the authenticity of the identifier given by the entity requiring proof. This proof provided electronically is conventionally called a Credential. It can take the form of either a one-time or static password, an electronic signature, a response given to a random number provided by the system... It is most often a static password that is required, with an average of 5 passwords per Internet user to access different online accounts [DEL 13].

The advantage of an authenticated entity is that it allows the authority to impute an action to an entity and thus to take legal action in case of malicious acts. It is therefore necessary to pay attention to credential and identity thefts that may ensue. We should also note that, even under the guise of an unauthenticated identity, a malicious act (downloading of illegal content, cyber-attacks, etc.) will certainly be subject to more thorough investigation in order to trace it back to the responsible entity. This investigation can be made much simpler if the individual responsible has used their own cable line from their residence in order to carry out the criminal act.

Various software solutions based on architectures, and various economic models share digital identity management today. This management includes the creation of these identities (enrollment) – a fundamental step which secures the more or less strong link between digital identity and the natural person or legal entity who owns the identity, authentication of these identities, management of associated personal data, the commitment contract concerning the management of this data (use, disclosure to third parties), and revocation of these identities.

If the field of digital identity is experiencing rapid growth in today's society, some major efforts are still required to manage these identities, to provide comfort to users while ensuring their safety and privacy. The diversity of uses associated with digital identity (see section 1.4.2), properties which are antithetical to their objectives and their implementation (see section 1.4.1), the range of solutions for identity management (see section 1.4.3), still unsatisfactory standards (see section 1.4.4) and significant risks in challenging privacy, and fraud (see section 1.4.5), make the field of digital identity exciting for researchers and industrials. More technically focused research tracks, particularly on the preservation of privacy, will be presented in Chapter 4.

1.4.1. *Important notions*

The notions and properties associated with digital identity are listed below. Some of these definitions are derived from the terminology of [ISO 11] and [WP 14].

– Identifier: an identifier is a set of attributes that allows an application domain to link the declared identity to a digital entity previously known to the system.

– Uniqueness: an identifier is unique within the naming space of an application domain (email inbox, mobile phone, etc.), thus enabling a direct linking to a single entity in the domain.

– Authentication: a digital identity proves by stating their identifier and digital proof of identity (Credential) that they are truely the declared identity.

– Anonymity: characteristic of information which cannot be used to directly or indirectly identify the individual to whom the information pertains.

– Unlinkability: inability to connect at least two separate pieces of information (messages, URLs, actions, identifiers) to one individual, or a group of individuals.

– Linkability: this is the opposite of unlinkability. It is particularly useful when tracing something back to the identity of a cybercriminal.

– Pseudonymity: information which is associated with a pseudonym. A pseudonym can be used to reference a digital identity in an application domain without knowing their true identity. In this way, unlike anonymity, linkability is possible.

– Trust: an application domain can test depending on the transaction, the honest or dishonest behavior of a digital identity, whether this entity is authenticated or not, and thus, assign it a level of trust. This trust reflects the application domain's perception of the entity, and not the perception of the other entities.

– Reputation: multiple digital identities can interact within the same application domain and, after transaction, rate each other to make the quality of the relationship and service provided public. This rating contributes to an entity's overall rating. Other entities will tend to favor entities with a good

reputation to obtain a service. Overall, this system encourages entities to adopt good behavior, but is vulnerable to Sybil attacks (see section 1.4.6). This reputation system is a transposition of word-of-mouth as practiced in the real world.

1.4.2. *The various digital identifiers*

Among the digital identifiers typically listed, we can list [IT 10]:

– Main identifier: an identifier associated with his/her own identity in the real world. It is common to have several main identifiers depending on the context of use: professional, family-related, childhood friends, etc.

– Pseudonym: alias or name assumed to conceal their identity. The same individual can have several pseudonyms.

– Alias: an identifier enabling its owner to benefit from the properties of anonymity, and unlinkability. The current tendency is to associate an alias with an avatar, that is to say, a more or less graphic self-representation highlighting certain aspects of their personality. The avatar may be similar to the world of role-playing games where everyone can have fun impersonating a real or imaginary person.

A pseudonym or alias may or may not be authenticated by an application domain. A pseudonym is often used within social networks (Twitter, Facebook), collaborative sites (Wikipedia), sites for classified ads (leboncoin.fr), sites which facilitate transactions and sales (eBay, Amazon) while an alias is more frequently reserved for social networks. Note that, to open an account under a pseudonym (Twitter, leboncoin.fr), most of the time, it is necessary to provide an email address that is mainly used to communicate content related to the service, to monitor the offered service and to issue a new password if it has been forgotten. It is the only identifier to be provided. It is not necessarily attached to their main identity (gmail) and can have a very short lifetime. On the other hand, for other accounts involving a financial transaction (Amazon, Quelle, etc.), whether you are a buyer or seller, you are often obliged, during a transaction, to provide a bank card number with a name (or Paypal identifier), and postal address to receive a package. Thus, pseudonymity is

preserved with other participants in the service (Amazon vendors), but not for the SP (Amazon).

Finally, note that an alias is generated by a natural person. A pseudonym can be generated by a person or by the federated identity management system in order to preserve the secret of the individual's real identity (see section 1.4.3.3). In the latter case, the pseudonym has a lifetime limited to the transaction. Another transaction involving the same stakeholders will result in the generation of another pseudonym, the aim being to protect themselves against SP linkability operations.

1.4.3. *Digital identity management*

Identity management systems have evolved significantly over the past 10 years. While the first generation of systems (isolated or silo model) for users consisted of managing in total isolation their identifiers and attributes according to the service, the next generation (centralized model) introduced centralized management thus providing users with ease of use. More recently, with the emergence of collaborative and distributed services, two new models have emerged: the federated model and the user-centered model. In this section, we intend to present these different models along with their advantages and disadvantages and a list of existing software solutions.

To unify the description of identity management models, we define the following entities:

– a user: a natural person with at least one digital identity wishes to conduct a transaction;

– an identity provider (IdP): an entity in charge of digital identity management and of the execution of the authentication mechanism. It enrolls any new user by registering their identifier(s) and some of their attributes. During enrollment, according to its policy, it may be necessary to verify the veracity of the identity provided with the help of an identity card, proof of residence, or even mere proof of receipt of an email;

– a service provider (SP): an entity providing users with a service usually a Web service, and relying on the IdP in order to verify the identity given by the user.

1.4.3.1. *Isolated or silo model*

In this historic model, the user must manage as many identifiers (ID) and credentials (for example passwords) as service providers SP1, SP2, SP3, etc. Note that in this model (see Figure 1.1), the attributes associated with each identifier are managed in isolation by each SP. Still today, a large number of Web services do operate in this way.

The big drawback of this model is the large number of logins and passwords to be memorized by the user. Therefore, there is a significant risk that the user will choose the same logins and passwords for several of their accounts, which reduces the level of security. In fact, a cyber-attacker may be more likely to attack servers known to be vulnerable to recover passwords, and then use these same passwords to access several user accounts hosted on more robust sites.

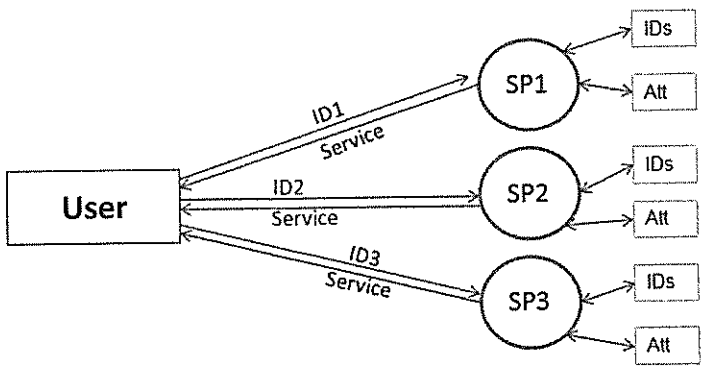

Figure 1.1. *Identity management – isolated or silo model*

1.4.3.2. *Centralized model*

This model introduces an IdP that centralizes digital identity management (see Figure 1.2). Thus the user can authenticate themselves with SPs with the same identity, the same credential and all this without having to repeat authentication for each new SP requested. We speak of a "Single Sign On" mechanism as a single instance of authentication grants access to all SPs depending on the same IdP. OpenAM (successor to Open SSO) [OPE 14a] in its simple version offers an open source software solution to this model. If

ease of use is undeniable with regard to the isolated model, the centralized model is vulnerable as disclosure of one identifier with the associated credential (provided it is static) is sufficient for giving at once unauthorized access to all services. Furthermore the centralized aspect of this model does not make it suitable for a large number of users or SPs.

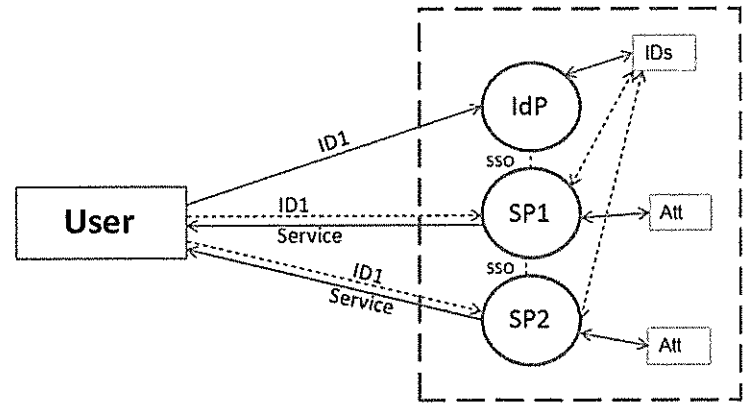

Figure 1.2. *Identity management – centralized model*

1.4.3.3. *Federated model*

The model illustrated in Figure 1.3 assumes that the IdP and SPs group together to form a federation of identities and are bound by relations of trust due to commercial agreements and a common technology platform (OpenID Connect [OPE 14b], Shibboleth [SHI 14], WS-Federation [WS 07], OpenAM [OPE 14a]). This federation is called a Circle of Trust (CoT). Just like in the centralized model, SSO mechanisms can be implemented so that the user can authenticate himself/herself a single time with the IdP to access the services of SPs that are members of the CoT. On the other hand, the user who accesses an SP is then referenced by the SP with the help of a pseudonym. In fact, all exchanges between SPs and IdP that are related to a user are done on the basis of these pseudonyms. This model is suitable for a large number of users and SPs. It is particularly interesting within the context of distributed and collaborative services. As in the previous model, the user assigns their attributes and identifier to the IdP and SPs and they are forced to trust them to respect their privacy.

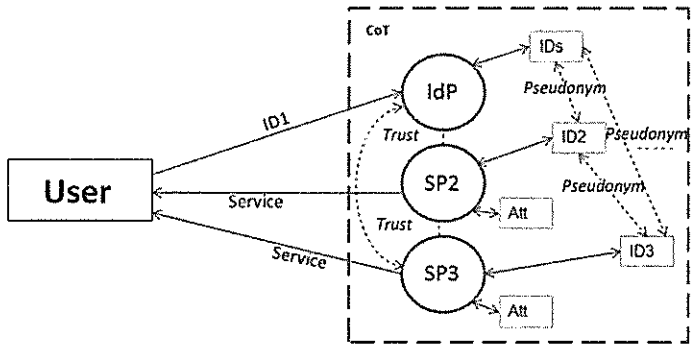

Figure 1.3. *Identity management – federated model*

1.4.3.4. *user-centric model*

Only the model presented in Figure 1.4 allows the user to have complete control over their personal attributes. From their workstation, in either a local or remote fashion on the IdP of their choice, they have a portfolio of electronic identities and sometimes an identity selector. At the request of the services and SPs being accessed, they can select an identity and decide whether to issue certain attributes. U-Prove [PAQ 13] is a software solution of this type involving an IdP responsible for signing a token proving the validity of the user's attributes. Note that the SPs act individually in this model and can, albeit not without difficulty, offer collaborative services. SPs are increasingly inclined to propose authentication of user by leaving them to decide on the choice of IdP. This is, for example, the case with Yahoo who offers the possibility of authenticating users with their Facebook or Google account.

The web community (W3C – World Wide Web Consortium [W3C 14]) is currently specifying a user-centric solution, but with different properties. This approach is known as both names: WebID for the approach used to identify a user and WebID-TLS for the protocol used to authenticate them [WEB 14]. WebID allows a user (or even an organization) to be uniquely identified by a Uniform Resource Identifier (URI) [BER 05] and to manage their profile in an online storage space at the same URI location, management being under their full control. The user's profile is defined on the basis of a vocabulary defined by Friend of a Friend (FOAF) [FOA 14] and is enriched with the user's electronic public key and an electronic

signature (potentially self-signed) for their WebID-TLS authentication. Thus, user authentication to an SP is pretty much like TLS protocol as the user transmits their certificate and an electronic signature. The difference with TLS is the form of certificate and the verification of the certificate by the SP. In fact, with WebID-TLS, the certificate carries the identifier corresponding to the profile's URI location and their verification consists of ensuring that the certificate received is the same as the one stored at the URI. Thus the WebID-TLS authentication has the sole purpose of verifying that the requesting user is the owner of the URI. The advantage of the WebID approach is that it leaves profile management up to the owner. On the other hand, by leaving all attributes of the personal space without access control, it does not so far offer any means for protecting user privacy.

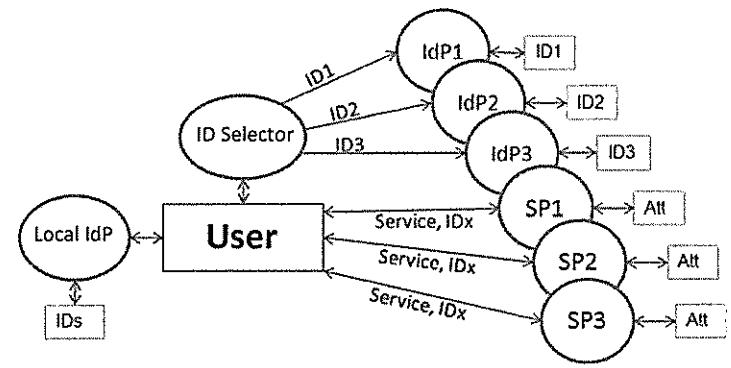

Figure 1.4. *Identity management – user-centric model*

1.4.5. *The norms*

The majority of identity management systems are based on web standards, in order to achieve entity authentication, exchange attributes, express particular policies regarding privacy. It is possible to split these standards into two categories:

– the language supporting authentication and exchange of attributes;

– the language for expressing privacy policies.

Security Assertion Markup Language (SAML 2.0) [MIS 05] was standardized in 2005 by OASIS (Committee of the Organization for the

Advancement of Structured Information Standards) [OAS 14]. It defines a structure based on XML to allow for the exchange of authentication and authorization data, and of attributes between an IdP and an SP, this structure can itself be secured by the Simple Object Access Protocol (SOAP) protocol [SOA 14] which can implement encryption and electronic signature mechanisms. SAML is very flexible with regard to the content of the exchanges between IdP, SP and user and identifies the different exchange patterns in the form of profiles (for example Web browser SSO, Single Logout, Basic Attribute Profile). The profile is selected according to the use scenario. Finally OAuth 2.0 [OAU 14] provides the user with the advantage of delegating rights to an application which will act on behalf of the user. Readers interested in this area may wish to refer to Chapter 2 for more details.

In compliance with regulations [EC 14], it has become necessary for an SP to publish their privacy management practices online, and for the user to express their preferences with regard to the protection of their privacy. Several languages have been designed for this purpose, most notably "Platform for Privacy Preferences" (P3P) [P3P 14] and "A P3P Preference Exchange Language" (APPEL) [APP 14]. P3P and APPEL were standardized by the consortium W3C [W3C 14] in 2002 to allow Websites to communicate their practices with regard to the collection, use and distribution of attributes received from users, and to allow the user to specify their privacy preferences. P3P refers to an XML syntax which can be understood by computer programs so that a browser can compare the practices of a Website with a user's preferences before continuing with the transaction. Note that P3P and APPEL are like a set of responses to multiple choice questions and do not make it easy to specify certain combinations, particularly what is acceptable. Readers interested in the technical and legal aspects associated with privacy in networks today may wish to refer to Chapter 4.

1.4.6. *The risks related to digital identity*

In the case of a main identity, just like in the real world, a cybercriminal may be interested in committing identity theft for profit or even in order to access advantageous services or confidential information whether private or professional. This cyber-attack [LAU 11] can be achieved by theft of identifier and credential obtained by social engineering (for example you

receive a telephone call from your network administrator who convinces you to provide your login/password), via phishing (for example: an e-mail indicating a computer problem invites you to connect to a site with your login/password), or through the installation of a keylogger Trojan on the victims machine which can capture and record the keys struck on a keyboard. The motives for such offences are as diverse as in the real world: bank transfers, the sale of trade secrets to competitors, fraud, etc. However, with our increasing tendency to entrust more and more of our private life on our computer, the theft of intimate data (photos and correspondence) could go further and lead to an increase in harassment, blackmail, etc.

As we have seen, the risk may also come from the owner of a main identity. Through negligence, they may divulge certain personal information on a social network, blog, or on a large distribution network such as DailyMotion or YouTube. It will then be difficult for them to control the disclosure and replication of this data and to have the right to be forgotten.

Finally, it is easy for a cyber-attacker to evolve when they are anonymous or using a pseudonym. In fact nothing stops them from creating several identities and, once identified as undesirable, from changing and disturbing the service operations again (e.g. a social network). This Sybil attack where the attacker has several identities is problematic for maintaining the quality of service (QoS) and for reputation systems that can see their rating system distorted. For instance, a cyber-crook can wear the hat of both the buyer and the seller so as to give positive reviews of completely fictitious sales transactions and thus favorably increase their seller rating.

1.5. Conclusions

This chapter has highlighted the multidisciplinary nature of the digital identity field. Researchers have proposed their own understanding and analysis about the self-representation in the digital world, the new economic models and challenges, and existing technical solutions along with their limitations.

Digital identities are the result of the major technological advancements that we have experienced during the last 20 years. They have given way to a virtual world in which individuals, groups and businesses need to find their place. The emerging area of digital identities is constantly and rapidly

changing and growing. As such, it brings together a community of scientists, philosophers, lawyers, industrials, educators, politicians, etc., who attempt to provide answers in order to build a regulated economically viable, secure and trusted digital world that society can appropriate.

1.6. Bibliography

[ACQ 10] ACQUISTI A., "The economics and behavioral economics of privacy: a note", *Proceedings of the Third International Conference on Ethics and Policy of Biometrics and International Data Sharing*, Hong Kong, Springer LNCS, vol. 60059, 2010.

[ACS 13] ACSEL (Association de l'Economie Numérique), relation numérique de confiance; des enjeux des identités numériques, June 2013. Available at:http://www.acsel.asso.fr/2013/version-numerique-du-cahier-la-relation-numerique-de-confiance-des-enjeux-des-identites.

[AGU 09] AGUITON C., *et al.*, "Does showing off help to make friends? Experimenting a sociological game on self-exhibition and social networks", *Proceedings of the 3rd International ICWSM Conference*, San Jose, pp. 10–17, 2009.

[AKE 00] AKERLOF G.A., KRANTON R.E., "Economics and identity", *Quarterly Journal of Economics*, vol. 115, no. 3, pp. 715–753, 2000.

[ALL 03] ALLARD L., VANDENBERGHE F., "Express yourself! Les pages perso. Entre légitimation technopolitique de l'individualisme expressif et authenticité réflexive peer to peer", *Réseaux*, vol. 21, no. 117, pp. 191–220, 2003.

[ALL 07] ALLARD L., "Blogs, podcast, tags, mashups, cartographies, locative medias : le tournant expressiviste du web", *Médiamorphoses*, no. 21, pp. 57–62, 2007.

[ALL 09] ALLARD L., "Pragmatique de l'Internet mobile. Technologies de soi et culture du transfert", in DERVIN F., ABBAS Y., (eds.), *Technologies numériques du soi et (co-)constructions identitaires,* coll. Questions contemporaines, L'Harmattan, Paris, pp. 59–82, 2009.

[APP 14] APPEL, *APPEL 1.0: A P3P Preference Exchange Language 1.0*, http://www.w3.org/TR/P3P-preferences/, 2014.

[BEA 99] BEAUDOUIN V., VELKOVSKA J., "Constitution d'un espace de communication sur Internet (forums, pages personnelles, courrier électronique...)", *Réseaux*, vol. 17, no. 97, pp. 121–178, 1999.

[BER 05] BERNERS-LEE T., MASINTER L., BERNERS-LEE R., Uniform Resource Identifier (URI): GenericSyntax, Standards Track, RFC 3986, January 2005.

[BOU 11] BOUNIE D., EANG B., SIRBU M., *et al.*, "Une analyse empirique de la dispersion des prix sur Internet", *Revue Française d'Economie*, vol. 25, pp. 121–145, 2011.

[BOY 08] BOYD D., ELLISON N., "Social Network Sites: definition, history and scholarship", *Journal Of Computer Mediated Communication*, vol. 13, no. 1, pp. 210–230, 2008.

[BRO 07] BRÖCKLING U., *Das unternehmerische Selbst. Soziologieeiner Subjektivierungsform*, Frankfurt, Suhrkamp, 2007.

[CAB 10] CABRAL L., HORTASU A., "The dynamics of seller reputation: theory and evidence from eBay", *Journal of Industrial Economics*, vol. 58, pp. 54–78, 2010.

[CAB 12] CABRAL L., "Reputation on the internet", *Oxford Handbook of the Digital Economy*, Oxford University Press, 2012.

[CAR 08] CARDON D., "Le design de la visibilité. Un essai de cartographie du Web 2.0", *Réseaux*, vol. 26, no. 152, pp. 93–137, 2008.

[CAR 09] CARDON D., "L'identité comme stratégie relationnelle", *Hermès*, no. 53, pp. 61–66, 2009.

[CAS 03] CASTRONOVA E., Theory of the Avatar, CESIfo Working paper 863, 2003.

[CHE 07] CHESTER A., BRETHERTON D., "Impression management and identity online", in JOINSON A., MCKENNA K., POSTMES T., *et al.* (eds.), *Oxford Handbook of Internet Psychology*, Oxford: Oxford University Press, pp. 223–236, 2007.

[COU 10] COUTANT A., STENGER T., *Processus identitaire et ordre de l'interaction sur les réseaux socionumériques*, Mime, 2010.

[COU 11] COUTANT A., STENGER T., (eds.), "Ces réseaux numériques dits sociaux", *Hermès*, no. 59, 2011.

[DAV 10] DAVIS J.B., *Individuals and Identity in Economics*, Cambridge University Press, 2010.

[DEL 13] DELOITTE, Technology, Media & Telecommunications Predictions, 2013. Available at http://www2.deloitte.com/content/dam/Deloitte/global/ Documents/ Technology-Media-Telecommunications/gx-TMT-Predictions2013-Final.pdf

[DEN 11] DENOUËL J., GRANJON F., (eds.), *Communiquer à l'ère numérique. Regards croisés sur la sociologie des usages*, Paris, coll. Sciences sociales, Presses de l'école des Mines, 2011.

[DIN 07] DINI F., SPAGNOLO G., *Buying Reputation on eBay*, QuadernoConsip VIII, 2007.

[DOR 02] DÖRING N., "Personal home pages on the web: a review of research", *Journal of Computer-Mediated Communication*, vol. 7, no. 3, 2002.