they take note of their capabilities, in order to avoid stretching preconceived notions about identification and privacy that hold true in the physical world into the electronic world, where they no longer hold.