International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 12 (2020), pp. 4344-4362
© International Research Publication House. http://www.irphouse.com

Aken Infrastructure: The Abstract Model of a User Authentication and Attribute Sharing Infrastructure for the Cyberspace of the Future

Tibor Roskó*

University of Debrecen, Faculty of Informatics, Egyetem tér 1, 4032 Hungary.

Abstract

Nowadays, almost everything happens through the Internet, such as shopping, online banking, getting informed, or keeping the contact, which caused that the number of cybercrimes has extremely been increased in the last few years. Unfortunately, the behaviour of service providers does not help in the fight against these attacks because a significant part of them do not perform two-factor authentication; if so, then SMS-based 2FA is the most common, which is considered as a do not use solution by the NIST. But the biggest problem is that in almost each registration process, users have to register local accounts with unique username and password pairs; and only less than ten percent of service providers support the Single Sign-on method. So, in the AKEN Infrastructure designed by us, each user has only one account to eliminate password memorization problems, but the Infrastructure forces all authentication providers to apply NIST AAL3-compliant authentication methods ensuring high-level security. Along ensuring high-level security, supporting those services which, beyond that they are also emphasized by the GDPR, are the key elements in such ecosystems was also in our mind. For example, automatized data erasure service, former data migration service, or shared attribute monitoring service. To implement these, we have designed several new methods, in part by extending the existing solutions, such as STORK, eduGAIN, Yoti ID, and Google Sig-in. The protocol of managing human DNA-profiles and the model of the UID are the two most important of them, which are the basis of the other services, such as identity verification, uniqueness checking, automatized data erasure, or former data migration.

Keywords: Global User Authentication, Human DNA-Based Identity Verification, Shared Attributes Monitoring, Identity Management, Eid

I. INTRODUCTION

Nowadays, almost everything happens through the Internet, such as shopping, online banking, getting informed, or keeping the contact, which caused that the number of cybercrimes has extremely been increased in the last few years. Based on the [1] report, human element attacks – such as identity theft and phishing – and human errors – such as information leakage – are the most significant elements.

Unfortunately, the behaviour of service providers does not help in the fight against these attacks.

However, there are several great methods for user authentication – such as STORK, eduGAIN, Yoti ID, or Google Sig-in –, there are two main common problems with these solutions. The first one is that users have to register local accounts with unique username and password pairs in almost each registration process (less than 20.00 percent of the service providers allow registration without a local username and password [34] resulting in that users have to manage a significantly huge number of passwords. Which might cause vulnerabilities – such as passwords in a post-it, regularly requested password resets, or use of untrusted password manager applications – due to about 80.00 percent of the users have found memory limitations in the case of using more than three passwords – such as forgetting or mixing them – highlighted by the results of the paper [3]. The second one is that less than ten percent of the inspected services perform two-factor authentication – based on our study [34], in which we inspected one hundred big and small online services – which did not meet the requirements of NIST AAL3 because they applied SMS-based or mobil application-based 2FA; SMS-based 2FA is considered as a non-recommended (do not use) method by the NIST. And in some cases, the SSO-bypassing appears as a very dangerous/vulnerable practice because some local services force the user to locally enter login data instead of redirecting the user to the SSO endpoint for a login. In this case, a malicious application might steal the login data.

To develop better solutions and avoid these problems, we might increase the number of SSO-supporting local services which prefer centralized logins instead of local authentications or local login data storing, increase the security with applying NIST AAL3-compliant 2FA solutions, such as tokens, chip cards, or any hardware security keys.

All of these reasons led us to design the abstract model of a user authentication and attribute sharing infrastructure which defines a unified protocol to support the collaboration among existing systems without creating unnecessary new authentication services. So, in the AKEN Infrastructure designed by us, each user has only one account to eliminate password memorization problems, but the Infrastructure forces all authentication providers to apply AAL3-compliant authentication methods ensuring high-level security. Along ensuring high-level security, supporting those services which, beyond that they are also emphasized by the GDPR, are the key elements in such ecosystems was also in our mind. For example, automatized data erasure service, former data migration service, or shared attribute monitoring service. To implement these, we have designed several new methods, in part by extending the existing solutions, such as STORK, eduGAIN, Yoti ID, and Google Sig-in. The protocol of managing human DNA-profiles and the model of the UID are the two most important of them, which are the basis of the other services, such as identity verification, uniqueness checking, automatized data erasure, or former data migration. The Infrastructure can be applied to authenticate users in small webshops or even in high-level security solutions, such as government applications.

In subsequent sections, this paper gives a detailed description of the main methods – such as the protocol of managing human DNA-profiles, the model of the UID, identity verification, uniqueness checking, automatized data erasure, or former data migration services – of the Infrastructure developed by us with emphasis on the abstract model, and make guiding suggestions in the field of practical application only for key implementations, for instance, standards for facial image (International Civil Aviation Organization, 9303 part 3 and ISO 19794-5), date format (ISO 8601), and character set (International Civil Aviation Organization, 9303 part 3).

All the terms and abbreviations used in our paper are defined in Appendix A and Appendix B.

The goal of our fundamental research is to design the abstract model of a user authentication and attribute sharing infrastructure which defines a unified protocol to support the collaboration among existing systems. We aimed to design and describe all the functions, methods, and recommendations that build up the Infrastructure, such as the protocol of managing human DNA-profiles, the model of the UID, identity verification, uniqueness checking, automatized data erasure, or former data migration services. It is very important to emphasize that our Infrastructure is not a new X plus one authentication solution but a framework for cooperation among existing services.

II. MATERIALS AND METHODS

In this paper, we applied mixed research methods, like qualitative and quantitative methods as well. As the first step, we made observations to analyze the mechanisms of action of the existing solutions – such as STORK, eduGAIN, Yoti ID, and Google Sig-in –, and identified the deficiencies of them, such as missing NIST AAL3-compliant 2FA, identifiers leaking personal data, forced local login data, or bypassed SSO. And, we analyzed existing standards to build the recommendations to the Infrastructure, such as requirements of user portraits, the GDPR, and NIST AAL3 or the recommendations for a common writing and date system.

In our paper [34], we made an analysis to meet how many services require obligatory use of 2FA and how many of them meet the requirements of NIST AAL3. In which, we inspected randomly selected one hundred websites (from small webshops to big banks) in the first half of 2019.

To validate the need for such an infrastructure and its security, we inspected the results of the survey in the paper [3] and using these results we proved our hypotheses detailed in Section 9.

Reducing the vulnerability of centralized authentication, avoiding the SSO-bypass, and reducing the number of passwords are the most important cases that require robust solutions. Besides, to solve the problem of insecure third-party participants – such as service providers – is also an important question.

As mentioned above, in developing the Infrastructure the best properties of the existing systems have been integrated as much as possible. Three projects should be described here, the approaches of which have been integrated into our Infrastructure.

The objective of the STORK European personal identification platform is to promote the cross-border cooperation of acceding, primarily European, countries. The project developed the model of electronic authentication by focusing on practical implementation. A given person has a single identifier, which identifies the person in the Federation; its format is "NC/NC/xxxx...". The identifier consists of three components: the first NC is the identifier of the person's country of origin, the second NC is the identifier of the country visited, while the third element is the person's unique identifier. In the framework of the eduGAIN project, authentication providers in the fields of research and education have been connected into a common system. As a result, connected services can be used by users of authentication providers available in the eduGAIN system.

In terms of practical implementation, both projects are built on SAML 2.0 SSO, which is to ensure user authentication and the sharing of attributes.

The cooperation of these two projects was made possible by the model developed by researchers at the University of Murcia (Spain). In this cooperation, STORK and eduGAIN users can both use services pertaining to the other platform. Interoperability is ensured by an intermediate point, which converts the attributes of the user managed on the STORK platform into a format that can be used on the eduGAIN platform. Figure 1 illustrates the collaboration enabled by the intermediate point between STORK and eduGAIN platforms. If passage between the two platforms is necessary, then it is done through the intermediate point, and when using an in-platform service without the intermediate point, then it is performed 'in-house'.

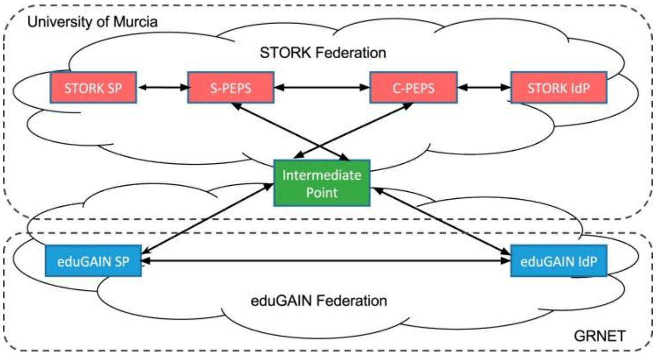


Figure 1. Model of cooperation between STORK and eduGAIN [5]

Yoti ID project is a digital identity solution developed in 2014 and is currently being used in the United Kingdom and the State of Jersey for the verification of identity [33]. The application cooperates with an Indian dating app called TrulyMadly to verify the users' identity [42]. Registering users based on identity documents such as passports, personal identity cards, driving licenses, and biometric identifiers allows the creation of an authentic user profile [44]. The registration process is demonstrated in Figure 2.

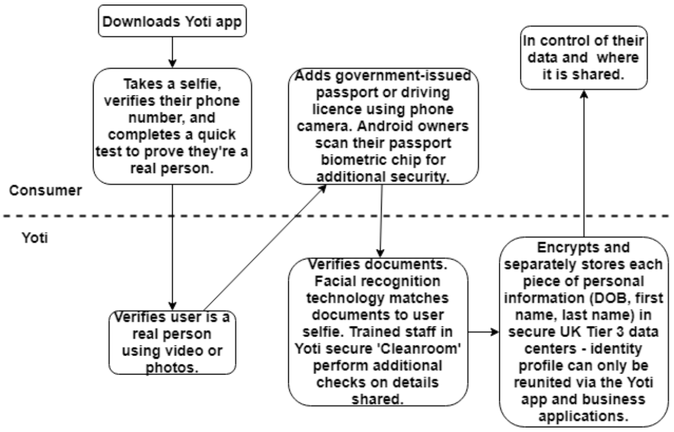

Downloads Yoti app
Takes a selfie, verifies their phone number, and completes a quick test to prove they're a real person.
Adds government-issued passport or driving licence using phone camera. Android owners scan their passport biometric chip for additional security.
In control of their data and where it is shared.
Consumer
Yoti
Verifies user is a real person using video or photos.
Verifies documents. Facial recognition technology matches documents to user selfie. Trained staff in Yoti secure 'Cleanroom' perform additional checks on details shared.
Encrypts and separately stores each piece of personal information (DOB, first name, last name) in secure UK Tier 3 data centers - identity profile can only be reunited via the Yoti app and business applications.

Figure 2. Yoti registration process [44]

Having downloaded the Yoti application, the user enters his/her telephone number and uploads his/her profile picture. After confirming the telephone number entered, the Yoti central system verifies if registration was initiated by a real person. Then, the user uploads a picture of his/her identity document, which is verified by the Yoti central system using algorithms and human resources, and, upon approval, the user profile is confirmed. The scope of documents accepted by the Yoti system is primarily limited to the passport, but leaders of some countries (e.g. Croatia, Hungary, Hong Kong) also accept driving licenses and identity cards [43]. The Yoti center protects stored information using AES-256 encryption, the decryption key is encrypted with the user's PIN code, hence the user can only have access to these data [45]. The process of user authentication and attribute sharing is demonstrated in Figure 3.

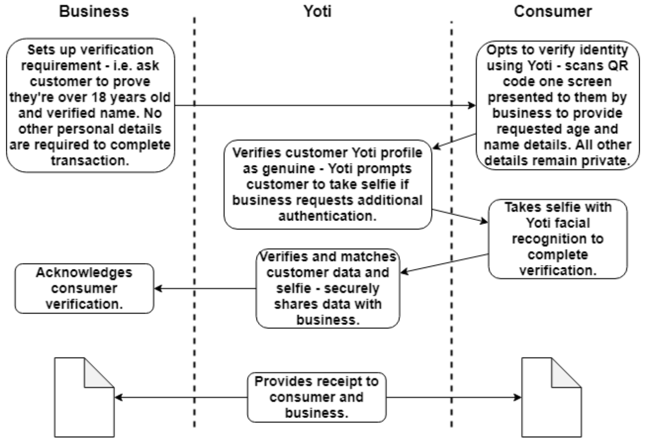

Business
Yoti
Consumer
Sets up verification requirement - i.e. ask customer to prove they're over 18 years old and verified name. No other personal details are required to complete transaction.
Opts to verify identity using Yoti - scans QR code one screen presented to them by business to provide requested age and name details. All other details remain private.
Verifies customer Yoti profile as genuine - Yoti prompts customer to take selfie if business requests additional authentication.
Takes selfie with Yoti facial recognition to complete verification.
Verifies and matches customer data and selfie - securely shares data with business.
Acknowledges consumer verification.
Provides receipt to consumer and business.

Figure 3. The process of user authentication and attribute sharing in Yoti [10]

The process of authentication is built on the general authentication scenario with the participation of three actors: user, service provider, and Yoti authentication provider. The Service Provider creates a confirmation request, for which the user can respond through the mobile application after scanning the QR code. The Yoti system requests the user to authenticate himself/herself by entering his/her PIN code or using facial image-based identification, and if the authentication was successful, then the Service Provider can gain access to the requested attributes and authenticate the user. Both the user and the service provider receives a confirmation on the delivery of attributes, which the user can access from his/her Yoti account, and it contains what attributes were shared when and with whom [10].

The service provided by Google is similar to the ones described above, and it combines centralized authentication with attribute sharing and monitoring. The only deficiency of the system is that registered user accounts are not authentic. The platform developed by Google uses the OAuth 2.0 protocol as opposed to the two solutions above that use the SAML 2.0 protocol. In addition to authentication, OAuth protocol manages access to resources, meaning that it also performs authorization. In this case, OAuth controls the access rights of service providers to get attributes stored in a Google account. The process of authentication and attribute sharing, which is demonstrated in Figure 4 (https://developers.google.com/identity/protocols/OAuth2).

1. The local Service Provider redirects the user to the Google Sign-in interface.
2. The user signs into his/her Google account.
1. If necessary, the user permits the application to have access to his/her attributes.
3. The user is redirected to the local Service Provider with an access token.
4. The local Service Provider redeems the access token in the Google system and gains access to the user's shared attributes.

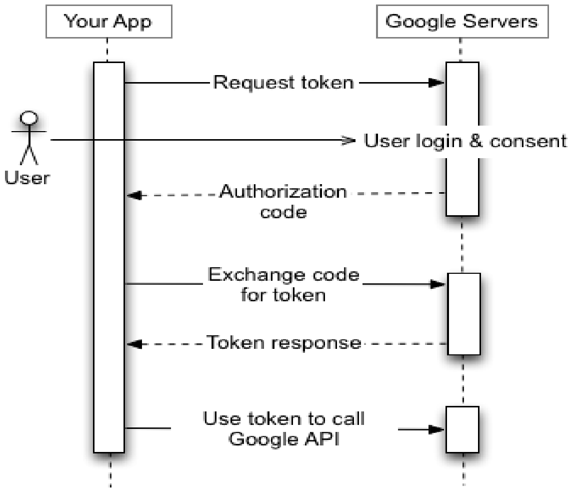


Figure 4. Google Sign-in OAuth authentication [7]

The above-mentioned solutions show that the centralization of user authentication and the creation of an authentic user account are not new ideas, but as mentioned before, the primary issue is that these developments only implement single components separately out of the policies defined by us in Section 6.1. The concept of centralization is almost always limited to some local area, hence it cannot be ensured for the user that services are used with only one user account, instead, separate user accounts need to be used for each area. By integrating the prominent properties of existing developments into our Infrastructure, we strove to ensure that the user only needs one single user account to use services, regardless of the focus area, such as government, finances, webshop, education, or community portal. These prominent properties are summarized in Table 1. It is important to emphasize that existing solutions in themselves are absolutely correct developments in a specific environment, however, as noted above, they can only partially fulfill or not at all fulfill the requirements set out above (also that meet today's expectations). But, it is also important to be highlighted that the STORK user identifier carries additional information about the user, such as the identifier of the user's country of origin, the identifier of the country visited virtually, and the user's unique identifier which is unique in the country of origin. This can be as problematic as the Hungarian personal identification number which is covered in detail in our paper [14].

Table 1. The summary of prominent properties of the existing solutions

	STORK-eduGAIN	Yoti ID	Google Sign-in
document-based identity verification	x	x	
biometric identity verification		x	
attribute sharing is monitored		x	x
access to a shared attribute can be withdrawn			x
encrypted data storage		x	

III. RESULTS AND DISCUSSION

In this section, we describe our results divided into seven main subsections, such as requirements, standard and method recommendations, the process of a user account registration, the process of authentication and attribute sharing, the process of user attributes update, the process of a user account deletion, and the process of former data import. The main results and impacts are summarized in a single section below the detailed descriptions.

To connect to the Infrastructure, all participants must fulfill the following requirements which are also expanded with some standard recommendations

All participants, such as Authentication Provider, Service Provider, National Identity Verifier, DNA laboratories, and the Infrastructure must have an Extended Validation SSL certificate.

A National Identity Verifier must fulfill the following criteria:

1. It has to be able to answer the following questions:
   i. Does a person with the given 4N attributes and DNA-profile exist?
   ii. Is the given DNA-profile unique?
   iii. Does a person with the given 4N attributes exist?
2. It must have a method that returns the portrait of the person to whom the given 4N attributes belong (if it exists).

An Authentication Provider must fulfill the following criteria:

1. for the user account registration
   i. Fulfills the requirements of the NIST AAL3.
   ii. Fulfills the standard recommendations described in Section 6.2.
   iii. Meets the principles of the NIST 800-63B on Memorized Secrets.
2. It must have functions to answer the following questions – in the case of uniqueness ck. –
   i. Does the given UID exist?
   ii. Does the given DNA-profile hash exist?
   iii. Returns all portraits that belong to the given DNA-profile hash.
3. for the user account
   1. It must keep a detailed list of shared attributes (active and archive) to ensure the user following what attributes shared with whom.
   2. It cooperates with the automatized data erasure services of specific service providers.

      For the user account deletion, it must have functions to do the following:
   3. generate a former data migration certificate
   4. register this certificate
   5. perform former data import

A Service Provider must fulfill the following criteria:

1. It must use a UID converter table.
2. It must ensure/perform the former and current UID association.
3. It must perform the automatized data erasure service management in cooperation with specific authentication providers.

In this section, we are going to give some common standard recommendations for mistakeless cooperation among all parties. It is important to emphasize that these are not strict requirements but only recommendations for possible common frameworks to avoid incompatibilities, such as different dates, methods, or spellings.

In consideration of the international cooperation, we wish to propose a recommendation for the textual attributes of personal information, such as name, country name based on standard ICAO 9303, which defines the standard issuance of international machine-readable travel documents [4]. The primary objective of the recommendation is to establish a common writing system for all participants: user, Authentication Provider, and Service Provider. To standardize differences between various languages and character sets, ICAO recommends the use of the Latin alphabet (A-Z) and Arabic numbers 0-9. The standard specifically recommends the transliteration of the national Latin-based alphabet to the basic Latin alphabet, which does not contain accents or special characters, such as Đ, Ł, ä, ő.

In the course of transliterating the national Latin-based alphabet, accented characters are omitted and converted into non-accented characters, which is displayed as some sort of categorization, for instance, A=A, Á, Â, A. The loss of information resulting from this conversion may substantially limit the unequivocal identification of persons, which is demonstrated in the following example (Table 2).

Table 2. Example of accented and non-accented character sets

first name and surname at birth	márton dávid	marton david	marton dávid
place of birth	vas hun	vas hun	vás hun
date of birth	1955-10-05	1955-10-05	1955-10-05
first name and surname of the mother at birth	szücs izabella	szucs izabella	szűcs izabella

In reference to the example (Table 2), the use of the Latin-based national alphabet is recommended for creating the Infrastructure, since it can establish a common foundation for substituting differing international characters with the benefit of avoiding loss of information resulting from, e.g. the omission of accents. This solution is also permitted and accepted under the ICAO standard in the Visual Inspection Zone field of the document [4]. It is recommended that textual information should only be upper case.

For the name of the user, the first name and surname of the user are recommended to be managed in separate fields, where prefixes, suffixes, and numbers entered with Roman numbers should be written into the first name field.

In terms of standardized and uniform information management, we recommend the use of ISO 3166-1 alpha-3 country code for the designation of the country of birth [13]. The introduction of this enables the unequivocal identification of the country of birth independent from the writing system, which lays the foundation for the verification of the user's identity. In the course of cooperation between the Identity Verifier and the National Identity Verifier systems covered in detail in Section 6.3, the Identity Verifier system will be capable of selecting – based on the country code entered – the National Identity Verifier system that can verify the specific user's identity according to the country of birth. For the textual designation of the country name, standard ICAO 9303 defines the Latin-based writing in the national language or English [4]. Example: AUS Australia, HUN Hungary, JAM Jamaica (source: iso.org OBP).

One of the most important consensuses in the area of date management is the definition of the format in consideration of international cooperation because most countries use different date formats. This problem inevitably sets the precondition of using a uniform format when creating the user's unique identifier with a centralized algorithm. According to the standard, the format defined by ISO 8601 is recommended for use (YYYY-MM-DD), in accordance with standard ICAO 9303 and the date set by the Georgian calendar, e.g. 1955-07-05 [13].

In addition to personal information, the user account records the user's portrait, which will be significant for the identical twins' identity verification process and services that can be used with the digital identity account connected to, for instance, dating, social media, or LinkedIn profiles. Based on standard ICAO 9303, the facial image can be accepted as a portrait that can be used in an authentic document by satisfying several strict requirements, such as a solid colour background without strange things, appropriate brightness and contrast, a person without face-covering things like sunglasses, or the whole person must be in the picture [4]. Standard ICAO 9303 provides for the use of a portrait that complies with standard ISO 19794-5 [15].

To ensure compliance with the quality and technical specifications set out in the standard, it is more practical to take the portrait from a National Identity Verifier system, than leaving it to the user to take the photograph. One of the reasons for this is the issue of compliance when it comes to devices capable of taking photographs of varying resolution and quality, and users cannot be obligated to use cameras of identical types, and, we presume that the average user is inexperienced in taking photographs as the standard requires. Typical problems might be that home background is used, such as plants, things, or other people are in the background or the image is too light or too dark.

One important tool in promoting data protection is the calculation of indecipherable hash values of personal data, then storing these instead of personal information. Decisive elements in selecting the algorithm generating the hash value are described in the paper [19] such as:

1. preimage resistance,
2. 2nd preimage resistance,
3. collision resistance,
4. Chosen Target Forced Prefix preimage resistance.

Currently, the SHA-2 algorithm is used for creating checksums and imprints when creating an electronic signature [12]. In the practical implementation of the infrastructure, the use of the SHA-2 and the new generation, SHA-3 algorithms is recommended [24; 20].

SHA3-256 hash algorithm is only used in this paper as an example method.

Forensic DNA-based criminal identification is a widely used method all over the world, its basis is the CODIS system that enables borderless cooperation among law enforcement agencies to identify people. Nowadays, disaster victim identification (DVI) is another upgrowing application field of human DNA. Despite these uses, human DNA is not a widely used method for person's identity verification not as facial recognition or fingerprint methods. In this section, we are going to describe the specific steps of our method from collecting DNA-samples to storing generated DNA-profiles based on our recent work [35]. These steps can give a possible solution for all cases which are inevitable to perform human DNA-based identity verification in the Infrastructure.

It is the first and most important step of the human DNA-based identity verification because if the collecting process is not secure enough, the DNA-sample can be modified or exchanged with another one. To build trusted DNA-profile databases for identity verification, we have to use trusted delivery devices and protocols. In this subsection, we are going to present a possible solution for collecting DNA-samples, especially human saliva. It is important to emphasize that most of the people who will give DNA-samples are not technical users, they require simple solutions. By this, our method applies simple but secure techniques, such as RFID chips and tamper-evident indicators.

The sample collector cartridge is made from plastic with an embedded RFID chip, and it is transparent to help content checking. The unique identifier of the cartridge is built up from the identifier of the cartridge generated by the manufacturer – which is encrypted with the secret key of the manufacturer and it can be decrypted with the public key of the manufacturer based on the IETF RFC 5280 Internet X.509 Public Key Infrastructure standard described in [11] – and the unique identifier of the manufacturer generated by the Central DNA laboratory and cartridge register separated with a hashtag. The unique identifier of the cartridge is printed on the cartridge in plain text for human use, and it is stored in the embedded chip for machine use. This property ensures that cartridges cannot be exchanged accidentally or intentionally. To avoid further misuse, the unique identifier of each manufactured cartridge is registered in the database of the Central DNA laboratory and cartridge register; the details of official manufacturers, such as the name, location, and public key are also registered in which system for further public key use.

Using unreopenable caps ensures that the sample cannot be modified, for example, after the person who gave the saliva sample closed the cap of the cartridge, it cannot be reopened for exchanging the sample. Giving a saliva sample in the same way as it can be seen in this video (https://www.youtube.com/watch?v=3oTaydRPm3w) can reduce the unauthorized DNA-sample use, such as using saliva left on glass edge because the process needs as much saliva that cannot be got without the donor's knowledge (unnoticed).

Using tamper-evident bags can improve the protection of cartridges.

The second step of building up a DNA-based identity verifier architecture is to generate the DNA-profile from the presented DNA-sample. This method requires fully equipped laboratory and human resources as well. When we talk about DNA-profiles in this paper, we think of STR DNA-profiles. These STR DNA-profiles can be generated with fully automatic devices, in this paper, we are not going to give a deep description of this process but only technical information about cartridge identification and the outcome DNA-profile scheme.

Figure 5 presents the course of the DNA-profile generation, which is the detailed image of the sub-process (C.2) in Figure 6, from the sample entering process to the final DNA-profile result which can be all the applied STR markers concatenated into one string without any delimiters, such as comma, decimal separator, or space. This concatenation can ensure that person's diseases or medical symptoms cannot be derived from the DNA-profile. Mixing the order of the STR-markers in the concatenation also can reduce the possibility of guessing the original STR-marker values.

An example process can be seen in our Zenodo repository [36] which also contains the original DNA-profile and the concatenated string.

For long-term storage, the form of storage is a key point because it has to ensure that data do not represent too much information about a person but are still usable. To reach that, we have to precisely define what we want to do with the specific data. In the case of the Infrastructure, our primary goal is to use the DNA-profile for verifying people's identity but we would like to leave the door open for some similar cases, such as DVI or forensic identification. The simplest way to store a DNA-profile and a portrait in relation with the 5N attributes is to build up a database from these data and encrypting them with a symmetric encryption algorithm, but in this paper, we suppose that each country has own personal register. So, DNA-profiles can be stored in a separated register which only refers to the personal register records with an identifier. The identifier of a record from the personal register can be a hash generated from the concatenated 5N attributes similar to the UID generation method, but it is important to emphasize that the generated record identifier must only be dynamically generated and it must not be stored in the personal register to rise the safety of the separated DNA-profile register; the DNA-profile can also be stored as a hash value (the DNA-profile hash generation method of the Identity Verifier might be applied, in Section 6.3.3). A possible solution is available in our Zenodo repository [37].

Human DNA-profile management is the key element of the Infrastructure. Nevertheless, we are not going to preface strict constraints on storing DNA-profiles but a National Identity Verifier system must be able to answer the following questions of the Identity Verifier system:

1. Does a person with the given 4N attributes and DNA-profile exist?

2. Is the given DNA-profile unique?
3. Return the portrait of the person to whom the given 4N attributes belong (if it exists).

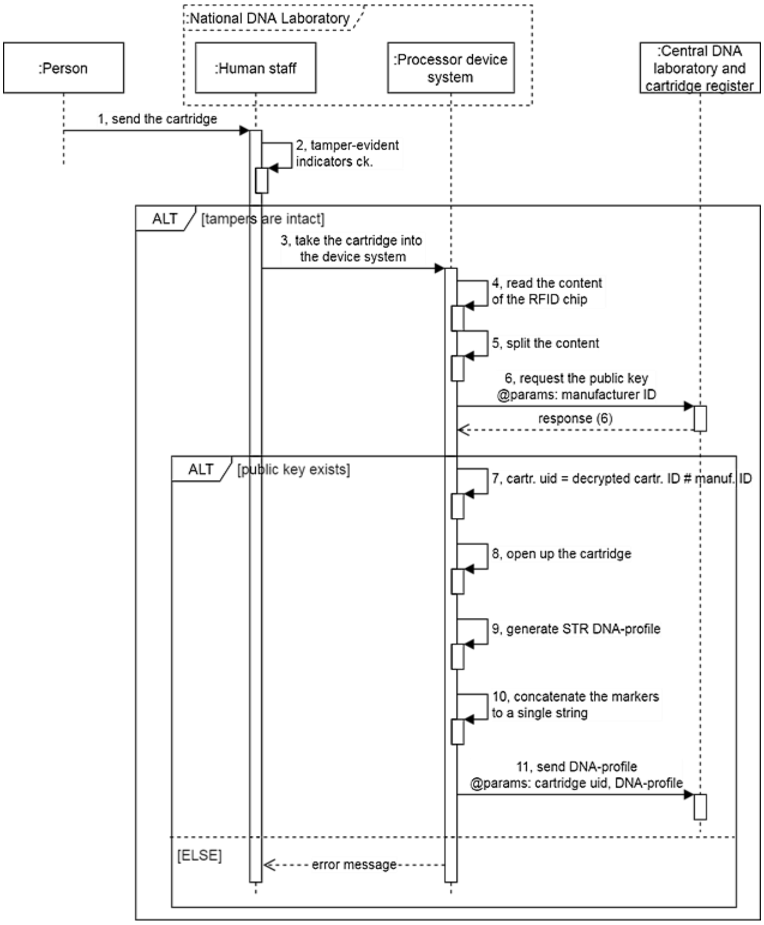

:National DNA Laboratory
:Person
:Human staff
:Processor device system
:Central DNA laboratory and cartridge register
1, send the cartridge
2, tamper-evident indicators ck.
ALT [tampers are intact]
3, take the cartridge into the device system
4, read the content of the RFID chip
5, split the content
6, request the public key @params: manufacturer ID
response (6)
ALT [public key exists]
7, cartr. uid = decrypted cartr. ID # manuf. ID
8, open up the cartridge
9, generate STR DNA-profile
10, concatenate the markers to a single string
11, send DNA-profile @params: cartridge uid, DNA-profile
[ELSE]
error message

Figure 5. The course of the DNA-profile generation

As we also emphasized in our paper [34] there are a huge amount of unsafe, malicious websites, applications. Most of them cannot be caught in time because the identity of the developers without strict control cannot be verified. For example, a malicious version of the Amazon Alexa application could appear in the Apple App Store and several people installed as a valid version of it [21; 9].

Using Extended Validation certificates, as a trusted source of the required information, can be a good practice to avoid untrusted service providers connecting to the Infrastructure. Because the extended validation certificate contains all the information, such as organization name, service endpoint URL that an Authentication Provider or a Service Provider needs to connect to the Infrastructure, the connection process can be performed without human interaction [8]. For example, the endpoint URL can also be a unique identifier of a connected provider in the Infrastructure. Requirements listed in Section 6.1 on the connection protocol can also be automatically checked. The possibility of malicious applications connecting to the Infrastructure can be further reduced by verifying the source codes of the applications. Along with this, educating users to carefully read registration terms, required permissions, and privacy and cookies policies is also recommended.

To use services via the Infrastructure as a user, you have to register a user account at the chosen authentication provider connected to the Infrastructure. Attributes you have to enter are divided into required and optional. The following attributes are required to create an account:

1. unique user identifier (UID),
2. human DNA-profile hash value,
3. portrait in 2D photograph format,
4. current name (last name, first name),
5. maiden name (last name, first name),
6. date of birth,
7. place of birth (name of the city, name of the country, and ISO 3166-1 alpha-3 country code),
8. one of the parents' maiden name (last name, first name / based on the application of the local registration authority).

These are the attributes of a basic user digital profile (account) which can be extended with optional attributes for wider use, such as ordering from a webshop, electronic banking, or using utility websites. Optional attributes can be almost every personal information, such as e-mail address, phone number, address, or bank account number. Figure 6 presents the user account registration process and the sources of the required attributes in detail. In Phase B, a passive user account will be created which cannot be used to authenticate a user. After identity proofing, it will be extended to a complete user account that can be used to authenticate a user (Phase D).

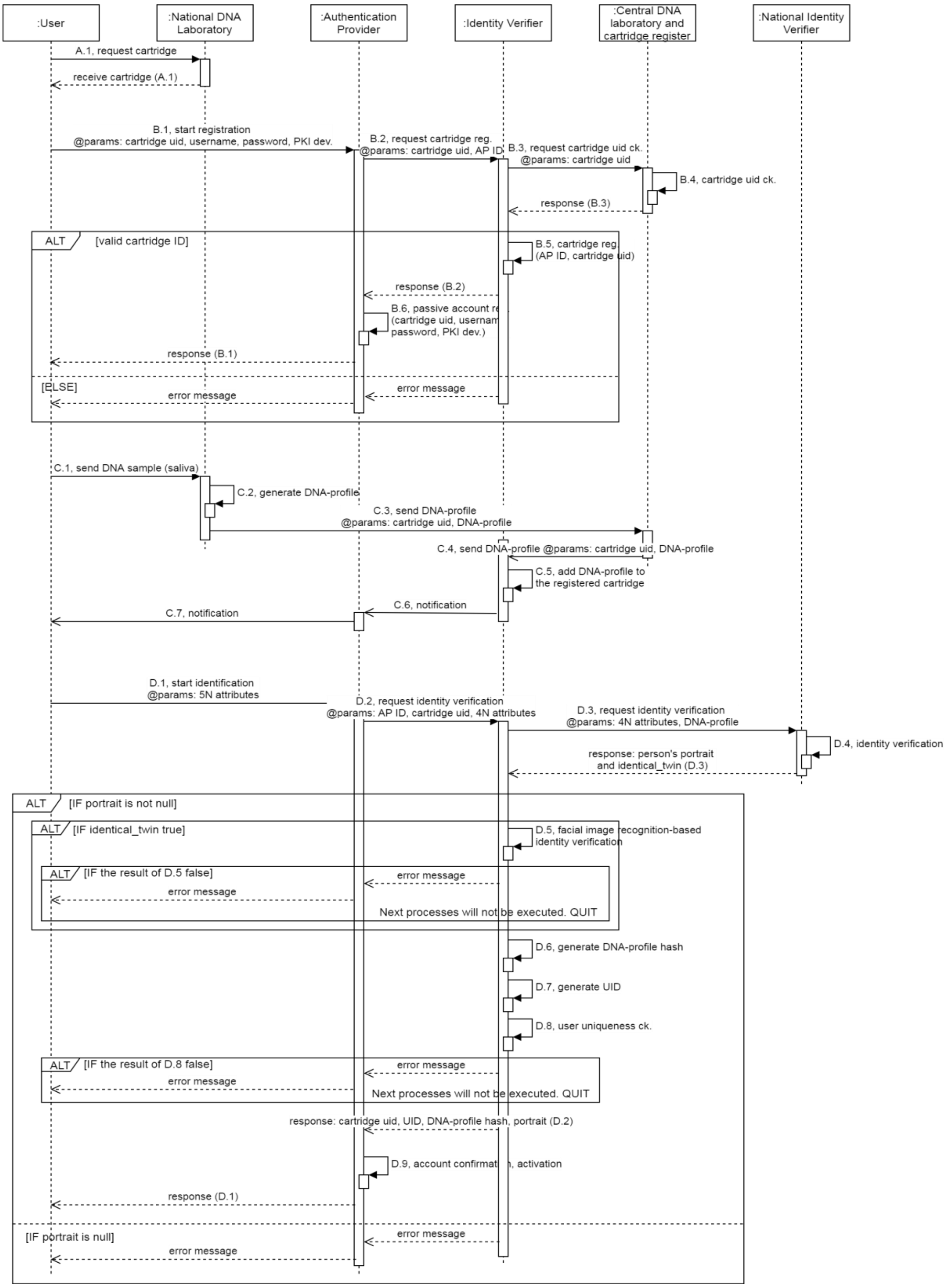

:User
:National DNA Laboratory
:Authentication Provider
:Identity Verifier
:Central DNA laboratory and cartridge register
:National Identity Verifier
A.1, request cartridge
receive cartridge (A.1)
B.1, start registration
@params: cartridge uid, username, password, PKI dev.
B.2, request cartridge reg.
@params: cartridge uid, AP ID
B.3, request cartridge uid ck.
@params: cartridge uid
B.4, cartridge uid ck.
response (B.3)
ALT
[valid cartridge ID]
B.5, cartridge reg.
(AP ID, cartridge uid)
response (B.2)
B.6, passive account re
(cartridge uid, usernam
password, PKI dev.)
response (B.1)
[ELSE]
error message
error message
C.1, send DNA sample (saliva)
C.2, generate DNA-profile
C.3, send DNA-profile
@params: cartridge uid, DNA-profile
C.4, send DNA-profile @params: cartridge uid, DNA-profile
C.5, add DNA-profile to
the registered cartridge
C.6, notification
C.7, notification
D.1, start identification
@params: 5N attributes
D.2, request identity verification
@params: AP ID, cartridge uid, 4N attributes
D.3, request identity verification
@params: 4N attributes, DNA-profile
D.4, identity verification
response: person's portrait
and identical_twin (D.3)
ALT
[IF portrait is not null]
ALT
[IF identical_twin true]
D.5, facial image recognition-based
identity verification
ALT
[IF the result of D.5 false]
error message
error message
Next processes will not be executed. QUIT
D.6, generate DNA-profile hash
D.7, generate UID
D.8, user uniqueness ck.
ALT
[IF the result of D.8 false]
error message
error message
Next processes will not be executed. QUIT
response: cartridge uid, UID, DNA-profile hash, portrait (D.2)
D.9, account confirmation, activation
response (D.1)
[IF portrait is null]
error message
error message

Figure 6. The registration process of a user account

The Identity Verifier system is the key element of the registration process, the scheme of which is based on the scheme of the intermediate point designed by the researchers of Murcia University, which keeps connected those systems that take part of the identity verification, such as a National DNA Laboratory and a National Identity Verifier service, and performs priority operations, such as UID generation, DNA-profile hash generation, and uniqueness checking.

Participants of the registration process, such as Authentication Providers, National Identity Verifier systems, DNA laboratories, and the Identity Verifier system exchange and store the public key of their Extended Validation SSL certificates with each other to authenticate each other. And, the Identity Verifier system uses the ISO 3166-1 alpha-3 country code to call the verifier method of the user's country of birth (D.4).

Undoubtedly verifying the identity of the user is a key step in the registration process. To achieve that the Infrastructure to be trusted, we used the NIST SP 800-63 Digital Identity Guidelines package in the design phase. For designing the registration process, we used the principles of Identity Assurance Level 3 which requires in-person or supervised remote in-person identity proofing with checking pieces of superior evidence, such as identity documents that contain biometric template and electronic information protected with a PKI-based method. Supervised remote in-person proofing requires a live operator who performs the identity verification process via a tamper-resistant kiosk which provides a continuous high-resolution video transmission and is equipped with all necessary sensors, such as scanner or camera [26].

Nowadays, identity document-based verification is the most common method to verify somebody's identity. It is no different at the Yoti ID application either because they also use this method extended with live remote video chat to verify the applicant's identity [44]. Although this is a commonly used solution, its use might cause many problems. In the following, we summarize all advantages and disadvantages of using identity document-based proofing.

Advantages

1. Identity documents are certified and protected against forgery using several security measures, such as watermark, hologram, or fluorescent fiber [30].

Disadvantages

1. Not everybody has an identity document that contains a biometric template and electronic information protected with PKI-based method, for example, passport or identity card.

2. The quality of scanning, image resolution, and the vast number of document types are critical sources of errors in the Optical Character Recognition process. There is a critical number of different identity document types to be managed based on the PRADO registry, for example, only in Hungary, there are 20 different document types in use, such as passport, personal identity card, or driving license; and solutions for scanning and taking photographs probably differ by each user.

3. The PRADO security features (e.g. fluorescent fiber or watermark) of an identity document cannot be examined in each case on scanned pictures [30].

4. The accuracy and expiry of the information on the identity document, if there is no live connection with the document issuer, cannot be verified.

5. The submitted identity document might contain more attributes than it is necessary, for example, restrictions of driving license. To withhold surplus information, the user has to edit the scanned image of the document, for instance, by covering the extra information.

Because applying identity document-based proofing has several dangerous disadvantages, and it also requires using kiosks, instead of using that, we have designed a human DNA-based verification solution. The registration process is made up of four separate parts, as recommended in Section 9.2 of NIST SP 800-63A, because processing a DNA sample might require a long waiting time [27]. The National Identity Verifier uses the user's 4N attributes and DNA-profile to perform the identity verification. As we described in Section 6.2.6.3 on Storing DNA-profiles, the system has to check that the person exists with the given data and that the given DNA-profile is unique or not, and it returns the portrait that belongs to the person (if who exists with the given data). If the specific person is an identical twin then the Identity Verifier system also performs a facial image recognition-based identity verification, method of which is described below.

Table 3 presents the comparison of commonly used identification methods, and it summarizes the properties used to select the appropriate one.

Table 3. The properties of commonly used identification methods

properties	fingerprint	facial image	human DNA
sampling option	not everybody can	everybody can	everybody can
accuracy (EER, FNIR)	FNIR >= 0.0009 (at FPIR = 0.001, using ten-finger IDFlats)	FNIR >= 0.068 (at FPIR = 0.002)	no FNIR, EER
special hardware to collect	yes, fingerprint scanner	yes, high-resolution camera	no
special hardware to process	no	no	yes, DNA sequencer / PCR
keeping up to date	yes, sometimes (accident)	yes, continuously	no
on identical twins	different	different, but EER 17.4%	non-different

The biggest advantages of human DNA are that it has no false negative or false positive (FNIR: False Negative Identification Rate, FPIR: False Positive Identification Rate) measures, everybody can give samples, and the DNA-profile does not change; but the non-difference on identical twins is the biggest problem of which. Using facial image recognition as an additional method can solve the identical twins' identity verification problem. Although fingerprint technology performs better accuracy, it requires the ten-finger IDFlats method, which captures all fingers, left slap, right slap, and two thumbs simultaneously, to achieve 0.0009 FNIR value; and using ordinary one index finger capturing performs 0.019 FNIR at 0.001 FPIR [2]. And unfortunately, fingerprint capturing requires a special scanner which is not as widely available as a high-resolution camera required for facial image recognition. Both two technologies are dependent on changing personal characteristics, such as changing face or finger damages. Fingerprint capturing might fail if the user's finger damaged, dirty, or missing. While, the accuracy (FNIR) of facial image recognition depends on the user's age, like it is 0.008 for people over age 55, 0.027 for young (19-30-year-old) people, 0.29 for younger (8-13 year old) people, and 0.4 for kids at 0.005 FPIR [29].

In summary, these are the reasons why we chose human DNA for identity verification.

Facial image recognition-based identity verification is an additional method to verify the identity of identical twins after DNA-based verification. It is a required method because human DNA cannot be used to differentiate two identical twins (of course, who are each other's siblings). The algorithm requests a portrait from the user and compares it to the user's portrait received from the National Identity Verifier, if they match, then the same person wants to register as who belongs to the given 4N attributes. Unfortunately, facial image recognition also has limitations in differentiating identical twins as the paper 9 [17]. emphasized that the Equal Error Rate was 17.40 percent with the best performing algorithm, and the average EER (Equal Error Rate) was 42.70 percent. All EERs were measured with matching one image with controlled and one image with uncontrolled illumination; and the images were taken one year apart. Nevertheless, facial image recognition can be an applicable method to perform identical twins' identity verification; and the process can also be extended with live operators for more precise performance. The user uniqueness check algorithm also uses this method to filter out account redundancies.

To minimize the risk of misuse of stored DNA-profiles, we recommend storing the hash of the DNA-profiles instead. Adding SALT to the string that will be hashed can rise the safety of DNA-profile storing like Recital 28 of the GDPR recommends (https://eur-lex.europa.eu/legal-content/HU/TXT/ELI/?eliuri=eli:reg:2016:679:oj. 2016). Also, the stored DNA-profile hash can only be used by the system that owns the SALT, for example, as we mentioned in Section 6.2.6.3, only the National register can identify people with their DNA-profiles which are stored in the register because only that system owns the SALT used for the hash generation. In other words, if the database is compromised then nobody except the owner can resolve the DNA-profile hash values. The method of the process is described in Pseudocode 1.

```
DNA-profile hash generation method
@params: DNA-profile string
@return DNA-profile hash string
READ SALT from the secure storage into variable temp
concA = CONCAT @params with temp
DNA-profile hash = generate a hash from concA
return DNA-profile hash
```

Pseudocode 1: The method of DNA-profile hash generation

Minimizing the number of required attributes is one of the key messages of the GDPR, which means that try to request only those attributes that are inevitable to identify the individual or to provide the requested service (https://eur-lex.europa.eu/legal-content/HU/TXT/ELI/?eliuri=eli:reg:2016:679:oj). A unique identifier is a key to achieve the goal of avoiding requesting several attributes just to unequivocally identify the user, for example, leaving a comment, in the ordinary case, requires name, e-mail address, or phone number, and to prove such services as automatized data erasure or attribute sharing and monitoring. However, there are several personal identifiers all over the World, none of them can be used as a global identifier because all of them are only local and not unified; and caused by these, they are not unique. But a more worrying problem is that most of these identifiers contain and leak personal information, such as gender, birth date, ethnic, or issuer identifier [39]. To avoid these issues, we have invented a new unique identifier model for the Infrastructure, which can be globally unique based on the 4N attributes; Pseudocode 2 describes the method of which.

```
UID generation method
@params: 4N attributes
@return string
READ SALT from the secure storage into variable temp
concA = CONCAT @params with temp
UID = generate a hash from concA
return UID
```

Pseudocode 2: The method of the UID generation

Generating a hash value is not a big deal but finding the appropriate components is that. So, we focused on the following criteria when we designed our UID model:

1. Each user has only one UID.
2. Each UID must be unique.
3. The UID must not leak personal information, for example, when the gender or date of birth can be derived from a user identifier – such as in the case of the Hungarian personal identifier –
4. The UID must be related to the user in such a way as to ensure that when a user changes authentication provider or creates a new account, but the 4N attributes remain unchanged, then get the same identifier.
5. The UID should be changeable/replaceable when it is necessary, for example, in the case of witness protection or domestic violence.

Using the 4N attributes as the basis of the UID generation fulfills all the criteria above; and using a hash algorithm with the SALT ensures that the UID does not leak personal information, and it can be verified that the UID was generated by the Infrastructure.

In addition, there are circumstances when changing the identity or the UID is inevitably necessary. For example, identity change in witness protection programs is a dominant one out of all the circumstances. In the course of the procedure, by completely changing the known 5N attributes of a specific person, a completely new identity is given, but when the root cause is not sought for in such a drastic solution, even then it is possible that a user wishes to gain a new identity by changing any of his/her 5N attributes, for instance, due to domestic violence someone changes his/her name, moves to another city, and does not want to reveal his/her former life.

To satisfy these requirements, the unique user identifier is based on the 4N attributes and it can be changed, and as a result of this change, the person appears as a new user in the Infrastructure.

Ensuring the uniqueness of a user account, similar to the uniqueness of a UID, is a key element of the Infrastructure. So, the algorithm filters out the anomalies originating from changing the 4N attributes using physical properties to prevent that more than one virtual account belongs to a specific physical user, and ensures that authentication providers are disjoint on a specific user. As we mentioned previously, human DNA cannot be used to differentiate identical twins, however, facial image recognition also has deficiencies, it can be an applicable alternate. Pseudocode 3 describes the method of uniqueness checking.

```
uniqueness checking
@params: UID, DNA-profile hash, portrait, identical_twin
@return bool
FOR each Authentication Provider DO
    IF UID is unique
        IF DNA-profile hash is not unique
            IF identical_twin is true
                SELECT all portraits belong to the given DNA-profile hash
                    FOR each portrait DO
                        IF portrait match with the given portrait
                            return false
    ELSE
                return false
    ELSE
        return false
return true
```

Pseudocode 3: The method of uniqueness checking

The Router, the scheme of which is based on the scheme of the intermediate point designed by the researchers of Murcia

University, is the key element of the user authentication and attribute sharing presented in Figure 7 because it coordinates the cooperation between authentication providers and service providers based on the Extended Validation SSL certificate standard described in [8]. This cooperation includes the Service Provider registering into the system of an Authentication Provider, user authentication and attribute sharing, and automatized data erasure service management. In these cases, all participants also use Extended Validation SSL certificates to authenticate each other. Pseudocode 4 describes the frame of a Service Provider registration process. While, the method of user authentication and attribute sharing is based on the concept of Google Sign-in service extended with a detailed list of shared attributes – active and archive as well – allowing the user to follow what shared with whom in accordance with Article 15 of the GDPR (https://eur-lex.europa.eu/legal-content/HU/TXT/ELI/?eliuri=eli:reg:2016:679:oj. 2016). The user can check what attributes are shared with the service providers without directly interacting with them. And, by the way of the attribute transfer, which is performed in case of each authentication, occurring in the point 4 of Figure 7, the Infrastructure enables the local Service Provider to maintain the correctness and consistency of data required under Paragraph 1/d of Article 5 of the GDPR (https://eur-lex.europa.eu/legal-content/HU/TXT/ELI/?eliuri=eli:reg:2016:679:oj. 2016).

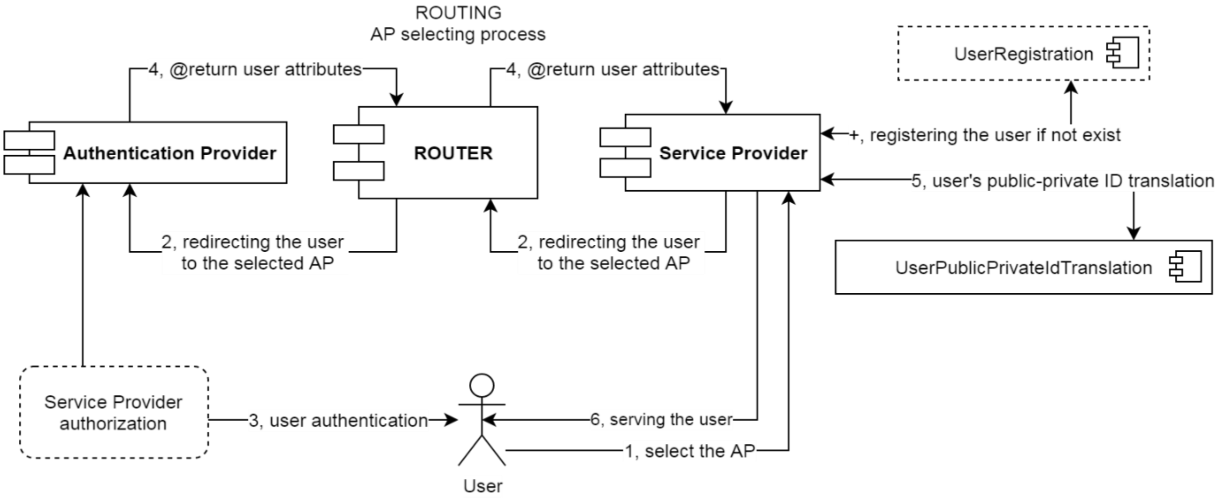

ROUTING
AP selecting process
4, @return user attributes
4, @return user attributes
UserRegistration
Authentication Provider
ROUTER
Service Provider
+, registering the user if not exist
5, user's public-private ID translation
2, redirecting the user to the selected AP
2, redirecting the user to the selected AP
UserPublicPrivateIdTranslation
Service Provider authorization
3, user authentication
6, serving the user
1, select the AP
User

Figure 7. User authentication and attribute sharing

```
Service Provider connection to the Infrastructure
@params: Extended Validation SSL certificate
Service Provider starts the registration with entering its Extended Validation SSL certificate
Router for each Authentication Provider DO
        REQUEST Service Provider registration using the public key of the Service Provider
                Authentication Provider REGISTER the Service Provider
                        --the Authentication Provider will use the endpoint URL of the Router which was entered earlier
Router REGISTER the certificate of the Service Provider
```

Pseudocode 4: The method of a Service Provider connection to the Infrastructure

And, the automatized data erasure service management performed by the Router allows the user to request from a Service Provider to delete all information stored about him/her. The user can call the automatized data erasure service of the specific Service Provider via his/her user account without any direct interaction with the Service Provider, and the service automatically performs data erasure; Figure 8 presents the procedure. The automatized data erasure service primarily provides support to the Service Provider in fulfilling requests relating to the user's right to erasure under Article 17 of the GDPR(https://eur-lex.europa.eu/legal-content/HU/TXT/ELI/?eliuri=eli:reg:2016:679:oj. 2016).

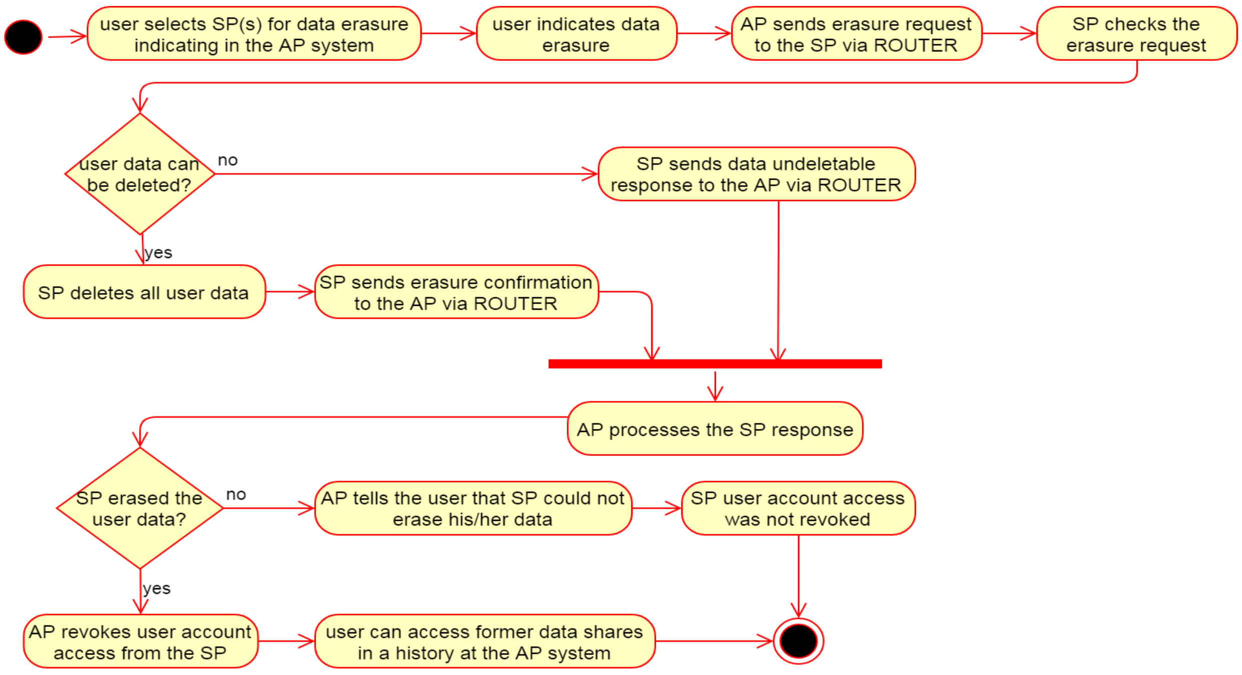
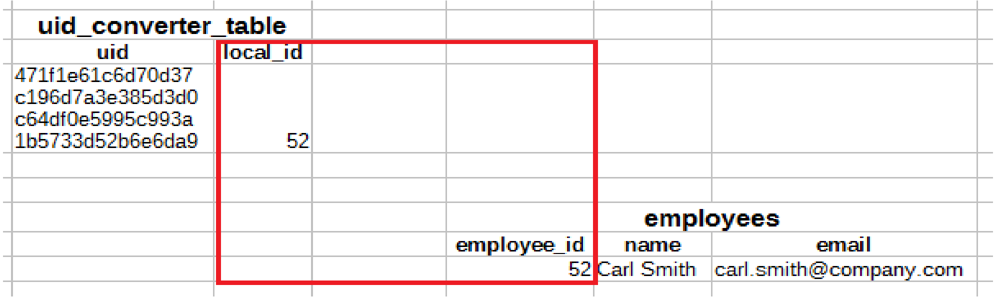

user selects SP(s) for data erasure indicating in the AP system
user indicates data erasure
AP sends erasure request to the SP via ROUTER
SP checks the erasure request
user data can be deleted?
no
SP sends data undeletable response to the AP via ROUTER
yes
SP deletes all user data
SP sends erasure confirmation to the AP via ROUTER
AP processes the SP response
SP erased the user data?
no
AP tells the user that SP could not erase his/her data
SP user account access was not revoked
yes
AP revokes user account access from the SP
user can access former data shares in a history at the AP system

Figure 8. The process of the automatized data erasure service

User's base attributes sometimes need updates which require special cases, such as the 4N attributes can be updated by a new account creation, the portrait can be updated automatically from a National Identity Verifier system via the Identity Verifier system using the 4N attributes, and the DNA-profile hash must remain unchanged.

Updating the 4N attributes, such as name, data of birth, or parent's name causes the change of the UID; we will describe a method to export and import data of former account in Section 6.7, which ensures that a user may import the former UID and the formerly shared attributes information. Instead of updating the 4N attributes, we require that the user to delete the current account and register a new one to avoid the extreme complexity of the uniqueness check algorithm that could cause its vulnerability. For example, the algorithm should be able to differentiate a redundant account and an account being updated. While, using UID converter tables ensures that a changed UID can be related to a former UID stored in the database of a Service Provider to avoid anomalies and inconsistencies. In the case of the traditional model, a user registration function has been implemented in each system, in general, to locally manage the users' data. In our view, people should not be registered as users, but according to roles, such as employee or client. The role of the UID converter table, in this case, is to keep a record of who has access to a specific system and to associate the entity identifier created in the given role to the entity's UID, which is illustrated in Figure 9; an example is available in our Zenodo repository [40].


uid_converter_table

uid	local_id
471f1e61c6d70d37	
c196d7a3e385d3d0	
c64df0e5995c993a	
1b5733d52b6e6da9	52

employees

employee_id	name	email
52	Carl Smith	carl.smith@company.com

Figure 9. An example of the UID converter table

While, the permanent online connection with National Identity Verifier systems, which also ensures automatic portrait update, allows the specific Authentication Provider via the Identity Verifier system to periodically check the validity of the 4N attributes in a user account. This is a great advantage over identity document-based proofing because it ensures that the validity of the 4N attributes to be checked without disturbing the user or depending on the expiration date, for example, might as well every day or every week; and if the Authentication Provider finds that the validity of the 4N attributes was expired then it might set the user account to passive state.

The day of a user account deletion once comes in every system, but in the case of the Infrastructure, it is a little bit more special because of user data migration. Because, in support of exercising the user's right under Article 20 of the GDPR, as emphasized in the introductory part of our paper, the Infrastructure enables the import of the list describing the attributes already shared with services when the authentication provider is changed(https://eur-lex.europa.eu/legal-content/HU/TXT/ELI/?eliuri=eli:reg:2016:679:oj). Which list will be generated by the user account deletion algorithm; Pseudocode 5 describes the method of which. The service ensures data consistency. An example former data migration certificate and the scheme of which is available in our Zenodo repository [41].

```
user account deletion
@return file or void
user REQUEST his/her account deletion from the Authentication Provider
IF the user does not want to keep his/her attribute shares
    Authentication Provider FOR each linked Service Provider DO
        REQUEST automatized data erasure
        IF the process FAILS
            Authentication Provider INFORM the user about failed data erasure
            Authentication Provider CANCEL the account deletion
```

```
Authentication Provider GENERATE a certificate of the user's all information (former UID, DNA-profile hash, portrait, shared attributes)
Authentication Provider REGISTER the certificate generated above
Authentication Provider REQUEST the user to download the certificate created in the previous step
Authentication Provider DELETE the user's account
Authentication Provider GENERATE a certificate of the user's all information (former UID, DNA-profile hash, portrait, shared attributes)
```

Authentication Provider REGISTER the certificate generated above Authentication Provider REQUEST the user to download the certificate created in the previous step Authentication Provider DELETE the user's account Pseudocode 5: The method of a user account deletion.

Former data import function is the extension of the user account deletion service because it performs the import of the exported data using the former data migration certificate generated by the user account deletion service; Pseudocode 6 describes the method of which.

```
former data import
@params: former data migration certificate file
@return bool or string
user REQUEST former data import
IF current Authentication Provider certificate verification requested from the former Authentication Provider via the Identity Verifier is true
    IF the former UID and the current UID are equal
        current Authentication Provider imports the list of the user's formerly shared attributes (both active and archive)
    ELSE
        IF the former DNA-profile hash and the current DNA-profile hash are equal
            IF portrait verification requested from the Identity Verifier is true
                current Authentication Provider imports the former UID and list of the user's formerly shared attributes (both active and archive)
                current Authentication Provider generates a human-readable list of Service Providers which require user activity to update user's data
                    --user has to manually update his/her data in the systems of former Service Providers using his/her former UID
        ELSE
            return false
    ELSE
            return false
ELSE
    return false
```

Pseudocode 6: The method of the former data import function

In this case, adding the user's former UID to the current account enables that the formerly shared attributes are still available for the service providers, but the user has to manually request the association between the former and current UID via the UID update service of the service providers: the Service Provider requires access to the former UID and it updates the UID stored in its database with the current UID.

The AKEN Infrastructure combines the essential functions of the existing systems, such as the scheme of an intermediate point, listing of shared attributes, or facial image recognition-based identity verification with our robust methods, such as human DNA-profile management, automatized data erasure, or former data migration. This section summarizes and emphasizes these unique values.

Our protocol for human DNA-profile management covers the whole process from sample collection, through DNA profile generation, to DNA profile storage. Which is the basis of the identity verification and uniqueness checking. To handle the identification trouble of identical twins, we extended our human DNA protocol with facial image recognition. In our protocol, human saliva samples are collected with a transparent plastic cartridge; the properties of which are the following:

1. Made from transparent plastic with an embedded RFID chip.

2. The chip stores the cartridge identifier encrypted with the secret key of the manufacturer and the identifier of the cartridge manufacturer separated with a hashtag, which ensures that cartridges cannot be exchanged accidentally or intentionally.

3. To avoid further misuse, the unique identifier of each manufactured cartridge is registered at the Central DNA laboratory and cartridge register, and the details of official manufacturers, such as the name, location, and public key are also registered in which system.

4. Using unreopenable caps and tamper-evident bags protects against samples modification.

And, giving saliva sample in the same way as it can be seen in this video (https://www.youtube.com/watch?v=3oTaydRPm3w) can reduce the unauthorized DNA-sample use, such as using saliva left on glass edge because the process needs as much saliva which cannot be got without the donor's knowledge (unnoticed).

Figure 5 presents the course of DNA-profile generation from sample entering process to final DNA-profile result which can be all the applied STR markers concatenated into one string without any delimiters, such as comma, decimal separator, or space. This concatenation can ensure that person's diseases or medical symptoms cannot be derived from the DNA-profile.

In this paper, we recommended a possible method to permanently store DNA-profiles in such a way that DNA-profiles can be stored in a separated register which only refers to the personal register records with an identifier. The identifier of a record from the personal register can be a hash generated from the concatenated 5N attributes similar to the UID generation method, but it is important to emphasize that the generated record identifier must only be dynamically generated and it must not be stored in the personal register to rise the safety of the separated DNA-profile register; the DNA-profile can also be stored as a hash value (the DNA-profile hash generation method of the Identity Verifier might be applied, in Section 6.3.3). Human DNA-profile management is the key element of the Infrastructure. Nevertheless, we are not going to preface strict constraints on storing DNA-profiles but a National Identity Verifier system must be able to answer the following questions of the Identity Verifier system:

1. Does a person with the given 4N attributes and DNA-profile exist?

2. Is the given DNA-profile unique?

3. Return the portrait of the person to whom the given 4N attributes belong (if it exists).

While the UID generation method is the basis of such other services as automatized data erasure and complete former data migration. The model of our identifier is based on the following criteria:

1. Each user has only one UID.

2. Each UID must be unique.

3. The UID must not leak personal information, for example, when the gender or date of birth can be derived from a user identifier – such as in the case of the Hungarian personal identifier –

4. The UID must be related to the user in such a way as to ensure that when a user changes authentication provider or creates a new account, but the 4N attributes remain unchanged, then get the same identifier.

5. The UID should be changeable/replaceable when it is necessary, for example, in the case of witness protection or domestic violence.

Generating a salted hash as a UID from the concatenated 4N attributes ensures that the identifier fulfills all of the listed criteria. And, it can fulfill the data minimization principle of the GDPR (Paragraph 1/c of Article 5), for example, there is no need for other identifier attributes, such as email address or the 5N attributes but the UID to perform the automatized data erasure service (https://eur-lex.europa.eu/legal-content/HU/TXT/ELI/?eliuri=eli:reg:2016:679:oj. 2016).

The list below summarizes the services of the Infrastructure, and it highlights those GDPR Articles which are supported by the service (https://eur-lex.europa.eu/legal-content/HU/TXT/ELI/?eliuri=eli:reg:2016:679:oj):

1. automatized data erasure: Article 17

The automatized data erasure service management performed by the Router allows the user to request from a Service Provider to delete all information stored about him/her. The user can call the automatized data erasure service of the specific Service Provider via his/her user account without any direct interaction with the Service Provider, and the service automatically performs data erasure; presented in Figure 8. The service ensures data consistency in the Infrastructure.

2. former data migration: Article 20

1. former data migration certificate generation

The certificate is generated by the user account deletion algorithm; Pseudocode 5 describes the method of which. An example former data migration certificate and the scheme of that is available in our Zenodo repository [37].

2. former data import service

It performs the import of the exported data using the former data migration certificate generated by the user account deletion method; Pseudocode 6 describes the method of which.

3. former and current UID association

In this case, adding the user's former UID to the current account enables that the formerly shared attributes are still available for the service providers, but the user has to manually request the association between the former and current UID via the UID update service of the service providers: the Service Provider requires access to the former UID and it updates the UID stored in its database with the current UID.

4. the model of the UID converter table

Using UID converter tables ensures that a changed UID can be related to a former UID stored in the database of a Service Provider to avoid anomalies and inconsistencies. So, the role of the UID converter table is to keep a record of who has access to a specific system and to associate the entity identifier created in the given role to the entity's UID; presented in Figure 9.

3. Shared attributes monitoring service: Article 15

The method of user authentication and attribute sharing is based on the concept of Google Sign-in service extended with a detailed list of shared attributes – active and archive as well – allowing the user to follow what shared with whom in accordance with Article 15 of the GDPR. The user can check what attributes are shared with service providers without directly interacting with them. And, by way of the attribute transfer, which is performed in case of each authentication, occurring in the point 4 of Figure 7, the Infrastructure enables the local service provider to maintain the correctness and consistency of data required under Paragraph 1/d of Article 5 of the GDPR.

IV. CONCLUSION

In our Infrastructure design, we were about to give a solution which might be a unified framework of cooperation among existing providers without forcing users to use an uncountable set of access data, and which solution includes and expands the essential services, such as former data migration, shared attribute monitoring, or automatized data erasure. In addition, we have paid special attention to reducing the vulnerabilities caused by the deficiencies of existing solutions described in the Introduction.

As the main solution, in the Infrastructure, each user has only one user account to reduce the vulnerability caused by that human password memorization is limited and the use of a NIST AAL3-compliant 2FA is mandatory to ensure the high-level security. In addition, the human DNA-based identity verification and uniqueness checking further increase the security by eliminating those vulnerabilities that might result in identity crimes due to applying document-based proofing, such as PRADO security features cannot be examined in scanned images or the accuracy and expiry of identity information cannot be verified without a live connection to the authorities.

While, SSO-bypassing and insecure service provider identity verification also need treatment, as these are also serious vulnerabilities that can further weaken the security. Mandatory use of Extended Validation SSL certificates and performing user authentication via the Router can treat these vulnerabilities.

In spite of the fact that the Infrastructure might seem to be unrealizable, it is very important to emphasize that all of its components, except the protocol of human DNA-profile management, are completely available; these only need some modifications or extensions, for example, Google Sign-in, STORK infrastructure, Yoti ID, or Auth0.com service. In addition, the bases of human DNA management have already been laid, they only require some new viewpoints and devices.

Before starting the design, we formulated some hypotheses to validate the need for such an infrastructure. This section presents the validation of these hypotheses.

Compared to authentication solutions which are used nowadays (when every user has M sets of access data) a globally centralized solution (when every user has only one set of access data) results in a significantly lower security risk using the National Institute of Standards and Technology Authenticator Assurance Level 3 (NIST AAL3), and M > 1.

User authentication can be selected into three groups: local, locally centralized (SSO), globally centralized. Local authentication is when every user has M=N sets of access data, opposite this, globally centralized authentication is when every user has only one set of access data (M=1), and locally centralized is between them because in this case, every user has M<=N sets of access data but at least M=2. For example, eduID, STORK, or login.gov are locally centralized authentication solutions because a user cannot use all the services with only one of them.

Nowadays, local authentication is the most common way, only 6.30 percent of the inspected 912,206 websites supported SSO [23]. In our study [36]. We also inspected the SSO use of 100 websites (from little webshops to big banks) and found that none of them applied NIST AAL3-compliant 2FA. These deficiencies rise the possibility/risk of the most common attacks become successful, such as Phishing, Identity Theft, or Information Leakage. Another significant problem is that the users' password memorization is limited and it causes problems, such as forgetting or mixing passwords, and might cause vulnerabilities, such as passwords in a post-it, regularly requested password resets, or use of untrusted password manager applications based on that the paper [3], highlighted that 54.75 percent of the respondents made physical notes about their passwords while users had at least 5 passwords and 0.7 percent of these passwords met the recommendations of NIST.

Applying NIST AAL3-compliant 2FA significantly reduces the chance that attacks become successful in all the cases of user authentication. The higher degree of centralization in user authentication significantly reduces the number of access data (M<=N), but M=1 can only be reached with global centralization. In our Infrastructure, it does not mean that there is only one authentication provider but an unbounded number of authentication providers might connect to the Infrastructure; the restriction is that every user has only one user account. In the case of the NIST AAL3 level, the access data must include a hardware security key, which, if the unique password scheme is followed for each account, should also be unique (K=M). Which also causes the same problems as the password memorization. When M=1 with one hardware security device, the user only needs to manage three Memorized Secrets, such as one username, one password, and one PIN code for a hardware security key.

Denise Ranghetti and her colleagues found that the rate of users' password memorization problems is 53.10 percent for 1-3 passwords, 80.70 percent for 4-6 passwords, and 84.00 percent for 7-9 passwords. It means that if a user has only 1-3 passwords, then the security risk is definitely lower than using several user accounts with more than one set of access data at NIST AAL3.

Introducing a global UID with a globally centralized authentication can significantly reduce resource requirements and the number and type of required attributes in the case of data protection principles implementation than applying local or locally centralized authentication.

The Infrastructure developed by us requires an authentication provider or a service provider to implement all the functions that perform such services as shared attributes monitoring, former data migration, or automatized data erasure.

Without shared attributes monitoring, the user has to manually list all attribute shares, but any of the authentication providers connected to the Infrastructure solves it instead of the user. For example, STORK or eduID services do not provide a list to the user about shared attributes.

Without globally unique UID, the automatized data erasure service has to require additional attributes which can undoubtedly verify the user, such as name, date of birth, place of birth, or parent's name – for example, the e-mail address cannot be appropriate because a user might have more than one –. And of course, without shared attributes monitoring the user does not know where data erasure should be requested.

Using hash values generated with secure officially approved hash functions instead of storing plain-text information ensures that even if the database is compromised, hashed values cannot be decrypted or guessed based on the following statement of [6].

"With a well-designed cryptographic hash function, it is not feasible to construct or find a message that will produce a given hash value (pre-image resistance), nor is it feasible to find two messages that produce the same hash value (collision resistance)."

The Infrastructure introduced in our paper can be a reliable source of the necessary user information for service providers; possibilities lying in the service are highlighted in the following.

Based on the [16], statistical data of Facebook, 8,200 million user accounts were deleted from the system in 2019, and, as a result of that, the ratio of fictitious accounts could be around 5 percent (M. Armstron. 2020). Facebook can request an official identity document to verify the user's identity, but it is not part of the registration, which might cause fake accounts. Beyond the fact of misuse of others' personal information, this can lead to manipulation in matters of great importance, such as United States presidential elections, Hungarian elections, or the migrant crisis [25; 18]. While, in the case of dating portals, fake accounts can be sources of such problems as using others' images.

Considering that social media platforms – such as Facebook, Instagram, or Twitter – act as primary sources of opinion, and dating portals are the primary sources of meeting new partners, it will be inevitable in the future that authentic identity verification is performed during user registrations. In this process, the Infrastructure might ensure that users have only one account in these sites; and that only real people can register accounts with their own attributes only.

In the case of the employee market, we can also draw up a possible application. For example, LinkedIn is one of the biggest platforms for employees for sharing their professional achievements with companies and partners. Based on personal experiences and relying on statistics published on social media platforms, checking the professional background of a future employee could mean considerable headaches for a company. The reason of this is the use of divided and multiple user account platforms for sharing professional achievements. Linking scattered information is not solved in all cases, if the person does not use unique identifiers, such as email address or publication identifier, then it is virtually impossible to associate the datasets of various platforms with absolute certainty. To tackle this issue, the unique identifier of the Infrastructure we developed could be used well, which can unequivocally authenticate (find) the specific person in all isolated databases.

Based on the abstract infrastructure models developed we aim to carry out a practical implementation in a pilot project, which would be a fine opportunity for the testing and assessment of the Infrastructure in a practical environment. The results of the analysis would be used in our publication on practical implementation.

Some special functions are planned, such as National Identity Verifiers might notify the person who belongs to the given 4N attributes during the registration process, for example, via SMS or e-mail messages to avoid unauthorized registrations. While, the gigantic task of centrally managing the Infrastructure is similar to monitoring the Domain Name System by the Internet Corporation for Assigned Names and Numbers. As envisaged, the Infrastructure we developed would also be operated under an organization founded by participating countries, with particular attention to a georedundant design. The solution for the management of processes carried out between linked island-like services is provided by the OASIS Web Services Transaction (WS-TX) standard to avoid inconsistent processes, for example, the payment was successful and the costs were transferred, but the webshop canceled the delivery of the purchased product [28].

V. CONFLICT OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

VI. ACKNOWLEDGEMENT

Thank you for your support: Peter G. Gyarmati (1941, software engineer), József Tick (Óbuda University), Gyöngyi Bujdosó (Debrecen University), Balázs Egyed (SYNLAB, ELTE), Ferenc Fazakas (Debrecen University), Yoti ID team, Attila Tamás Adamkó (Debrecen University). This work was supported by the construction of EFOP-3.6.3-VEKOP-16-2017-00002. The project was supported by the European Union, co-financed by the European Social Fund.

REFERENCES

[1] ENISA Threat Landscape Report 2018. 2019. https://doi.org/https://doi.org/10.2824/622757.

[2] Watson CI, Fiumara GP, Tabassi E, Salamon WJ, Flanagan PA. Fingerprint Vendor Technology Evaluation, Gaithersburg, MD. 2014. https://doi.org/10.6028/NIST.IR.8034.

[3] Pilar DR, Jaeger A, Gomes CF, Stein LM. Passwords usage and human memory limitations: A survey across age and educational background. PloS one. 2012 Dec 5;7(12):e51067.

[4] Doc 9303 Machine Readable Travel Documents Part 3: Specifications Common to all MRTDs, 7th ed., International Civil Aviation Organization. 2015. https://www.icao.int/publications/Documents/9303_p3_cons_en.pdf.

[5] Torroglosa E, Ortiz J, Skarmeta A. Matching federation identities, the eduGAIN and STORK approach. Future Generation Computer Systems. 2018 Mar 1;80:126-38.

[6] Barker E, Barker W, Burr W, Polk W, Smid M. NIST special publication 800-57. NIST Special publication. 2007 Mar;800(57):1-42.

[7] Google Sign-in OAuth, (n.d.). https://developers.google.com/identity/protocols/OAuth2.

[8] Guidelines for The Issuance and Management of Extended Validation Certificates, v.1.7.2., 2020. https://cabforum.org/wp-content/uploads/CA-Browser-Forum-EV-Guidelines-v1.7.2.pdf.

[9] Chung H, Iorga M, Voas J, Lee S. Alexa, can I trust you?. Computer. 2017 Sep 22;50(9):100-4.

[10] How does Yoti work: authentication flow, (n.d.). https://www.yoti.com/business/how-does-yoti-work/.

[11] IETF RFC 5280 Internet X.509 Public Key Infrastructure standard. 2008. https://tools.ietf.org/html/rfc5280.

[12] IETF RFC 5754 Using SHA2 Algorithms with Cryptographic Message Syntax standard, 2010. https://tools.ietf.org/html/rfc5754.

[13] ISO 3166 Country Codes alpha-3. 2013. https://www.iso.org/iso-3166-country-codes.html.

[14] ISO 8601 Date and time format. 2004. https://www.iso.org/iso-8601-date-and-time-format.html.

[15] ISO/IEC 19794-5:2011 Part 5: Face image data. 2011. https://www.iso.org/standard/50867.html.

[16] Clement J. Global number of fake accounts taken action on by Facebook from 4th quarter 2017 to 1st quarter 2020, 2020. https://www.statista.com/statistics/1013474/facebook-fake-account-removal-quarter/.

[17] Paone JR, Flynn PJ, Philips PJ, Bowyer KW, Bruegge RW, Grother PJ, Quinn GW, Pruitt MT, Grant JM. Double trouble: Differentiating identical twins by face recognition. IEEE Transactions on Information forensics and Security. 2014 Jan 2;9(2):285-95.

[18] Juhász A, Szicherle P. The political effects of migration-related fake news, disinformation and conspiracy theories in Europe. Friedrich Ebert Stiftung, Political Capital Policy Research & Consulting Institute, Budapest. 2017 May.

[19] Rjaško M. Properties of cryptographic hash functions. Mikulášska Kryptobesıdka. 2008 Jun 9:53-62.

[20] Dworkin MJ. SHA-3 standard: Permutation-based hash and extendable-output functions. 2015 Aug 4.

[21] Locklear M. Fake Alexa setup app is topping Apple’s App Store charts. 2018. https://live.engadget.com/2018/12/27/fake-alexa-app-topping-apple-app-store-charts.

[22] Armstrong M. 16% of All Facebook Accounts Are Fake or Duplicates. 2020. https://www.statista.com/chart/20685/duplicate-and-false-facebook-accounts/.

[23] Ghasemisharif M, Ramesh A, Checkoway S, Kanich C, Polakis J. O single sign-off, where art thou? an empirical analysis of single sign-on account hijacking and session management on the web. In27th {USENIX} Security Symposium ({USENIX} Security 18) 2018 (pp. 1475-1492).

[24] Chandrana NR, Manuelb EM. Performance Analysis of Modified SHA-3. Procedia Technology. 2016 Jan 1;24:904-10.

[25] Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 US presidential election. Science. 2019 Jan 25;363(6425):374-8.

[26] Grassi PA, Lefkovitz NB, Fenton JL, Danker JM, Choong YY, Greene K, Theofanos MF. Digital Identity Guidelines: Enrollment and Identity Proofing Requirements [including updates as of 12-01-2017]. 2017 Dec 1.

[27] Grassi PA, Lefkovitz NB, Fenton JL, Danker JM, Choong YY, Greene K, Theofanos MF. Digital Identity Guidelines: Enrollment and Identity Proofing Requirements [including updates as of 12-01-2017]. 2017 Dec 1.

[28] Grassi PA, Fenton JL. NIST Special Publication 800-63-3. Digital Identity Guidelines. National Institute of Standards and Technology. 2017.

[29] Grother P, Ngan M. Face Recognition Vendor Test (FRVT), Gaithersburg, MD. 2014.

https://doi.org/10.6028/NIST.IR.8009.

[30] PRADO Glossary: technical terms related to security features and to security documents in general (v.8269.en.17+c3+add3). 2020. https://www.consilium.europa.eu/prado/en/prado-glossary/prado-glossary.pdf.

[31] Casado R, Tuya J, Younas M. A family of test criteria for web services transactions. Procedia Computer Science. 2012 Jan 1;10:880-7.

[32] Regulation (EU) No 2016/679 of The European Parliament and of The Council. 2016. https://eur-lex.europa.eu/legal-content/HU/TXT/ELI/?eliuri=eli:reg:2016:679:oj.

[33] States of Jersey: How we're using Yoti, (n.d.). https://www.gov.je/government/publicsectorreform/digitalid/pages/aboutdigitalid.aspx.

[34] Roskó T. A központosított felhasználó azonosítás jelene és jövője: biztonságos infrastruktúra vagy időzített bomba?. Információs Társadalom. 2019 Dec 17;19(2):52-85.

[35] Roskó T, Adamkó A. The human DNA can be the bridge between the Human and its data set in the Future. in: A 15 Éves PEME XVII. PhD - Konf. Előadásai, Professzorok az Európai Magyarországért Egyesület.2018:128–138.

[36] Roskó T. AKEN Infrastructure: an example STR DNA-profile, Zenodo Repos. 2020. https://doi.org/10.5281/ZENODO.3732031.

[37] Roskó T. AKEN Infrastructure: an example Personal and DNA-profile Register, Zenodo Repos. 2020. https://doi.org/10.5281/ZENODO.3732033.

[38] Roskó T. AKEN Infrastructure: an example SALT value, Zenoodo Repos. 2020. https://dio.org/10.5281/ZENODO.3731962.

[39] Roskó T, Adamkó A. Global personal identifier: advantage or disadvantage?, in: ADA 2018. 2018. (accepted for publication)

[40] Roskó T. AKEN Infrastructure: an example UID converter table and an example UID, Zenodo Repos. 2020. https://doi.org/10.5281/ZENODO.3876617.

[41] Roskó T. AKEN Infrastructure: an example former data migration certificate, Zenodo Repos. 2020. https://doi.org/10.5281/ZENODO.3876329.

[42] Truly Madly and Yoti build a safer community of online daters, (2018). https://www.yoti.com/blog/trulymadly-and-yoti-build-a-safer-community-of-online-daters/.

[43] Where is Yoti available? Which ID documents can be used?, (n.d.). https://yoti.zendesk.com/hc/en-us/articles/209273869-Where-is-Yoti-available-Which-ID-documents-can-be-used-.

[44] Yoti registration workflow, (n.d.). https://www.yoti.com/personal/create-yoti/.

[45] Yoti: Keep your data safe, (n.d.). https://www.yoti.com/personal/security/.